DOUGLAS PAPER 5205
JANUARY 1969

AD 681430

DDC
RECEIVED
FEB 6 1969
B

EXTENDED ITERATIVE WEIGHTED LEAST SQUARES: ESTIMATION OF A LINEAR MODEL IN THE PRESENCE OF COMPLICATIONS

A.F. GOODMAN

Appears in
Computer Science and Statistics:
Partners in Progress

This document has been approved for public release and sale; its distribution is unlimited

MCDONNELL DOUGLAS ASTRONAUTICS COMPANY
WESTERN DIVISION
MCDONNELL DOUGLAS CORPORATION

Best Available Copy

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va. 22151

EXTENDED ITERATIVE WEIGHTED LEAST SQUARES: ESTIMATION OF A LINEAR MODEL IN THE PRESENCE OF COMPLICATIONS*

A. F. GOODMAN

Senior Technical Staff to Vice President
Information Systems Subdivision
McDonnell Douglas Astronautics Company—Western Division

ABSTRACT

This paper introduces an extended iterative weighted least squares procedure, denoted by EIWLS, for solution of a classical problem in analysis of scientific and technical information: estimation in the presence of complications, of the coefficients for linearly independent component signals in a linear model, from observations on the component signals and a composite signal containing the linear model plus noise which is nonstationary and/or correlated with unknown covariance matrix. An iterative weighted least squares procedure, denoted by IWLS, is developed for estimation in the absence of complications. Then IWLS is extended to perform the estimation subject to: (1) estimators being consistent with a priori information describing the random variation of coefficients over all possible states of nature (e.g., all systems of a specified type from a production process); (2) utilization of data from all pertinent channels in the estimation of coefficients which appear in the linear models for more than one data channel; and (3) replacement of the linear model by a new linear model containing only representative component signals which are highly-descriptive, but not highly-related, when there is a large number of component signals in the linear model and some of them are highly-related. FORTRAN computer programs have been written to implement IWLS and EIWLS on the IBM 7094 for the case of nonstationary and uncorrelated noise.

---

*The paper forms a portion of Chapter 2 in Computer Science and Statistics: Partners in Progress, a forthcoming volume edited by A. F. Goodman and N. R. Mann. It represents a current and revised version of the author's "Estimation of Coefficients in a Linear Model by Extended Iterative Weighted Least Squares," Autonetics Publication X4-1290/32, North American Rockwell Corporation, August 1964. Section 1 has been revised, Section 7 and the References have been expanded and brought up to date, and Section 8 has been summarized.

1. INTRODUCTION AND SUMMARY

Since the early 19th century, estimation of a linear model from data subject to error has been a classical problem in analysis of scientific and technical information. The least squares procedure for solution of the problem, formulated by Gauss in 1802 and published by Legendre in 1806, is essentially the first statistical technique developed for analysis of information (Ref 41). However, the effectiveness of least squares and related procedures mainly depends upon characteristics of the error.

This paper introduces an extended iterative weighted least squares procedure, denoted by EIWLS, for solution of the classical problem in the presence of complications. A complete description of the problem, which may be termed a generalized statistical regression problem, is: estimation in the presence of complications, of the coefficients for linearly independent component signals in a linear model, from observations on the component signals and a composite signal containing the linear model plus noise with unknown characteristics. Component signals are sometimes called input or independent variables, the composite signal is sometimes called an output or dependent variable, and the noise is sometimes called a residual or random error. Since EIWLS was developed originally for error analysis of an inertial navigation system, the signal-and-noise terminology is employed.

Pertinent characteristics of the noise are contained in the square array of noise variances and covariances, called the noise covariance matrix. Noise is said to be stationary and uncorrelated when that matrix is a constant multiple of the identity matrix, nonstationary and uncorrelated when it is a diagonal matrix, stationary and correlated when each row of it is a proper arrangement of elements in the first row, and nonstationary and correlated otherwise.

Consider estimation in the absence of complications. If the noise covariance matrix is known, then the optimum estimators of the coefficients are the weighted least squares estimators determined by its inverse. If the noise covariance matrix is not known, its estimation appears to be a reasonable step toward estimation of the coefficients.

Estimation of the linear model is, itself, required to estimate the noise covariance matrix. Goodman (Ref 1) presented an iterative weighted least squares procedure, denoted by IWLS, to accomplish the estimation when the noise is nonstationary and/or correlated (i.e., not stationary and uncorrelated--in which case, the least squares estimators are optimum) with unknown covariance matrix. Briefly, IWLS:

1. Obtains the least squares estimators of the coefficients.

2. Calculates an estimator of the noise covariance matrix by using the composite signal and its least squares estimator, based upon the least squares estimators of the coefficients, to estimate the necessary noise variances and covariances.

3. Obtains the weighted least squares estimators of the coefficients which are determined by the inverse of this matrix estimator.

4. Iteratively repeats 2 and 3, with the least squares estimators of the coefficients replaced by the latest set of weighted least squares estimators of the coefficients, and obtains a new estimator of the noise covariance matrix and a new set of weighted least squares estimators of the coefficients.

5. Continues the iteration in 4 until a preassigned level of stability is attained.

The complications and their treatment were also summarized in general terms by Goodman (Ref 1). This paper is an extension of Ref 1 and introduces an improved and extended IWLS. Improvement of IWLS, as presented in Ref 1, involves improved estimation of the noise covariance matrix. In the following paragraphs, each complication and the corresponding extension of IWLS to EIWLS is briefly discussed.

Coefficients in the linear model are constant for a particular state of nature (e.g., a particular system of a specified type from a production process). However, they may vary randomly from one state of nature to another. A priori information describing the coefficients' random variation, over all possible states of nature (e.g., all systems of that specified type from the production process), may exist from previous analysis; and the estimators ought to be consistent with it. To insure this, a modification of IWLS permits the incorporation into the procedure of a priori information concerning the means and covariance matrix of the coefficients.

Data may exist from several data channels, and a coefficient may appear in the linear models for more than one data channel. The data from all pertinent channels should be utilized in the estimation of that coefficient. This may be accomplished by properly arranging the coefficients and data from all channels into a form suitable for the application of IWLS.

There may be a large number of component signals in the linear model and some of them, though linearly independent, may be highly-related. Component signals are called highly-related in this paper when they possess a high degree of linear dependence. For accuracy and ease of computation, it is frequently desirable to replace the linear model by a new linear model containing only representative component signals which are highly-descriptive, but not highly-related, and to estimate the coefficients of the representative component signals in the new linear model. To accomplish this, the set of component signals is partitioned into subsets of highly-related ones and the appropriate weighted average of a subset is selected to be the representative component signal for that subset in the new linear model. The coefficients in the new linear model may then be estimated by IWLS. In addition, an estimator for the coefficient of the representative component signal for a subset is apportioned among the coefficients of component signals in the subset, via the weighting scheme used in the selection of that representative component signal.

Additional operations, such as pre-editing of data and inclusion of estimates based upon previous sets of data into the a priori information, may be added without too much difficulty.

Although EIWLS is applicable when the noise is nonstationary and/or correlated, it has been programmed for the computer only in the case of nonstationary and uncorrelated noise. Two FORTRAN computer programs, the basic one (Ref 2) and another (Ref 3) using Efroymson's technique (Ref 4) to preselect the representative component signals, have been written to implement IWLS as presented in Ref 1; and a FORTRAN computer program (Ref 5) has been written to implement EIWLS.

Iterative statistical procedures such as EIWLS extract considerably more information from the data than do noniterative, closed-form procedures such as least squares. In view of recent computer hardware and software development, iterative procedures have been feasible to implement and evaluate for some time and continue to become more so with the passage of time. It is therefore "penny-wise and pound-foolish" not to utilize EIWLS, when dictated by theory, statistical tests such as the one given in Ref 42, or examination of the data. Indeed, EIWLS is even more appropriate today than at the time of its development.

Since iterative statistical procedures bridge the gap between the two extremes of noniterative, closed-form and optimum statistical procedures, it is somewhat surprising that the development of a meaningful theory for iterative statistics has been essentially neglected in favor of the continued--and almost academic--characterization and comparison of the two extremes. This is well illustrated by the survey of related literature in Section 7. It is noteworthy that the quite-recent Ref 42 contains an iterative procedure which is closely-related to IWLS, as well as an approach to the construction of confidence intervals and statistical tests, when the noise is nonstationary and uncorrelated.

Ref 41 recommends the consideration of four questions regarding an iterative statistical procedure:

1. Under what conditions does the iterative procedure converge?

2. How rapidly does the iterative procedure converge?

3. Under what conditions does the iterative procedure converge to the proper solution?

4. To what extent does the iterative procedure improve upon a noniterative, closed-form procedure?

Partial answers to all four questions are provided by Sections 2-8.

Those interested in only the essence of EIWLS may confine themselves to Section 1. Sections 6, 7 and 8 augment Section 1 with a discussion, a

comprehensive survey of related literature, and an illustrative example. In Sections 2 through 5, the statistical details of EIWLS (whose comprehension may require careful reading by the nonstatistician) are displayed, with a minimum of development and amplification, along with some reasonable alternatives which are listed in footnotes.

2. ITERATIVE WEIGHTED LEAST SQUARES

Suppose that t represents time or some other auxiliary variable. At time t: $X_1(t)$, $X_2(t)$, ..., $X_p(t)$ denote linearly independent component signals; β_1, β_2, ..., β_p denote unknown coefficients; $Y(t)$ denotes a composite signal; and $\epsilon(t)$ denotes noise with mean zero, variance $\sigma^2(t)$ and covariance $\sigma(t, t')$ between $\epsilon(t)$ and $\epsilon(t')$. Then the linear model* is $\sum_{j=1}^{p} \beta_j X_j(t)$ and the representation of $Y(t)$ as containing the linear model plus noise is

$$Y(t) = \sum_{j=1}^{p} \beta_j X_j(t) + \epsilon(t). \tag{1}$$

Given the explicit observations X_{ji} of $X_j(t)$ for $j=1, 2, \ldots, p$ and Y_i of $Y(t)$, which imply the implicit observation $\epsilon_i = Y_i - \sum_{j=1}^{p} \beta_j X_{ji}$ of $\epsilon(t)$, at time t_i for $i = 1, 2, \ldots, n > p$, the generalized statistical regression problem without the complicating restrictions is the estimation of β_1, β_2, ..., β_p when ϵ_1, ϵ_2, ..., ϵ_n have mean zero and unknown n-by-n covariance matrix Σ of variances $\sigma_{ii} = \sigma_i^2 = \sigma^2(t_i)$ and covariances $\sigma_{hi} = \sigma(t_h, t_i)$.

Complicated expressions in this and subsequent paragraphs may be written in compact form by the introduction of matrix notation. Let

$$\underline{X}_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jn} \end{bmatrix} \text{ for } j = 1, 2, \ldots, p,$$

$$X = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p),$$

*A constant term may be included in the linear model by setting $X_1(t)$ identically equal to one.

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix},$$

$$\underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},$$

$$\underline{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad \text{and}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_{nn} \end{bmatrix}.$$

The matrix alter ego of Equation (1) is

$$\underline{Y} = \sum_{j=1}^{p} \beta_j \underline{X}_j + \underline{\epsilon} = X\underline{\beta} + \underline{\epsilon}\ ; \qquad (2)$$

and the matrix alter ego of the generalized statistical regression problem without the complicating restrictions is the estimation of $\underline{\beta}$ given X and $\underline{Y}$, when Σ is unknown.

The weighted least squares estimators of $\beta_1, \beta_2, \ldots, \beta_p$ which are determined by the n-by-n matrix

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{12} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{1n} & w_{2n} & \cdots & w_{nn} \end{bmatrix}$$

of weights w_{hi} (or any matrix that is a constant multiple of W) are those minimizing the quadratic form

$$Q_W = \sum_{h=1}^{n} \sum_{i=1}^{n} w_{hi} \left(Y_h - \sum_{j=1}^{p} \beta_j X_{jh}\right)\left(Y_i - \sum_{j=1}^{p} \beta_j X_{ji}\right)$$

$$= (\underline{Y} - X\underline{\beta})' W (\underline{Y} - X\underline{\beta}).$$

It may be shown that these estimators are given by

$$\underline{\hat{\beta}}_W = (\hat{\beta}_{W1}, \hat{\beta}_{W2}, \ldots, \hat{\beta}_{Wp})' = (X' W X)^{-1} X' W \underline{Y}$$

and have covariance matrix $(X' W X)^{-1} X' W \Sigma W X (X' W X)^{-1}$. They are called the least squares estimators if W is the n-by-n identity matrix I_n.

In the sense of possessing minimum variance among all unbiased linear estimators, of providing an estimator of $\sum_{j=1}^{p} \beta_j X_{ji}$ which has minimum variance among all unbiased linear estimators of it and of being maximum likelihood estimators when the noise is normally distributed, the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ are the weighted least squares estimators which are determined by Σ^{-1} (Ref 6, Chapter 14 and Ref 7, Sections 1.3-1.5). The least squares estimators are, therefore, optimum only if the noise is stationary and uncorrelated, and Σ is a constant multiple of the identity matrix.

