Complex Linear Least Squares
Author(s): Kenneth S. Miller
Source: *SIAM Review*, Vol. 15, No. 4 (Oct., 1973), pp. 706-726
Published by: Society for Industrial and Applied Mathematics
Stable URL: http://www.jstor.org/stable/2028727
Accessed: 17-05-2018 13:55 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

Society for Industrial and Applied Mathematics is collaborating with JSTOR to digitize, preserve and extend access to *SIAM Review*

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

SIAM REVIEW
Vol. 15, No. 4, October 1973

COMPLEX LINEAR LEAST SQUARES*

KENNETH S. MILLER†

Abstract. Certain elementary properties of the theory of least squares are presented from the point of view of complex stochastic processes. The development parallels the real case. We consider the least squares estimate (LSE), the best linear unbiased estimate (BLUE), the Markov estimate (ME), and the relationships among them. In particular, we prove the Gauss–Markov theorem and give necessary and sufficient conditions that the LSE and ME be identical. The efficiency of the LSE is defined and a lower bound is obtained for the efficiency of a certain class of models. Estimation of the mean vector and variance for a complex normal population leads to the maximum likelihood estimates (MLE). We prove that the MLE of the mean vector is identical with the LSE, and deduce other analogous properties concerning the distribution of the MLE.

Introduction. The subject matter of least squares is classic. Some indication of the vast literature that exists in this field may be gleaned from the Bibliography at the end of this paper. Most of our references are restricted to the last two decades. In more recent years it has been observed that some statistical problems as well as many physical phenomena may be conveniently formulated using *complex* Gaussian processes. We have listed some papers dealing with such applications in the References. Our purpose, in this present work, is to discuss some of the elementary properties of linear least squares theory from the point of view of complex stochastic processes. The results, of course, parallel the real case.

We illustrate the basic idea (for the real case) in § 1. Since our analysis rests heavily on matrix notation, we introduce, in § 2, the concept of differentiation of a scalar with respect to a complex vector. In § 3 we derive the *normal equation* satisfied by the *least squares estimate* (LSE). After a discussion of various equivalent least squares models, we introduce the *best linear unbiased estimate* (BLUE) and the *Markov estimate* (ME). We next prove the Gauss–Markov theorem which expresses the relationship between the LSE and the ME (for a certain class of covariance functions), and then give a necessary and sufficient condition that the LSE and ME be identical.

The efficiency of the LSE is defined in § 6. We use the Kantorovich inequality to obtain a lower bound for the efficiency of a certain class of models. Next we turn to the problem of estimating the mean vector and the variance of a sample from a normal population. This leads to the *maximum likelihood estimate* (MLE) and the fact that the MLE of the mean is identical with the LSE. We also show that the maximum likelihood estimators form a set of sufficient statistics. Finally, in § 8, we show that the maximum likelihood estimators of the mean vector and variance are independently distributed, and then explicitly determine these distributions.

A word on notation: Vectors and matrices will be indicated by boldface letters; and primes will denote transposes. In particular, unprimed vectors will be column vectors.

* Received by the editors December 17, 1971, and in final revised form September 29, 1972.

† Riverside Research Institute, New York, New York 10023.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

1. The basic idea. Anyone who has taken a laboratory course in the physical sciences has faced the problem of attempting to interpret empirical data. For example, at a particular value of an independent variable, say x_t, one might observe a physical quantity with numerical value, say z_t. If a finite number, say T, of these observations are made, one may represent the data graphically as in Fig. 1. A typical problem might be to find the "best" straight line that fits the data (see Fig. 2). By "best" straight line one generally means a line with equation $z = \beta_1 x + \beta_2$, where β_1 and β_2 are so chosen that the sum

(1.1) $$v = \sum_{t=1}^{T} [z_t - (\beta_1 x_t + \beta_2)]^2$$

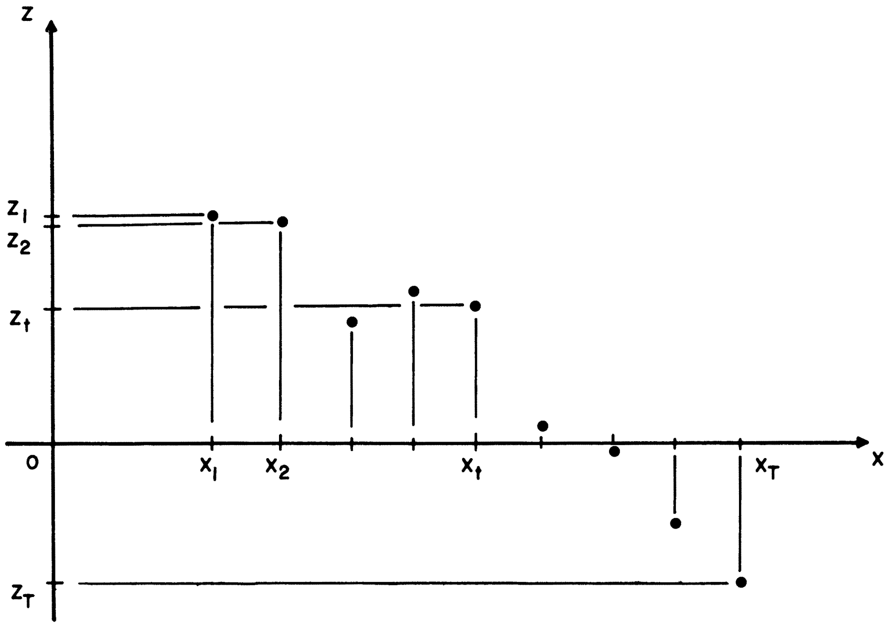

FIG. 1

is minimized. If one computes $\partial v/\partial\beta_1$ and $\partial v/\partial\beta_2$ and sets these derivatives equal to zero, there results

(1.2) $$\boldsymbol{\beta} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{z},$$

where $\boldsymbol{\beta}' = \{\beta_1, \beta_2\}$, $\mathbf{z}' = \{z_1, \cdots, z_T\}$ and

(1.3) $$\mathbf{H}' = \begin{bmatrix} x_1 & x_2 & \cdots & x_T \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

If not all the x_t, $1 \leqq t \leqq T$, are identical, then $\det \mathbf{H}'\mathbf{H} > 0$. In the matrix notation just introduced, (1.1) may be written as

(1.4) $$v = |\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2,$$

where the bars indicate the Euclidean norm of the vector.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

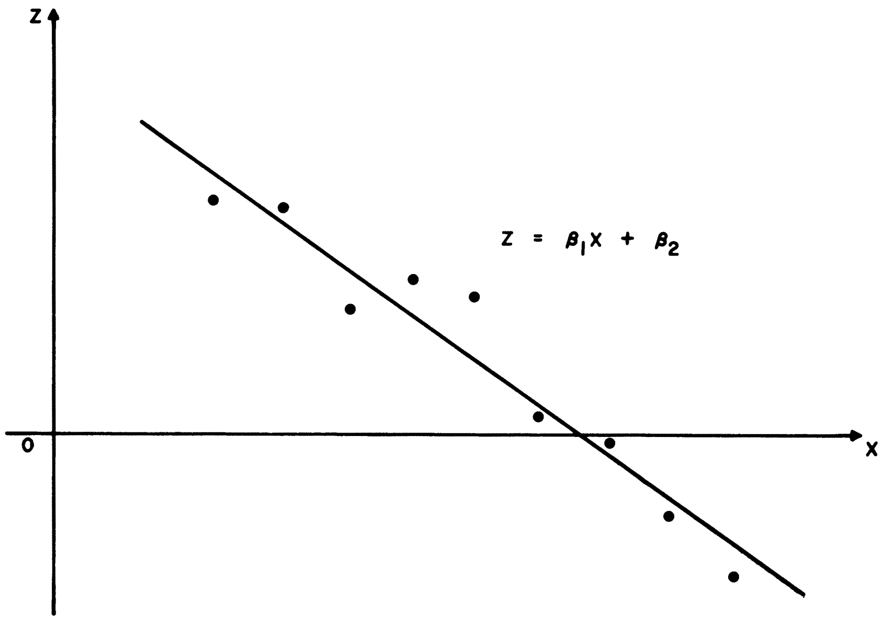

FIG. 2

We wish to generalize this problem to complex quantities and to vectors $\boldsymbol{\beta}$ of arbitrary dimension. To carry out this program most efficiently we shall use matrix notation. In particular, the concept of differentiation with respect to a vector is a useful tool. We shall therefore make a slight digression in the next section to explain this technique, and then return to our main theme in § 3.

2. Differentiation with respect to a vector. Let $\mathbf{x}' = \{x_1, \cdots, x_n\}$ be a real vector and let $g(\mathbf{x}) = g(x_1, \cdots, x_n)$ be a scalar-valued differentiable function of $\mathbf{x}$. By the *derivative* of $g(\mathbf{x})$ with respect to the vector $\mathbf{x}$ we shall mean the vector

$$\frac{dg}{d\mathbf{x}} = \begin{bmatrix} \dfrac{\partial g}{\partial x_1} \\ \vdots \\ \dfrac{\partial g}{\partial x_n} \end{bmatrix},$$

where $x_1, \cdots, x_n$ are to be considered as n independent variables.

If $g(\boldsymbol{\xi})$ is a real scalar-valued function of the complex vector $\boldsymbol{\xi}' = \{\xi_1, \cdots, \xi_n\}$, the situation is more complicated. For example, if $z = x + iy$ is a complex scalar and $g(z) = |z|^2$, the derivative of $g(z)$ with respect to z exists only at $z = 0$, although $\partial g/\partial x$ and $\partial g/\partial y$ exist for all z. To overcome this difficulty let us write

$$g(\boldsymbol{\xi}) = g(x_1, \cdots, x_n, y_1, \cdots, y_n),$$

where $x_k = \operatorname{Re} \xi_k$, $y_k = \operatorname{Im} \xi_k$, $1 \leqq k \leqq n$. Alternatively we may write

$$g(\boldsymbol{\xi}) = g(\tfrac{1}{2}(\xi_1 + \bar{\xi}_1), \cdots, \tfrac{1}{2}i(\bar{\xi}_n - \xi_n)).$$

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

Thus we may consider $g(\boldsymbol{\xi})$ as a function of the $2n$ real independent variables $x_1, \cdots, x_n, y_1, \cdots, y_n$, or as a function of the $2n$ independent variables $\xi_1, \cdots, \xi_n, \bar{\xi}_1, \cdots, \bar{\xi}_n$.

Now let $g(u_1, \cdots, u_{2n})$ be a differentiable function of the $2n$ indeterminates $u_1, \cdots, u_{2n}$ in some region of $2n$-dimensional Euclidean space. Let us denote the partial derivative of g with respect to u_j by g_j:

$$\frac{\partial g}{\partial u_j} = g_j, \qquad 1 \leqq j \leqq 2n. \tag{2.1}$$

Then, in particular,

$$\frac{\partial g}{\partial x_k} = g_k, \qquad \frac{\partial g}{\partial y_k} = g_{k+n}, \qquad 1 \leqq k \leqq n, \tag{2.2}$$

and

$$\frac{\partial g}{\partial \boldsymbol{\xi}} = \mathbf{0} \tag{2.3}$$

if and only if $\partial g/\partial \mathbf{x} = \mathbf{0}$ and $\partial g/\partial \mathbf{y} = \mathbf{0}$, where $\mathbf{x} = \operatorname{Re} \boldsymbol{\xi}$ and $\mathbf{y} = \operatorname{Im} \boldsymbol{\xi}$.

