The Annals of Statistics
1982, Vol. 10, No. 1, 256–265

MAXIMUM LIKELIHOOD AND LEAST SQUARES ESTIMATION IN LINEAR AND AFFINE FUNCTIONAL MODELS[1]

BY C. VILLEGAS

Simon Fraser University

In a linear (or affine) functional model the principal parameter is a subspace (respectively an affine subspace) in a finite dimensional inner product space, which contains the means of n multivariate normal populations, all having the same covariance matrix. A relatively simple, essentially algebraic derivation of the maximum likelihood estimates is given, when these estimates are based on single observed vectors from each of the n populations and an independent estimate of the common covariance matrix. A new derivation of least squares estimates is also given.

1. Introduction. The problem of making inferences concerning linear relations when all the observed vectors are contaminated with errors or fluctuations has a long history and important applications, particularly in Factor Analysis and Econometrics. Traditionally, two different versions of this problem have been considered. In one version, called by Malinvaud (1970) the functional model, the corresponding "true" vectors are considered as unknown parameters, whereas in the other version, called the structural model, they are considered as independent random vectors.

More recently, Villegas (1976) and Villegas and Rennie (1976) introduced another version in which the true vectors constitute an autoregressive process. From a Bayesian viewpoint, the differences among these versions are not so important and perhaps it might be better if all of them were called linear functional models. The maximum likelihood (ML) estimation of linear functional models was considered in the general multivariate case by Anderson (1951) and Nussbaum (1976), using differential calculus. Villegas (1961) gave a relatively simple, essentially algebraic derivation for the case in which the linear relation is a hyperplane, i.e., an affine $(p-1)$-dimensional subspace of a p-dimensional space. An algebraic derivation for a general model has been given by Healy (1980). Least squares (LS) estimation of functional models has been considered recently by Eckart and Young (1936), Rao (1964) and Höschel (1978).

In the present paper, relatively simple, essentially algebraic derivations of LS and ML estimators are given for the general case in which the linear (or affine) relation is an r-dimensional subspace (resp. affine subspace) of a p-dimensional space.

2. The model. In the present paper, R^p is the space of p-dimensional row vectors. Assume that we are given n observed vectors $y^{(i)} \in R^p$, $i = 1, \cdots, n$, independently drawn from n multivariate normal distributions having the same $p \times p$ covariance matrix Σ and unknown mean vectors $x^{(i)}$ belonging to an unknown r-dimensional subspace H. (Unless otherwise specified, we use superscripts to distinguish between vectors $x^{(i)}$ and the components x_i of a single vector x.) The $p \times p$ triangular deviation Λ is defined as the positive lower triangular square root of Σ, i.e., the lower triangular matrix with positive

---

Received February 1981; revised June 1981.

[1] Research supported by the Natural Sciences and Engineering Research Council of Canada Grant A8097.

AMS 1970 *subject classifications.* Primary 62A05; secondary 62F15.

Key words and phrases. Bayesian inference, logical priors, inner statistical inference, invariance, conditional confidence, multivariate analysis.

diagonal elements uniquely defined by

$$\Sigma = \Lambda\Lambda'. \tag{2.1}$$

The linear functional model is then, symbolically,

$$\mathscr{L}\{\Lambda^{-1}(y^{(i)} - x^{(i)})' : i = 1, \cdots, n\} = \mathscr{N}_p^n, \tag{2.2}$$

$$x^{(i)} \in H, \qquad (i = 1, \cdots, n), \tag{2.3}$$

where H is an unknown r-dimensional subspace and $\mathscr{N}_p^n$ denotes the joint distribution of n independent p-dimensional standard normal vectors. In the affine functional model, the only difference is that the vectors $x^{(i)}$ are assumed to belong to an affine subspace.

In the usual case where Σ (and therefore Λ) is unknown, we assume that we are also given a $p \times p$ matrix W drawn from a Wishart distribution with covariance matrix Σ and ν degrees of freedom. This matrix W may be obtainable from a preliminary statistical analysis, which is possible, for example, when replications are available, or when certain econometric models are analyzed (Anderson, 1976). In other cases, as in factor analysis for instance, the matrix W may still be available as a fictitious data matrix representing personal, subjective knowledge.

Let Y and ΔY be the $n \times p$ matrices whose ith rows are, respectively, $y^{(i)}$ and $\Delta y^{(i)} = y^{(i)} - \bar{y}$, where $\bar{y} = n^{-1} \sum_{i=1}^{p} y^{(i)}$. If the model is linear, we shall assume that $Y'Y$ is nonsingular, and, if the model is affine, we shall assume that $\Delta Y' \Delta Y$ is nonsingular. These assumptions imply that, in the linear case, $n \geq p$, and, in the affine case, $n - 1 \geq p$ (in which case the above assumptions will be satisfied with probability 1).

