ON THE ASYMPTOTIC EFFICIENCY OF LEAST SQUARES ESTIMATORS

By C. Villegas

Instituto de Matemática y Estadística, Montevideo

0. Summary. The problem of estimating a linear transformation between two finite dimensional vector spaces is considered, when the observed vectors in both spaces are subject to error, and there is an indefinitely increasing number of replications of a fixed number of treatments. A general class of ordinary estimators is defined, and it is shown that, in the case of homogeneity of variances, the simple least squares estimators are asymptotically efficient within the class of ordinary estimators, in the sense that they minimize, within that class, the asymptotic mean square error of prediction. When the error variances are unequal, however, the asymptotically efficient estimators are weighted least squares estimators, whose weights are based on preliminary estimators of the linear transformation and the error variances.

1. Introduction. Let $T: \mathfrak{X} \rightarrow \mathfrak{Y}$ be a linear transformation between two finite dimensional vector spaces $\mathfrak{X}$, $\mathfrak{Y}$. Suppose that, in order to estimate T, we have performed an experiment with n replications, and that, from a preliminary statistical analysis of the data obtained, we have derived $2k$ estimators $\mathbf{x}_{in}$, $\mathbf{y}_{in}(i = 1, \cdots, k)$ such that:

(1) for any given i, $\mathbf{x}_{in}$ is a random vector in $\mathfrak{X}$ converging in probability to $\boldsymbol{\xi}_i \varepsilon \mathfrak{X}$ when n, the number of replications, tends to infinity:

(2) similarly, $\mathbf{y}_{in}$ is a random vector in $\mathfrak{Y}$ which converges in probability to $\boldsymbol{\eta}_i \varepsilon \mathfrak{Y}$ when n tends to infinity:

(3) the unknown sure (non-random) vectors $\boldsymbol{\xi}_i$, $\boldsymbol{\eta}_i$ satisfy the relation $\boldsymbol{\eta}_i = T\boldsymbol{\xi}_i$, $(i = 1, \cdots, k)$, or, equivalently, the unknown random vectors $\mathbf{e}_{in}$ defined by $\mathbf{y}_{in} = T\mathbf{x}_{in} + \mathbf{e}_{in}$ converge in probability to zero when n tends to infinity:

(4) we assume also that the joint distribution of the random vectors $n^{\frac{1}{2}}\mathbf{e}_{1n}, \cdots, n^{\frac{1}{2}}\mathbf{e}_{kn}$ converges, when n tends to infinity, to the joint distribution of k random vectors $\mathbf{d}_1, \cdots, \mathbf{d}_k$ (usually normally distributed, with mean value zero).

In order to frame our problem in a more compact notation, consider an auxiliary vector space $\mathfrak{L}$ with an orthonormal basis $\mathbf{w}_1, \cdots, \mathbf{w}_k$ and define the linear transformations X_n, $X: \mathfrak{L} \rightarrow \mathfrak{X}$; Y_n, $Y: \mathfrak{L} \rightarrow \mathfrak{Y}$ by

$$X_n\boldsymbol{\lambda} = \textstyle\sum_i \lambda_i\mathbf{x}_{in}, \qquad X\boldsymbol{\lambda} = \sum \lambda_i\boldsymbol{\xi}_i; \tag{1.1}$$

$$Y_n\boldsymbol{\lambda} = \textstyle\sum_i \lambda_i\mathbf{y}_{in}, \qquad Y\boldsymbol{\lambda} = \sum \lambda_i\boldsymbol{\eta}_i, \tag{1.2}$$

where $\boldsymbol{\lambda} = \sum \lambda_i\mathbf{w}_i$. Then our previous assumptions may be summarized a

Received 15 November 1965; revised 24 May 1966.

follows:

(i) $X_n \lambda \rightarrow_P X\lambda$,

(ii) $Y_n \lambda \rightarrow_P Y\lambda$,

(iii) $Y = TX$, or, equivalently, if $E_n : \mathfrak{L} \rightarrow \mathfrak{Y}$ is defined by

$$Y_n = TX_n + E_n , \tag{1.3}$$

then $E_n \lambda \rightarrow_P \mathbf{0}$.