To illustrate the use of (2.3) let $\mathbf{A}$ be an $m \times n$ complex matrix, $\mathbf{a}$ an m-dimensional complex vector, and let it be required to find an n-dimensional (complex) vector $\boldsymbol{\xi}$ which minimizes

$$g(\boldsymbol{\xi}) = |\mathbf{a} - \mathbf{A}\boldsymbol{\xi}|^2. \tag{2.4}$$

Now

$$\frac{\partial g}{\partial \boldsymbol{\xi}} = -\mathbf{A}'(\bar{\mathbf{a}} - \bar{\mathbf{A}}\bar{\boldsymbol{\xi}}).$$

Setting this derivative equal to the n-dimensional zero vector and taking conjugates, we obtain

$$\bar{\mathbf{A}}'\mathbf{A}\boldsymbol{\xi} = \bar{\mathbf{A}}'\mathbf{a}. \tag{2.5}$$

Thus if $\bar{\mathbf{A}}'\mathbf{A}$ is nonsingular,

$$\boldsymbol{\xi} = (\bar{\mathbf{A}}'\mathbf{A})^{-1}\bar{\mathbf{A}}'\mathbf{a}. \tag{2.6}$$

Of course, in this example, it is evident from an inspection of (2.4) that if the equation $\mathbf{a} - \mathbf{A}\boldsymbol{\xi} = \mathbf{0}$ admits a solution $\boldsymbol{\xi}$, then $\mathbf{A}\boldsymbol{\xi} = \mathbf{a}$ is the desired solution. But this equation (when multiplied by $\bar{\mathbf{A}}'$) is (2.5).

3. The normal equation. We now consider a generalization of the regression problem discussed in § 1. Let $\boldsymbol{\beta}' = \{\beta_1, \cdots, \beta_p\}$ be a p-dimensional (complex) vector parameter. Let $\mathbf{H}$ be a $T \times p$ matrix of rank p. The elements of $\mathbf{H}$ are known, fixed complex numbers. By analogy with (1.3) we call these elements *independent variables*. Let $\mathbf{z}' = \{z_1, \cdots, z_T\}$ be the T-dimensional (complex) vector of observables. By analogy with (1.1) we call the z_t *dependent variables*. Let $\mathbf{b}^*$ be a p-dimensional (complex) vector with the property that when $\boldsymbol{\beta}$ is replaced by $\mathbf{b}^*$ in $|\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2$, this expression is minimized. We call $\mathbf{b}^*$ the *least*

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

squares estimate (LSE) of $\boldsymbol{\beta}$. In expanded notation, $\mathbf{b}^{*\prime} = \{b_1^*, \cdots, b_p^*\}$ is that value of $\boldsymbol{\beta}$ which minimizes

$$v = \sum_{t=1}^{T} \left| z_t - \sum_{j=1}^{p} H_{tj}\beta_j \right|^2 \tag{3.1}$$

(where $\mathbf{H} = \|H_{tj}\|$).

It is not difficult to find $\mathbf{b}^*$. We first write (3.1) in matrix form as $v = |\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2$. Then, as in the previous section, we deduce that

$$\mathbf{b}^* = (\overline{\mathbf{H}}'\mathbf{H})^{-1}\overline{\mathbf{H}}'\mathbf{z} \tag{3.2}$$

is the LSE of $\boldsymbol{\beta}$ (cf. (1.2) and (2.6)).

The matrix $\overline{\mathbf{H}}'\mathbf{H}$ in (3.2) is nonsingular. This follows from the more general result that if the rank of $\mathbf{H}$ is p and if $\boldsymbol{\Sigma}$ is a $T \times T$ positive definite matrix, then $\overline{\mathbf{H}}'\boldsymbol{\Sigma}\mathbf{H}$ is nonsingular.

If we let $\mathbf{A} = \overline{\mathbf{H}}'\mathbf{H}$ and $\mathbf{c} = \overline{\mathbf{H}}'\mathbf{z}$, then (3.2) may be written

$$\mathbf{b}^* = \mathbf{A}^{-1}\mathbf{c} \tag{3.3}$$

(and $\mathbf{A}\mathbf{b}^* = \mathbf{c}$). We call (3.3) the *normal equation.*

4. The Gauss–Markov theorem. Suppose, as above, $\mathbf{z}' = \{z_1, \cdots, z_T\}$ is a T-dimensional observation vector, $\boldsymbol{\beta}' = \{\beta_1, \cdots, \beta_p\}$ a p-dimensional parameter vector, and $\mathbf{H} = \|H_{tj}\|$ a $T \times p$ constant matrix of rank p. Let

$$\begin{gathered} \mathscr{E}\mathbf{z} = \mathbf{H}\boldsymbol{\beta}, \\ \mathscr{E}(\mathbf{z} - \mathbf{H}\boldsymbol{\beta})(\bar{\mathbf{z}} - \overline{\mathbf{H}}\bar{\boldsymbol{\beta}})' = \boldsymbol{\Sigma}, \end{gathered} \tag{4.1}$$

where $\boldsymbol{\Sigma}$ is a $T \times T$ positive definite Hermitian covariance matrix. We shall refer to such a model as $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$.

Another way of describing $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$ is as follows: Let $\{z_t | t = \cdots, -1, 0, 1, \cdots\}$ be a complex stochastic process with discrete parameter t. Let $z_1, \cdots, z_T$ be observations from $\{z_t\}$ and let the vector of observations $\mathbf{z}' = \{z_1, \cdots, z_T\}$ have a mean value $\mathbf{H}\boldsymbol{\beta}$ and positive definite covariance matrix $\boldsymbol{\Sigma}$.

Still another description of $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$ is furnished by a consideration of

$$z_t = \sum_{j=1}^{p} H_{tj}\beta_j + u_t, \qquad 1 \leqq t \leqq T, \tag{4.2}$$

where the H_{tj}, $1 \leqq t \leqq T$, $1 \leqq j \leqq p$, are constants, the β_j, $1 \leqq j \leqq p$, are parameters to be estimated, and the u_t, $1 \leqq t \leqq T$, are random variables with means zero and positive definite covariance matrix $\boldsymbol{\Sigma}$. One may write (4.2) in vector notation as

$$\mathbf{z} = \mathbf{H}\boldsymbol{\beta} + \mathbf{u}, \tag{4.3}$$

where $\mathbf{u}' = \{u_1, \cdots, u_T\}$ and $\mathscr{E}\mathbf{u} = \mathbf{0}$, $\mathscr{E}\mathbf{u}\bar{\mathbf{u}}' = \boldsymbol{\Sigma}$. We call $\mathbf{H}\boldsymbol{\beta}$ the *regression function* of $\mathbf{z}$ on $\boldsymbol{\beta}$ and call $\mathbf{u}$ the *residual.*

The fundamental problem is to find the "best" estimate of the parameter $\boldsymbol{\beta}$ based on the observables $z_1, \cdots, z_T$. To be more precise, let $\boldsymbol{\gamma}$ be an arbitrary but fixed p-dimensional vector. We shall find a T-dimensional vector $\mathbf{g}$ such that $\mathbf{g}'\mathbf{z}$ is a minimum variance, unbiased estimate of $\boldsymbol{\gamma}'\boldsymbol{\beta}$. Since $\mathbf{g}'\mathbf{z}$ is a linear combination

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

of the observations $z_1, \cdots, z_T$, we see that $\mathbf{g}'\mathbf{z}$ will be a minimum variance, *linear*, unbiased estimate. We call such an estimate a *best linear unbiased estimate* (BLUE).

Let us find $\mathbf{g}$. If $\mathbf{g}'\mathbf{z}$ is to be unbiased, we must have

$$\mathscr{E}\mathbf{g}'\mathbf{z} = \mathbf{g}'\mathbf{H}\boldsymbol{\beta} = \boldsymbol{\gamma}'\boldsymbol{\beta}$$

which implies that

$$\mathbf{g}'\mathbf{H} = \boldsymbol{\gamma}'. \tag{4.4}$$

Hence, for any p-dimensional vector $\boldsymbol{\lambda}$,

$$\mathbf{g}'\mathbf{H}\bar{\boldsymbol{\lambda}} + \bar{\mathbf{g}}'\bar{\mathbf{H}}\boldsymbol{\lambda} = \boldsymbol{\gamma}'\bar{\boldsymbol{\lambda}} + \bar{\boldsymbol{\gamma}}'\boldsymbol{\lambda}$$

is a real scalar. Since the variance of $\mathbf{g}'\mathbf{z}$ is

$$\operatorname{var}\mathbf{g}'\mathbf{z} = \mathscr{E}(\mathbf{g}'\mathbf{z} - \mathbf{g}'\mathbf{H}\boldsymbol{\beta})(\bar{\mathbf{g}}'\bar{\mathbf{z}} - \bar{\mathbf{g}}'\bar{\mathbf{H}}\bar{\boldsymbol{\beta}})' = \mathbf{g}'\boldsymbol{\Sigma}\bar{\mathbf{g}},$$

we must minimize

$$v = \mathbf{g}'\boldsymbol{\Sigma}\bar{\mathbf{g}} - (\mathbf{g}'\mathbf{H}\bar{\boldsymbol{\lambda}} + \bar{\mathbf{g}}'\bar{\mathbf{H}}\boldsymbol{\lambda}),$$

where $\boldsymbol{\lambda}$ plays the role of a Lagrange multiplier. That is, we have a constrained minimization problem. Setting the derivative

$$\frac{dv}{d\mathbf{g}} = \boldsymbol{\Sigma}\bar{\mathbf{g}} - \mathbf{H}\bar{\boldsymbol{\lambda}}$$

equal to the T-dimensional zero vector leads to

$$\mathbf{g}' = \boldsymbol{\lambda}'\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}, \tag{4.5}$$

and by (4.4),

$$\boldsymbol{\gamma}' = \mathbf{g}'\mathbf{H} = \boldsymbol{\lambda}'(\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H}) \quad \text{or} \quad \boldsymbol{\lambda}' = \boldsymbol{\gamma}'(\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}.$$

Hence (4.5) implies that

$$\mathbf{g}'\mathbf{z} = \boldsymbol{\gamma}'(\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{z} \tag{4.6}$$

is the BLUE of $\boldsymbol{\gamma}'\boldsymbol{\beta}$.

Let $\mathbf{b}' = \{b_1, \cdots, b_p\}$ be the vector with the property that b_j, $1 \leqq j \leqq p$, is the BLUE of β_j. Then we call $\mathbf{b}$ the *Markov estimate* (ME) of $\boldsymbol{\beta}$. Now suppose, in particular, that $\boldsymbol{\gamma} = \boldsymbol{\delta}_k$, where $\boldsymbol{\delta}_k$ is a p-dimensional vector with one in the kth position and zeros elsewhere. That is, $\boldsymbol{\delta}_k' = \{\delta_{1k}, \cdots, \delta_{pk}\}$, where δ_{jk} is the Kronecker delta. Then $\beta_k = \boldsymbol{\delta}_k'\boldsymbol{\beta}$ and from (4.6), the BLUE of $\boldsymbol{\delta}_k'\boldsymbol{\beta}$ (= the kth component of $\boldsymbol{\beta}$) is $\boldsymbol{\delta}_k'(\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$ (= the kth component of $(\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1} \cdot \bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$). Thus

$$\mathbf{b} = (\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\bar{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{z} \tag{4.7}$$

is the ME of $\boldsymbol{\beta}$. We also see that the BLUE of $\boldsymbol{\gamma}'\boldsymbol{\beta}$ is $\boldsymbol{\gamma}'\mathbf{b}$, where $\mathbf{b}$ is the ME of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ is an arbitrary, known, p-dimensional vector.