For nonstationary and/or correlated noise, it is not unreasonable to presume that estimation of Σ and then estimation of $\beta_1, \beta_2, \ldots, \beta_p$ by weighted least squares is superior to estimation of $\beta_1, \beta_2, \ldots, \beta_p$ by least squares. Estimation of $\sum_{j=1}^{p} \beta_j X_{ji}$ for $i = 1, 2, \ldots, n$ is, however, required to estimate Σ. This may be accomplished via estimation of $\beta_1, \beta_2, \ldots, \beta_p$. Hence, iterative estimation of $\beta_1, \beta_2, \ldots, \beta_p$ and then Σ, as accomplished by IWLS, should provide improvement over least squares estimation of $\beta_1, \beta_2, \ldots, \beta_p$.

Stated in symbols, IWLS:

1. Obtains the least squares estimators $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots, \hat{\beta}_p^{(0)}$ from

$$\underline{\hat{\beta}}^{(0)} = (\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots, \hat{\beta}_p^{(0)})' = (X'X)^{-1} X' \underline{Y} . \quad (3)$$

2. Calculates an estimator $\hat{\Sigma}^{(1)}$, by using Y_i and its least squares estimator

$$\hat{Y}_i^{(0)} = \sum_{j=1}^{p} \hat{\beta}_j^{(0)} X_{ji} . \quad (4)$$

for $i = 1, 2, \ldots, n$ in an appropriate estimation scheme to calculate the necessary $\hat{\sigma}_i^{(1)2}$ and $\hat{\sigma}_{hi}^{(1)}$

3. Obtains the weighted least squares estimators $\hat{\beta}_1^{(1)}, \hat{\beta}_2^{(1)}, \ldots, \hat{\beta}_p^{(1)}$ which are determined by $(\hat{\Sigma}^{(1)})^{-1}$ from

$$\underline{\hat{\beta}}^{(1)} = (\hat{\beta}_1^{(1)}, \hat{\beta}_2^{(1)}, \ldots, \hat{\beta}_p^{(1)})'$$

$$= \left[X' (\hat{\Sigma}^{(1)})^{-1} X \right]^{-1} X' (\hat{\Sigma}^{(1)})^{-1} \underline{Y} . \quad (5)$$

4. Iteratively repeats 2 and 3, with $\hat{\beta}_1^{(0)}, \hat{\beta}_2^{(0)}, \ldots, \hat{\beta}_p^{(0)}$ and $\hat{Y}_i^{(0)}$ for $i = 1, 2, \ldots, n$ replaced by the latest set of weighted least squares estimators $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ and

$$\hat{Y}_i^{(c)} = \sum_{j=1}^{p} \hat{\beta}_j^{(c)} X_{ji} \quad (6)$$

for i = 1, 2, ..., n, and obtains a new estimator $\hat{\Sigma}^{(c+1)}$ and a new set of weighted least squares estimators $\hat{\beta}_1^{(c+1)}$, $\hat{\beta}_2^{(c+1)}$, ..., $\hat{\beta}_p^{(c+1)}$ from

$$\hat{\underline{\beta}}^{(c+1)} = (\hat{\beta}_1^{(c+1)}, \hat{\beta}_2^{(c+1)}, \ldots, \hat{\beta}_p^{(c+1)})'$$

$$= \left[X' (\hat{\Sigma}^{(c+1)})^{-1} X\right]^{-1} X' (\hat{\Sigma}^{(c+1)})^{-1} \underline{Y} . \qquad (7)$$

5. Continues the iteration in 4 for c = 1, 2, ..., c*-1, where c* either is determined when a valid measure of the change in $\hat{\beta}_1^{(c)}$, $\hat{\beta}_2^{(c)}$, ..., $\hat{\beta}_p^{(c)}$ becomes less than a preassigned constant or is, itself, a preassigned constant. An estimator for the covariance matrix of $\hat{\beta}_1^{(c)}$, $\hat{\beta}_2^{(c)}$, ..., $\hat{\beta}_p^{(c)}$ is given by $\left[X' (\hat{\Sigma}^{(c)})^{-1} X\right]^{-1} X' (\hat{\Sigma}^{(c)})^{-1} \hat{\Sigma}^{(c+1)} (\hat{\Sigma}^{(c)})^{-1} X \left[X' (\hat{\Sigma}^{(c)})^{-1} X\right]$, which simplifies to $\left[X' (\hat{\Sigma}^{(c)})^{-1} X\right]$ when $\hat{\Sigma}^{(c+1)} = \hat{\Sigma}^{(c)}$, for $\hat{\Sigma}^{(0)} = I_n$ and c = 0, 1, ..., c*.

The selection of an appropriate estimation scheme to use in 2 is influenced not only by the type of noise and the data, but also by statistical considerations.

When the noise is nonstationary and uncorrelated,

$$\Sigma = \text{diag}\,(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2).$$

An estimator of σ_i^2 is provided by

$$\hat{\sigma}_i^{(c)2} = (Y_i - \hat{Y}_i^{(c-1)})^2 \text{ for } c = 1, 2, \ldots, c^* \text{ and } i = 1, 2, \ldots, n. \qquad (8)$$

These estimators are somewhat unsatisfactory because each of them is based upon only one observation; and they should be combined to produce more satisfactory estimators. In most applications, it is reasonable to assume a linear model for the variation of $\sigma^2(t)$. Let $T_1(t)$, $T_2(t)$, ..., $T_P(t)$ be known functions of t and ν_1, ν_2, ..., ν_P be unknown coefficients with p + P < n; and assume

$$\sigma^2(t) = \sum_{k=1}^{P} \nu_k T_k(t).$$

A simple version of $\sum_{k=1}^{P} \nu_k T_k(t)$, which is employed by the computer program (Ref 5), is the polynomial $\sum_{k=1}^{P} \nu_k t^{k-1}$. Suppose that $T_{ki} = T_k(t_i)$ for $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, P$,

$$\underline{T}_k = \begin{bmatrix} T_{k1} \\ T_{k2} \\ \vdots \\ T_{kn} \end{bmatrix} \quad \text{for } k = 1, 2, \ldots, P,$$

$$T = (\underline{T}_1, \underline{T}_2, \ldots, T_P),$$

$$\underline{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_P \end{bmatrix} \quad \text{and}$$

$$\underline{\hat{\sigma}}(c)2 = \begin{bmatrix} \hat{\sigma}_1(c)2 \\ \hat{\sigma}_2(c)2 \\ \vdots \\ \hat{\sigma}_n(c)2 \end{bmatrix} \quad \text{for } c = 1, 2, \ldots, c^*.$$

Then an appropriate estimation scheme* to use in 2 is

$$\underline{\hat{\nu}}^{(1)} = (\hat{\nu}_1^{(1)}, \hat{\nu}_2^{(1)}, \ldots, \hat{\nu}_P^{(1)})' = (T'T)^{-1} T' \underline{\hat{\sigma}}^{(1)2}, \tag{9}$$

$$\underline{\hat{\nu}}^{(c)} = (\hat{\nu}_1^{(c)}, \hat{\nu}_2^{(c)}, \ldots, \hat{\nu}_P^{(c)})'$$

$$= \left\{ T' \left[(\hat{\Sigma}^{(c-1)})^{-1} \right] T \right\}^{-1} T' \left[(\hat{\Sigma}^{(c-1)})^{-1} \right] \underline{\hat{\sigma}}^{(c)2} \tag{10}$$

for $c = 2, 3, \ldots, c^*$ and

$$\hat{\Sigma}^{(c)} = \operatorname{diag}\left(\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1}, \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2}, \ldots, \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kn} \right) \tag{11}$$

for $c = 1, 2, \ldots, c^*$. It might be observed that the use of Equations (9)-(11) produces a simple iterative procedure for obtaining a

*This estimation scheme is preferable to

$$\tilde{\sigma}_i^{(c)2} = \begin{cases} \dfrac{1}{M+i} \displaystyle\sum_{h=-i+1}^{M} \hat{\sigma}_h^{(c)2} & \text{for } i=1,2,\ldots,M \text{ and } c=1,2,\ldots,c^* \\ \dfrac{1}{2M+1} \displaystyle\sum_{h=-M}^{M} \hat{\sigma}_h^{(c)2} & \text{for } i=M+1, M+2, \ldots, n-M \text{ and } c=1,2,\ldots,c^* \\ \dfrac{1}{n+M-i+1} \displaystyle\sum_{h=-M}^{n-i} \hat{\sigma}_h^{(c)2} & \text{for } i=n-M+1, n-M+2, \ldots, n \text{ and } c=1,2,\ldots,c^* \end{cases},$$

which is the suitably truncated running average of 2M+1 estimators $\hat{\sigma}_h^{(c)2}$ that was suggested by Ref 1, when it is reasonable to assume a linear model for the variation of $\sigma^2(t)$; and the estimation scheme suggested by Ref 1 is preferable, when such an assumption is not reasonable.

solution to the maximum likelihood equations*, whose accuracy may be checked, for nonstationary, uncorrelated noise which is normally distributed with $\sigma^2(t) = \sum_{k=1}^{P} \nu_k T_k(t)$. To establish an upper limit for all weights, all estimators $\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{ki}$ of σ_i^2 must be bounded away from zero by a lower limit.

Consider the case of stationary, correlated noise, and let the observations be taken at different and equally-spaced times with $t_i = t_0 + i\ \Delta t$. Then $\sigma(t, t') = \sigma(\Delta)$ is a function of only the separation time $\Delta = |t-t'|$, $\sigma_{hi} = \sigma(r)$ is a function of only $r = |h-i|$ and

$$\Sigma = \begin{bmatrix} \sigma(0) & \sigma(1) & \sigma(2) \cdots \sigma(n-1) \\ \sigma(1) & \sigma(0) & \sigma(1) \cdots \sigma(n-2) \\ \sigma(2) & \sigma(1) & \sigma(0) \cdots \sigma(n-3) \\ \vdots & \vdots & \vdots \quad \vdots \\ \sigma(n-1) & \sigma(n-2) & \sigma(n-3) \ \ldots \ \sigma(0) \end{bmatrix} .$$

An estimator** of $\sigma(r)$ is

$$\hat{\sigma}^{(c)}(r) = \frac{1}{n-r} \sum_{i=1}^{n-r} (Y_i - \hat{Y}_i^{(c-1)})(Y_{i+r} - \hat{Y}_{i+r}^{(c-1)}) \qquad (12)$$

for $c = 1, 2, \ldots, c^*$ and $r = 0, 1, \ldots, n-1$

*The investigation which yielded Equations (9)-(11) was partially prompted by the conjecture of Dr. T. L. Gunckel that IWLS, as presented in Ref 1, might provide an iterative solution to the maximum likelihood equations.

**An alternative estimator, which takes the estimation of $\beta_1, \beta_2, \ldots, \beta_p$ explicitly into account, divides the sum by $n-p-r$ and limits r to $r = 0, 1, \ldots, n-p-1$.

It is usually reasonable to assume a linear model for the variation of $\sigma(\Delta)$*. Suppose that $T_1(\Delta)$, $T_2(\Delta)$, ..., $T_P(\Delta)$ are known functions of Δ, $\nu_1, \nu_2, \ldots, \nu_P$ are unknown coefficients with $p + P < n$ and

$$\sigma(\Delta) = \sum_{k=1}^{P} \nu_k T_k(\Delta),$$

whose simple version is

$$\sigma(\Delta) = \sum_{k=1}^{P} \nu_k \Delta^{k-1}.$$

If $T_{ki} = T_k((i-1)\,\Delta t)$ for $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, P$,

$$\underline{T}_k = \begin{bmatrix} T_{k1} \\ T_{k2} \\ \vdots \\ T_{kn} \end{bmatrix} \quad \text{for } k = 1, 2, \ldots, P,$$

$$T = (\underline{T}_1, \underline{T}_2, \ldots, \underline{T}_p),$$

$$\underline{\nu} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_P \end{bmatrix},$$

*Either of these estimators may be used, as is, when it is not reasonable to assume a linear model for the variation of $\sigma(\Delta)$.

$$\underline{\hat{\sigma}}^{(c)} = \begin{bmatrix} \hat{\sigma}^{(c)}_{(0)} \\ \hat{\sigma}^{(c)}_{(1)} \\ \vdots \\ \hat{\sigma}^{(c)}_{(n-1)} \end{bmatrix} \quad \text{and}$$

$N = \text{diag}\,(n, n-1, \ldots, 1)$,

then an appropriate estimation scheme to use in 2 is given by

$$\underline{\hat{\nu}}^{(c)} = (\hat{\nu}_1^{(c)}, \hat{\nu}_2^{(c)}, \ldots, \hat{\nu}_P^{(c)})' = (T' N T)^{-1} T' N \underline{\hat{\sigma}}^{(c)} \qquad (13)$$

and

$$\sum(c) = \begin{bmatrix} \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k3} \cdots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kn} \\ \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} \cdots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,n-1} \\ \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k3} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} \cdots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,n-2} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kn} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,n-1} & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k,n-2} \cdots & \sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1} \end{bmatrix} \qquad (14)$$

for $c = 1, 2, \ldots, c^*$.

The weight matrix N is used in Equation (13) because it constitutes a simple means of reflecting the increase of statistical dependability in $\hat{\sigma}^{(c)}(r)$ with $n-r$. It is necessary to require that all estimators $\left|\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{ki}\right|$ of $\sigma^{(c)}(r)$ be bounded above by the estimator

$\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kl}$ of $\sigma^{(c)}(0)$. A common engineering model for the variation of $\sigma(\Delta)$ is

$$\sigma(\Delta) = \sigma_0 e^{-\nu\Delta},$$

which is nonlinear in the unknown coefficients σ_0 and ν. This exponential variation of $\sigma(\Delta)$ may be treated* by transforming the model to the linear one,

$$\log_e \sigma(\Delta) = \log_e \sigma_0 - \nu\Delta = \nu_0 - \nu\Delta,$$

and proceeding as above with $\hat{\sigma}^{(c)}(i)$, T_{1i}, T_{2i}, ν_1, and ν_2 being replaced by $\log_e \hat{\sigma}^{(c)}(i)$, 1, $(i-1)\Delta t$, ν_0 and $-\nu$, respectively.