(iv) There is a random linear transformation $D: \mathfrak{L} \rightarrow \mathfrak{Y}$ such that, for any $\lambda \varepsilon \mathfrak{L}$, $n^{\frac{1}{2}} E_n \lambda$ converges in distribution to $D\lambda$ when n tends to infinity.

We shall assume in addition that:

(v) X is a surjective transformation (i.e., a transformation from $\mathfrak{L}$ *onto* $\mathfrak{X}$), and, with probability 1, X_n is also a surjective transformation;

(vi) If $D' : \mathfrak{Y} \rightarrow \mathfrak{L}$ is the adjoint of D, then the expected value of $D'D\lambda$, denoted by $\mathcal{E}D'D\lambda$, exists and is equal to $\Sigma\lambda$, where $\Sigma: \mathfrak{L} \rightarrow \mathfrak{L}$ is a positive definite transformation;

(vii) There is an estimator S of Σ which, with probability 1, is a positive definite transformation, and converges in probability to Σ in the sense that, for any $\lambda \varepsilon \mathfrak{L}$, $S\lambda \rightarrow_P \Sigma\lambda$.

Note that, since $\mathbf{d}_i = D\mathbf{w}_i$, the inner product $(\Sigma \mathbf{w}_i , \mathbf{w}_j)$ is equal to $\mathcal{E}(\mathbf{d}_i , \mathbf{d}_j)$. Therefore, in the usual case in which the random vectors $\mathbf{d}_i$ are independent, the matrix of Σ with respect to the given basis is the diagonal matrix diag $(\sigma_1^2, \cdots, \sigma_k^2)$ whose diagonal elements are $\sigma_i^2 = \mathcal{E} \|\mathbf{d}_i\|^2$. Obviously, in this case the matrix of S will be diag $(s_1^2, \cdots, s_k^2)$, where s_i^2 are consistent estimators of σ_i^2. It should be remarked that, in general, the computation of s_i^2 involves the use of a consistent preliminary estimator of T.

This model, in which both variables, $\mathbf{x}$ and $\mathbf{y}$, are subject to error, includes as a special case the usual linear regression model in which the independent variable is known without error. As in the linear regression theory, the case in which the linear transformation is non-homogeneous may be brought back to the homogeneous case by introducing a new component of $\mathbf{x}$ which is always equal to 1.

For a review of the literature on the statistical analysis of linear transformations when both variables are subject to error, the reader is referred to [5] (and, for more recent contributions, to [2] and [7], which also contain additional bibliography).

2. Ordinary estimators. Let $L_n : \mathfrak{X} \rightarrow \mathfrak{L}$ be a random linear transformation. The transformation $T_n : \mathfrak{X} \rightarrow \mathfrak{Y}$ defined by

$$T_n = Y_n L_n \tag{2.1}$$

will be called an *ordinary estimator* of T, if, with probability 1,

$$X_n L_n = I, \tag{2.2}$$

where I is the identity transformation.

Obviously, if p is the dimension of $\mathfrak{X}$, then $\mathfrak{R}(L_n)$, the range of L_n, is a subspace of $\mathfrak{L}$ of dimension p, and therefore we must have $k \geqq p$. Let $\mathbf{w}_{1n}, \cdots, \mathbf{w}_{pn}$ be a basis in $\mathfrak{R}(L_n)$, and let

$$\mathbf{x}_{hn}^* = X_n \mathbf{w}_{hn}, \qquad \mathbf{y}_{hn}^* = Y_n \mathbf{w}_{hn}, \qquad (h = 1, \cdots, p). \tag{2.3}$$

Then the p vectors $\mathbf{x}_{hn}^*$ are a basis in $\mathfrak{X}$, and the transformation T_n is defined by

$$T_n \mathbf{x}_{hn}^* = \mathbf{y}_{hn}^*, \qquad (h = 1, \cdots, p). \tag{2.4}$$