The LSE of $\boldsymbol{\beta}$ (see (3.2)),

$$\mathbf{b}^* = (\bar{\mathbf{H}}'\mathbf{H})^{-1}\bar{\mathbf{H}}'\mathbf{z}, \tag{4.8}$$

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

is also a linear unbiased estimate of $\boldsymbol{\beta}$. (The form of (4.8) demonstrates its linearity in $\mathbf{z}$, and $\mathscr{E}\mathbf{b}^* = (\overline{\mathbf{H}}'\mathbf{H})^{-1}\overline{\mathbf{H}}'\mathscr{E}\mathbf{z} = (\overline{\mathbf{H}}'\mathbf{H})^{-1}\overline{\mathbf{H}}'\mathbf{H}\boldsymbol{\beta} = \boldsymbol{\beta}$.) The covariance of $\mathbf{b}^*$ is

$$\mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})(\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})' = (\overline{\mathbf{H}}'\mathbf{H})^{-1}\overline{\mathbf{H}}'\boldsymbol{\Sigma}\mathbf{H}(\overline{\mathbf{H}}'\mathbf{H})^{-1}. \tag{4.9}$$

The covariance matrix of the ME of $\boldsymbol{\beta}$ is

$$\mathscr{E}(\mathbf{b} - \boldsymbol{\beta})(\overline{\mathbf{b}} - \overline{\boldsymbol{\beta}})' = (\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}. \tag{4.10}$$

Since $\boldsymbol{\gamma}'\mathbf{b}$ is the BLUE of $\boldsymbol{\gamma}'\boldsymbol{\beta}$ for arbitrary $\boldsymbol{\gamma}$, we infer that

$$\operatorname{var} \boldsymbol{\gamma}'\mathbf{b} \leqq \operatorname{var} \boldsymbol{\gamma}'\mathbf{b}^*.$$

But this, in turn, is equivalent to saying that the matrix

$$(\overline{\mathbf{H}}'\mathbf{H})^{-1}(\overline{\mathbf{H}}'\boldsymbol{\Sigma}\mathbf{H})(\overline{\mathbf{H}}'\mathbf{H})^{-1} - (\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1} \tag{4.11}$$

is nonnegative definite.

If in particular $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $T \times T$ identity matrix, then (cf. (4.7) and (4.8)) $\mathbf{b} = \mathbf{b}^*$. That is, the LSE and ME are identical. This last sentence is a statement of the *Gauss–Markov theorem*, which is now formally stated.

THEOREM 4.1 (Gauss–Markov theorem). *Consider the model $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$. Let $\mathbf{b}^*$ be the least squares estimate of $\boldsymbol{\beta}$. Then the components of $\mathbf{b}^*$ are the minimum variance linear unbiased estimates (among all linear unbiased estimates) of the corresponding components of $\boldsymbol{\beta}$.*

COROLLARY 4.1. *Consider the model $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$. Let $\mathbf{b}^*$ be the least squares estimate of $\boldsymbol{\beta}$. Let $\boldsymbol{\gamma}$ be an arbitrary, but fixed p-dimensional vector. Then the minimum variance linear unbiased estimate of $\boldsymbol{\gamma}'\boldsymbol{\beta}$ is $\boldsymbol{\gamma}'\mathbf{b}^*$.*

Trivially, we note that when $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, the covariance matrix of $\mathbf{b}$ $(= \mathbf{b}^*)$ becomes

$$\sigma^2(\overline{\mathbf{H}}'\mathbf{H})^{-1}. \tag{4.12}$$

5. Equivalence of least squares and Markov estimates. If in the model $\mathscr{M}_{T_p}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$ the covariance matrix $\boldsymbol{\Sigma}$ is a scalar multiple of the identity matrix, $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, then the ME $\mathbf{b}$ and LSE $\mathbf{b}^*$ are identical. In this section we wish to establish necessary and sufficient conditions under which $\mathbf{b}$ equals $\mathbf{b}^*$. The results are given in Theorem 5.1 below. However, before proving this theorem we need the following standard lemma from matrix theory. This lemma is also used in the next section.

LEMMA 5.1. *Let $\mathbf{G}$ be a positive definite $n \times n$ Hermitian matrix and let $\boldsymbol{\Gamma}$ be a nonnegative definite $n \times n$ Hermitian matrix. Then there exists a nonsingular $n \times n$ matrix $\mathbf{Q}$ such that*

$$\overline{\mathbf{Q}}'\mathbf{G}\mathbf{Q} = \mathbf{I} \tag{5.1}$$

and

$$\overline{\mathbf{Q}}'\boldsymbol{\Gamma}\mathbf{Q} = \mathbf{M}, \tag{5.2}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix and $\mathbf{M} = \|\mu_j\delta_{jk}\|$ is a diagonal matrix where the μ_j, $1 \leqq j \leqq n$, are the roots of the characteristic equation $\det(\boldsymbol{\Gamma} - \mu\mathbf{G}) = 0$. *Furthermore, if $\mathbf{G} - \boldsymbol{\Gamma}$ is nonnegative definite, then*

$$\det \mathbf{G} - \det \boldsymbol{\Gamma} \geqq 0. \tag{5.3}$$

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

THEOREM 5.1. *Consider the model* $\mathcal{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$. *Let* $\mathbf{b}$ *be the Markov estimate of* $\boldsymbol{\beta}$ *and* $\mathbf{b}^*$ *the least squares estimate of* $\boldsymbol{\beta}$. *Then a necessary and sufficient condition that* $\mathbf{b} = \mathbf{b}^*$ *identically in the observation vector* $\mathbf{z}$ *is that* $\mathbf{H} = \mathbf{VC}$, *where* $\mathbf{C}$ *is a nonsingular* $p \times p$ *matrix and* $\mathbf{V}$ *is a* $T \times p$ *matrix of characteristic vectors of* $\boldsymbol{\Sigma}$.

Proof. Sufficiency. Let $\mathbf{H} = \mathbf{VC}$, where $\mathbf{V}$ and $\mathbf{C}$ are as defined in the statement of the theorem. Then $\boldsymbol{\Sigma}\mathbf{V} = \mathbf{VD}$, where $\mathbf{D}$ is a $p \times p$ diagonal matrix of characteristic numbers of $\boldsymbol{\Sigma}$ (corresponding to the characteristic vectors which are the columns of $\mathbf{V}$). Thus

$$\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1} = \overline{\mathbf{C}}'\overline{\mathbf{D}}^{-1}\overline{\mathbf{C}}'^{-1}\overline{\mathbf{H}}'. \tag{5.4}$$

Now the ME of $\boldsymbol{\beta}$ is

$$\mathbf{b} = (\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{z}$$

(see (4.7)). If we replace $\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}$ in $\mathbf{b}$ by (5.4), there results

$$\mathbf{b} = (\overline{\mathbf{C}}'\overline{\mathbf{D}}^{-1}\overline{\mathbf{C}}'^{-1}\overline{\mathbf{H}}'\mathbf{H})^{-1}\overline{\mathbf{C}}'\overline{\mathbf{D}}^{-1}\overline{\mathbf{C}}'^{-1}\overline{\mathbf{H}}'\mathbf{z} = (\overline{\mathbf{H}}'\mathbf{H})^{-1}\overline{\mathbf{H}}'\mathbf{z}$$

—which is $\mathbf{b}^*$, the LSE of $\boldsymbol{\beta}$ (see (4.8)).

Necessity. Let $\mathbf{b} = \mathbf{b}^*$ identically in the observation vector $\mathbf{z}$. Then

$$(\overline{\mathbf{H}}'\mathbf{H})(\overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})^{-1}\overline{\mathbf{H}}' = \overline{\mathbf{H}}'\boldsymbol{\Sigma}. \tag{5.5}$$

Let $\mathbf{G} = \overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H}$ and $\boldsymbol{\Gamma} = \overline{\mathbf{H}}'\mathbf{H}$. From (4.10) (see also the comments following (3.2)) we infer that $\mathbf{G}$ and $\boldsymbol{\Gamma}$ are positive definite matrices. Thus, in the notation of Lemma 5.1,

$$\boldsymbol{\Gamma}\mathbf{G}^{-1} = (\overline{\mathbf{Q}}'^{-1}\boldsymbol{\Delta}\mathbf{Q}^{-1})(\mathbf{Q}\overline{\mathbf{Q}}') = \overline{\mathbf{Q}}'^{-1}\boldsymbol{\Delta}\overline{\mathbf{Q}}',$$

where $\boldsymbol{\Delta}$ is a $p \times p$ diagonal matrix and $\mathbf{Q}$ is nonsingular. The transpose of (5.5) may therefore be written as $\boldsymbol{\Sigma}\mathbf{H} = \mathbf{H}(\mathbf{Q}\boldsymbol{\Delta}\mathbf{Q}^{-1})$ and multiplying on the right by $\mathbf{Q}$,

$$\boldsymbol{\Sigma}(\mathbf{HQ}) = (\mathbf{HQ})\boldsymbol{\Delta}.$$

Now let $\mathbf{V} = \mathbf{HQ}$. Then $\boldsymbol{\Sigma}\mathbf{V} = \mathbf{V}\boldsymbol{\Delta}$. Since $\boldsymbol{\Delta}$ is diagonal, $\mathbf{V}$ is a $T \times p$ matrix of characteristic vectors of $\boldsymbol{\Sigma}$, and the diagonal elements of $\boldsymbol{\Delta}$ are characteristic numbers of $\boldsymbol{\Sigma}$. Since $\mathbf{Q}$ is nonsingular, we may let $\mathbf{C} = \mathbf{Q}^{-1}$. Then

$$\mathbf{H} = \mathbf{VQ}^{-1} = \mathbf{VC},$$

as we desired to prove.

6. Efficiency of the least squares estimate. Consider the model $\mathcal{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$ of § 4. Let $\mathbf{b}^*$ be the LSE of $\boldsymbol{\beta}$. Then we define the *efficiency* of $\mathbf{b}^*$ as

$$\operatorname{Eff} \mathbf{b}^* = \frac{\det \mathscr{E}(\mathbf{b} - \boldsymbol{\beta})(\overline{\mathbf{b}} - \overline{\boldsymbol{\beta}})'}{\det \mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})(\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'},$$

where $\mathbf{b}$ is the ME of $\boldsymbol{\beta}$. From (4.10) and (4.9),

$$\operatorname{Eff} \mathbf{b}^* = \frac{\det^2 \overline{\mathbf{H}}'\mathbf{H}}{(\det \overline{\mathbf{H}}'\boldsymbol{\Sigma}\mathbf{H})(\det \overline{\mathbf{H}}'\boldsymbol{\Sigma}^{-1}\mathbf{H})}. \tag{6.1}$$

We assert that $\operatorname{Eff} \mathbf{b}^* \leqq 1$.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

THEOREM 6.1. *In the model* $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\Sigma}]$ *let* $\mathbf{b}^*$ *be the least squares estimate of* $\boldsymbol{\beta}$. *Then*

$$\operatorname{Eff} \mathbf{b}^* \leqq 1.$$

Proof. By (4.9), (4.10) and (4.11),

$$\mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})(\bar{\mathbf{b}}^* - \bar{\boldsymbol{\beta}})' - \mathscr{E}(\mathbf{b} - \boldsymbol{\beta})(\bar{\mathbf{b}} - \bar{\boldsymbol{\beta}})'$$

is nonnegative definite while $\mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})(\bar{\mathbf{b}}^* - \bar{\boldsymbol{\beta}})'$ and $\mathscr{E}(\mathbf{b} - \boldsymbol{\beta})(\bar{\mathbf{b}} - \bar{\boldsymbol{\beta}})'$ are both positive definite. Hence by (5.3) of Lemma 5.1,

$$\det \mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})(\bar{\mathbf{b}}^* - \bar{\boldsymbol{\beta}})' \geqq \det \mathscr{E}(\mathbf{b} - \boldsymbol{\beta})(\bar{\mathbf{b}} - \bar{\boldsymbol{\beta}})'$$

or $\operatorname{Eff} \mathbf{b}^* \leqq 1$.