A proper modification of Equations (12) - (14) would eliminate the restriction of observations to different and equally-spaced times. In addition, an appropriate combination of the techniques used for nonstationary, uncorrelated noise and for stationary, correlated noise may be employed in the event of nonstationary, correlated noise. The calculations would then become more complicated and, perhaps in some cases, prohibitive.

Inversion of $\hat{\Sigma}^{(c)}$ in 3 must now be accomplished. If the noise is nonstationary and uncorrelated and $\sigma^2(t) = \sum_{k=1}^{P} \nu_k T_k(t)$, $(\hat{\Sigma}^{(c)})^{-1}$ may be written in closed form as

$$(\hat{\Sigma}^{(c)})^{-1} = \operatorname{diag}\left(\frac{1}{\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k1}}, \frac{1}{\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{k2}}, \ldots, \frac{1}{\sum_{k=1}^{P} \hat{\nu}_k^{(c)} T_{kn}} \right)$$

for $c = 1, 2, \ldots, c^*$. (15)

---

*If preferred, $\sigma_0 e^{-\nu\Delta}$ may be approximated by $\sum_{k=1}^{P} \nu_k \Delta^{k-1}$ and $\nu_1, \nu_2, \ldots, \nu_P$ may be estimated as in Equation (13); or σ_0 and ν may be estimated by those (nonlinear) estimators that minimize $\sum_{i=1}^{n} (n-i+1) \left[\hat{\sigma}^{(c)}(i-1) - \sigma_0 e^{-\nu(i-1)\Delta t} \right]^2$, but are difficult to compute.

The performance of this inversion has not yet been investigated for correlated noise.

A valid, yet simple, measure* of the change in $\hat{\beta}_1(c), \hat{\beta}_2(c), \ldots, \hat{\beta}_p(c)$ for 5 is $\max_{j=1,2,\ldots,p} \left| \frac{\hat{\beta}_j(c) - \hat{\beta}_j(c-1)}{\hat{\beta}_j(c)} \right|$.

*Depending upon the application, $\max_{i=1,2,\ldots,n} \left| \frac{\hat{Y}_i(c) - \hat{Y}_i(c-1)}{\hat{Y}_i(c)} \right|$, $\sum_{j=1}^{p} \left| \frac{\hat{\beta}_j(c) - \hat{\beta}_j(c-1)}{\hat{\beta}_j(c)} \right|$ or $\sum_{i=1}^{n} \left| \frac{\hat{Y}_i(c) - \hat{Y}_i(c-1)}{\hat{Y}_i(c)} \right|$ may be a more desirable measure of this change.

3. INCORPORATION OF A PRIORI INFORMATION

Let $\beta_1, \beta_2, \ldots, \beta_p$ vary randomly over all possible states of nature with means $\beta_1^*, \beta_2^*, \ldots, \beta_p^*$ and covariance matrix Γ of variances $\gamma_{j^2} = \gamma_{jj}$ and covariances γ_{jk}. Introducing matrix notation yields

$$\underline{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_p^* \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \ldots & \gamma_{1p} \\ \gamma_{12} & \gamma_{22} & \ldots & \gamma_{2p} \\ \vdots & \vdots & \ldots & \vdots \\ \gamma_{1p} & \gamma_{2p} & & \gamma_{pp} \end{bmatrix}$$

The a priori information, which exists from previous analysis, is knowledge of $\underline{\beta}^*$ and Γ.

Two quite reasonable methods to ensure that estimators of $\beta_1, \beta_2, \ldots, \beta_p$ will be consistent with the a priori information are: incorporation of $\underline{\beta}^*$ and Γ into the quadratic form to be minimized by the estimators; and employment of a linear combination $\Lambda\, \hat{\underline{\beta}}^{(c^*)} + (I_p - \Lambda)\, \underline{\beta}^*$ (a special case of which is $\lambda \hat{\underline{\beta}}^{(c^*)} + (1 - \lambda)\, \underline{\beta}^*$, with $\Lambda = \lambda I_p$ and $C < \lambda < 1$) as the vector of estimators, where Λ is a matrix which measures one's relative confidence in the data (as represented by $\hat{\underline{\beta}}^{(c^*)}$) and the a priori information (as represented by $\underline{\beta}^*$). Suppose

$$Q^* = (\underline{Y} - X\underline{\beta})' \, (\hat{\Sigma}^{(c^*)})^{-1} \, (\underline{Y} - X\underline{\beta}) + (\underline{\beta} - \underline{\beta}^*)' \Gamma^{-1} \, (\underline{\beta} - \underline{\beta}^*)$$

is the quadratic form that incorporates $\underline{\beta}^*$ and Γ and is to be minimized*. Then it may be observed that the corresponding estimators $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*$ which incorporate $\underline{\beta}^*$ and Γ are given by

$$\begin{aligned}
\hat{\underline{\beta}}^* &= (\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*)' \\
&= \left[X'(\hat{\Sigma}^{(c*)})^{-1}X + \Gamma^{-1}\right]^{-1} \left[X'(\hat{\Sigma}^{(c*)})^{-1}\,\underline{Y} + \Gamma^{-1}\,\underline{\beta}^*\right] \\
&= \left[X'(\hat{\Sigma}^{(c*)})^{-1}X + \Gamma^{-1}\right]^{-1} \left[X'(\hat{\Sigma}^{(c*)})^{-1}X\,\hat{\underline{\beta}}^{(c*)} + \Gamma^{-1}\,\underline{\beta}^*\right] \\
&= \hat{\Lambda}\,\hat{\underline{\beta}}^{(c*)} + (I_p - \hat{\Lambda})\,\underline{\beta}^*,
\end{aligned} \tag{16}$$

---

*If $0 < \hat{\lambda} < 1$, then two other quadratic forms that incorporate $\underline{\beta}^*$ and Γ and might be minimized are:

$$Q\textcircled{*} = \hat{\lambda}(\underline{Y} - X\underline{\beta})'(\hat{\Sigma}^{(c*)})^{-1}(\underline{Y} - X\underline{\beta})$$

$$+ (1 - \hat{\lambda})(X\underline{\beta} - X\underline{\beta}^*)'(\hat{\Sigma}^{(c*)})^{-1}(X\underline{\beta} - X\underline{\beta}^*),$$

with the corresponding estimators $\hat{\beta}_1\textcircled{*}, \hat{\beta}_2\textcircled{*}, \ldots, \hat{\beta}_p\textcircled{*}$ that incorporate $\underline{\beta}^*$ and Γ being yielded by

$$\hat{\underline{\beta}}\textcircled{*} = (\hat{\beta}_1\textcircled{*}, \hat{\beta}_2\textcircled{*}, \ldots, \hat{\beta}_p\textcircled{*})' = \hat{\lambda}\hat{\underline{\beta}}^{(c*)} + (1 - \hat{\lambda})\,\underline{\beta}^*; \text{ and}$$

$$Q\triangle = \hat{\lambda}(\underline{Y} - X\underline{\beta})'(\hat{\Sigma}^{(c*)})^{-1}(\underline{Y} - X\underline{\beta}) + (1 - \hat{\lambda})(\underline{\beta} - \underline{\beta}^*)'\Gamma^{-1}(\underline{\beta} - \underline{\beta}^*),$$

with the corresponding estimators $\hat{\beta}_1\triangle, \hat{\beta}_2\triangle, \ldots, \hat{\beta}_p\triangle$ that incorporate $\underline{\beta}^*$ and Γ being yielded by

$$\hat{\underline{\beta}}\triangle = (\hat{\beta}_1\triangle, \hat{\beta}_2\triangle, \ldots, \hat{\beta}_p\triangle)' = \left[\hat{\lambda}X'(\hat{\Sigma}^{(c*)})^{-1}X + (1 - \hat{\lambda})\Gamma^{-1}\right]^{-1}\left[\hat{\lambda}X'(\hat{\Sigma}^{(c*)})^{-1}\underline{Y} + (1 - \hat{\lambda})\Gamma^{-1}\,\underline{\beta}^*\right]$$

One such $\hat{\lambda}$ is

$$\hat{\lambda} = \|\cdot\| \,/\, (\|\hat{\Lambda}\| + \|I_p - \hat{\Lambda}\|)$$

for $\|\hat{\Lambda}\|$ being a norm of $\hat{\Lambda}$ (e.g., $\left(\sum_{j=1}^{p}\sum_{k=1}^{p}\hat{\lambda}_{jk}^{\,2}\right)^{1/2}$).

where $\hat{\Lambda}$ is the p-by-p matrix of elements $\hat{\lambda}_{jk}$ defined as

$$\hat{\Lambda} = \left[X' (\hat{\Sigma}^{(c*)})^{-1} X + \Gamma^{-1}\right]^{-1} \left[X' (\hat{\Sigma}^{(c*)})^{-1} X\right]. \qquad (17)$$

Note that $\hat{\underline{\beta}}^{(c*)}$ and $\hat{\Sigma}^{(c*)}$ may be replaced by $\hat{\underline{\beta}}^{(0)}$ and I_n, if it is desired to incorporate the a priori information into least squares rather than IWLS.

It might easily be shown that $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*$ would be the maximum likelihood estimators if Σ and $\hat{\Sigma}^{(c*)}$ were equal for both the noise and the coefficients being normally distributed.

4. PROPER ARRANGEMENT OF COEFFICIENTS AND DATA

To denote the kth data channel, for k = 1, 2, ..., q, prefix a k to the subscripts of the previously-introduced notation and obtain, in particular: $X_{k1}(t), X_{k2}(t), \ldots, X_{kp_k}(t), \beta_{k1}, \beta_{k2}, \ldots, \beta_{kp_k}, Y_k(t)$ and $\epsilon_k(t)$; $X_{kji} = X_{kj}(t_{ki})$ for $j = 1, 2, \ldots, p_k$, $Y_{ki} = Y_k(t_{ki})$ and $\epsilon_{ki} = \epsilon_k(t_{ki}) = Y_{ki} - \sum_{j=1}^{p_k} \beta_{kj} X_{kji}$ for $i = 1, 2, \ldots, n_k$; and

$$\underline{X}_{kj} = \begin{bmatrix} X_{kj1} \\ X_{kj2} \\ \vdots \\ X_{kjn_k} \end{bmatrix} \qquad \text{for } j = 1, 2, \ldots, p_k,$$

$$X_k = (\underline{X}_{k1}, \underline{X}_{k2}, \ldots, X_{kp_k}),$$

$$\underline{\beta}_k = \begin{bmatrix} \beta_{k1} \\ \beta_{k2} \\ \vdots \\ \beta_{kp_k} \end{bmatrix}$$

$$\underline{Y}_k = \begin{bmatrix} Y_{k1} \\ Y_{k2} \\ \vdots \\ Y_{kn_k} \end{bmatrix}$$

$$\underline{\epsilon}_k = \begin{bmatrix} \epsilon_{k1} \\ \epsilon_{k2} \\ \vdots \\ \epsilon_{kn_k} \end{bmatrix}$$

and

$$\Sigma_{kk} = \begin{bmatrix} \sigma_{k11} & \sigma_{k12} & \cdots & \sigma_{k1n_k} \\ \sigma_{k12} & \sigma_{k22} & \cdots & \sigma_{k2n_k} \\ \cdot & \vdots & & \vdots \\ \sigma_{k1n_k} & \sigma_{k2n_k} & \cdots & \sigma_{kn_k n_k} \end{bmatrix}$$

The analogues of Equations (1) and (2) for the kth channel are now

$$Y_k(t) = \sum_{j=1}^{P_k} \beta_{kj} X_{kj}(t) + \epsilon_k(t) \text{ and} \tag{18}$$

$$\underset{\sim}{Y}_k = \sum_{j=1}^{P_k} \beta_{kj} \underset{\sim}{X}_{kj} + \underset{\sim}{\epsilon}_k = X_k \underset{\sim}{\beta}_k + \underset{\sim}{\epsilon}_k \text{ for } k = 1, 2, \ldots, q. \tag{19}$$

Let

$$X = \begin{bmatrix} X_1 & 0 & 0 & \cdots & 0 \\ 0 & X_2 & 0 & \cdots & 0 \\ 0 & 0 & X_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & X_q \end{bmatrix}$$

$$\underline{\beta} = \begin{bmatrix} \underline{\beta}_1 \\ \underline{\beta}_2 \\ \vdots \\ \underline{\beta}_q \end{bmatrix}$$

$$\underline{Y} = \begin{bmatrix} \underline{Y}_1 \\ \underline{Y}_2 \\ \vdots \\ \underline{Y}_q \end{bmatrix}$$

and

$$\underline{\epsilon} = \begin{bmatrix} \underline{\epsilon}_1 \\ \underline{\epsilon}_2 \\ \vdots \\ \underline{\epsilon}_q \end{bmatrix}$$

which has the effect of stacking the data and coefficients associated with channels 2, 3, ..., q under the data and coefficients associated with channel 1. If σ_{hkij} is the covariance between ϵ_{hi} and ϵ_{kj} and the n_h-by-n_k matrix Σ_{hk} is

$$\Sigma_{hk} = \begin{bmatrix} \sigma_{hk11} & \sigma_{hk12} & \cdots & \sigma_{hk1n_k} \\ \sigma_{hk21} & \sigma_{hk22} & \cdots & \sigma_{hk2n_k} \\ \vdots & \vdots & & \vdots \\ \sigma_{hkn_h1} & \sigma_{hkn_h2} & \cdots & \sigma_{hkn_hn_k} \end{bmatrix}$$

then the covariance matrix of $\epsilon_{11}, \epsilon_{12}, \ldots, \epsilon_{1n_1}, \epsilon_{21}, \epsilon_{22}, \ldots, \epsilon_{2n_2}, \ldots, \epsilon_{q1}, \epsilon_{q2}, \ldots, \epsilon_{qn_q}$ is