Let l_{hin} be random variables defined by

$$\mathbf{w}_{hn} = \sum_i l_{hin} \mathbf{w}_i . \tag{2.5}$$

Then, from (2.3) it follows that

$$\mathbf{x}_{hn}^* = \sum_i l_{hin} \mathbf{x}_{in}, \qquad \mathbf{y}_{hn}^* = \sum_i l_{hin} \mathbf{y}_{in} . \tag{2.6}$$

Conversely, assume that l_{hin} are random variables such that, with probability 1, the random vectors $\mathbf{x}_{hn}^*$ defined by (2.6) are a basis of $\mathfrak{X}$, and define T_n by (2.4), where $\mathbf{y}_{hn}^*$ is defined by (2.6). Let L_n be the linear transformation defined by $L_n \mathbf{x}_{hn}^* = \mathbf{w}_{hn}$, where $\mathbf{w}_{hn}$ is defined by (2.5). Then, obviously, (2.1) and (2.2) are satisfied. Hence, an ordinary estimator of T may be defined as the unique linear transformation $T_n : \mathfrak{X} \rightarrow \mathfrak{Y}$ which satisfies (2.4), where the random vectors $\mathbf{x}_{hn}^*$, $\mathbf{y}_{hn}^*$ are defined by (2.6) as random linear combinations (i.e., linear combinations with random coefficients) of the observed vectors $\mathbf{x}_{in}$, $\mathbf{y}_{in}$, such that, with probability 1, the vectors $\mathbf{x}_{hn}^*$ are a basis of $\mathfrak{X}$.

In the important case in which $\mathfrak{X}$ and $\mathfrak{Y}$ have dimension 1, the transformation T is defined by $T\xi = \alpha\xi$, where α is a real number, and L_n is defined by

$$L_n \xi = \xi \sum_i l_{in} \mathbf{w}_i / \sum_i l_{in} x_{in},$$

where l_{in} are random variables, which, in general, are functions of the observed values x_{in}, y_{in}. Then the estimator T_n is defined by $T_n \xi = a_n \xi$, where

$$a_n = \sum_i l_{in} y_{in} / \sum_i l_{in} x_{in} \tag{2.7}$$

is an estimator of α. Estimators of this type have been considered by Geary [3] in situations in which no replications are available.

It can be seen that the class of ordinary estimators is indeed a very large class of estimators. If, in (2.7), we choose $l_{in} = x_{in}$, then a_n is the classical least squares estimator. Estimators of the type (2.7) with constant coefficients $l_{in} = \lambda_i$ will be called *weighting estimators*. Simple estimators of this kind are the *grouping estimators* [5]. In the classical Gauss-Markov theory of linear estimation, the vectors $\mathbf{x}_{in}$ are not subject to error, and $\mathcal{E}\mathbf{y}_{in} = \mathfrak{n}_i$. In that case, a linear unbiased estimator of T is a transformation defined by (2.1) in which L_n is a non-random linear transformation such that $\mathcal{E}T_n\xi = T\xi$ for any $\xi \varepsilon \mathfrak{X}$, whatever be the linear transformation T. Since this condition implies (2.2), it follows that the linear unbiased estimators considered in the Gauss-Markov theory are also ordinary estimators.

3. Some theorems on linear transformations and convergence in distribution.

THEOREM 3.1. *The sequence of random vectors* $\{\mathbf{x}_n : n = 1, 2, \cdots\}$ *of a finite dimensional vector space* $\mathcal{V}$, *converges in distribution to a random vector* $\mathbf{x}$, *if and only if the inner product* $(\mathbf{x}_n, \mathbf{v})$ *converges in distribution to* $(\mathbf{x}, \mathbf{v})$ *for any* $\mathbf{v} \in \mathcal{V}$.

PROOF. See [4], p. 340, Proposition 7.1.

THEOREM 3.2. *If* $\mathcal{V}$ *is a finite dimensional vector space, f is a continuous, real function defined in* $\mathcal{V}$, *and* $\mathbf{x}_n$ *is a random vector in* $\mathcal{V}$ *which converges in distribution to a random vector* $\mathbf{x}$, *then the random variable* $f(\mathbf{x}_n)$ *converges in distribution to* $f(\mathbf{x})$.

PROOF. See [1], Theorem 2.1 (iv).