We may express $\operatorname{Eff} \mathbf{b}^*$ in a more canonical form. Let $\mathbf{W}$ be a unitary matrix with the property that $\overline{\mathbf{W}}'\boldsymbol{\Sigma}\mathbf{W} = \mathbf{L}$, where $\mathbf{L}$ is the diagonal matrix $\|\lambda_j \delta_{jk}\|$ of the characteristic zeros of $\boldsymbol{\Sigma}$. Let $\mathbf{K} = \overline{\mathbf{W}}'\mathbf{H}$. Then $\overline{\mathbf{K}}'\mathbf{K} = \overline{\mathbf{H}}'\mathbf{H}$ and (6.1) may be written as

$$\operatorname{Eff} \mathbf{b}^* = \frac{\det^2 \overline{\mathbf{K}}'\mathbf{K}}{(\det \overline{\mathbf{K}}'\mathbf{L}\mathbf{K})(\det \overline{\mathbf{K}}'\mathbf{L}^{-1}\mathbf{K})}.$$

In particular, if $p = 1$, then $\mathbf{K}' = \mathbf{k}' = \{k_1, \cdots, k_T\}$ is a T-dimensional vector and

$$\operatorname{Eff} b^* = \frac{|\mathbf{k}|^4}{(\bar{\mathbf{k}}'\mathbf{L}\mathbf{k})(\bar{\mathbf{k}}'\mathbf{L}^{-1}\mathbf{k})}.$$

In this case (that is, $p = 1$), a lower bound for $\operatorname{Eff} b^*$ in terms of the characteristic zeros of $\boldsymbol{\Sigma}$ may be found. In fact, we have the following theorem.

THEOREM 6.2. *In the model* $\mathscr{M}_{T1}[\mathbf{z}, \beta, \mathbf{H}, \boldsymbol{\Sigma}]$ *let* b^* *be the least squares estimate of* β. *Let* λ_1 *be the smallest and* λ_T *the largest characteristic zero of* $\boldsymbol{\Sigma}$. *Then*

$$\operatorname{Eff} b^* \geqq \frac{4\lambda_1\lambda_T}{(\lambda_1 + \lambda_T)^2}. \tag{6.2}$$

Theorem 6.2 is an immediate consequence of *Kantorovich's inequality*, which is: Let $\alpha_t \geqq 0$, $1 \leqq t \leqq T$, and $\sum_{t=1}^{T} \alpha_t = 1$. Let $0 < \lambda_1 \leqq \lambda_2 \leqq \cdots \leqq \lambda_T < \infty$. Then

$$\left(\sum_{t=1}^{T} \alpha_t \lambda_t\right)\left(\sum_{t=1}^{T} \alpha_t \lambda_t^{-1}\right) \leqq \frac{(\lambda_1 + \lambda_T)^2}{4\lambda_1\lambda_T}. \tag{6.3}$$

If we let

$$\alpha_t = \frac{|k_t|^2}{\sum_{s=1}^{T} |k_s|^2}, \qquad 1 \leqq t \leqq T,$$

the truth of Theorem 6.2 becomes apparent.

Suppose now we consider the model $\mathscr{M}_{T1}[\mathbf{z}, \beta, \mathbf{H}, \sigma^2\mathbf{I}]$, that is, the case where the covariance matrix $\boldsymbol{\Sigma}$ is a multiple of the identity matrix. Then $\lambda_1 = \lambda_T$ and (6.2) becomes

$$\operatorname{Eff} b^* = 1.$$

This, of course, corresponds to the case where the ME is identical with the LSE.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

7. Maximum likelihood estimates. We recall (see, for example, [11, p. 550]) that if $\mathbf{z}' = \{z_1, \cdots, z_n\}$ is an n-dimensional complex random vector with mean $\mathbf{c}' = \{c_1, \cdots, c_n\}$ and positive definite covariance matrix $\boldsymbol{\Psi}$, then we say that $\mathbf{z}$ has a joint n-dimensional complex Gaussian distribution if

$$f(\mathbf{z}) = \frac{1}{\pi^n \det \boldsymbol{\Psi}} \exp\{-(\bar{\mathbf{z}} - \bar{\mathbf{c}})'\boldsymbol{\Psi}^{-1}(\mathbf{z} - \mathbf{c})\}$$

is the density function of $\mathbf{z}$. It follows that

$$\mathscr{E}(\mathbf{z} - \mathbf{c})(\mathbf{z} - \mathbf{c})' = \mathbf{0},$$

and from this fact and the equation $\boldsymbol{\Psi} = \mathscr{E}(\mathbf{z} - \mathbf{c})(\bar{\mathbf{z}} - \bar{\mathbf{c}})'$ we see that

$$\mathscr{E}(\mathbf{x} - \mathbf{a})(\mathbf{x} - \mathbf{a})' = \tfrac{1}{2}\operatorname{Re}\boldsymbol{\Psi} = \mathscr{E}(\mathbf{y} - \mathbf{b})(\mathbf{y} - \mathbf{b})',$$
$$\mathscr{E}(\mathbf{x} - \mathbf{a})(\mathbf{y} - \mathbf{b})' = -\tfrac{1}{2}\operatorname{Im}\boldsymbol{\Psi} = -\mathscr{E}(\mathbf{y} - \mathbf{b})(\mathbf{x} - \mathbf{a})',$$

where $\mathbf{x} = \operatorname{Re}\mathbf{z}$, $\mathbf{y} = \operatorname{Im}\mathbf{z}$, $\mathbf{a} = \operatorname{Re}\mathbf{c}$, $\mathbf{b} = \operatorname{Im}\mathbf{c}$. In particular, if $\boldsymbol{\Psi} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix, then

$$\mathscr{E}(x_k - a_k)^2 = \mathscr{E}(y_k - b_k)^2 = \tfrac{1}{2}\sigma^2, \qquad 1 \leqq k \leqq n,$$
$$\mathscr{E}(x_k - a_k)(y_j - b_j) = 0, \qquad 1 \leqq j, \quad k \leqq n,$$

where $x_k = \operatorname{Re} z_k$, $y_k = \operatorname{Im} z_k$, $a_k = \operatorname{Re} c_k$, $b_k = \operatorname{Im} c_k$, $1 \leqq k \leqq n$. Thus f represents the product of the density functions of $2n$ (real) independent Gaussian variates $x_1, \cdots, x_n, y_1, \cdots, y_n$, each with the same variance, $\frac{1}{2}\sigma^2$:

$$f(\mathbf{z}) = \frac{1}{(2\pi)^n(\frac{1}{2}\sigma^2)^n} \exp\left\{-\frac{1}{2}\frac{1}{(\frac{1}{2}\sigma^2)}\sum_{k=1}^{n}[(x_k - a_k)^2 + (y_k - b_k)^2]\right\}.$$

Consider now the model $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$, where σ^2 as well as $\boldsymbol{\beta}$ is to be considered as an unknown parameter. We shall assume that $\mathbf{z}$ is normal $(\mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. Then the likelihood function of $\mathbf{z}$ is

$$l(\mathbf{z}) = \frac{1}{\pi^T\sigma^{2T}} e^{-|\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2/\sigma^2}. \tag{7.1}$$

We wish to determine the *maximum likelihood estimate* (MLE) of $\boldsymbol{\beta}$ and σ^2.

Setting $(\partial/\partial\boldsymbol{\beta}) \log l(\mathbf{z})$ equal to the p-dimensional zero vector leads to

$$\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\bar{\mathbf{H}}'\mathbf{z}$$

as the MLE of $\boldsymbol{\beta}$. But [see (3.2)], this estimate is identical with the LSE $\mathbf{b}^*$ of $\boldsymbol{\beta}$. Thus we conclude that $\hat{\boldsymbol{\beta}} = \mathbf{b}^*$ when $\mathbf{z}$ is normally distributed with covariance matrix a multiple of $\mathbf{I}$.

To determine $\hat{\sigma}^2$, the MLE of σ^2, we differentiate the logarithm of the likelihood function with respect to σ^2 and set this derivative equal to zero to obtain

$$\hat{\sigma}^2 = \frac{1}{T}|\mathbf{z} - \mathbf{H}\hat{\boldsymbol{\beta}}|^2.$$

Thus we have shown the following result.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

THEOREM 7.1. *Consider the model* $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$, *where* $\boldsymbol{\beta}$ *and* σ^2 *are unknown parameters. Let* $\mathbf{z}$ *be normal* $(\mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. *Then*

$$\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{z} \tag{7.2}$$

and

$$\hat{\sigma}^2 = \frac{1}{T}|\mathbf{z} - \mathbf{H}\hat{\boldsymbol{\beta}}|^2 \tag{7.3}$$

are the maximum likelihood estimates of $\boldsymbol{\beta}$ *and* σ^2 *respectively, where* $\mathbf{A} = \overline{\mathbf{H}}'\mathbf{H}$. *Furthermore,* $\hat{\boldsymbol{\beta}} = \mathbf{b}^*$, *where* $\mathbf{b}^*$ *is the least squares estimate of* $\boldsymbol{\beta}$.

The residual $\mathbf{z} - \mathbf{H}\boldsymbol{\beta}$ may be written

$$\mathbf{z} - \mathbf{H}\boldsymbol{\beta} = \mathbf{z} - \mathbf{H}\mathbf{b}^* + \mathbf{H}\mathbf{b}^* - \mathbf{H}\boldsymbol{\beta}.$$

Thus the exponent in the likelihood function (7.1) becomes

$$\begin{aligned}|\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2 = {} & |\mathbf{z} - \mathbf{H}\mathbf{b}^*|^2 + (\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{b}^* - \boldsymbol{\beta}) \\ & + (\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\overline{\mathbf{H}}'(\mathbf{z} - \mathbf{H}\mathbf{b}^*) + (\overline{\mathbf{z}} - \overline{\mathbf{H}}\overline{\mathbf{b}}^*)'\mathbf{H}(\mathbf{b}^* - \boldsymbol{\beta}).\end{aligned}$$

But

$$\overline{\mathbf{H}}'(\mathbf{z} - \mathbf{H}\mathbf{b}^*) = \overline{\mathbf{H}}'\mathbf{z} - \mathbf{A}\mathbf{b}^* = \overline{\mathbf{H}}'\mathbf{z} - \mathbf{A}(\mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{z}) = \mathbf{0}.$$

Thus

$$|\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2 = T\hat{\sigma}^2 + (\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{b}^* - \boldsymbol{\beta}) \tag{7.4}$$

and the likelihood function $l(z)$ may be written

$$l(\mathbf{z}) = (\pi\sigma^2)^{-T}\exp[-T\hat{\sigma}^2/\sigma^2]\exp[-\sigma^{-2}(\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{b}^* - \boldsymbol{\beta})].$$

An examination of the above formula leads to the following theorem.

THEOREM 7.2. *Consider the model* $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$, *where* $\boldsymbol{\beta}$ *and* σ^2 *are unknown parameters. Let* $\mathbf{z}$ *be normal* $(\mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. *Then* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$, *the maximum likelihood estimates of* $\boldsymbol{\beta}$ *and* σ^2 *respectively, form a set of sufficient statistics for estimating* $\boldsymbol{\beta}$ *and* σ^2.