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1q} \\ \Sigma'_{12} & \Sigma_{22} & \cdots & \Sigma_{2q} \\ \vdots & \vdots & & \vdots \\ \Sigma'_{1q} & \Sigma'_{2q} & \cdots & \Sigma_{qq} \end{bmatrix}$$

Suppose that $\beta_{k_1j_1}, \beta_{k_2j_2}, \ldots, \beta_{k_rj_r}$ are the same (i.e., the same coefficient appears in the linear models for channels $k_1, k_2, \ldots, k_r$). All but one of them, for example $\beta_{k_1j_1}$, are superfluous and should be eliminated from consideration; and the pertinent columns of $X_{k_1}, X_{k_2}, \ldots, X_{k_r}$ and $\underline{Y}_{k_1}, \underline{Y}_{k_2}, \ldots, \underline{Y}_{k_r}$ (i.e., the pertinent data from channels $k_1, k_2, \ldots, k_r$) should be utilized in the estimation of $\beta_{k_1j_1}$. This may be accomplished in a simple (though not simple-appearing) manner, by deleting $\beta_{k_2j_2}, \beta_{k_3j_3}, \ldots, \beta_{k_rj_r}$ from $\underline{\beta}$; replacing column number $\sum_{k=1}^{k_1-1} p_k + j_1$ in X by the sum of column numbers $\sum_{k=1}^{k_1-1} p_k + j_1, \sum_{k=1}^{k_2-1} p_k + j_2, \ldots, \sum_{k=1}^{k_r-1} p_k + j_r$ in X (which has

the effect of alining* the j_2th column of X_{k_2}, the j_3th column of X_{k_3}, ..., the j_rth column of X_{k_r} under the j_1th column of X_{k_1} in the formation of the $\left(\sum_{k=1}^{k_1-1} p_k + j_1\right)$th column of X); and deleting column numbers $\sum_{k=1}^{k_2-1} p_k + j_2$, $\sum_{k=1}^{k_3-1} p_k + j_3$, ..., $\sum_{k=1}^{k_r-1} p_k + j_r$ from X. If more convenient, one of $\beta_{k_1 j_1}$, $\beta_{k_2 j_2}$, ..., $\beta_{k_r j_r}$ other than $\beta_{k_1 j_1}$ may be used as the focal point for the alinement.

After all duplications among the β_{kj}'s have been treated as indicated, the coefficients and data from all channels are properly arranged; and a properly arranged analogue of Equation (2) is given, in the same form as before, by

$$\underline{Y} = X\underline{\beta} + \underline{\epsilon}. \tag{20}$$

A proper arrangement of the data from all channels has to be considered in determining the form of Σ and in estimating its necessary variances and covariances for IWLS. In particular, the requirement of nonstationary, uncorrelated noise for this proper arrangement means that not only ϵ_{ki} and ϵ_{kj} (i.e., observations of the noise at any two times within any channel) must be uncorrelated, but also that ϵ_{hi} and ϵ_{kj} (i.e., observations of the noise at any two times in any two channels) must be uncorrelated.

One may view the $p < \sum_{k=1}^{q} p_k$ resulting columns of X as p vectors $\underline{X}_j$ of $n = \sum_{k=1}^{q} n_k$ observations X_{ji} on the p component signals $X_j(t)$, the p elements of $\underline{\beta}$ as the p coefficients β_j of $X_j(t)$, the n elements of $\underline{Y}$ as n observations Y_i on the composite signal Y(t), the n elements of $\underline{\epsilon}$ as n observations ϵ_i on the noise $\epsilon(t)$, the n^2 elements of Σ as the n variances σ_i^2 of ϵ_i and n^2-n covariances σ_{hi} of ϵ_h and ϵ_i and Equation (20) as the matrix alter ego of Equation (1).

*Dr. J. C. Pinson proposed the alinement as a simple way to utilize the pertinent columns of X_{k_1}, X_{k_2}, ..., X_{k_r}.

5. PARTITIONING, SELECTING AND APPORTIONING*

For exposition purposes, it is desirable: to change the number of linearly independent component signals and corresponding unknown coefficients in the linear model to m; to redesignate them by $Z_1(t)$, $Z_2(t)$, ..., $Z_m(t)$ and α_1, α_2, ..., α_m; and to transform $Z_k(t)$, via multiplication by a positive scale factor, so that the units of $Z_k(t)$ will become the units of $Y(t)$ and α_k will become unit-less for k = 1, 2, ..., m. This transformation will be compensated for in the technique described.

Let $Z_{ki} = Z_k(t_i)$ for i = 1, 2, ..., n and k = 1, 2, ..., m; α_1, α_2, ..., α_m vary randomly over the ensemble of all possible states of nature with means α_1^*, α_2^*, ..., α_m^* and covariance matrix Γ^* of variances $\gamma_k^{*2} = \gamma_{kk}^*$ and covariances γ_{hk}^*; and

$$\underline{Z}_k = \begin{bmatrix} Z_{k1} \\ Z_{k2} \\ \vdots \\ Z_{kn} \end{bmatrix} \qquad \text{for } k = 1, 2, \ldots, m,$$

$$Z = (\underline{Z}_1, \underline{Z}_2, \ldots, \underline{Z}_m),$$

$$\underline{\alpha} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{bmatrix}$$

*The suggestion to investigate the feasibility of, and devise an analysis for, such a technique was made by Mr. H. J. Goldfisher and Mr. L. H. Pinson.

$$\underline{\alpha}^* = \begin{bmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_m^* \end{bmatrix}$$

and

$$\Gamma^* = \begin{bmatrix} \gamma_{11}^* & \gamma_{12}^* & \cdots & \gamma_{1m}^* \\ \gamma_{12}^* & \gamma_{22}^* & \cdots & \gamma_{2m}^* \\ \vdots & \vdots & & \vdots \\ \gamma_{1m}^* & \gamma_{2m}^* & \cdots & \gamma_{mm}^* \end{bmatrix}$$

Then the linear model, its matrix alter ego and the corresponding a priori information become $\sum_{k=1}^{m} \alpha_k Z_k(t)$, $\sum_{k=1}^{m} \alpha_k Z_k = Z\underline{\alpha}$ and $\underline{\alpha}^*$ and Γ^*.

When m is large and/or some of $Z_1(t), Z_2(t), \ldots, Z_m(t)$ are highly-related consideration of accuracy and ease of computation frequently dictates that $\sum_{k=1}^{m} \alpha_k Z_k(t)$ be replaced by a new linear model containing only representative component signals which are highly-descriptive, but not highly-related. Suppose that the $p < m$ representative component signals, whose units are the units of $Y(t)$, and corresponding unit-less unknown coefficients are $X_1(t), X_2(t), \ldots, X_p(t)$ and $\beta_1, \beta_2, \ldots, \beta_p$; $X_{ji} = X_j(t_i)$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$; $\beta_1, \beta_2, \ldots, \beta_p$ vary randomly over the ensemble of all possible states of nature with means $\beta_1^*, \beta_2^*, \ldots, \beta_p^*$ and covariance matrix Γ of variances $\gamma_j^2 = \gamma_{jj}$ and covariances γ_{hj}; and

$$\underline{X}_j = \begin{bmatrix} X_{j1} \\ X_{j2} \\ \vdots \\ X_{jn} \end{bmatrix} \quad \text{for } j = 1, 2, \ldots, p,$$

$$X = (\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p)$$

$$\underline{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$\underline{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \\ \vdots \\ \beta_p^* \end{bmatrix}$$

and

$$\Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1p} \\ \gamma_{12} & \gamma_{22} & \cdots & \gamma_{2p} \\ \vdots & \vdots & & \vdots \\ \gamma_{1p} & \gamma_{2p} & \cdots & \gamma_{pp} \end{bmatrix}$$

Hence the new linear model, its matrix alter ego and the corresponding a priori information are $\sum_{j=1}^{p} \beta_j X_j(t)$, $\sum_{j=1}^{p} \beta_j \underline{X}_j = X\underline{\beta}$ and $\underline{\beta}^*$ and Γ

One manner of replacing $\sum_{k=1}^{m} \alpha_k Z_k(t)$ by $\sum_{j=1}^{p} \beta_j X_j(t)$, and in particular $Z\underline{\alpha}$ by $X\underline{\beta}$, is to partition $Z_1(t)$, $Z_2(t)$, ..., $Z_m(t)$ into subsets S_1, S_2, ..., S_p of highly-related $Z_k(t)$'s and to select an $X_j(t)$ to represent S_j (i.e., each $Z_k(t)$ in S_j) for $j = 1, 2, \ldots, p$. In addition, it is desirable to be able to apportion an estimator of each β_j among the α_k's which correspond to the $Z_k(t)$'s in S_j.

A measure of the degree of linear dependence between $Z_h(t)$ and $Z_k(t)$ is needed to accomplish this partitioning. One such measure, based upon $\underline{Z}_h$ and $\underline{Z}_k$, is the cosine

$$d_{hk} = d_{kh} = \frac{\sum_{i=1}^{n} Z_{hi} Z_{ki}}{\left[\left(\sum_{i=1}^{n} Z_{hi}^2\right)\left(\sum_{i=1}^{n} Z_{ki}^2\right)\right]^{1/2}} = \frac{\underline{Z}_h' \underline{Z}_k}{\left[(\underline{Z}_h' \underline{Z}_h)(\underline{Z}_k' \underline{Z}_k)\right]^{1/2}} \quad (21)$$

of the angle $D_{hk} = D_{kh}$ between $\underline{Z}_h$ and $\underline{Z}_k$. If

$$\overline{Z}_k = \frac{1}{n} \sum_{i=1}^{n} Z_{ki} \quad (22)$$

and

$$\underline{\bar{Z}}_k = \begin{bmatrix} \bar{Z}_k \\ \bar{Z}_k \\ \vdots \\ \bar{Z}_k \end{bmatrix} \qquad \text{for } k = 1, 2, \ldots, m, \tag{23}$$

then the sample correlation coefficient

$$r_{hk} = r_{kh} = \frac{\sum_{i=1}^{n} (Z_{hi} - \bar{Z}_h)(Z_{ki} - \bar{Z}_k)}{\left\{\left[\sum_{i=1}^{n} (Z_{hi} - \bar{Z}_h)^2\right]\left[\sum_{i=1}^{n} (Z_{ki} - \bar{Z}_k)^2\right]\right\}^{1/2}} \tag{24}$$

$$\frac{(\underline{Z}_h - \underline{\bar{Z}}_h)'(\underline{Z}_k - \underline{\bar{Z}}_k)}{\left\{\left[(\underline{Z}_h - \underline{\bar{Z}}_h)'(\underline{Z}_h - \underline{\bar{Z}}_h)\right]\left[(\underline{Z}_k - \underline{\bar{Z}}_k)'(\underline{Z}_k - \underline{\bar{Z}}_k)\right]\right\}^{1/2}}$$

between $\underline{Z}_h$ and $\underline{Z}_k$ is the cosine of the angle between $\underline{Z}_h - \underline{\bar{Z}}_h$ and $\underline{Z}_k - \underline{\bar{Z}}_k$ and also a measure of the degree of linear dependence between $Z_h(t)$ and $Z_k(t)$. Observe that the multiplication of $Z_h(t)$ by c_h and $Z_k(t)$ by c_k does not change the magnitude of d_{hk} and r_{hk} and changes the sign of d_{hk} and r_{hk} only for c_h and c_k having different signs. It might be demonstrated that any three of the d_{hk}'s and any three of the r_{hk}'s satisfy the inequalities

$$-1 < d_{hi}\, d_{ki} - \left[(1 - d_{hi}^2)(1 - d_{ki}^2)\right]^{1/2} \leq d_{hk} \leq d_{hi}\, d_{ki} + \left[(1 - d_{hi}^2)(1 - d_{ki}^2)\right]^{1/2} < 1 \tag{25}$$

and

$$-1 < r_{hi}\, r_{ki} - \left[(1 - r_{hi}^2)(1 - r_{ki}^2)\right]^{1/2} \leq r_{hk} \leq r_{hi}\, r_{ki} + \left[(1 - r_{hi}^2)(1 - r_{ki}^2)\right]^{1/2} < 1. \tag{26}$$

Consequently, it is feasible to utilize either d_{hk} or r_{hk} in a partitioning procedure for obtaining $S_1, S_2, \ldots, S_p$ from $Z_1(t), Z_2(t), \ldots, Z_m(t)$. The pertinent one to use is d_{hk} when a constant term is not included in the linear model and r_{hk} when a constant term is included in the linear model. Without loss of generality, d_{hk} will be used in the text; but it should be replaced by r_{hk} where appropriate.