THEOREM 3.3. *If* $\mathfrak{X}$, $\mathcal{Y}$ *are finite dimensional vector spaces, and* $\mathbf{x}_n \in \mathfrak{X}$, $\mathbf{y}_n \in \mathcal{Y}$ *are random vectors which converge in distribution to a constant* $\boldsymbol{\alpha}$ *and to a random vector* $\mathbf{y}$ *respectively, then the pair* $\{\mathbf{x}_n, \mathbf{y}_n\}$, *considered as a random vector in the product space* $\mathfrak{X} \times \mathcal{Y}$ *converges in distribution to the random vector* $\{\boldsymbol{\alpha}, \mathbf{y}\}$.

PROOF. It is sufficient to note that, if $\mathbf{u} \in \mathfrak{X}$, $\mathbf{v} \in \mathcal{Y}$, then

$$(\{\mathbf{x}_n, \mathbf{y}_n\}, \{\mathbf{u}, \mathbf{v}\}) = (\mathbf{x}_n, \mathbf{u}) + (\mathbf{y}_n, \mathbf{v})$$

converges in distribution to $(\boldsymbol{\alpha}, \mathbf{u}) + (\mathbf{y}_n, \mathbf{v})$.

REMARK. This theorem holds more generally in metric spaces [6].

THEOREM 3.4. *Let* $\mathfrak{X}$, $\mathcal{Y}$ *be two finite dimensional vector spaces, and let* $A_n : \mathfrak{X} \to \mathcal{Y}$ *be a random linear transformation, which converges in distribution to a random linear transformation* $A: \mathfrak{X} \to \mathcal{Y}$ *in the sense that, for any* $\boldsymbol{\xi} \in \mathfrak{X}$, $A_n\boldsymbol{\xi}$ *converges in distribution to* $A\boldsymbol{\xi}$. *If* $\mathbf{x}_n$ *is a random vector in* $\mathfrak{X}$ *which converges in probability to a sure vector* $\boldsymbol{\xi}$, *then* $A_n\mathbf{x}_n$ *converges in distribution to* $A\boldsymbol{\xi}$.

PROOF. By the definition of the adjoint A_n': $\mathcal{Y} \to \mathfrak{X}$ of A_n, $(A_n\boldsymbol{\xi}, \mathfrak{n}) = (\boldsymbol{\xi}, A_n'\mathfrak{n})$ for any $\boldsymbol{\xi} \in \mathfrak{X}$ and any $\mathfrak{n} \in \mathcal{Y}$. Hence, by Theorem 3.1, it follows that $A_n'\mathfrak{n}$ converges in distribution to $A'\mathfrak{n}$. Then, by Theorem 3.3, the pair $\{\mathbf{x}_n, A_n'\mathfrak{n}\}$, considered as a random vector in the product space $\mathfrak{X} \times \mathfrak{X}$, converges in distribution to $\{\boldsymbol{\xi}, A'\mathfrak{n}\}$. Therefore, by Theorem 3.2, the inner product $(\mathbf{x}_n, A_n'\mathfrak{n})$ converges in distribution to $(\boldsymbol{\xi}, A'\mathfrak{n})$, or equivalently, $(A_n\mathbf{x}_n, \mathfrak{n})$ converges in distribution to $(A\boldsymbol{\xi}, \mathfrak{n})$ and the conclusion follows immediately by Theorem 3.1.

THEOREM 3.5. *If* $A: \mathfrak{X} \to \mathcal{Y}$ *is a surjective linear transformation, and* $A': \mathcal{Y} \to \mathfrak{X}$ *is the adjoint transformation, then* AA' *is an invertible transformation.*

PROOF. It is well known that $\mathfrak{N}(A)$, the null-space of A (i.e., the set of all vectors $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{0}$) is orthogonal to the range $\mathfrak{R}(A')$ of A', and that, if $\mathfrak{X}$ is finite dimensional, then $\mathfrak{R}(A') + \mathfrak{N}(A) = \mathfrak{X}$. Taking images under A on both sides of this equality, we have $\mathfrak{R}(AA') = \mathcal{Y}$, which means that AA' is a surjective transformation. Finally, we have to show that AA' is an injective transformation (i.e., $\mathfrak{N}(AA') = \mathbf{0}$). This follows immediately from the fact that, if A is a surjective transformation, then A' is an injective transformation, and that, if A' is injective, so is AA'.