Since $\mathscr{E}\mathbf{z} = \mathbf{H}\boldsymbol{\beta}$, equation (7.2) implies that

$$\mathscr{E}\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\overline{\mathbf{H}}'\mathscr{E}\mathbf{z} = \mathbf{A}^{-1}\mathbf{A}\boldsymbol{\beta} = \boldsymbol{\beta},$$

that is, that $\hat{\boldsymbol{\beta}}$ is an unbiased estimate of $\boldsymbol{\beta}$. (This we already knew since $\mathbf{b}^* = \hat{\boldsymbol{\beta}}$ and the LSE is unbiased.) To compute $\mathscr{E}\hat{\sigma}^2$ we take the expectation of both sides of (7.4) to obtain

$$\operatorname{tr}\sigma^2\mathbf{I}_T = T\mathscr{E}\hat{\sigma}^2 + \operatorname{tr}\sigma^2\mathbf{I}_p,$$

where the subscripts T and p on the identity matrix indicate its dimension. Hence

$$T\mathscr{E}\hat{\sigma}^2 = \operatorname{tr}\sigma^2\mathbf{I}_{T-p}$$

and

$$\mathscr{E}\hat{\sigma}^2 = \frac{T-p}{T}\sigma^2.$$

Thus we have the following theorem.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

THEOREM 7.3. *Consider the model* $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$, *where* $\boldsymbol{\beta}$ *and* σ^2 *are unknown parameters. Let* $\mathbf{z}$ *be normal* $(\mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. *Let* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$ *be the maximum likelihood estimates of* $\boldsymbol{\beta}$ *and* σ^2 *respectively. Then* $\hat{\boldsymbol{\beta}}$ *is an unbiased estimate of* $\boldsymbol{\beta}$ *and* $s^2 = (T/(T-p))\hat{\sigma}^2$ *is an unbiased estimate of* σ^2.

8. Distributions of the maximum likelihood statistics. Consider the model $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$, where $\boldsymbol{\beta}$ and σ^2 are unknown parameters. Let $\mathbf{z}$ be normal $(\mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I})$. Let $\mathbf{b}^* = \hat{\boldsymbol{\beta}}$ and $s^2 = (T/(T-p))\hat{\sigma}^2$, where $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are the MLE of $\boldsymbol{\beta}$ and σ^2 respectively. We have seen in Theorem 7.3 that $\mathbf{b}^*$ and s^2 are unbiased. What we wish to do now is to show that $\mathbf{b}^*$ is normally distributed and that s^2 is chi-square distributed (see Theorem 8.1).

The first part is easy. Since $\mathbf{b}^* = \mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{z}$ is a linear combination of $\mathbf{z}$, and $\mathbf{z}$ is normal, it follows that $\mathbf{b}^*$ is normal with mean $\mathscr{E}\mathbf{b}^* = \boldsymbol{\beta}$ and covariance

$$\mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})(\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})' = \sigma^2\mathbf{A}^{-1}.$$

We shall now demonstrate that $(T-p)s^2/(\frac{1}{2}\sigma^2)$ is $\chi^2_{2(T-p)}$. Towards this end, choose a nonsingular $p \times p$ matrix $\mathbf{F}$ such that

$$\overline{\mathbf{F}}'\mathbf{A}\mathbf{F} = \mathbf{I}_p,$$

where $\mathbf{I}_p$ is the $p \times p$ identity matrix. Since $\mathbf{A} = \overline{\mathbf{H}}'\mathbf{H}$, we may write the above equation as

$$(\overline{\mathbf{H}}\overline{\mathbf{F}})'(\mathbf{H}\mathbf{F}) = \mathbf{I}_p.$$

Let $\mathbf{U}_1 = \mathbf{H}\mathbf{F}$ and let $\mathbf{U}$ be a $T \times T$ unitary matrix whose first p columns are $\mathbf{U}_1$. We write

$$\mathbf{U} = \|\mathbf{U}_1 \quad \mathbf{U}_2\|,$$

where $\mathbf{U}_2$ is $T \times (T-p)$.

Now let $\mathbf{v} = \overline{\mathbf{U}}'(\mathbf{z} - \mathbf{H}\boldsymbol{\beta})$. Then

$$\overline{\mathbf{v}}'\mathbf{v} = |\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2 \tag{8.1}$$

and

$$\begin{aligned}(\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{b}^* - \boldsymbol{\beta}) &= (\overline{\mathbf{A}}^{-1}\mathbf{H}'\overline{\mathbf{z}} - \overline{\mathbf{A}}^{-1}\mathbf{H}'\overline{\mathbf{H}}\overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{z} - \mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{H}\boldsymbol{\beta}) \\ &= (\overline{\mathbf{v}}'\overline{\mathbf{U}}')(\mathbf{U}_1\mathbf{F}^{-1})(\overline{\mathbf{F}}'^{-1}\overline{\mathbf{U}}_1'\mathbf{U}_1\mathbf{F}^{-1})^{-1}(\overline{\mathbf{F}}'^{-1}\overline{\mathbf{U}}_1')(\mathbf{U}\mathbf{v})\end{aligned}$$

since $\mathbf{H} = \mathbf{U}_1\mathbf{F}^{-1}$ by definition of $\mathbf{U}_1$. Using the orthogonality of the columns of $\mathbf{U}$, the above expression reduces to

$$(\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{b}^* - \boldsymbol{\beta}) = \overline{\mathbf{v}}'\begin{bmatrix}\mathbf{I}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{v}. \tag{8.2}$$

If we let $\mathbf{v}^{*\prime} = \{v_{p+1}, \cdots, v_T\}$ be the last $T-p$ components of $\mathbf{v}$, then from (8.1) and (8.2),

$$|\mathbf{z} - \mathbf{H}\boldsymbol{\beta}|^2 - (\overline{\mathbf{b}}^* - \overline{\boldsymbol{\beta}})'\mathbf{A}(\mathbf{b}^* - \boldsymbol{\beta}) = |\mathbf{v}^*|^2.$$

But from (7.4),

$$T\hat{\sigma}^2 = |\mathbf{v}^*|^2.$$

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

Since $\mathbf{z}$ is normal $(\mathbf{H}\boldsymbol{\beta}, \sigma^2\mathbf{I})$ and $\mathbf{v} = \overline{\mathbf{U}}'(\mathbf{z} - \mathbf{H}\boldsymbol{\beta})$ is a linear combination of $\mathbf{z}$, it follows that $\mathbf{v}$ is normal with mean

$$\mathscr{E}\mathbf{v} = \overline{\mathbf{U}}'\mathscr{E}(\mathbf{z} - \mathbf{H}\boldsymbol{\beta}) = \mathbf{0}$$

and covariance matrix

$$\mathscr{E}\mathbf{v}\overline{\mathbf{v}}' = \overline{\mathbf{U}}'\mathscr{E}(\mathbf{z} - \mathbf{H}\boldsymbol{\beta})(\overline{\mathbf{z}} - \overline{\mathbf{H}}\overline{\boldsymbol{\beta}})'\mathbf{U} = \sigma^2\overline{\mathbf{U}}'\mathbf{I}_T\mathbf{U} = \sigma^2\mathbf{I}_T.$$

Thus $\mathbf{v}^*$ is normal $(\mathbf{0}, \sigma^2\mathbf{I}_{T-p})$ and $|\mathbf{v}^*|^2/\frac{1}{2}\sigma^2$ is $\chi^2_{2(T-p)}$ since $|\mathbf{v}^*|^2$ is the sum of the squares of $2(T - p)$ independent *real* Gaussian variates, each normal $(0, \frac{1}{2}\sigma^2)$. Hence

$$\frac{(T-p)s^2}{\frac{1}{2}\sigma^2} = \frac{T\hat{\sigma}^2}{\frac{1}{2}\sigma^2} = \frac{|\mathbf{v}^*|^2}{\frac{1}{2}\sigma^2}$$

is chi-square distributed with $2(T - p)$ degrees of freedom. Summarizing:

THEOREM 8.1. *Consider the model* $\mathscr{M}_{Tp}[\mathbf{z}, \boldsymbol{\beta}, \mathbf{H}, \sigma^2\mathbf{I}]$, *where* $\boldsymbol{\beta}$ *and* σ^2 *are unknown parameters. Let* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\sigma}^2$ *be the maximum likelihood estimates of* $\boldsymbol{\beta}$ *and* σ^2 *respectively. Then* $\hat{\boldsymbol{\beta}}$ *is normally distributed* $(\boldsymbol{\beta}, \sigma^2\mathbf{A}^{-1})$, *where* $\mathbf{A} = \overline{\mathbf{H}}'\mathbf{H}$, *and* $T\hat{\sigma}^2/\frac{1}{2}\sigma^2$ *is chi-square distributed with* $2(T - p)$ *degrees of freedom. Furthermore,* $\hat{\boldsymbol{\beta}}$ *and* $T\hat{\sigma}^2/\frac{1}{2}\sigma^2$ *are independent.*

We have proved, above, all but the last statement. Since $\mathbf{b}^* = \hat{\boldsymbol{\beta}}$ and $\mathbf{v}^*$ (where $|\mathbf{v}^*|^2 = T\hat{\sigma}^2$) are normal, we need show only that $\mathbf{b}^*$ and $\mathbf{v}^*$ are uncorrelated. Now

$$\begin{aligned}\mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})\overline{\mathbf{v}}^{*\prime} &= \mathscr{E}(\mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{z} - \mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{H}\boldsymbol{\beta})[\mathbf{U}_2'(\overline{\mathbf{z}} - \overline{\mathbf{H}}\overline{\boldsymbol{\beta}})]' \\ &= \sigma^2\mathbf{A}^{-1}\overline{\mathbf{H}}'\mathbf{U}_2.\end{aligned}$$

But by definition of $\mathbf{U}_1$,

$$\overline{\mathbf{H}}' = \overline{\mathbf{F}}'^{-1}\overline{\mathbf{U}}_1'.$$

Thus

$$\mathscr{E}(\mathbf{b}^* - \boldsymbol{\beta})\overline{\mathbf{v}}^{*\prime} = \sigma^2\mathbf{A}^{-1}\overline{\mathbf{F}}'^{-1}(\overline{\mathbf{U}}_1'\mathbf{U}_2) = \mathbf{0}$$

since $\mathbf{U}$ is a unitary matrix. Hence $\mathbf{b}^*$ and s^2 are independently distributed.

Acknowledgment. The author wishes to thank the various referees for making a number of clarifying observations and instructive comments.

REFERENCES

[1] R. ARENS, *Complex processes for envelopes of normal noise*, IRE Trans. Information Theory, IT-3 (1957), pp. 204–207.

[2] N. GIRI, *On the complex analogues of T^2- and R^2-tests*, Ann. Math. Statist., 36 (1965), pp. 664–670.

[3] N. R. GOODMAN, *Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)*, Ibid., 34 (1963), pp. 152–177.

[4] T. L. GRETTENBERG, *A representation theorem for complex normal processes*, IEEE Trans. Information Theory, IT-11 (1965), pp. 305–306.

[5] R. P. GUPTA, *Some extensions of the Wishart and multivariate t-distributions in the complex case*, J. Indian Statist. Assoc., 2 (1964), pp. 131–136.