A reasonable partitioning procedure should not be affected by a renumbering of $Z_1(t), Z_2(t), \ldots, Z_m(t)$. The procedure also ought to yield $S_1, S_2, \ldots, S_p$ with a relatively high degree of linear dependence existing between any two $Z_k(t)$'s in the same S_j and a relatively low degree of linear dependence existing between any two $Z_k(t)$'s in different S_j's. Finally, it should provide some control over the size of p.

The partitioning procedure employed by EIWLS:*

1. Computes

$$d_k = \sum_{h=1}^{k-1} |d_{hk}| + \sum_{h=k+1}^{m} |d_{hk}| = \sum_{h=1}^{m} |d_{hk}| - 1 \qquad (27)$$

for $k = 1, 2, \ldots, m$.

2. Selects that $Z_k(t)$, say $Z_{k_1}(t)$, whose d_k is smallest.

3. Puts $Z_k(t)$ into S_1 if and only if $|d_{kk_1}| \geq d$, where $\cos 45^\circ = 0.70711 \leq d < 1$ is a preassigned constant which should be selected to reflect the desire for the existence of a relatively high degree of linear dependence between any two $Z_k(t)$'s in the same S_j and a relatively low degree of linear dependence between any two $Z_k(t)$'s in different S_j's and the desire for a relatively small p.

4. Selects that $Z_k(t)$ not in S_1, say $Z_{k_2}(t)$, whose d_k is smallest among the $Z_k(t)$'s not in S_1.

5. For $Z_k(t)$ not in S_1, puts it into S_2 if and only if $|d_{kk_2}| > d$, and for $Z_k(t)$ in S_1, removes it from S_1 and puts it into S_2 if and only if $|d_{kk_2}| > |d_{kk_1}|$.

---

*Mr. P. L. Hsu developed the details of this procedure.

6. Continues in the same manner and selects that $Z_k(t)$ not in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}, say $Z_{k_j}(t)$, whose d_k is smallest among the $Z_k(t)$'s not in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1} .

7. For $Z_k(t)$ not in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}, puts it into S_j if and only if $|d_{kk_j}| \geq d$; and for $Z_k(t)$ in $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1}, removes it from $S_1, S_2, \ldots, S_{j-2}$ or S_{j-1} and puts it into S_j if and only if $|d_{kk_j}| > \max_{h=1,2,\ldots,j-1} |d_{kk_h}|$.

8. Continues in the same manner until selecting a $Z_k(t)$ not in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1}, say $Z_{k_p}(t)$, for which $|d_{kk_p}| \geq d$ for all the $Z_k(t)$'s not in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1} and, thereby, defining p .

9. For $Z_k(t)$ not in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1}, puts it into S_p; and for $Z_k(t)$ in $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1}, removes it from $S_1, S_2, \ldots, S_{p-2}$ or S_{p-1} and puts it into S_p if and only if $|d_{kk_p}| > \max_{j=1,2,\ldots,p-1} |d_{kk_j}|$.

10. Inspects the resulting $S_1, S_2, \ldots, S_p$ to determine if $|d_{hk}|$ is sufficiently large for all $Z_h(t)$ and $Z_k(t)$ in the same S_j, if $|d_{hk}|$ is sufficiently small for all $Z_h(t)$ and $Z_k(t)$ in different S_j's and if p is sufficiently small .

11. Modifies $S_1, S_2, \ldots, S_p$ and d until 10 is satisfied to a reasonable extent.

It is notationally convenient to define $0 = m_0 < m_1 < m_2 < \ldots < m_{p-1} < m_p = m$ and renumber $Z_1(t), Z_2(t), \ldots, Z_m(t)$ with $Z_{m_{j-1}+1}(t), Z_{m_{j-1}+2}(t), \ldots, Z_{m_j}(t)$ being in S_j for $j = 1, 2, \ldots, p$. Using Equation (25), it might also be demonstrated that each $Z_k(t)$ in S_j may be transformed, via multiplication by $d_{m_j k}/|d_{m_j k}|$, so that d_{hk} becomes positive when $Z_h(t)$ and $Z_k(t)$ are in S_j for $j = 1, 2, \ldots, p$. This transformation will also be compensated for in the ensuing discussion.

The replacement of $\sum_{k=1}^{m} \alpha_k Z_k(t)$ by $\sum_{j=1}^{p} \beta_j X_j(t)$ means that $X_1(t), X_2(t), \ldots, X_p(t)$ should be selected to satisfy

$$\sum_{j=1}^{p} \beta_j X_j(t) = \sum_{k=1}^{m} \alpha_k Z_k(t). \qquad (28)$$

In addition, the representation of S_j by $X_j(t)$ suggests that $X_j(t)$ should be selected to satisfy

$$\beta_j X_j(t) = \sum_{k=m_{j-1}+1}^{m_j} \alpha_k Z_k(t) \text{ and} \tag{29}$$

$$X_j(t) = \sum_{k=m_{j-1}+1}^{m_j} v_k Z_k(t), \tag{30}$$

where $v_{m_{j-1}+1}, v_{m_{j-1}+2}, \ldots, v_{m_j}$ are non-negative weights whose sum is one, for $j = 1, 2, \ldots, p$. Simple analytic consequences of the validity of Equations (29) and (30) for all values of t are

$$\alpha_k = v_k \beta_j \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j \text{ and} \tag{31}$$

$$\beta_j = \sum_{k=m_{j-1}+1}^{m_j} \alpha_k \text{ for } j = 1, 2, \ldots, p. \tag{32}$$

Regardless of the true relationships (or lack of them) among $\alpha_{m_{j-1}+1}, \alpha_{m_{j-1}+2}, \ldots, \alpha_{m_j}$, the selecting of $X_j(t)$ to represent S_j (i.e., Equations (29) and (30)) induces analytic relationships among $\alpha_{m_{j-1}+1}, \alpha_{m_{j-1}+2}, \ldots, \alpha_{m_j}$ and β_j (i.e., Equations (31) and (32)) which cause them to become analytically and probabilistically indistinguishable (i.e., yield all of them once any one of them is determined) for $j = 1, 2, \ldots, p$. The selecting, in turn, induces a web of analytic relationships, which are not difficult to derive, into the structure of the a priori information:

$$\alpha_k^* = v_k \beta_j^* \text{ for } k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j; \tag{33}$$

$$\beta_j^* = \sum_{k=m_{j-1}+1}^{m_j} \alpha_k^*; \tag{34}$$

$$\gamma_k^{*2} = \gamma_{kk}^{*} = v_k^2 \gamma_j^2 \text{ for } k = m_{j-1}+1,\ m_{j-1}+2,\ \ldots,\ m_j\ ; \qquad (35)$$

$$\gamma_{hk}^{*} = v_h v_k \gamma_j^2 = \gamma_h^{*} \gamma_k^{*} \text{ for } k = h+1,\ h+2, \ldots, m_j \text{ and}$$

$$h = m_{j-1}+1,\ m_{j-1}+2,\ \ldots,\ m_j-1\ ; \text{ and} \qquad (36)$$

$$\gamma_j^2 = \left(\sum_{k=m_{j-1}+1}^{m_j} \gamma_k^{*} \right)^2 \text{ for } j = 1,\ 2,\ \ldots,\ p. \qquad (37)$$

Consideration of this structure produces*

$$v_k = \gamma_k^{*} \Big/ \sum_{h=m_{j-1}+1}^{m_j} \gamma_h^{*} = \gamma_k^{*}/\gamma_j \text{ for}$$

$$k = m_{j-1}+1,\ m_{j-1}+2,\ \ldots,\ m_j \text{ and } j = 1,\ 2,\ \ldots,\ p \qquad (38)$$

as the natural set** of weights to use in Equations (30) and (31). It follows immediately that

$$\frac{\alpha_k - \alpha_k^{*}}{\gamma_k^{*}} = \frac{\beta_j - \beta_j^{*}}{\gamma_j}\ ;$$

and so α_k and β_j occur with equal probability when α_k, hence β_j, is normally distributed for $k = m_{j-1}+1,\ m_{j-1}+2,\ \ldots,\ m_j$ and $j = 1,\ 2,\ \ldots,\ p$.

---

*These weights were proposed by Mr. C. M. Shipplett.

**An alternative set, which assigns equal weight to $Z_{m_{j-1}+1}$, $Z_{m_{j-1}+2}$, $\ldots$, Z_{m_j} in Equation (30) and to $\alpha_{m_{j-1}+1}$, $\alpha_{m_{j-1}+2}$, $\ldots$, α_{m_j} in Equation (31), is given by $v_k = \dfrac{1}{m_j - m_{j-1}}$ for $k = m_{j-1}+1$, $m_{j-1}+2,\ \ldots,\ m_j$ and $j = 1,\ 2,\ \ldots,\ p$.

Hence, the selecting of $X_1(t)$, $X_2(t)$, ..., $X_p(t)$ from $Z_1(t)$, $Z_2(t)$, ..., $Z_m(t)$ is obtained by

$$X_j(t) = \sum_{k=m_{j-1}+1}^{m_j} \left(\frac{\gamma_k^*}{\sum_{h=m_{j-1}+1}^{m_j} \gamma_h^*} \right) Z_k(t) \text{ for } j = 1, 2, \ldots, p; \quad (39)$$

the selecting of $\underline{X}_1$, $\underline{X}_2$, ..., $\underline{X}_p$ from $\underline{Z}_1$, $\underline{Z}_2$, ..., $\underline{Z}_m$ is accomplished by

$$X_{ji} = \sum_{k=m_{j-1}+1}^{m_j} \left(\frac{\gamma_k^*}{\sum_{h=m_{j-1}+1}^{m_j} \gamma_h^*} \right) Z_{ki} \text{ for}$$

$$i = 1, 2, \ldots, n \text{ and } j = 1, 2, \ldots, p, \quad (40)$$

in particular; and the apportioning of an estimator $\hat{\beta}_j$ (e.g., $\hat{\beta}_j(0)$, $\hat{\beta}_j(c^*)$ or $\hat{\beta}_j^*$) of β_j among $\alpha_{m_{j-1}+1}$, $\alpha_{m_{j-1}+2}$, ..., α_{m_j} is provided by

$$\hat{\alpha}_k = \left(\frac{\gamma_k^*}{\sum_{h=m_{j-1}+1}^{m_j} \gamma_h^*} \right) \hat{\beta}_j \text{ for}$$

$$k = m_{j-1}+1, m_{j-1}+2, \ldots, m_j \text{ and } j = 1, 2, \ldots, p. \quad (41)$$

The combination of the two previous transformations to each $Z_k(t)$ in S_j (i.e., its multiplication by the product of the original $d_{m_j k}/|d_{m_j k}|$ and the corresponding positive scale factor) may now be compensated for by performing the same combination of transformations to the corresponding $\hat{\alpha}_k$ (i.e., its multiplication by the product of the original $d_{m_j k}/|d_{m_j k}|$ and the corresponding positive scale factor). Indeed, it might be proved that these transformations may

be eliminated by replacing v_k in Equations (30), (31), (33), (35), (36), (39), (40) and (41) with the product of the original $d_{m_jk}/|d_{m_jk}|$, the corresponding positive scale factor and v_k.

6. DISCUSSION

A procedure has been described for estimation of the coefficients for linearly independent component signals in a linear model from observations on the component signals and a composite signal containing the linear model plus noise which is nonstationary and/or correlated with unknown covariance matrix, when the coefficients may vary randomly over all possible states of nature and they may appear in the linear models for more than one data channel, and when there may be a large number of component signals in the linear model and some of them may be highly-related. The procedure, denoted by EIWLS: properly arranges the coefficients and data from all channels; partitions the resulting set of component signals into subsets of highly-related ones and selects a representative component signal from each subset to include in the new linear model; obtains the least squares estimators of the coefficients in the new linear model, and then iteratively calculates an estimator of the noise covariance matrix and obtains the weighted least squares estimators of these coefficients which are determined by the inverse of the matrix estimator; incorporates the a priori information into estimators of the coefficients in the new linear model; and apportions an estimator for the coefficient of the representative component signal for a subset among the coefficients of component signals in that subset.

During the development of EIWLS, the difficulty of the generalized statistical regression problem and the requirement for an operational analysis within a reasonable period of time combined to dictate that the security of rigorous theoretical proof be occasionally sacrificed for the expedience of apparent theoretical implication plus intuitive justification. The computer provided a feasible means to implement the procedure and to empirically test it on examples, when the noise is nonstationary and uncorrelated. Empirical testing of the procedure and its parts has included not only the analysis of fabricated examples (e.g., those in Section 8 and Ref 1), but also the successful evaluation of inertial navigation systems via the analysis of real field-test data. Therefore, EIWLS is presented as a theoretically promising, intuitively appealing and empirically tested (to a limited extent) solution for the generalized statistical regression problem.

The feasibility of IWLS depends upon the practical inversion of the estimators $\hat{\Sigma}^{(c)}$ of the noise covariance matrix Σ. This inversion has to be investigated in the case of correlated noise. When the noise is stationary and correlated and the observations are taken at different and equally-spaced times, the form of $\hat{\Sigma}^{(c)}$ is particularly simple and inversion may not be too difficult. A simplifying assumption, which is tenable in a large number of scientific problems, is that the noise covariance $\sigma(t, t')$ becomes zero as the time difference $|t-t'|$ increases beyond a certain limit. Then $\hat{\Sigma}^{(c)}$ has blocks of zero elements and inversion by partitioning may be practical.