REMARK. This theorem holds also for infinite dimensional Hilbert spaces.

4. Consistency and asymptotic distribution.

THEOREM 4.1. *In order that the ordinary estimator* T_n *defined by* (2.1) *be a con-*

sistent estimator of T, *it is sufficient that there exists a linear transformation* $L: \mathfrak{X} \to \mathfrak{L}$ *such that, for any* $\xi \in \mathfrak{X}$,

$$L_n\xi \to_P L\xi. \tag{4.1}$$

PROOF. Taking probability limits in (2.1) we have

$$T_n\xi \to_P YL\xi = TXL\xi.$$

On the other hand, taking probability limits in (2.2), we have

$$XL = I, \tag{4.2}$$

and therefore the conclusion follows immediately.

REMARK. Clearly, a sufficient condition for the consistency of the ordinary estimator T_n defined by (2.4) is that the random variables l_{hin} converge in probability to limits λ_{hi} and that the vectors

$$\mathbf{x}_h^* = \sum_i \lambda_{hi}\mathbf{x}_i \tag{4.3}$$

be a basis of $\mathfrak{X}$. Obviously, a necessary condition in order that the vectors $\mathbf{x}_h^*$ be a basis of $\mathfrak{X}$, is that the vectors $\mathbf{x}_i$ do not lie on any proper subspace of $\mathfrak{X}$.

THEOREM 4.2. *Under the hypothesis* (4.1), *the error of prediction of the ordinary estimator* T_n, *defined by* $\|(T_n - T)\xi\|$ *is asymptotically distributed as* $n^{-\frac{1}{2}}\|D\lambda\|$, *and the asymptotic mean square error of prediction is*

$$n^{-1}\mathcal{E}\|D\lambda\|^2 = (\lambda, \Sigma\lambda)/n, \tag{4.4}$$

where

$$\lambda = L\xi. \tag{4.5}$$

PROOF. By (2.1) and (2.2) we have

$$\|(T_n - T)\xi\| = \|(Y_n - TX_n)L_n\xi\| = \|E_nL_n\xi\|. \tag{4.6}$$

Since $n^{\frac{1}{2}}E_n$ converges in distribution to D and $L_n\xi$ converges in probability to $L\xi$, by Theorem 3.4 it follows that $n^{\frac{1}{2}}E_nL_n\xi$ converges in distribut ion to $D\lambda$. Then, by Theorem 3.2, it follows that $n^{\frac{1}{2}}\|E_nL_n\xi\|$ converges in distribution to $\|D\lambda\|$, and the first part of the theorem follows immediately from (4.6). Finally, equation (4.4) follows from $\mathcal{E}D'D\lambda = \Sigma\lambda$ and

$$\|D\lambda\|^2 = (D\lambda, D\lambda) = (\lambda, D'D\lambda).$$

5. Least squares. The simple least squares estimator is the linear transformation $\tilde{T}_n$ which minimizes the double norm of $Y_n - TX_n$, defined by

$$|||Y_n - TX_n|||^2 = \sum_{i=1}^k \|(Y_n - TX_n)\mathbf{w}_i\|^2 = \sum_{i=1}^k \|\mathbf{y}_i - T\mathbf{x}_i\|^2. \tag{5.1}$$

It is well known that the double norm of a linear transformation is independent of the particular orthonormal basis which is used in its definition, but is dependent on the norms which are used in the spaces $\mathfrak{L}$ and $\mathfrak{Y}$. Let $S: \mathfrak{L} \to \mathfrak{L}$ be a positive definite transformation, and define an S-inner product in $\mathfrak{L}$ by

$$(\lambda', \lambda'')_S = (\lambda', S\lambda''). \tag{5.2}$$

Then an S-orthonormal basis in $\mathcal{L}$ is $S^{-\frac{1}{2}}\mathbf{w}_i$, $i = 1, \cdots, k$. Instead of a simple least squares estimator, we shall consider an S-least squares estimator defined as the linear transformation $\tilde{T}_n$ which minimizes the S-double norm of $Y_n - TX_n$, defined by

$$|||Y_n - TX_n|||_S^2 = \sum_{i=1}^k \|(Y_n - TX_n)S^{-\frac{1}{2}}\mathbf{w}_i\|^2. \tag{5.3}$$