[6] D. G. KABE, *Complex analogues of some classical noncentral multivariate distributions*, Austral. J. Statist., 8 (1966), pp. 99–103.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

[7] ———, *Complex analogues of some results in classical normal multivariate regression theory*, Ibid., 8 (1966), pp. 87–91.
[8] ———, *On the distribution of the complex analog of Rao's U statistic*, J. Indian Statist. Assoc., 4 (1966), pp. 189–194.
[9] C. G. KHATRI, *Classical statistical analysis based on a certain multivariate complex Gaussian distribution*, Ann. Math. Statist., 36 (1965), pp. 98–114.
[10] V. I. LADOHIN, *Generalization of some theorems of probability theory to the case of complex distributions*, Kazan. Gos. Univ. Uchen. Zap., 125 (1965), pp. 94–100. (In Russian.)
[11] K. S. MILLER, *Complex Gaussian processes*, this Review, 11 (1969), pp. 544–567.
[12] K. S. MILLER AND M. M. ROCHWARGER, *On estimating spectral moments in the presence of colored noise*, IEEE Trans. Information Theory, IT-16 (1970), pp. 303–309.
[13] ———, *A covariance approach to spectral moment estimation*, Ibid., IT-18 (1972), pp. 588–596.
[14] W. F. MCGEE, *Complex Gaussian noise moments*, Ibid., IT-17 (1971), pp. 149–157.
[15] I. S. REED, *On a moment theorem for complex Gaussian processes*, IRE Trans. Information Theory, IT-8 (1962), pp. 194–195.
[16] W. Y. TAN, *Some distribution theory associated with complex Gaussian distribution*, Tamkang J., 7 (1968), pp. 263–302.
[17] G. L. TURIN, *The characteristic function of Hermitian quadratic forms in complex normal variables*, Biometrika, 47 (1960), pp. 199–201.
[18] R. A. WOODING, *The multivariate distribution of complex normal variables*, Ibid., 43 (1956), pp. 212–215.

BIBLIOGRAPHY

A. P. J. ABRAHAMSE AND J. KOERTS, *New estimators of disturbances in regression analysis*, J. Amer. Statist. Assoc., 66 (1971), pp. 71–74.

A. P. J. ABRAHAMSE AND A. S. LOUTER, *On a new test for autocorrelation in least squares regression*, Biometrika, 58 (1971), pp. 53–60.

D. J. AIGNER, *Errors of measurement and least squares estimation in a simple recursive model of dynamic equilibrium*, Econometrica, 34 (1966), pp. 424–432.

S. A. AĬVAZJAN AND JU. A. ROZANOV, *Some remarks on asymptotically efficient linear estimates of regression coefficients*, Trudy Mat. Inst. Steklov., 71 (1964), pp. 3–16.

H. AKAIKE, *Statistical predictor identification*, Ann. Inst. Statist. Math., 22 (1970), pp. 203–217.

A. ALBERT AND R. W. SITTLER, *A method for computing least squares estimators that keep up with the data*, SIAM J. Control, 3 (1965), pp. 384–417.

T. W. ANDERSON, *Least squares and best unbiased estimates*, Ann. Math. Statist., 33 (1962), pp. 266–272.

———, *The Statistical Analysis of Time Series*, John Wiley, New York, 1971.

F. J. ANSCOMBE, *Topics in the investigation of linear relations fitted by the method of least squares*, J. Roy. Statist. Soc. Ser. B, 29 (1967), pp. 1–52.

F. J. ANSCOMBE AND J. W. TUKEY, *The examination and analysis of residuals*, Technometrics, 5 (1963), pp. 141–160.

M. ATIQULLAH, *The estimation of residual variance in quadratically balanced least-squares problems and the robustness of the F-test*, Biometrika, 49 (1962), pp. 83–91.

———, *On a restricted least squares estimator*, J. Amer. Statist. Assoc., 64 (1969), pp. 964–968.

R. H. BARHAM AND W. DRANE, *An algorithm for least squares estimation of nonlinear parameters when some of the parameters are linear*, Technometrics, 14 (1972), pp. 757–766.

G. A. BARNARD, *The logic of least squares*, J. Roy. Statist. Soc. Ser. B, 25 (1963), pp. 124–127.

M. S. BARTLETT AND P. D. M. MACDONALD, "*Least-squares*" *estimation of distribution mixtures*, Nature, 217 (1968), pp. 195–196.

D. W. BEHNKEN AND N. R. DRAPER, *Residuals and their variance patterns*, Technometrics, 14 (1972), pp. 101–111.

A. R. BERGSTROM, *The exact sampling distributions of least squares and maximum likelihood estimators of the marginal propensity to consume*, Econometrica, 30 (1962), pp. 480–490.

J. BERKSON, *Estimation by least squares and by maximum likelihood*, Proc. Third Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, Calif. 1954/55), vol. I, Univ. of California Press, Berkeley and Los Angeles, 1956, pp. 1–11.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

M. Bertaud, *The least squares straight line*, Statistica, 22 (1962), pp. 503–517.

I. Bhanot and C. G. Khatri, *Standard errors of the least square estimates of the constants in Makeham's second modification*, J. Maharaja Sayajirao Univ. Baroda, 10 (1961), pp. 15–19.

P. K. Bhattacharya, *Some properties of the least squares estimator in regression analysis when the predictor variables are stochastic*, Ann. Math. Statist., 33 (1962), pp. 1365–1374.

R. D. Bock, *Note on the least squares solution for the method of successive categories*, Psychometrika, 22 (1957), pp. 231–240.

H. W. Bode and C. E. Shannon, *A simplified derivation of linear least squares smoothing and prediction theory*, Proc. IRE, 38 (1950), pp. 417–425.

D. C. Bowden, *Simultaneous confidence bands for linear regression models*, J. Amer. Statist. Assoc., 65 (1970), pp. 413–421.

G. E. P. Box, *Use of statistical methods in the elucidation of basic mechanisms*, Bull. Inst. Internat. Statist., 36 (1958), pp. 215–225.

E. Breitenberger, *Remarks on the least-squares reduction of angular correlation data*, Proc. Phys. Soc. Sect. A., 69 (1956), pp. 489–491.

G. N. Brovkovič and Ju. V. Linnik, *Some remarks on estimation by means of confidence regions in the "least squares method"*, Teor. Verojatnost. i Primenen., 8 (1963), pp. 217–218.

T. M. Brown, *Simultaneous least-squares: a distribution-free method of equation system structure estimation*, Internat. Econ. Rev., 1 (1960), pp. 173–191.

P. G. Carlson, *A least squares interpretation of the bivariate line of organic correlation*, Skand. Aktuarietidskr., 39 (1956), pp. 7–10.

J. J. Chai, *Correlated measurement errors and the least squares estimator of the regression coefficient*, J. Amer. Statist. Assoc., 66 (1971), pp. 478–483.

J. S. Chipman, *On least squares with insufficient observations*, Ibid., 59 (1964), pp. 1078–1111.

J. S. Chipman and M. M. Rao, *The treatment of linear restrictions in regression analysis*, Econometrica, 32 (1964), pp. 198–209.

G. C. Chow, *A theorem on least squares and vector correlation in multivariate linear regression*, J. Amer. Statist. Assoc., 61 (1966), pp. 413–414.

C. F. Christ, *Simultaneous equation estimation: any verdict yet?*, Econometrica, 28 (1960), pp. 835–845.

J. F. Claerbout, *A summary, by illustrations, of least-squares filters with constraints*, IEEE Trans. Information Theory, IT-14 (1968), pp. 269–272.

E. R. Cohen, *The basis for the criterion of least squares*, Rev. Modern Physics, 25 (1953), pp. 709–713.

———, *Standard errors of the residues in a least-squares analysis*, Phys. Rev. (2), 101 (1956), pp. 1641–1642.

C. H. Cooke, *An efficiency property of the exponentially weighted least squares estimators*, SIAM J. Appl. Math., 19 (1970), pp. 769–777.

J. A. Cornell, *A note on the equality of least squares estimates using second-order equiradial rotatable designs*, Biometrika, 59 (1972), pp. 686–687.

D. R. Cox and D. V. Hinkley, *A note on the efficiency of least-squares estimates*, J. Roy. Statist. Soc. Ser. B, 30 (1968), pp. 284–289.

H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, N.J., 1946.

C. F. Crouse, *On a point arising in polynomial regression fitting*, Biometrika, 51 (1964), pp. 501–502.

T. Cunia, *Least squares estimates and parabolic regression with restricted location for the stationary point*, J. Amer. Statist. Assoc., 59 (1964), pp. 564–571.

———, *Weighted least squares method and construction of volume tables*, Bull. Inst. Internat. Statist., 40 (1964), pp. 1062–1063.

F. N. David and J. Neyman, *Extension of the Markoff theorem on least squares*, Statist. Res. Mem. London, 2 (1938), pp. 105–116.

G. W. Diederich, S. J. Messick and L. R. Tucker, *A general least squares solution for successive intervals*, Psychometrika, 22 (1957), pp. 159–173.

J. L. Doob, *Stochastic Processes*, John Wiley, New York, 1953.

F. Downton, *Least-squares estimates using ordered observations*, Ann. Math. Statist., 25 (1954), pp. 303–316.

———, *A note on ordered least-squares estimation*, Biometrika, 40 (1953), pp. 457–458.

N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley, New York, 1966.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

H. DRYGAS, *Consistency of the least squares and Gauss–Markov estimators in regression models*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 17 (1971), pp. 309–326.

———, *The Coordinate-Free Approach to Gauss–Markov Estimation*, Springer-Verlag, Berlin, 1970.

J. DURBIN, *Efficient fitting of linear models for continuous stationary time series from discrete data*, Bull. Inst. Internat. Statist., 38 (1961), pp. 273–282.

———, *Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables*, Econometrica, 38 (1970), pp. 410–421.

J. DURBIN AND G. S. WATSON, *Testing for serial correlation in least squares regression*, Biometrika, 37 (1950), pp. 409–428, and 38 (1951), pp, 159–178, and 58 (1971), pp. 1–19.

P. S. DWYER, *Generalizations of a Gaussian theorem*, Ann. Math. Statist., 29 (1958), pp. 106–117.

———, *Pearsonian correlation coefficients associated with least squares theory*, Ibid., 20 (1949), pp. 404–416.

P. S. DWYER AND M. S. MACPHAIL, *Symbolic matrix derivatives*, Ibid., 19 (1948), pp. 517–534.

M. L. EATON, *Gauss–Markov estimation for multivariate linear models: A coordinate free approach*, Ibid., 41 (1970), pp. 528–538.

F. EICKER, *Asymptotic normality and consistency of the least squares estimators for families of linear regressions*, Ibid., 34 (1963), pp. 447–456.

P. D. FINCH, *Linear least squares prediction in non-stochastic time series*, Advances in Appl. Probability, 1 (1969), pp. 111–122.

F. M. FISHER, *The place of least squares in econometrics: a comment*, Econometrica, 30 (1962), pp. 565–567.

R. FLETCHER, *Generalized inverse methods for the best least squares solution of systems of non-linear equations*, Comput. J., 10 (1968), pp. 392–399.

R. J. FREUND, R. W. VAIL AND C. W. CLUNIES-ROSS, *Residual analysis*, J. Amer. Statist. Assoc., 56 (1961), pp. 98–104.

B. FRIEDLAND, *Least squares filtering and prediction of nonstationary sampled data*, Information and Control, 1 (1958), pp. 297–313.

M. N. GHOSH, *Linear estimation of parameters in the case of unequal variances*, Calcutta Statist. Assoc. Bull., 10 (1961), pp. 117–130.

M. N. GHOSH AND P. K. SEN, *Sequential confidence interval for the regression coefficient based on Kendall's tau*, Ibid., 20 (1971), pp. 23–36.

A. S. GOLDBERGER, *Best linear unbiased prediction in the generalized linear regression model*, J. Amer. Statist. Assoc., 57 (1962), pp. 369–375.

———, *Stepwise least squares: residual analysis and specification error*, Ibid., 56 (1961), pp. 998–1000.