Consider the IWLS estimators $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ of the coefficients $\beta_1, \beta_2, \ldots, \beta_p$ for $c = 1, 2, \ldots, c^*$. In order to provide estimators of $\beta_1, \beta_2, \ldots, \beta_p$ of sufficient stability for a particular application, $\hat{\beta}_1^{(c)}, \hat{\beta}_2^{(c)}, \ldots, \hat{\beta}_p^{(c)}$ need not converge to $\hat{\beta}_1^{(c^*)}, \hat{\beta}_2^{(c^*)}, \ldots, \hat{\beta}_p^{(c^*)}$, but only possess a valid measure of change (e.g., $\max_{j=1,2,\ldots,p} \left| (\hat{\beta}_j^{(c)} - \hat{\beta}_j^{(c-1)})/\hat{\beta}_j^{(c)} \right|$) which becomes less than a particular preassigned constant for $c = c_o, c_o+1, \ldots, c^*$. Suppose that $\left| (\hat{\beta}_j^{(c)} - \hat{\beta}_j^{(c-1)})/\hat{\beta}_j^{(c)} \right| \leq b_j$ for $c = c_j, c_j+1, \ldots, c^*$. Then $\hat{\beta}_j^{(c_j)}, \hat{\beta}_j^{(c_j+1)}, \ldots, \hat{\beta}_j^{(c^*)}$ appear to be approaching or oscillating in a band (whose width depends upon b_j) about an asymptotic value and could probably be combined, by some technique (e.g., extrapolation), to yield an improved estimator of β_j. This improved estimation of β_j should be investigated. A measure of the inherent accuracy in the estimation of β_j, which depends upon the matrix X of component signal observations and the vector $\underline{Y}$ of composite signal observations (and, implicitly, the vector $\underline{\epsilon}$ of noise observations) is provided by b_j and c_j. Note that there may be a considerable variation among the inherent accuracy in the estimation of $\beta_1, \beta_2, \ldots, \beta_p$.

At each iteration, the guiding principle of IWLS is to provide the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ and then Σ, based upon the available information at that iteration concerning Σ and then $\beta_1, \beta_2, \ldots, \beta_p$. This principle is similar to Bellman's principle of optimality in dynamic programming (Ref 8, page 83), which states that "an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." In addition, IWLS extracts considerably more information from the data than does least squares and may be termed an "estimation servomechanism" or an "adaptive estimation procedure."

It is not at all unreasonable to conjecture that IWLS will produce, except perhaps under pathological conditions, near-optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ and Σ. However, the stability and estimation properties of IWLS should be further investigated, both theoretically and empirically.

It is informative to elaborate upon two previously mentioned characteristics of the optimum estimators for $\beta_1, \beta_2, \ldots, \beta_p$. The first characteristic is that they minimize the appropriate statistical distance $(Q_{\Sigma^{-1}})^{1/2}$ between the data vector $\underline{Y}$ and the estimator of the matrix alter ego for the linear model $X\underline{\beta}$. One has no knowledge whatsoever of the difference between the linear model $\sum_{j=1}^{p} \beta_j X_{ji}$ at time t_i and its resulting estimator or of the difference between β_j and its estimator, but only knowledge of the probabilistic properties of these differences. The second characteristic is that for the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$, each of these differences has mean zero and minimum variance among all corresponding differences resulting from unbiased linear estimators of $\beta_1, \beta_2, \ldots, \beta_p$. In other words, the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ do not always produce an estimator which is closest to the quantity being estimated; but they do produce an estimator whose probability distribution is more tightly spread about this quantity than the probability distribution of any estimator resulting from unbiased linear estimators (e.g., the least squares estimators) of $\beta_1, \beta_2, \ldots, \beta_p$.

The three sets of estimators of $\beta_1, \beta_2, \ldots, \beta_p$ that incorporate the a priori information, one set given by Equation (16) and the other two sets listed in the footnote on page 18, have been empirically tested on an example for which $\beta_1, \beta_2, \ldots, \beta_p$ had mean zero and covariance matrix $\Gamma = \text{diag}\,(\gamma_1^2, \gamma_2^2, \ldots, \gamma_p^2)$ of variances $\gamma_1^2, \gamma_2^2, \ldots, \gamma_p^2$ that varied. Although this testing provided empirical justification for the use of Equation (16), the results were somewhat inconclusive; since the version of the scalar weight presented in that footnote was too insensitive to the variation in $\gamma_1^2, \gamma_2^2, \ldots, \gamma_p^2$ and required an extreme variation in them to itself vary from zero to one. Additional empirical testing of the three sets of estimators that incorporate the a priori information, utilizing a more satisfactory version of λ, is needed. The improved estimators of $\beta_1, \beta_2, \ldots, \beta_p$ and the appropriately modified version of $\hat{\Sigma}^{(c^*)}$, rather than $\hat{\beta}_1^{(c^*)}, \hat{\beta}_2^{(c^*)}, \ldots, \hat{\beta}_p^{(c^*)}$ and $\hat{\Sigma}^{(c^*)}$, should be used to calculate these estimators.

Since the coefficients $\alpha_{m_{j-1}+1}, \alpha_{m_{j-1}+2}, \ldots, \alpha_{m_j}$ of the component signals in the j^{th} subset S_j become analytically and probabilistically indistinguishable from the coefficient β_j of the representative component signal X_j for S_j for $j = 1, 2, \ldots, p$ during selecting and apportioning, the procedure used to partition the component signals $Z_1(t), Z_2(t), \ldots, Z_m(t)$ into subsets $S_1, S_2, \ldots, S_p$ is very important. Hence, the possibility of improving upon the partitioning procedure employed by EIWLS ought to be investigated. One approach would be to define a measure of the quality of a partition and to choose that procedure which maximized this measure.

7. SURVEY OF RELATED LITERATURE

In order to place EIWLS in proper perspective, a survey of related literature is now presented. This literature is grouped, for convenience, into seven categories of articles: those treating the properties of the least squares estimators, a set of weighted least squares estimators or the optimum estimators (Ref 9-16); those treating estimation of residuals or iterative estimation procedures (Ref 17-26); those treating estimation procedures for correlated noise (Ref 27-30); those treating estimation procedures which incorporate a priori information (Ref 31-36); and those treating procedures for replacing the linear model by a new linear model and estimating the coefficients in the new linear model (Ref 4, 21, 34, 35, 37, 38, 39 and 40).

Grenander and Rosenblatt (Ref 9, Sections 7.1 - 7.4) derive important asymptotic properties of the least squares estimators and the optimum estimators of the coefficients $\beta_1, \beta_2, \ldots, \beta_p$ when the noise is stationary. The relationships among the columns of the component signal matrix X and the eigenvectors of the noise covariance matrix Σ and the conditions on the eigenvalues of Σ, required for the least squares estimators and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ to be equal, are obtained by Muller and Watson (Ref 10). Using the eigenvalues of certain matrices, Magness and McGuire (Ref 11) compare the covariance matrices of the least squares estimators and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$; and Golub (Ref 12) compares the covariance matrices of a set of weighted least squares estimators and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$.

Zyskind (Ref 13) and Watson (Ref 14) discuss the estimation problem in general; and necessary and sufficient conditions for the least squares and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ to be equal, or effectively equal, in particular. The contribution of errors in estimating weights to the variances of $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p$, for nonstationary and uncorrelated noise and for a special case of correlated noise, is treated by Williams (Ref 15). Both Williams (Ref 15) and McElroy (Ref 16) state necessary and sufficient conditions for the least squares and the optimum estimators of $\beta_1, \beta_2, \ldots, \beta_p$ to be equal. The above asymptotic properties, relationships, conditions and comparisons should be useful in the investigation of procedures such as IWLS.

Material relevant to the estimation of regression residuals, and thereby Σ, is presented by Thiel (Ref 17 and 18) and Koerts (Ref 19). When the noise is nonstationary and uncorrelated with variance $\sigma^2(t) = \left[\sum_{j=1}^{p} \beta_j X_j(t) \right]^r \sigma^2$, Prais and Aitchison (Ref 20) propose an iterative estimation procedure that is IWLS with the appropriate estimator $\hat{\Sigma}^{(c)}$ of the noise covariance matrix Σ; and Fisher (Ref 21) proposes two iterative solutions (modified forms of Newton's method of approximation) to the maximum likelihood equations resulting from normally distributed noise. Also analogous to IWLS is an iterative estimation procedure described by Turner, Monroe and Lucas (Ref 22) for the case of the linear model being replaced by the quotient of two polynomials in the component signal X(t) and the noise being stationary and uncorrelated. Mandel (Ref 23) develops an iterative estimation procedure, closely related to IWLS, to treat the linear model $\beta_1 + \beta_2 X(t)$ with nonstationary, uncorrelated noise. The convergence properties of $\hat{\underline{\beta}}$ are treated by Telser (Ref 24).

Iterative estimation procedures which iterate over time, combining past data and estimators with new data, have appeared for some time in the engineering literature. Two early examples of such procedures are contained in Ref 25 and 26.

An excellent survey of articles concerned with estimation (and hypothesis testing) procedures for correlated noise is presented by Anderson (Ref 27). In the event of a linear model $\beta_1 + \beta_2 X(t)$ and stationary, correlated noise that is normally distributed with

$$\Sigma = \begin{bmatrix} \sigma^2 & \sigma^2\rho & \sigma^2\rho^2 & \cdots & \sigma^2\rho^{n-1} \\ \sigma^2\rho & \sigma^2 & \sigma^2\rho & \cdots & \sigma^2\rho^{n-2} \\ \sigma^2\rho^2 & \sigma^2\rho & \sigma^2 & \cdots & \sigma^2\rho^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \sigma^2\rho^{n-1} & \sigma^2\rho^{n-2} & \sigma^2\rho^{n-3} & \cdots & \sigma^2 \end{bmatrix}$$

Murthy (Ref 28) discusses an iterative solution to the maximum likelihood equations and an explicit criterion for its convergence. Goodman (Ref 29) suggests a noniterative procedure, which performs the estimation in the frequency domain rather than in the time domain, when the noise is stationary and correlated. The estimation problem, with a system of linear models for several data channels and noise correlated over both channels and time, is covered by Parks (Ref 30).

From two different points of view, Raiffa and Schlaifer (Ref 31, Sections13.2 - 13.7) and Theil (Ref 32) cover the incorporation of a priori information into the estimation procedure for stationary, uncorrelated noise which is normally distributed with variance σ^2 either known or unknown. Gunckel (Ref 33) incorporates a priori information into estimators that reduce to the estimators $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*$ of $\beta_1, \beta_2, \ldots, \beta_p$ in Equation (16) when the noise has (mean zero and) covariance matrix $\hat{\Sigma}(c^*)$. The incorporation of a priori information described by Drucker (Ref 34) is in essentially the same form as $\hat{\underline{\beta}}^* = (\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_p^*)'$. Chipman (Ref 35) presents material concerning the properties of $\hat{\underline{\beta}}^*$ (plus perhaps $\hat{\underline{\beta}}^{\circledast}$ and $\hat{\underline{\beta}}^{\blacktriangle}$), as well as to partitioning, selecting and apportioning. Judge and Takayama (Ref 36) treat inequality restrictions, for incorporating a priori information into the estimation procedure.

Efroymson (Ref 4) proposes a stepwise selection of representative component signals via selection of: the best component signal $Z_k(t)$ to use as the representative component signal $X_1(t)$ in the linear model $\beta_1 X_1(t)$; the best $Z_k(t)$ to use with $X_1(t)$ as the representative component signal $X_2(t)$ in the linear model $\beta_1 X_1(t) + \beta_2 X_2(t)$; . . . ; and the best $Z_k(t)$ to use with $X_1(t)$, $X_2(t)$, . . . , $X_{p-1}(t)$ as the representative component signal $X_p(t)$ in the linear model $\sum_{j=1}^{p} \beta_j X_j(t)$. A selection of representative component signals via an iterative nomination, of the best subset of p-p' $Z_k(t)$'s to use with a subset $X_1(t)$, $X_2(t)$, . . . , $X_{p'}(t)$ of p' $Z_k(t)$'s as the representative component signals $X_{p'+1}(t)$, $X_{p'+2}(t)$, . . . , $X_p(t)$ in the linear model $\sum_{j=1}^{p} \beta_j X_j(t)$ and then of the best subset of p' $Z_k(t)$'s to use with that subset $X_{p'+1}(t)$, $X_{p'+2}(t)$, . . . , $X_p(t)$ of p-p' $Z_k(t)$'s as the representative component signals $X_1(t)$, $X_2(t)$, . . . , $X_{p'}(t)$ in the linear model $\sum_{j=1}^{p} \beta_j X_j(t)$ until two subsets renominate each other, is suggested by Villone, McCornack and Wood (Ref 37). Both of these procedures simultaneously provide least squares estimators of β_1, β_2, . . . , β_p. However, the two procedures which are most similar to partitioning, selecting and apportioning are the procedure for approximating by representative component signals of Fisher (Ref 21) and the procedure for grouping, selecting and apportioning of Drucker (Ref 34). Massy (Ref 38), Fortier and Solomon (Ref 39) and King (Ref 40) also are related to partioning, selecting and apportioning.

With the exception of Ref 4 and 37, this related literature became known to the author only after the development of EIWLS. One may, nevertheless, note the anticipation of: the iterative estimation procedures of Ref 1 (IWLS) and the subsequent Ref 23, by Ref 20-22 and 28; the incorporation of a priori information into the estimation procedures of Ref 1 (EIWLS in summarized form) and the subsequent Ref 34, by Ref 31-33; and the procedure for partitioning, selecting and apportioning of Ref 1 (EIWLS in summarized form) and the subsequent Ref 34, by Ref 21.