3. Least squares. The LS (least squares) estimates $\hat{x}^{(i)}$ are the vectors that minimize the sum of squares

$$\Sigma \| x^{(i)} - y^{(i)} \|^2 \tag{3.1}$$

under the restriction $x^{(i)} \in H$. This restriction is equivalent to $X \in \mathscr{X}_r$, where X is the matrix whose ith row is $x^{(i)}$ and $\mathscr{X}_r$ is the set of all $n \times p$ matrices which have rank at most r.

According to the Singular Value Decomposition (SVD) Theorem,

$$Y = \sum_{i=1}^{p} \lambda_i u_i' v_i \tag{3.2}$$

where the u_i are orthonormal vectors in R^n, the v_i are orthonormal vectors in R^p, and $\lambda_1 \geq \cdots \geq \lambda_p > 0$. Furthermore, $\lambda_1^2, \cdots, \lambda_p^2$ are the eigenvalues and $v_1, \cdots, v_p$ are the eigenvectors of $Y'Y$.

According to a result of Eckart and Young (1936) (see also Marshall and Olkin, 1979; Rao, 1964, 1979; and problem 10, page 70 in Rao, 1973), the LS estimate of X is

$$\hat{X} = \sum_{i=1}^{r} \lambda_i u_i' v_i. \tag{3.3}$$

Obviously, the LS estimate of H is

$$\hat{H} = \operatorname{span}\{v_1, \cdots, v_r\}. \tag{3.4}$$

In other words, the least squares estimate of H is the subspace spanned by the r eigenvectors of $Y'Y$ corresponding to the r larger eigenvalues.

An alternative, more geometric derivation of the LS estimators is as follows. Let $\| y^{(i)}, H \|$ denote the orthogonal distance from the point $y^{(i)}$ to a subspace H. Use the Pythagorean Theorem to show that the LS estimate $\hat{H}$ is the subspace that minimizes $\sum_{i=1}^{n} \| y^{(i)}, H \|^2$, and that the LS estimates $\hat{x}^{(i)}$ are the orthogonal projections of the points $y^{(i)}$ on $\hat{H}$. Let B be a $q \times p$ matrix whose q rows are an orthonormal basis in $H^{\perp}$, the orthogonal complement of H. Then

$$\sum_{i=1}^{n} \| y^{(i)}, H \|^2 = \sum_{i=1}^{n} \| B y^{(i)\prime} \|^2 = \operatorname{tr} BY'YB' = \sum_{j=1}^{q} e_j^*,$$

where $e_1^* \geq \cdots \geq e_q^*$ are the eigenvalues of $BY'YB'$, indexed in decreasing order. Let e_1

$\geq \cdots \geq e_p$ be the eigenvalues of $Y'Y$, also indexed in decreasing order. By the Cauchy-Poincaré Separation Theorem (Beckenbach and Bellman, 1965, page 76), $e_{p-j+1} \leq e^*_{q-j+1}$ for $j = 1, \cdots, q$. Therefore

$$\textstyle\sum_{i=1}^{n} \| y^{(i)}, H\|^2 \geq \sum_{j=1}^{q} e_{p-j+1},$$

and the minimum is reached iff

$$e_{p-j+1} = e^*_{q-j+1}, \qquad (j = 1, \cdots, q). \tag{3.5}$$

We shall consider only the case, which happens with probability 1, in which $e_r > e_{r+1}$. Then, by Theorem 1 of Appendix A, the equalities (3.5) hold iff $\hat{H}$ is the subspace spanned by the q eigenvectors of $Y'Y$ corresponding to the q smaller eigenvalues, or, equivalently, iff $\hat{H}$ is given by (3.4).

Given a positive definite symmetric matrix W, we can define, as in Eaton (1970), a W-inner product $(.,.)_W$ by $(x, y)_W = (x^1, W^{-1}y^1)$ *and a* W-norm $\| \cdot \|_W$ by $\| x \|_W^2 = (x, x)_W$. The estimators that minimize $\sum_{i=1}^{n} \| x^{(i)} - y^{(i)} |_W^2$, subject to the restrictions $x^{(i)} \in H$, $(i = 1, \cdots, n)$, are called generalized least squares (GLS) estimators.

The problem of finding GLS estimators may be reduced to the previous problem by a preliminary linear transformation with matrix L^{-1}, where L is the positive lower triangular matrix uniquely defined by the triangular decomposition $W = LL'$. It follows that the GLS estimate $\hat{H}$ is the image under L of the subspace spanned by the eigenvectors corresponding to the r largest eigenvalues of $L^{-1}Y'YL'^{-1}$, and that the GLS estimates $\hat{x}^{(i)}$ are the W-projections on $\hat{H}$ of the points $y^{(i)}$. Note that, if we denote by $|M|$ the determinant of a matrix M, the eigenvalues of $L^{-1}Y'YL'^{-1}$ are the solutions of the equation $| Y'Y - eW| = 0$ in the scalar variable e.