A. S. GOLDBERGER AND D. B. JOCHEMS, *Note on step-wise least squares*, Ibid., 56 (1961), pp. 105–110.

G. B. GOLDSTEIN AND P. SWERLING, *Bootstrapping the generalized least-squares estimator in colored Gaussion noise with unknown covariance parameters*, IEEE Trans. Information Theory, IT-16 (1970), pp. 385–392.

G. H. GOLUB, *Comparison of variance of minimum variance and weighted least squares regression coefficients*, Ann. Math. Statist., 34 (1963), pp. 984–991.

F. A. GRAYBILL, *An Introduction to Linear Statistical Models*, vol. I, McGraw-Hill, New York, 1961.

U. GRENANDER AND M. ROSENBLATT, *Statistical Analysis of Stationary Time Series*, John Wiley, New York, 1957.

P. G. GUEST, *Estimation of the error at a point on a least-squares curve*, Austral. J. Sci. Research Ser. A, 3 (1950), pp. 173–182.

E. J. HANNAN, *Least-squares efficiency for vector time series*, J. Roy. Statist. Soc. Ser. B, 30 (1968), pp. 490–498.

———, *Time Series Analysis*, John Wiley, New York, 1960.

E. J. HANNAN AND R. D. TERRELL, *Testing for serial correlation after least squares regression*, Econometrica, 36 (1968), pp. 133–150.

B. HARKIN, *The expected error of a least-squares solution of location from direction-finding equipment*, Austral. J. Appl. Sci., 7 (1956), pp. 263–272.

H. L. HARTER, *The method of least squares and some alternatives*, Rep. 72–0129, Aerospace Res. Labs., Wright-Patterson Air Force Base, Ohio, 1972.

H. O. HARTLEY, *The estimation of non-linear parameters by 'internal least squares'*, Biometrika, 35 (1948), pp. 32–45.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

———, *The modified Gauss–Newton method for the fitting of non-linear regression functions by least squares*, Technometrics, 3 (1961), pp. 269–280.

H. O. HARTLEY AND A. BOOKER, *Nonlinear least squares estimation*, Ann. Math. Statist., 36 (1965), pp. 638–650.

R. C. HENSHAW, *Testing single-equation least squares regression models for autocorrelated disturbances*, Econometrica, 34 (1966), pp. 646–660.

M. HESBOIS, *Comparisons of least-squares methods*, Ann. Soc. Sci. Bruxelles, 84 (1970), pp. 185–204.

C. HILDRETH, *Simultaneous equations: any verdict yet?*, Econometrica, 28 (1960), pp. 846–854.

C. HILDRETH AND J. P. HOUCK, *Some estimators for a linear model with random coefficients*, J. Amer. Statist. Assoc., 63 (1968), pp. 584–595.

B. M. HILL, *Foundations for the theory of least squares*, J. Roy. Statist. Soc. Ser. B, 31 (1969), pp. 89–97.

R. R. HOCKING, *The distribution of a projected least squares estimator*, Ann. Inst. Statist. Math., 17 (1965), pp. 357–362.

N. A. HUTTLY, *The fitting of regression curves with auto-correlated data*, Biometrika, 43 (1956), pp. 468–474.

J. A. JACQUEZ, F. J. MATHER AND C. R. CRAWFORD, *Linear regression with non-constant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. I*, Biometrics, 24 (1968), pp. 607–626.

R. I. JENNRICH, *Asymptotic properties of non-linear least squares estimators*, Ann. Math. Statist., 40 (1969), pp. 633–643.

D. E. JOHANSEN, *Linear mean square filtering and estimation when process statistics are undefined*, Proc. Symposium on System Theory (New York, 1965), Polytechnic Press, Polytechnic Inst. Brooklyn, Brooklyn, N.Y., 1965, pp. 247–284.

H. C. JOKSCH, *The reduction of the variance by least squares of polynomial approximations*, this Review, 8 (1966), pp. 211–219.

A. JONES, *Spiral—a new algorithm for non-linear parameter estimation using least squares*, Comput. J., 13 (1970), pp. 301–308.

A. M. KAGAN AND O. V. SALAEVSKIĬ, *The admissibility of least square estimates is an exclusive property of the normal law*, Mat. Zametki, 6 (1969), pp. 81–89.

N. C. KAKWANI, *The bias of the three-stage least-squares estimators in simultaneous equations*, Austral. J. Statist., 13 (1971), pp. 19–26.

H. H. KELEJIAN, *Two-stage least squares and econometric systems linear in parameters but nonlinear in the endogenous variables*, J. Amer. Statist. Assoc., 66 (1971), pp. 373–374.

C. G. KHATRI, *A further investigation on a method for fitting an exponential curve*, Vidya, 3 (1963), pp. 38–44.

———, *A theorem on least squares in multivariate regression*, J. Amer. Statist. Assoc., 62 (1967), pp. 1494–1495.

L. R. KLEIN, *Single equation versus equation system methods of estimation in economics*, Econometrica, 28 (1960), pp. 866–871.

A. N. KOLMOGOROV, *On the proof of the method of least squares*, Uspehi Mat. Nauk (N.S.), 1 (11) (1946), pp. 57–70.

D. D. KOSAMBI, *An extension of the least-squares method for statistical estimation*, Ann. Eugenics, 13 (1947), pp. 257–261.

———, *The method of least squares*, J. Indian Soc. Agric. Statist., 11 (1959), pp. 49–57.

W. KRUSKAL, *The coordinate-free approach to Gauss–Markov estimation, and its application to missing and extra observations*, Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, Calif.), vol. I, Univ. California Press, Berkeley, 1961, pp. 435–451.

———, *When are Gauss–Markov and least squares estimators identical? A coordinate-free approach*, Ann. Math. Statist., 39 (1968), pp. 70–75.

K. KUBIK, *On the efficiency of least squares estimators in non-linear models*, Statistica Neerlandica, 22 (1968), pp. 33–36.

T. LANCASTER, *Grouping estimators on heteroscedastic data*, J. Amer. Statist. Assoc., 63 (1968), pp. 182–191.

H. J. LARSON, *Least squares estimation of the components of a symmetric matrix*, Technometrics, 8 (1966), pp. 360–362.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

A. B. LEES, *Interpolation and extrapolation of sampled data*, IRE Trans. Information Theory, IT-2 (1956), pp. 12–17.
M. M. LENTNER, *Generalized least-squares estimation of a sub-vector of parameters in randomized fractional factorial experiments*, Ann. Math. Statist., 40 (1969), pp. 1344–1352.
V. LEVIS, *Mean square regression linear spaces*, Rend. Mat. Univ. Roma, 17 (1958), pp. 423–438.
T. O. LEWIS AND P. L. ODELL, *A generalization of the Gauss–Markov theorem*, J. Amer. Statist. Assoc., 61 (1966), pp. 1063–1066.
JU. V. LINNIK, *Method of Least Squares and Principles of the Theory of Observations*, Pergamon Press, New York, 1961.
———, *Some remarks on least squares in connection with direct and inverse location problems*, Teor. Verojatnost. i Primenen., 2 (1957), pp. 349–359.
E. H. LLOYD, *Least-squares estimation of location and scale parameters using order statistics*, Biometrika, 39 (1952), pp. 88–95.
S. P. LLOYD AND B. MCMILLAN, *Linear least squares filtering and prediction of sampled signals*, Proc. Symposium on Modern Network Synthesis (New York, 1955), Polytechnic Inst. Brooklyn, Brooklyn, N.Y., 1956, pp. 221–247.
J. W. LONGLEY, *An appraisal of least squares programs for the electronic computer from the point of view of the user*, J. Amer. Statist. Assoc., 62 (1967), pp. 819–841.
V. LOZANO LOPEZ, *Models of linear regression with heteroscedastic disturbances*, Estadíst. Española, 27 (1965), pp. 57–73.
———, *Models with autocorrelation in the random perturbations (generalized least squares method)*, Ibid., 27 (1965), pp. 5–37.
A. MADANSKY, *On the efficiency of three-stage least-squares estimation*, Econometrica, 32 (1964), pp. 51–56.
G. S. MADDALA, *Generalized least squares with an estimated variance covariance matrix*, Ibid., 39 (1971), pp. 23–33.
T. A. MAGNESS AND J. B. MCGUIRE, *Comparison of least squares and minimum variance estimates of regression parameters*, Ann. Math. Statist., 33 (1962), pp. 462–470.
T. L. MALEVIČ, *The asymptotic effectiveness of least squares estimates of the regression coefficients*, Teor. Verojatnost. i Primenen., 14 (1969), pp. 179–181.
H. B. MANN AND P. B. MORANDA, *On the efficiency of the least squares estimates of parameters in the Ornstein Uhlenbeck process*, Sankhyā, 13 (1954), pp. 351–358.
N. MANTEL, *Restricted least squares regression and convex quadratic programming*, Technometrics, 11 (1969), pp. 763–773.
D. W. MARQUARDT, *An algorithm for least-squares estimation of nonlinear parameters*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 431–441.
F. W. MCELROY, *A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased*, J. Amer. Statist. Assoc., 62 (1967), pp. 1302–1304.
D. A. MEETER, *On a theorem used in nonlinear least squares*, SIAM J. Appl. Math., 14 (1966), pp. 1176–1179.
K. S. MILLER, *Multidimensional Gaussian Distributions*, John Wiley, New York, 1964.
K. S. MILLER AND L. R. ABRAMSON, *Minimum variance estimation and prediction theory*, Internat. J. Engrg. Sci., 5 (1967), pp. 251–263.
S. K. MITRA AND C. R. RAO, *Conditions for optimality and validity of simple least squares theory*, Ann. Math. Statist., 40 (1969), pp. 1617–1624.
V. MUKHERJEE, *Note on the "asymptotic minimum variance property" of least squares estimators in time series regression*, J. Indian Soc. Agric. Statist., 20 (1968), pp. 71–74.
G. M. MULLET AND T. W. MURRAY, *A new method for examining rounding errors in least-squares regression computer programs*, J. Amer. Statist. Assoc., 66 (1971), pp. 496–498.
A. L. NAGAR AND Y. P. GUPTA, *The moment matrix of the two-stage least-squares estimator of coefficients in different equations of a complete system of simultaneous equations*, Econometrica, 38 (1970), pp. 39–49.
M. NAKAMURA, *A note on the consistency of simultaneous least squares estimation*, Internat. Econ. Rev., 1 (1960), pp. 192–197.
C. L. NARAYANA AND M. SUBRAHMANYAM, *On an alternative method to the principle of least squares in the theory of regression*, Sankhyā Ser. B, 31 (1969), pp. 427–434.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

R. NARAYANAN, *Computation of Zellner-Theil's three stage least squares estimates*, Econometrica, 37 (1969), pp. 298–306.

G. NILSSON, *Optimal stratification according to the method of least squares*, Skand. Aktuarietidskr., 50 (1967), pp. 128–136.

S. OBERLÄNDER, *The method of least squares for a three parameter exponential regression*, Z. Angew. Math. Mech., 43 (1963), pp. 493–506.

P. L. ODELL AND T. O. LEWIS, *Best linear recursive estimation*, J. Amer. Statist. Assoc., 66 (1971), pp. 893–896.

J. OGAWA, *Note on the Markoff's theorem on least squares*, Osaka J. Math., 2 (1950), pp. 145–150.

F. R. OLIVER, *Estimating the exponential growth function by direct least squares*, Appl. Statist., 19 (1970), pp. 92–100.

Z. PAWŁOWSKI, *On a special case when the variance of a least squares estimator does not tend to zero*, Zastos. Mat., 7 (1964), pp. 391–397.