8. ILLUSTRATIVE EXAMPLE*

An illustrative example concludes the paper. Although the example is described in detail by the paper's original version (see the footnote on page i), only its input and output are summarized now for simplicity. For the example:

1. The 272 observations on each of the three channel linear models, $M_1(t)$, $M_2(t)$ and $M_3(t)$, are constructed from simultaneous observations on 14 known channel component signals using known channel coefficients.

2. The 272 observations on each of the three independent channel noises, $\epsilon_1(t)$, $\epsilon_2(t)$ and $\epsilon_3(t)$, are generated to be uncorrelated with mean zero and known variance $\sigma^2(t)$.

3. The 272 observations on each of the three channel composite signals, $Y_1(t)$, $Y_2(t)$ and $Y_3(t)$, are constructed by adding the appropriate channel linear model and noise observations.

4. The coefficients and data from all three channels are properly arranged, component signals are partitioned into seven subsets, and seven representative component signals are selected.

5. The coefficients, β_j, in the new linear model are estimated via least squares ($\hat{\beta}_j^{(0)}$), seven iterations of IWLS ($\hat{\beta}_j^{(7)}$), incorporation of a priori information ($\hat{\beta}_j^{*}$), and the optimum weighted least squares (β_{0j}) for $j = 1, 2, \ldots, 7$.

6. The corresponding estimators, $\hat{Y}_k^{(0)}(t)$, $\hat{Y}_k^{(7)}(t)$, $\hat{Y}_k^{*}(t)$ and $\hat{Y}_{0k}(t)$, of $Y_k(t)$ and $M_k(t)$ are computed for $k = 1, 2, 3$.

Graphs of the three channel linear models and composite signals are presented in Figures 1, 2 and 3. Tables 1 and 2 show the actual coefficients in the channel and new linear models and their EIWLS estimators, and a summary of information regarding estimators of coefficients in the new linear model, respectively. The latter indicates that: (1) the seventh IWLS estimators are, mainly, both near-optimum and superior to the least squares estimators; (2) the majority of change in the IWLS estimators has occurred by the third iteration; and (3) the relative change between the sixth

*To implement EIWLS for the example, Mr. P. L. Hsu developed an experimental computer program of exceptional quality.

and seventh IWLS estimators is quite small. Actual comparisons of estimators of the three channel linear models are shown in Figures 4, 5, and 6. The superiority of the IWLS and EIWLS estimators over the least squares estimators is certainly corroborated by Figures 4, 5, and 6.

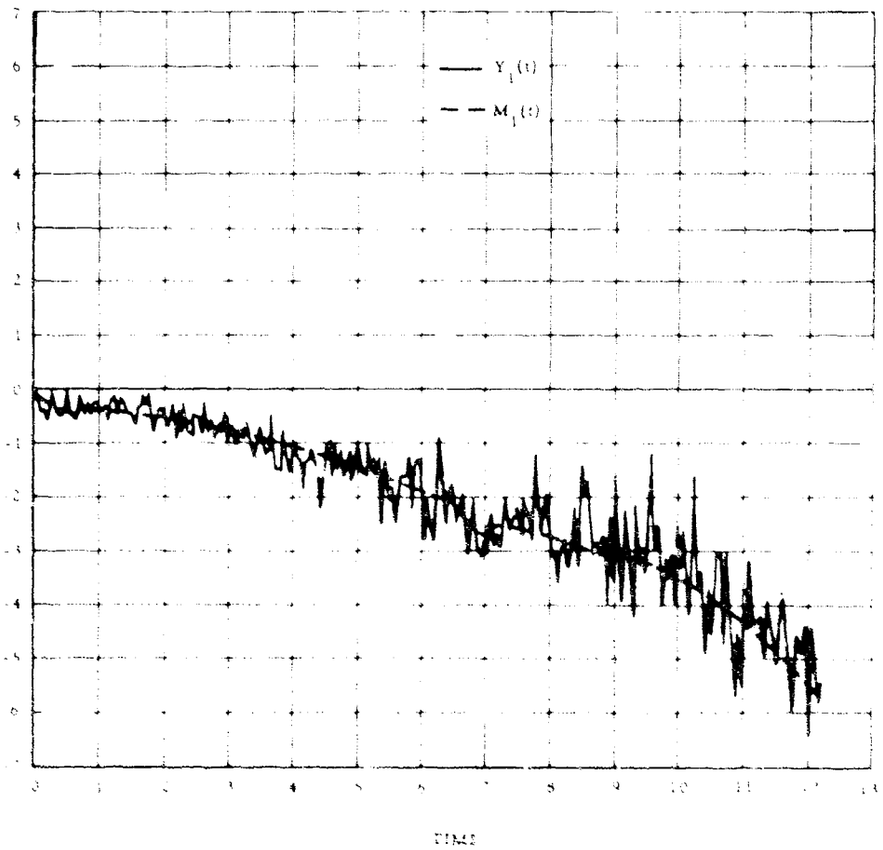

Figure 1. Composite Signal and Actual Linear Model for Channel 1

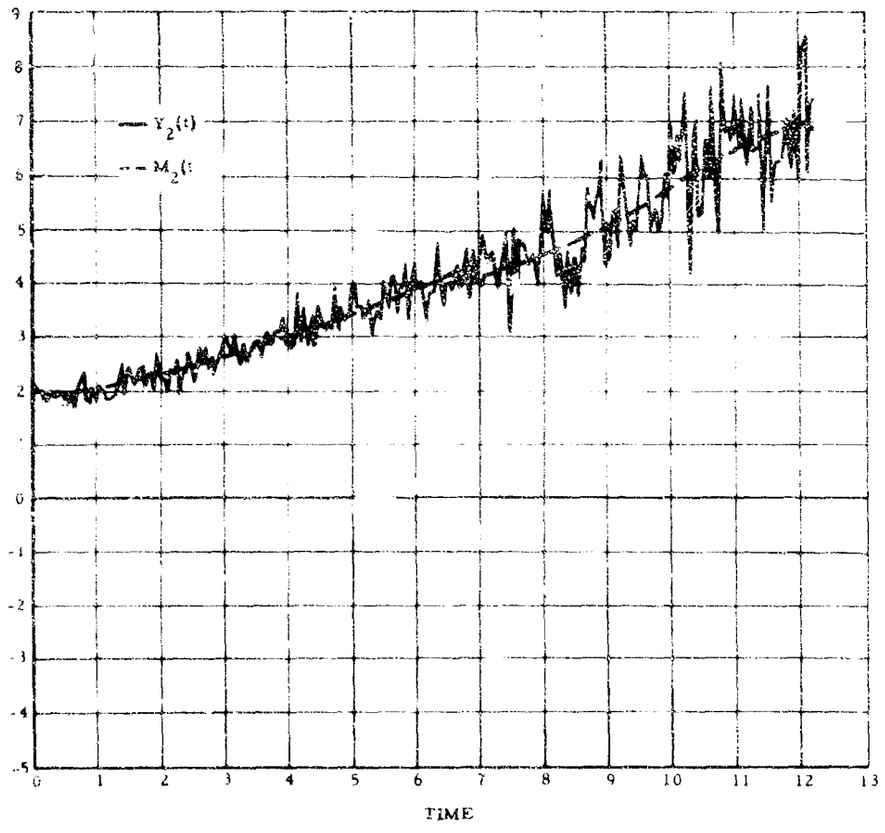

Figure 2. Composite Signal and Actual Linear Model for Channel 2

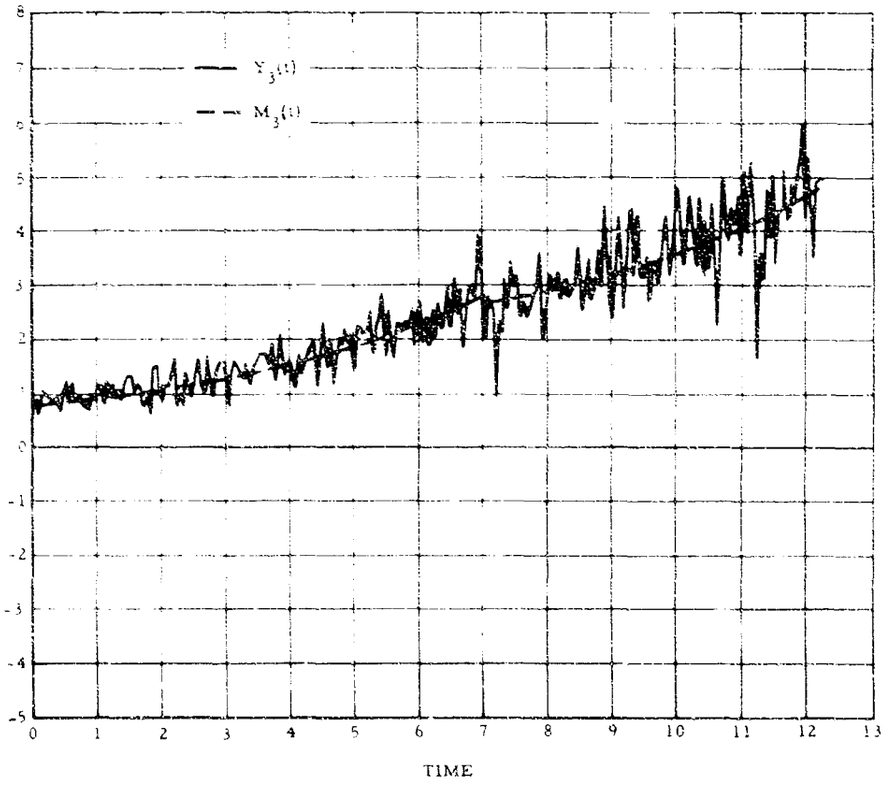

Figure 3. Composite Signal and Actual Linear Model for Channel 3

Table 1. Actual Coefficients in the Channel and New Linear Models and Their EIWLS Estimators

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
						Channel 1								
α_{1j}	-1.100	.095	-1.600	.450	.030	-.710	-.150	-.093	-.029	.060	.027	.014	.018	-2.100
$\hat{\alpha}_{1j}^*$	-7.321	.366	-.732	-.732	.073	-.732	-.058	-.070	-.004	.083	.035	.004	.004	-3.518
						Channel 2								
α_{2j}	-2.200	.020	-.600	-.700	-.030	-.920	-.150	-.093	-.029	.060	.027	.014	.018	-2.100
$\hat{\alpha}_{2j}^*$	-6.698	.340	-.670	-.670	-.034	-.670	-.058	-.070	-.004	.083	.035	.004	.004	-3.518
						Channel 3								
α_{3j}	.900	-.085	.300	.750	-.020	.550	-.150	-.093	-.029	.060	.027	.014	.018	-2.100
$\hat{\alpha}_{3j}^*$	5.846	-.292	.585	.585	-.029	.585	-.058	-.070	-.004	.083	.035	.004	.004	-3.518
						New [1]								
β_j	-2.835	-4.430	2.395	-.150	-2.166	.003	.060							
$\hat{\beta}_j^*$	-9.957	-9.075	7.922	-.058	-3.624	.012	.083							

[1] The association of α's with β's is as follows:

1. $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{16}$ with β_1
2. $\alpha_{21}, \alpha_{22}, \ldots, \alpha_{26}$ with β_2
3. $\alpha_{31}, \alpha_{32}, \ldots, \alpha_{36}$ with β_3
4. $\alpha_{17}, \alpha_{27}, \alpha_{37}$ with β_4
5. $\alpha_{18}, \alpha_{1,11}, \alpha_{1,14}, \alpha_{28}, \alpha_{2,11}, \alpha_{2,14}, \alpha_{38}, \alpha_{3,11}, \alpha_{3,14}$ with β_5
6. $\alpha_{19}, \alpha_{1,12}, \alpha_{1,13}, \alpha_{29}, \alpha_{2,12}, \alpha_{2,13}, \alpha_{39}, \alpha_{3,12}, \alpha_{3,13}$ with β_6
7. $\alpha_{1,10}, \alpha_{2,10}, \alpha_{3,10}$ with β_7

Table 2. Summary of Information Regarding Estimators of Coefficients in the New Linear Model

	β_1	β_2	β_3	β_4	β_5	β_6	β_7
Optimum Estimators $\hat{\beta}_{0j}$	-9.5643	-9.1254	6.9091	-.078207	-3.5744	.028344	.085588
Least Squares Estimators $\hat{\beta}_j^{(0)}$	-9.7123	-7.5915	4.3828	-.169373	-3.5790	.086772	.079524
First IWLS Estimators $\hat{\beta}_j^{(1)}$	-9.9432	-8.7641	7.4216	-.076029	-3.6201	.024685	.082372
Second IWLS Estimators $\hat{\beta}_j^{(2)}$	-10.0185	-8.7293	7.6326	-.072770	-3.6311	.022528	.082753
Third IWLS Estimators $\hat{\beta}_j^{(3)}$	-9.9987	-8.7606	7.6502	-.071504	-3.6295	.021563	.082898
Fourth IWLS Estimators $\hat{\beta}_j^{(4)}$	-10.0012	-8.7444	7.6377	-.072353	-3.6301	.022033	.082857
Fifth IWLS Estimators $\hat{\beta}_j^{(5)}$	-9.9947	-8.7604	7.6455	-.071691	-3.6292	.021609	.082920
Sixth IWLS Estimators $\hat{\beta}_j^{(6)}$	-9.9976	-8.7499	7.6393	-.072178	-3.6298	.021893	.082879
Seventh IWLS Estimators $\hat{\beta}_j^{(7)}$	-9.9943	-8.7590	7.6438	-.071790	-3.6293	.021656	.082915
Estimators $\hat{\beta}_j^*$ that Incorporate A Priori Information	-9.9569	-9.0751	7.9218	-.058035	-3.6235	.011539	.082897
$\left\|(\hat{\beta}_j^{(0)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right\|$	.0155	.1681	.3655	1.1657	.0013	2.0614	.0708
$\left\|(\hat{\beta}_j^{(7)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right\|$	.0450	.0402	.1063	.0820	.0154	.2360	.0312
$\left\|(\hat{\beta}_j^{(7)} - \hat{\beta}_j^{(6)})/\hat{\beta}_j^{(7)}\right\|$	.0003	.0010	.0006	.0054	.0001	.0109	.0004