To any r-dimensional subspace H associate an $r \times p$ matrix $\Gamma(H)$ whose rows are an orthonormal basis in $L^{-1}H$. Then $L\Gamma(H)'\Gamma(H)L^{-1}$ is the matrix of the W-projection on H. Therefore, the GLS estimator of X' is

$$\hat{X}' = L\Gamma(\hat{H})'\Gamma(\hat{H})L^{-1}Y'. \tag{3.6}$$

Note the similarities of this estimator with the Gauss-Markov estimators in the case of a fixed design matrix (Eaton, 1970).

4. The likelihood function. In the usual case when the positive definite covariance matrix Σ is unknown, we assume the availability of a matrix W that has been drawn from a Wishart distribution with parameter Σ and $\nu \geq p$ degrees of freedom. The probability density of W is proportional to

$$|\Sigma|^{-\nu/2}| W|^{(\nu-p-1)/2}\exp(- \tfrac{1}{2} \operatorname{tr} \Sigma^{-1}W), \tag{4.1}$$

when W is positive definite and zero otherwise.

The joint probability density of the observed vectors $y^{(i)}$ is proportional to

$$|\Sigma|^{-n/2}\exp\{- \tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(Y - X)'(Y - X)\}. \tag{4.2}$$

The likelihood function is proportional to the product of (4.1) and (4.2), or equivalently, to

$$|\Sigma|^{-N/2}\exp[- \tfrac{1}{2} \operatorname{tr} \Sigma^{-1}\{W + (Y - X)'(Y - X)\}], \tag{4.3}$$

where $N = n + \nu$

From Lemma 5.1.1 of Giri (1977, page 74), it follows that, for a given X, the value of Σ that maximizes (4.3) is

$$\hat{\Sigma}(X) = N^{-1}\{W + (Y - X)'(Y - X)\}. \tag{4.4}$$

The corresponding maximum value of (6.3) is proportional to

$$| W + (Y - X)'(Y - X)|^{-N/2}, \tag{4.5}$$

and, when considered as a function of X, is called the ML (maximum likelihood) function for X.

5. The ML function for H. The ML estimate of X is the value $\hat{X}$ that maximizes the ML function (4.5), or, equivalently, the value that minimizes

$$|W + (Y - X)'(Y - X)|. \tag{5.1}$$

We shall first minimize (5.1) subject to the conditions $x^{(i)} \in H$, for a given r-dimensional subspace H.

Let L be the positive lower triangular matrix uniquely defined by the triangular decomposition

$$W = LL' \tag{5.2}$$

and consider new vectors $x_*^{(i)}$, $y_*^{(i)}$ defined by

$$y_*^{(i)\prime} = L^{-1}y^{(i)\prime}, \qquad x_*^{(i)\prime} = L^{-1}x^{(i)\prime},$$

$i = 1, \cdots, n$. Then the problem of minimizing (5.1), subject to the conditions $x^{(i)} \in H$, is equivalent to the problem of minimizing

$$|I_p + \textstyle\sum_{i=1}^n (y_*^{(i)} - x_*^{(i)})'(y_*^{(i)} - x_*^{(i)})| \tag{5.3}$$

subject to the conditions $x_*^{(i)} \in L^{-1}H$, $i = 1, \cdots, n$. By Lemma 1 of Appendix B, the minimizing $x_*^{(i)}$ are the orthogonal projections of the points $y_*^{(i)}$ on the subspace $L^{-1}H$. Equivalently, the matrix $\hat{X}(H)$ that minimizes (5.1) for a given H, is the matrix whose ith row $\hat{x}^{(i)}(H)$ is the W-projection of $y^{(i)}$ on the subspace H.

Let $\beta_1, \cdots, \beta_q$ be an orthonormal basis in $(L^{-1}H)^\perp$, the orthogonal complement of $L^{-1}H$, and let B be the $q \times p$ matrix whose ith row is β_i. Then $B'B$ is the matrix of the orthogonal projection on $(L^{-1}H)^\perp$, and

$$L^{-1}\hat{x}^{(i)}(H)' = (I_p - B'B)L^{-1}y^{(i)\prime}.$$

Therefore

$$L^{-1}(Y - \hat{X}(H))' = B'BL^{-1}Y'$$

and the minimum value of (5.1), considered as a function of B, is proportional to

$$|I_p + B'BL^{-1}Y'YL'^{-1}B'B|. \tag{5.4}$$

Since $BL^{-1}Y'YL'^{-1}B'$ is a positive definite $q \times q$ matrix, it has a triangular decomposition