O. PERSSON, *On the solution of an over-determined system of equations by the method of least squares*, Nordisk Mat. Tidskr., 6 (1958), pp. 69–77, 95.

I. PETERSEN, *A comparison of the method of reproducing kernels with the method of least squares*, Eesti NSV Tead. Akad. Toimetised Füüs.-Mat., 18 (1969), pp. 403–409.

V. V. PETROV, *On the method of least squares and its extremal properties*, Uspehi Mat. Nauk (N.S.), 9 (1954), pp. 41–62.

D. A. PIERCE, *A duality between autoregressive and moving average processes concerning their least squares parameter estimates*, Ann. Math. Statist., 41 (1970), pp. 422–426.

———, *Least squares estimation in dynamic-disturbance time series models*, Biometrika, 59 (1972), pp. 73–78.

———, *Least squares estimation in the regression model with autoregressive-moving average errors*, Ibid., 58 (1971), pp. 299–312.

R. L. PLACKETT, *A historical note on the method of least squares*, Biometrika, 36 (1949), pp. 458–460.

———, *Principles of Regression Analysis*, Clarendon Press, Oxford, 1960.

———, *Some theorems in least squares*, Biometrika, 37 (1950), pp. 149–157.

———, *Studies in the history of probability and statistics XXIX. The discovery of the method of least squares*, Ibid., 59 (1972), pp. 239–251.

F. PROSCHAN, *Precision of least squares polynomial estimates*, this Review, 3 (1961), pp. 230–235.

M. H. QUENOUILLE, *Computational devices in the application of least squares*, J. Roy. Statist. Soc. Ser. B, 12 (1950), pp. 256–272.

J. B. RAMSEY, *Tests for specification errors in classical linear least squares regression analysis*, Ibid., 31 (1969), pp. 350–371.

C. R. RAO, *Least squares theory using an estimated dispersion matrix and its application to measurement of signals*, Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability (Berkeley, Calif., 1965/66), vol. I: Statistics, Univ. California Press, Berkeley, 1967, pp. 352–372.

———, *A note on a previous lemma in the theory of least squares and some further results*, Sankhyā Ser. A, 30 (1968), pp. 259–266.

———, *Some problems involving linear hypotheses in multivariate analysis*, Biometrika, 46 (1959), pp. 49–58.

———, *The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves*, Ibid., 52 (1965), pp. 447–458.

H. R. RICHARDSON, *Recursive regression analysis when the least-squares estimate is not asymptotically efficient*, SIAM J. Appl. Math., 19 (1970), pp. 178–190.

J. B. ROBERTSON, *On a problem in non-linear prediction theory*, Ann. Math. Statist., 36 (1965), pp. 1554–1555.

C. A. ROHDE AND J. R. HARVEY, *Unified least squares analysis*, J. Amer. Statist. Assoc., 60 (1965), pp. 523–527.

S. ROSENBAUM, *The variance of least-squares estimates under linear restraints*, J. Roy. Statist. Soc. Ser. B, 13 (1951), pp. 250–255.

H. C. RUTEMILLER AND D. A. BOWERS, *Estimation in a heteroscedastic regression mode*, J. Amer. Statist. Assoc., 63 (1968), pp. 552–557.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

O. V. ŠALAEVSKIĬ, *Equalization of observations with unknown weights by the method of least squares*, Trudy Mat. Inst. Steklov., 104 (1968), pp. 189–214.

———, *On the method of least squares with unknown weights*, Proc. All-Union Conference on the Theory of Probability and Mathematical Statistics (Erevan, 1958), Izdat. Akad. Nauk Armjan. SSR, Erevan, 1960, pp. 106–111. (In Russian.)

J. D. SARGAN, *Three-stage least-squares and full maximum likelihood estimates*, Econometrica, 32 (1964), pp. 77–81.

T. SAWA, *The exact sampling distribution of ordinary least squares and two-stage least squares estimators*, J. Amer. Statist. Assoc., 64 (1969), pp. 923–937.

H. SCHEFFÉ, *The Analysis of Variance*, John Wiley, New York, 1959.

S. L. SCLOVE, *Improved estimators for coefficients in linear regression*, J. Amer. Statist. Assoc., 63 (1968), pp. 596–606.

G. A. F. SEBER, *The Linear Hypothesis: A General Theory*, Charles Griffin, London, 1966.

J. SEELY, *Linear spaces and unbiased estimation*, Ann. Math. Statist., 41 (1970), pp. 1725–1748.

P. K. SEN, *Estimates of the regression coefficient based on Kendall's tau*, J. Amer. Statist. Assoc., 63 (1968), pp. 1379–1389.

W. P. SEWELL, *Least squares, conditional predictions, and estimator properties*, Econometrica, 37 (1969), pp. 39–43.

B. K. SHAH, *Fitting of logistic curve by least square method*, J. Maharaja Sayajirao Univ. Baroda, 9 (1960), pp. 11–14.

B. K. SHAH AND I. R. PATEL, *The least square estimates of the constants for the Makeham second modification of Gompertz's law*, Ibid., 9 (1960), pp. 1–10.

M. M. SIDDIQUI, *Covariances of least-squares estimates when residuals are correlated*, Ann. Math. Statist., 29 (1958), pp. 1251–1256.

H. W. SORENSON, *Least-squares estimation: from Gauss to Kalman*, Spectrum IEEE, 7 (1970), pp. 63–68.

P. SPRENT, *Fitting a polynomial to correlated equally spaced observations*, Biometrika, 52 (1965), pp. 275–276.

———, *A generalized least-squares approach to linear functional relationships (with discussion)*, J. Roy. Statist. Soc. Ser. B, 28 (1966), pp. 278–297.

J. N. SRIVASTAVA, *On the extension of Gauss–Markov theorem to complex multivariate linear models*, Ann. Inst. Statist. Math., 19 (1967), pp. 417–437.

———, *A multivariate extension of the Gauss–Markov theorem*, Ibid., 17 (1965), pp. 63–66.

V. K. SRIVASTAVA, *The efficiency of estimating seemingly unrelated regression equations*, Ibid., 22 (1970), pp. 483–493.

E. STREISSLER, *A reexamination of the Haavelmo-bias in the least squares estimation of the Keynesian consumption function*, Metrika, 14 (1969), pp. 164–182.

B. F. SWINDEL, *On the basis of some least-squares estimators of variance in a general linear model*, Biometrika, 55 (1968), pp. 313–316.

A. A. TEMPEL'MAN, *Linear regression estimates*, Dokl. Akad. Nauk SSSR, 191 (1970), pp. 772–775.

R. D. TERRELL AND N. E. TUCKWELL, *The efficiency of least squares in estimating a stable seasonal pattern*, J. Amer. Statist. Assoc., 66 (1971), pp. 354–362.

H. THEIL AND A. SCHWEITZER, *The best quadratic estimator of the residual variance in regression analysis*, Statistica Neerlandica, 15 (1961), pp. 19–23.

J. ULMO, *Geometric study of the multiple linear regression*, C. R. Acad. Sci. Paris, 254 (1962), pp. 61–63.

J. VAN IJZEREN, *The theoretical aspect of least squares*, Statistica, Rijswijk, 8 (1954), pp. 21–45.

K. N. VENKATARAMAN, *A note on the least square estimators of the parameters of a second order linear stochastic difference equation*, Calcutta Statist. Assoc. Bull., 16 (1967), pp. 15–28.

C. VILLEGAS, *On the asymptotic efficiency of least squares estimators*, Ann. Math. Statist., 37 (1966), pp. 1676–1683.

———, *On the least squares estimation of a linear relation*, Bol. Fac. Ingen. Agrimens. Montevideo, 8 (1962), pp. 47–63.

———, *On the least squares estimation of a linear relation*, Publ. Inst. Mat. Estadíst. Fac. Ingen. Agric. Montevideo, 3 (1963), pp. 189–204.

———, *On the least squares estimation of non-linear relations*, Ann. Math. Statist., 40 (1969), pp. 462–466.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms

H. M. WAGNER, *Non-linear regression with minimal assumptions*, J. Amer. Statist. Assoc., 57 (1962), pp. 572–578.

G. S. WATSON, *Linear least squares regression*, Ann. Math. Statist., 38 (1967), pp. 1679–1699.

———, *Linear regression on proportions*, Biometrics, 25 (1969), pp. 585–588.

———, *Prediction and the efficiency of least squares*, Biometrika, 59 (1972), pp. 91–98.

F. V. WAUGH, *The place of least-squares in econometrics*, Econometrica, 29 (1961), pp. 386–396; 30 (1962), pp. 568–569.

P. A. WAYMAN, *A least-squares solution for a linear relation between two observed quantities*, Nature, 184 (1959), pp. 77–78.

P. WHITTLE, *Prediction and Regulation by Linear Least-Square Methods*, The English Universities Press, London, 1963.

———, *On principal components and least square methods of factor analysis*, Skand. Aktuarietidskr., 35 (1952), pp. 223–239.

M. R. WICKENS, *The consistency and efficiency of generalized least squares in simultaneous equation systems with autocorrelated errors*, Econometrica, 37 (1969), pp. 651–659.

S. S. WILKS, *Mathematical Statistics*, John Wiley, New York, 1962.

E. J. WILLIAMS, *A note on regression methods in calibration*, Technometrics, 11 (1969), pp. 189–192.

———, *Regression Analysis*, John Wiley, New York, 1959.

H. WOLD, *On the consistency of least squares regression*, Sankhyā Ser. A, 25 (1963), pp. 211–215.

———, *Estimation of principal components and related models by iterative least squares*, Multivariate Analysis (Proc. Internat. Sympos., Dayton, Ohio, 1965), Academic Press, New York, 1966, pp. 391–420.

———, *Nonlinear estimation by iterative least squares procedures*, Research Papers in Statistics (Festschrift J. Neyman), John Wiley, London, 1966, pp. 411–444.

———, *On least square regression with auto-correlated variables and residuals*, Bull. Inst. Internat. Statist., 32 (1950), pp. 277–289.

H. WOLD AND F. LYTTKENS, *Nonlinear iterative partial least squares (NIPALS) estimation procedures: Group report*, Ibid., 43 (1969), pp. 29–51.

S. ZACKS, *On a complete class of linear unbiased estimators for randomized factorial experiments*, Ann. Math. Statist., 34 (1963), pp. 769–779.

———, *Generalized least squares estimators for randomized fractional replication designs*, Ibid., 35 (1964), pp. 696–704.

A. ZELLNER, *An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias*, J. Amer. Statist. Assoc., 57 (1962), pp. 348–368.

A. ZELLNER AND H. THEIL, *Three-stage least squares: simultaneous estimation of simultaneous equations*, Econometrica, 30 (1962), pp. 54–78.

G. ZYSKIND, *On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models*, Ann. Math. Statist., 38 (1967), pp. 1092–1109.

———, *Parametric augmentations and error structures under which certain simple least squares and analysis of variance procedures are also best*, J. Amer. Statist. Assoc., 64 (1969), pp. 1353–1368.

G. ZYSKIND, O. KEMPTHORNE, F. B. MARTIN, E. J. CARNEY AND E. N. WEST, *Linear models and analysis of variance research procedures*, ARL Tech. Rep. No. 68-0119, Aerospace Res. Labs., Wright-Patterson Air Force Base, Ohio, 1968.

G. ZYSKIND AND F. B. MARTIN, *On the best linear estimation and a general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure*, SIAM J. Appl. Math., 17 (1969), pp. 1190–1202.

This content downloaded from 128.176.217.46 on Thu, 17 May 2018 13:55:16 UTC
All use subject to http://about.jstor.org/terms