THEOREM 5.1. *If $X_n' : \mathcal{X} \to \mathcal{L}$ is the adjoint of X_n, then, with probability* 1, *$X_nS^{-1}X_n'$ is invertible and the S-least squares estimator is given by*

$$\tilde{T}_n = Y_nS^{-1}X_n'(X_nS^{-1}X_n')^{-1}. \tag{5.4}$$

PROOF. It is well known that the double norm of a linear transformation is equal to the double norm of its adjoint. Since the S-adjoint of $Y_n - TX_n$ is $S^{-1}(Y_n - TX_n)'$, we have

$$|||Y_n - TX_n|||_S^2 = |||S^{-1}(Y_n' - X_n'T')|||_S^2.$$

If $\{\mathbf{v}_j : j = 1, \cdots, q\}$ is an orthonormal basis in $\mathcal{Y}$ we have

$$|||S^{-1}(Y_n' - X_n'T')|||_S^2 = \sum_j \|S^{-1}(Y_n' - X_n'T')\mathbf{v}_j\|_S^2,$$

where the S-norm in $\mathcal{L}$ is defined by $\|\lambda\|_S = \|S^{\frac{1}{2}}\lambda\|$. Hence we have to find the vectors $\tilde{\mathbf{u}}_{jn} = \tilde{T}_n'\mathbf{v}_j$ which minimize $\sum_{j=1}^q \|S^{-\frac{1}{2}}(Y_n'\mathbf{v}_j - X_n'\mathbf{u}_j)\|^2$.

Equivalently, for each j we have to find the vector $\tilde{\mathbf{u}}_{jn} \in \mathcal{X}$ which minimizes $\|S^{-\frac{1}{2}}(Y_n'\mathbf{v}_j - X_n'\mathbf{u}_j)\|$.

Obviously $S^{-\frac{1}{2}}X_n'\tilde{\mathbf{u}}_{jn}$ is the projection of $S^{-\frac{1}{2}}Y_n'\mathbf{v}_j$ over the range of the linear transformation $S^{-\frac{1}{2}}X_n'$. Hence, $S^{-\frac{1}{2}}(Y_n'\mathbf{v}_j - X_n'\tilde{\mathbf{u}}_{jn})$ belongs to the orthogonal complement of $\mathcal{R}(S^{-\frac{1}{2}}X_n')$. Since this is the null space of $X_nS^{-\frac{1}{2}}$, we have

$$X_nS^{-1}(Y_n'\mathbf{v}_j - X_n'\tilde{\mathbf{u}}_{jn}) = \mathbf{0}.$$

Since, under our hypothesis, $A = X_nS^{-\frac{1}{2}}$ is a surjective transformation with probability 1, by Theorem 3.5, it follows that $X_nS^{-1}X_n'$ is invertible with probability 1. Therefore,

$$\tilde{\mathbf{u}}_{jn} = (X_nS^{-1}X_n')^{-1}X_nS^{-1}Y_n'\mathbf{v}_j,$$

and

$$\tilde{T}_n' = (X_nS^{-1}X_n')^{-1}X_nS^{-1}Y_n',$$

from which the conclusion follows immediately.