$$BL^{-1}Y'YL'^{-1}B' = TT', \tag{5.5}$$

where T is a uniquely defined $q \times q$ positive lower triangular matrix. Therefore, (5.4) is equal to

$$|I_p + B'TT'B| = |I_q + T'BB'T| = |I_q + T'T| = |I_q + TT'|,$$

or, equivalently, to

$$|B(I_p + L^{-1}Y'YL'^{-1})B'|. \tag{5.6}$$

This determinant depends on B only through $H^\perp$, the subspace spanned by the rows of B. Therefore the maximum value of (4.5), subject to the conditions $x^{(i)} \in H$, and considered as a function of H, is proportional to

$$|B(I_p + L^{-1}Y'YL'^{-1})B'|^{-N/2}. \tag{5.7}$$

This function of H, defined up to an arbitrary scalar factor by (5.7), is the ML function for H.

6. The ML estimates. The ML estimate of H is the r-dimensional subspace $\hat{H}$ that maximizes the corresponding ML function (5.7), or, equivalently, the subspace that minimizes (5.6).

Let $e_1 \geq \cdots \geq e_p$ be the eigenvalues of $L^{-1}Y'YL'^{-1}$, indexed in decreasing order. Assuming that $e_r > e_{r+1}$, by Theorem 2 of Appendix A, the orthogonal complement of $L^{-1}\hat{H}$ is the subspace spanned by the eigenvectors corresponding to the q smaller eigenvalues of $L^{-1}Y'YL'^{-1}$.

It follows therefore that the ML estimate $\hat{H}$ of H is the image under L of the subspace spanned by the eigenvectors corresponding to the r larger eigenvalues of $L^{-1}Y'YL'^{-1}$. A comparison with the results of Section 3 shows that $\hat{H}$ is the GLS estimate that minimizes the sum of squares of W-distances to the points $y^{(i)}$.

The ML estimate of X is the matrix $\hat{X} = \hat{X}(\hat{H})$ that minimizes (5.1) subject to the conditions $x^{(i)} \in \hat{H}$. Clearly $\hat{X}$ is the matrix whose ith row $\hat{x}^{(i)}$ is the W-projection of $y^{(i)}$ on $\hat{H}$. According to the results of Section 3, the ML estimate $\hat{X}$ is the GLS estimate corresponding to the matrix W, and is given by (3.6).

Finally, from (4.4) it follows that the ML estimate of Σ is

$$\hat{\Sigma} = N^{-1}\{W + (Y - \hat{X})'(Y - X)\} \tag{6.1}$$

and is therefore a weighted average of $\nu^{-1}W$ and $n^{-1}(Y - \hat{X})'(Y - \hat{X})$ with weights ν and n.

7. GLS estimation in the affine model. The image of an r-dimensional subspace H under a translation $v \to v + \alpha$ is an r-dimensional affine subspace $H + \alpha$. In general, the vector α is not uniquely determined by the affine subspace $H + \alpha$. However, the vector α is uniquely determined by $H + \alpha$ if it is known that α is W-orthogonal to H, for a given positive definite symmetric matrix W, and to emphasize this the affine subspace will be denoted by $H \oplus_W \alpha$ or by $H \oplus \alpha$ if $W = I$.

We assume as before that we have n observations $y^{(i)} \in R^p$ having multivariate distributions with means $\mu^{(i)}$ belonging to an unknown r-dimensional affine subspace $H \oplus_W \alpha$. We shall denote by $x^{(i)}$ the unknown W-projection of $\mu^{(i)}$ on H. Therefore, we can write

$$\mu^{(i)} = x^{(i)} \oplus_W \alpha. \tag{7.1}$$

The GLS estimates $\hat{\mu}^{(i)}$ are the vectors that minimize the sum of squares of W-distances

$$\textstyle\sum_{i=1}^{n} \|y^{(i)} - \mu^{(i)}\|_W^2, \tag{7.2}$$

subject to the restrictions (7.1) and $x^{(i)} \in H$.

Let $\|y^{(i)}, H \oplus_W \alpha\|_W$ denote the W-distance from the point $y^{(i)}$ to the affine subspace $H \oplus_W \alpha$. By the Pythagorean Theorem, the sum of squares of W-distances (7.2) can be partitioned as the sum of two components,

$$\textstyle\sum_{i=1}^{n} \|y^{(i)}, H \oplus_W \alpha\|_W^2 \tag{7.3}$$

and

$$\textstyle\sum_{i=1}^{n} \|x^{(i)} - \hat{x}^{(i)}(H)\|_W^2, \tag{7.4}$$

where $\hat{x}^{(i)}(H)$ is the W-orthogonal projection of $y^{(i)}$ on H. It is clear that, for a given $H \oplus_W \alpha$, the $\hat{x}^{(i)}(H)$ are the values that minimize (7.4), and the corresponding minimum of (7.2) is (7.3).

But (7.3) is the W-moment of inertia, with respect to $H \oplus_W \alpha$, of a system of unit masses placed at the points $y^{(i)}$. Therefore, the sum of squares (7.3) can be partitioned as the sum of two components. The first component,

$$\textstyle\sum_{i=1}^{n} \|y^{(i)}, H + \bar{y}\|_W^2, \tag{7.5}$$

is the W-moment of inertia of the same system of masses, with respect to $H + \bar{y}$, the affine subspace parallel to H that goes through the baricenter $\bar{y} = n^{-1} \sum y^{(i)}$. The second component,

$$n \| \bar{y}, H \oplus_W \alpha \|_W^2, \tag{7.6}$$

is the W-moment of inertia, with respect to $H \oplus_W \alpha$, of the total mass n placed at the baricenter $\bar{y}$. The second component (7.6) is also equal to

$$n \| \bar{y} - \alpha - P_W \bar{y} \|_W^2, \tag{7.7}$$

where P_W is the W-projection on H. It is clear, then, that for a given H, the value of α that minimizes (7.2) is

$$\hat{\alpha}(H) = \bar{y} - P_W \bar{y}, \tag{7.8}$$

or, in other words, the W-projection of $\bar{y}$ on the W-orthogonal complement of H, say $H_W^{\perp}$. The corresponding minimum of (7.2) is (7.5) or, equivalently,

$$\textstyle\sum_{i=1}^{n} \| y^{(i)} - \bar{y}, H \|_W^2. \tag{7.9}$$

The GLS estimate $\hat{H}$ is simply the r-dimensional subspace that minimizes (7.9). This problem was solved in Section 3, the only difference being that $y^{(i)} - \bar{y}$ has to be substituted for $y^{(i)}$. The GLS estimate $\hat{\alpha}$ is obtained by substituting $\hat{H}$ for H in (7.8), and the GLS estimates $\hat{x}^{(i)}$ are obtained by substituting $\hat{H}$ for H in $\hat{x}^{(i)}(H)$.

8. ML estimation in the affine model. We assume now that the observed vectors $y^{(i)}$ are independent, and have multivariate normal distributions with common covariance matrix Σ and means $\mu^{(i)}$ belonging to an r-dimensional affine subspace. We assume that Σ and the means $\mu^{(i)}$, as well as the affine subspace to which they belong, are unknown. We also assume, as in Section 4, that an independent positive definite symmetric matrix W is available, having a Wishart distribution with covariance matrix Σ and ν degrees of freedom. If μ is the $n \times p$ matrix whose ith row is $\mu^{(i)}$, the joint likelihood function is proportional to

$$|\Sigma|^{-N/2} \exp[-\tfrac{1}{2} \Sigma^{-1}\{W + (Y - \mu)'(Y - \mu)\}], \tag{8.1}$$

where $N = n + \nu$ and the $\mu^{(i)}$ are assumed to belong to an r-dimensional affine subspace. Because of this restriction, we have to choose convenient coordinates for the points $\mu^{(i)}$. We shall choose a coordinate system that depends on W. As in Section 7, we shall denote the affine subspace by $H \oplus_W \alpha$, where H is an r-dimensional subspace and α is a row vector W-orthogonal to H, and we shall write $\mu^{(i)} = x^{(i)} \oplus_W \alpha$, so that $x^{(i)}$ is the W-projection of $\mu^{(i)}$ on H.

Let X be the $n \times p$ matrix whose ith row is $x^{(i)}$. Then the likelihood function (10.1) may be written

$$|\Sigma|^{-N/2} \exp[-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}\{W + (Y - X - \mathbf{1}'\alpha)'(Y - X - \mathbf{1}'\alpha)\}], \tag{8.2}$$

where $\mathbf{1}$ is a row vector of ones. As in Section 4, for fixed X and α, the value of Σ that maximizes (8.2) is

$$\hat{\Sigma}(X, \alpha) = N^{-1}\{W + (Y - X - \mathbf{1}'\alpha)'(Y - X - \mathbf{1}'\alpha)\}, \tag{8.3}$$

and the corresponding maximum is proportional to

$$|W + (Y - X - \mathbf{1}'\alpha)'(Y - X - \mathbf{1}'\alpha)|^{-N/2}. \tag{8.4}$$

This is the ML function for X and α. The problem of maximizing (8.4) for a fixed H is equivalent to the problem of minimizing the determinant

$$|W + \textstyle\sum_{i=1}^{n} (y^{(i)} - \alpha - x^{(i)})'(y^{(i)} - \alpha - x^{(i)})|. \tag{8.5}$$

By the Corollary of Appendix B, the minimizing $\hat{\alpha}(H)$ is the W-projection of $\bar{y}$ on $H_W^{\perp}$, the W-orthogonal complement of H, and the minimizing $\hat{x}^{(i)}(H)$ is the W-projection of $y^{(i)}$ on H. Therefore, $y^{(i)} - \hat{\alpha}(H) - \hat{x}^{(i)}(H)$ is the W-projection of $\Delta y^{(i)} = y^{(i)} - \bar{y}$ on $H_W^{\perp}$.