REMARK. In the usual case in which the matrix of S is a diagonal matrix whose diagonal elements are $s_1^2, \cdots, s_k^2$, we have $S^{-\frac{1}{2}}\mathbf{w}_i = s_i^{-1}\mathbf{w}_i$ and the S-least squares estimator is the linear transformation which minimizes the weighted sum of squares of deviations

$$\sum_{i=1}^k \|\mathbf{y}_i - T\mathbf{x}_i\|^2/s_i^2. \tag{5.5}$$

Theorem 5.2. *The S-least squares estimator is an ordinary consistent estimator with asymptotic mean square error of prediction given by*

$$n^{-1}(\xi, (X\Sigma^{-1}X')^{-1}\xi). \tag{5.6}$$

Proof. From (5.4) it follows immediately that the S-least squares estimator is an ordinary estimator with

$$L_n = S^{-1}X_n'(X_nS^{-1}X_n')^{-1}. \tag{5.7}$$

Since obviously the condition (4.1) is satisfied with

$$L = \Sigma^{-1}X'(X\Sigma^{-1}X')^{-1}, \tag{5.8}$$

it follows that the S-least squares estimator is a consistent estimator and that the asymptotic mean square error of prediction is given by (4.4) and (4.5). By substitution of (5.8) into (4.4) and (4.5), the conclusion follows immediately.

6. Asymptotic efficiency. We shall find now the linear transformations L which minimize the asymptotic mean squaie error of prediction of an ordinary estimator, which is given by (4.4) and (4.5). From (4.5) and (4.2) it follows that

$$X\lambda = \xi. \tag{6.1}$$

We shall find the vector λ which minimizes $(\lambda, \Sigma\lambda)$ subject to the restriction (6.1). Consider the change of variables $\varphi = \Sigma^{\frac{1}{2}}\lambda$. The problem is now to minimize $\|\varphi\|^2$ subject to the condition

$$X\Sigma^{-\frac{1}{2}}\varphi = \xi. \tag{6.2}$$

Obviously, the solution is the projection $\tilde{\varphi}$ of the origin on the flat (6.2), or, in other words, it is the intersection of the flat (6.2) with the orthogonal complement of the null space of $X\Sigma^{-\frac{1}{2}}$. Since $\mathfrak{L}$ is a finite dimensional vector space, this orthogonal complement is the range of $\Sigma^{-\frac{1}{2}}X'$. Therefore we have $\tilde{\varphi} = \Sigma^{-\frac{1}{2}}X'\tilde{\xi}$ for some $\tilde{\xi} \varepsilon \mathfrak{X}$, and by substitution in (6.2) we get $\tilde{\xi} = (X\Sigma^{-1}X')^{-1}\xi$. Making the proper substitutions, we obtain finally, for the minimizing $\tilde{\lambda}$,

$$\lambda = \Sigma^{-1}X'(X\Sigma^{-1}X')^{-1}\xi.$$

Hence the asymptotically efficient ordinary estimators are those for which (5.8) holds, and therefore, the S-least squares estimators are asymptotically efficient, in the sense that they minimize, within the class of ordinary estimators, the asymptotic mean square error of prediction. Note that, if

$$\sigma_1^2 = \cdots = \sigma_k^2 = \sigma^2, \tag{6.3}$$

then $\Sigma = \sigma^2 I$, where I is the identity transformation. If we know that this is the case, then we can choose $S = s^2 I$, where s^2 is a consistent estimator of σ^2, and in this case the S-least squares estimator is simply the unweighted least squares estimator. Hence, only in the case (6.3) will the unweighted least squares estimator be asymptotically efficient.

REFERENCES

[1] BILLINGSLEY, P. (1956). The invariance principle for dependent random variables. *Trans. Am. Math. Soc.* **83** 250–268.

[2] DORFF, MARTIN and GURLAND, JOHN (1961). Estimation of the parameters of a linear functional relation. *J. Roy. Statist. Soc. Ser. B* **23** 160–170.

[3] GEARY, R. C. (1949). Determination of linear relations between systematic parts of variables with errors of observations the variances of which are unknown. *Econometrica* **17** 30–59.

[4] LOMNICKI, Z. A. and ZAREMBA, S. K. (1957). On some moments and distributions occurring in the theory of linear stochastic processes, I. *Monatsh. Math.* **61** 318–358.

[5] MADANSKY, ALBERT (1959). The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.* **54** 173–205.

[6] PRATT, JOHN W. On order relations and convergence in distribution. Notes based on lectures given by Herman Chernoff. Applied Math. and Statist. Lab., Stanford University.

[7] VILLEGAS, C. (1964). Confidence region for a linear relation. *Ann. Math. Statist.* **35** 780–788.