Let L be defined as in (5.2), and let us associate with any subspace H a $q \times p$ matrix $B = B(H)$ whose rows are an orthonormal basis in $(L^{-1}H)^{\perp}$, the orthogonal complement of $L^{-1}H$. Then $B'B$ is the matrix of the orthogonal projection on $(L^{-1}H)^{\perp}$, and the W-projection of $\Delta y^{(i)}$ on $H_W^{\perp}$ is $LB'BL^{-1}\Delta y^{(i)\prime}$.

The corresponding maximum value of (8.4) for a given H, when considered as a function of H, is proportional to

$$(8.6) \qquad |I_p + B'BL^{-1}\Delta Y'\Delta YL'^{-1}B'B|^{-N/2},$$

where ΔY is the $n \times p$ matrix whose ith row is $\Delta y^{(i)}$.

The ML estimate $\hat{H}$ is the r-dimensional subspace whose coordinate matrix $\hat{B}$ maximizes (8.6). The determinant in (8.6) is similar to (5.4) the only difference being that ΔY has to be substituted for Y. Therefore, as was shown in Section 5, the ML function for H is proportional to

$$(8.7) \qquad |B(I_p + L^{-1}\Delta Y'\Delta YL'^{-1})B'|^{-N/2}.$$

The problem of maximizing (8.7) was solved in Section 6. It follows that the ML estimate $\hat{H}$ is the image under L of the r-dimensional subspace spanned by the eigenvectors corresponding to the r largest eigenvalues of $\Delta Y'\Delta Y$.

The ML estimate $\hat{\alpha}$ is the W-projection of $\bar{y}$ on $\hat{H}_W^{\perp}$. The ML estimates $\hat{x}^{(i)}$ are the W-projections of the $y^{(i)}$ on $\hat{H}$, and the ML estimate $\hat{\Sigma}$ is obtained from (8.3) by substitution of the ML estimates of X and α.

A comparison with the results of Section 7 shows that the ML estimates of $H \oplus_W \alpha$ and X are also GLS estimates corresponding to the W-metric.

APPENDIX A

Let M be a q-dimensional subspace of R^p, and let P be the orthogonal projection on M. Let A be a positive definite transformation of R^p. Assume that the eigenvalues of A, $\lambda_1 \geq \cdots \geq \lambda_p$, are indexed in decreasing order. Clearly, if $\lambda_{p-q} > \lambda_{p-q+1}$ there is a unique subspace M_0 spanned by any orthonormal set of q eigenvectors of A corresponding to the q smaller eigenvalues.

Let $T: M \to M$ be a restriction of the transformation PA. For any vector $v \in M$,

$$(1) \qquad (Tv, v) = (Av, v),$$

and therefore T is positive definite. Let $\mu_1 \geq \cdots \geq \mu_q$ be the eigenvalues of T, indexed in decreasing order.

THEOREM 1. *If* $\lambda_{p-q} > \lambda_{p-q+1}$, *then*

$$(2) \qquad \lambda_{p-j+1} = \mu_{q-j+1}$$

for $j = 1, \cdots, q$, *if and only if* $M = M_0$.

PROOF. Suppose that $M = M_0$. Let $v_1, \cdots, v_q$ be an orthonormal basis in M_0, whose elements are eigenvectors of A corresponding to the q smaller eigenvalues. From (1) it follows that these vectors are also q orthonormal eigenvectors of T, and (2) follows immediately.

Conversely, suppose that (2) holds. Let $v_1, \cdots, v_q$ be an orthonormal set of eigenvectors of T, corresponding to the eigenvalues $\mu_1, \cdots, \mu_q$.

We shall prove by induction that they are also eigenvectors of A. Suppose that v_{q-j+1}, $j = 1, \cdots, k$, $(k < q)$, are eigenvectors of A, and let H_k be the orthogonal complement of the subspace spanned by these vectors. Then, because of (1), v_{q-k-1} minimizes $(v, Av)/\|v\|^2$ for all $v \in H_k$ and is therefore an eigenvector of A. □

Suppose that, to each q-dimensional subspace M, we have assigned a $q \times p$ matrix $X(M)$ such that q rows of $X(M)$ are an orthonormal basis for the subspace M.

Clearly $|X(M)A\, X(M)'|$ is independent of the particular choice of $X(M)$ and therefore there is a function f defined by

$$f(M) = |X(M)A\, X(M)'|. \tag{3}$$

THEOREM 2. *If $\lambda_{p-q} > \lambda_{p-q+1}$, then M_0 is the unique q-dimensional subspace that minimizes the function f defined by* (3).

PROOF. The q rows of $X(M)$ are an orthonormal basis for M. The matrix $X(M)A\, X(M)'$ is simply the matrix, with respect to this orthonormal basis, of the restriction $T: M \to M$ of the transformation PA. Therefore

$$|X(M)A\, X(M)'| = \textstyle\prod_{j=1}^{q} \mu_j,$$

where $\mu_1 \geq \cdots \geq \mu_q$ are the eigenvalues of T. By the Cauchy-Poincaré Separation Theorem (Beckenbach and Bellman, 1965, page 76)

$$\lambda_{p-j+1} \leq \mu_{q-j+1}$$

for $j = 1, \cdots, q$ and therefore

$$|X(M)A\, X(M)'| \geq \textstyle\prod_{j=1}^{q} \lambda_{p-j+1}.$$

(This inequality is due to Fan, 1950; see also Marshall and Olkin, 1979, page 512).

The conclusion follows immediately by Theorem 1.

APPENDIX B

Let H be a given r-dimensional subspace in R^p, and let $y_i \in R^p (i = 1, \cdots, n)$ be n given $1 \times p$ vectors. We shall find the $1 \times p$ vectors $\alpha \in H^\perp$ and $x_i \in H$ that minimize the determinant

$$|I_p + \textstyle\sum_{i=1}^{n} (y_i - \alpha - x_i)'(y_i - \alpha - x_i)|. \tag{1}$$

LEMMA 1. *If $\alpha \in H^\perp$ is fixed, then the determinant* (1) *is minimized, subject to the conditions $x_i \in H$, if and only if x_i is the orthogonal projection of y_i on H.*

PROOF. By an orthogonal transformation in R^p, the problem can be transformed into a similar problem with $H = H_1$, the subspace spanned by the first r elements of the canonical basis. Therefore we assume, without losing generality, that $H = H_1$.

The proof will be by induction in n. Suppose the result is true for $n - 1$. Let T be the positive upper triangular matrix uniquely defined by

$$(T'T)^{-1} = I_p + \textstyle\sum_{i=1}^{n-1} (y_i - \alpha - x_i)'(y_i - \alpha - x_i). \tag{2}$$

The determinant (1) is then equal to

$$|T'T|^{-1}|I_p + T(y_n - \alpha - x_n)'(y_n - \alpha - x_n)\, T'|,$$

or, equivalently, to

$$|T'T|^{-1}\{1 + \|T(y_n - \alpha - x_n)'\|^2\}. \tag{3}$$

Let $x_n(1)$, $y_n(1)$ be the $1 \times r$ vectors defined by the partitions

$$x_n = (x_n(1), 0), \qquad y_n = (y_n(1), y_n(2)),$$

and let α and T be partitioned accordingly as $\alpha = (0, \alpha(2))$ and

$$T = \begin{Bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{Bmatrix}.$$

Then

$$\| T(y_n - \alpha - x_n)' \| \geq \| T_{22}(y_n(2) - \alpha(2)') \| \tag{4}$$

and

$$(T_{22}' T_{22})^{-1} = I_q + \textstyle\sum_{i=1}^{n-1} \{y_i(2) - \alpha(2)\}'\{y_i(2) - \alpha(2)\},$$

where $q = p - r$. Therefore, the right side of (4) does not depend on the x_i and is a lower bound for $\| T(y_n - \alpha - x_n)' \|$. By the induction assumption, the first factor in (3), namely $| T'T |^{-1}$, is minimized if and only if x_i is the orthogonal projection of y_i for $i = 1, \cdots, n - 1$. This implies the vanishing of the first r components of the vectors $y_i - \alpha - x_i$, $i = 1, \cdots, n - 1$. Substitution in T_{11} and T_{12} gives $T_{11} = I_r$ and $T_{12} = 0$. Under these conditions, the second factor in (3) is

$$1 + \| y_n(1) - x_n(1) \|^2 + \| T_{22}\{y_n(2) - \alpha(2)\}' \|^2.$$

Therefore, the two factors of (3) are simultaneously minimized when the x_i are the orthogonal projections of the $y_i (i = 1, \cdots, n)$.

THEOREM 1. *Under the conditions $x_i \in H$, $\alpha \in H^\perp$, the determinant* (1) *is minimized if and only if α is the orthogonal projection of $\bar{y} = n^{-1} \Sigma y_i$ on the subspace $H^\perp$, and the x_i are the orthogonal projections of the y_i on H, $i = 1, \cdots, n$.*

PROOF. We start by fixing $\alpha \in H^\perp$. By Lemma 1, the determinant is then minimized if and only if the x_i are the orthogonal projections of the y_i on H. If P is the matrix of the orthogonal projection on H, we have $x_i' = Py_i'$. The corresponding minimum is

$$| I_p + \textstyle\sum_{i=1}^{n} \{(I_p - P)y_i' - \alpha'\}\{y_i(I_p - P) - \alpha\} |. \tag{5}$$

Let T be the positive upper triangular matrix defined by

$$(T'T)^{-1} = I_p + (I_p - P)\Delta Y' \Delta Y (I_p - P),$$

where ΔY is the $n \times p$ matrix whose ith row is $y_i - \bar{y}$. A bit of algebra shows that (5) is equal to

$$| T'T |^{-1} | I_p + n\, T\{\alpha' - (I_p - P)\bar{y}'\}\{\alpha - \bar{y}(I_p - P)\}T' |$$

or, equivalently, to

$$| T'T |^{-1} [1 + n \| T\{\alpha' - (I_p - P)\bar{y}'\} \|^2],$$

and the conclusion follows immediately.

COROLLARY. *If W is a positive definite symmetric matrix, then the determinant*

$$| W + \textstyle\sum_{i=1}^{n} (y_i - \alpha - x_i)'(y_i - \alpha - x_i) | \tag{6}$$

is minimized, subject to the conditions $x_i \in H$, $\alpha \in H_W^\perp$, if and only if α is the W-projection of $\bar{y}$ on the subspace $H_W^\perp$ (the W-orthogonal complement of H), and the x_i are the W-projections of the y_i on H.

PROOF. The problem of minimizing (6) is reduced to the problem of minimizing (1) by the transformation L^{-1}, where L is the positive lower triangular matrix uniquely defined by (5.3).

Acknowledgement. The author wants to thank an Associate Editor for his helpful comments.

REFERENCES

ANDERSON, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22** 327–351.

ANDERSON, T. W. (1976). Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics (with discussion). *J. Roy. Statist. Soc. B* **38** 1–36.

BECKENBACH, E. F. and BELLMAN, R. (1965). *Inequalities.* Springer-Verlag, New York.

EATON, M. L. (1970). Gauss-Markov estimation for multivariate linear models: a coordinate free approach. *Ann. Math. Statist.* **41** 528–538.

ECKART, C. and YOUNG, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika* **1** 211–218.

FAN, K. (1950). On a theorem of Weyl concerning eigenvalues of linear transformations II. *Proc. Nat. Acad. Sci. U.S.A.* **36** 31–35.

GIRI, N. C. (1977). *Multivariate Statistical Inference.* Academic, New York.

HEALY, J. D. (1980). Maximum likelihood estimation of a multivariate linear functional relationship. *J. Multivariate Anal.* **10** 243–251.

HÖSCHEL, H. P. (1978). Generalized least squares estimators of linear functional relations with known error-covariance. *Math. Operationsforsch. Statist., Ser. Statistics* **9** 9–26.

MALINVAUD, E. (1970). *Statistical Methods in Econometrics.* North Holland, Amsterdam.

MARSHALL, A. W. and OLKIN, I. (1979). *Inequalities: Theory of Majorization and Its Applications.* Academic, New York.

NUSSBAUM, M. (1976). Maximum likelihood and least squares estimation of linear functional relationships. *Math. Operationsforsch. Statist.* **7** 23–49.

RAO, C. R. (1964). The use and interpretation of principal component analysis in applied research. *Sankhyā A* **26** 329–358.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications, second edition.* Wiley, New York.

RAO, C. R. (1979). Separation theorems for singular values of matrices and their applications in Multivariate Analysis. *J. Multivariate Anal.* **9** 362–377.

VILLEGAS, C. (1961). Maximum likelihood estimation of a linear functional relationship. *Ann. Math. Statist.* **32** 1048–1062.

VILLEGAS, C. (1972). Bayesian inference in linear relations. *Ann. Math. Statist.* **43** 1767–1791.

VILLEGAS, C. (1976). Linear relations in time series models I. *Journ. Multivariate Anal.* **6** 21–45.

VILLEGAS, C. and RENNIE, ROBERT R. (1976). Linear relations in time series models II. *J. Multivariate Anal.* **6** 46–64.

DEPARTMENT OF MATHEMATICS
SIMON FRASER UNIVERSITY
BURNABY, B.C.
CANADA V5A 1S6