$\max_{j=1,2,\ldots,7} \left|(\hat{\beta}_j^{(0)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right| = 2.0614$ $\max_{j=1,2,\ldots,7} \left|(\hat{\beta}_j^{(7)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right| = .2360$ $\max_{j=1,2,\ldots,7} \left|(\hat{\beta}_j^{(7)} - \hat{\beta}_j^{(6)})/\hat{\beta}_j^{(7)}\right| = .0109$

$\sum_{j=1}^{7} \left|(\hat{\beta}_j^{(0)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right| = 3.8484$ $\sum_{j=1}^{7} \left|(\hat{\beta}_j^{(7)} - \hat{\beta}_{0j})/\hat{\beta}_{0j}\right| = .5561$ $\sum_{j=1}^{7} \left|(\hat{\beta}_j^{(7)} - \hat{\beta}_j^{(6)})/\hat{\beta}_j^{(7)}\right| = .0187$

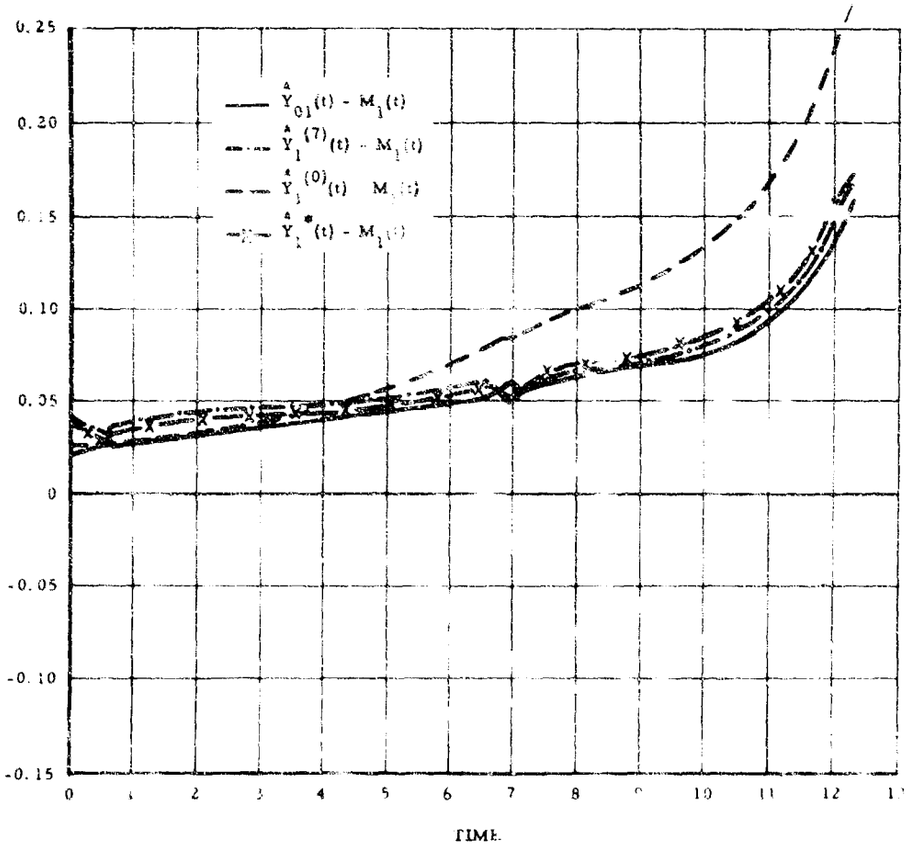

Figure 4. Actual Comparison of Estimators of the Linear Model for Channel 1

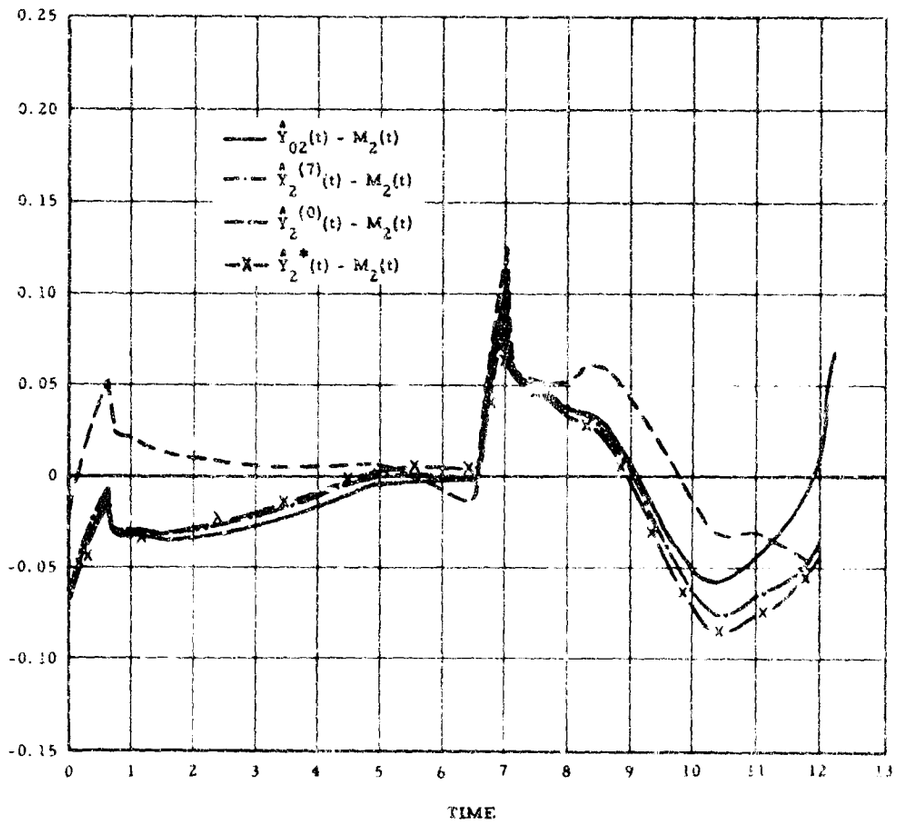

Figure 5. Actual Comparison of Estimators of the Linear Model for Channel 2

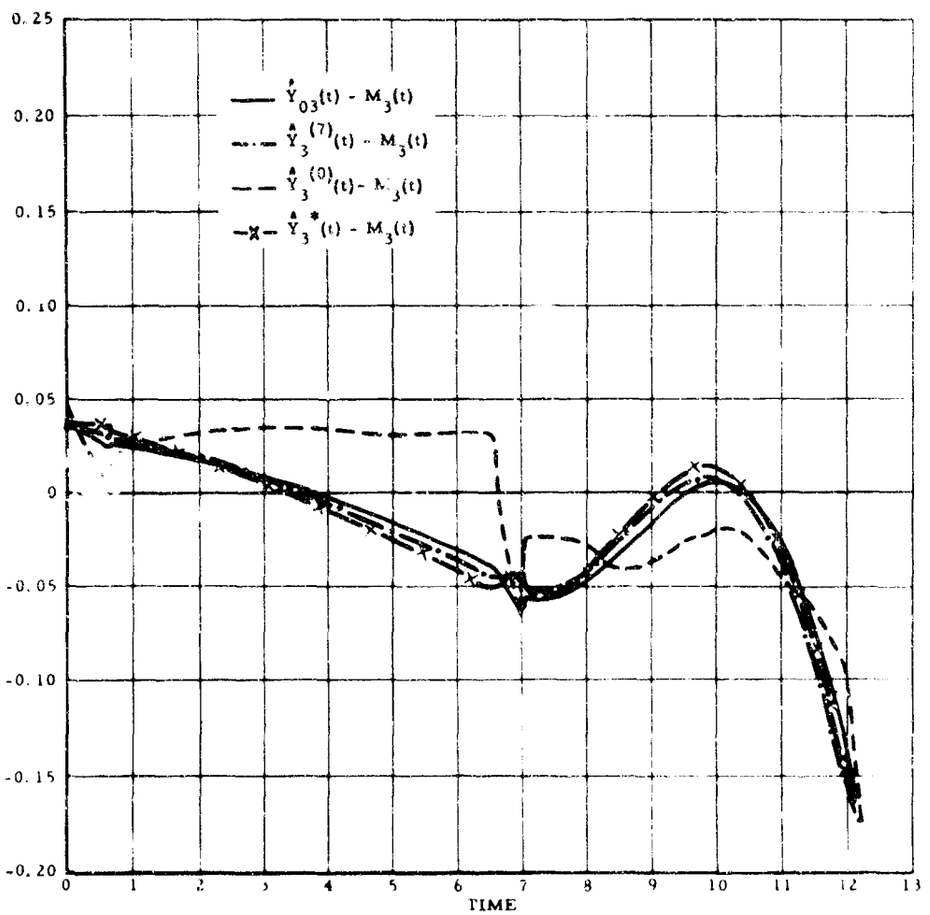

Figure 6. Actual Comparison of Estimators of the Linear Model for Channel 3

REFERENCES

1. W. Eisner and A. F. Goodman. Determination of Dominant Error Sources in an Inertial Navigation System by Iterative Weighted Least Squares. Autonetics Report T4-134/32, North American Rockwell Corporation, March 1963, and American Institute of Aeronautics and Astronautics Journal, Vol. 2 (1964), pages 722-727.

2. P. L. Hsu and K. V. Smith. NAWL1-North American Iterative Weighted Least Squares 7090 or 7094 FORTRAN Program. Autonetics Technical Memorandum 243-2-131, North American Rockwell Corporation, October 1962.

3. P. L. Hsu. NAWL2-North American Stepwise Iterative Weighted Least Squares 7090 or 7094 FORTRAN Program. Autonetics Technical Memorandum 243-2-157, North American Rockwell Corporation, August 1963.

4. M. A. Efroymson. Multiple Regression Analysis. Chapter 17 of Mathematical Methods for Digital Computers, edited by A. Ralston and H. S. Wilf. John Wiley and Sons, Inc., 1960.

5. S. H. Dalrymple. NAWL4-North American Extended Iterative Weighted Least Squares 7094 FORTRAN Program. Autonetics Informal Document, North American Rockwell Corporation.

6. R. L. Anderson and T. A. Bancroft. Statistical Theory in Research. McGraw-Hill Book Company, 1952.

7. H. Scheffe. The Analysis of Variance. John Wiley and Sons, Inc., 1959.

8. R. Bellman. Dynamic Programming. Princeton University Press, 1957.

9. U. Grenander and M. Rosenblatt. Statistical Analysis of Stationary Time Series. John Wiley and Sons, Inc., 1957.

10. M. E. Muller and G. S. Watson. Randomization and Linear Least Squares Estimation. Statistical Techniques Research Group Technical Report 32, Princeton University, July 1959.

11. T. A. Magness and J. B. McGuire. Comparison of Least Squares and Minimum Variance Estimates of Regression Parameters. Annals of Mathematical Statistics, Vol. 33 (1962), pages 462-470.

12. G. H. Golub. Comparison of the Variance of Minimum Variance and Weighted Least Squares Regression Coefficients. Annals of Mathematical Statistics, Vol. 34 (1963), pages 984-991.

13. G. Zyskind. On Canonical Forms, Non-Negative Covariance Matrices and Best and Simple Least Squares Linear Estimators in Linear Models. Annals of Mathematical Statistics, Vol. 38 (1967), pages 1092-1109.

14. G. S. Watson. Linear Least Squares Regression. Annals of Mathematical Statistics, Vol. 38 (1967), pages 1679-1699.

15. J. S. Williams. The Variance of Weighted Regression Estimators. Journal of the American Statistical Association, Vol. 62 (1967), pages 1290-1301.

16. F. W. McElroy. A Necessary and Sufficient Condition that Ordinary Least-Squares Estimators Be Best Linear Unbiased. Journal of the American Statistical Association, Vol. 62 (1967), pages 1302-1304.

17. H. Theil. The Analysis of Disturbances in Regression Analysis. Journal of the American Statistical Association, Vol. 60 (1965), pages 1067-1079.

18. H. Theil. A Simplification of the BLUS Procedure for Analyzing Regression Disturbances. Journal of the American Statistical Association, Vol. 63 (1968), pages 242-251.

19. J. Koerts. Some Further Notes on Disturbance Estimates in Regression Analysis. Journal of the American Statistical Association, Vol. 62 (1967), pages 169-183.

20. S. J. Prais and J. Aitchison. The Grouping of Observations in Regression Analysis. Review of the International Statistical Institute, Vol. 22 (1954), pages 1-22.

21. G. R. Fisher. Iterative Solutions and Heteroscedasticity in Regression Analysis. Review of the International Statistical Institute, Vol. 30 (1962), pages 153-158.

22. M. E. Turner, R. J. Monroe and H. L. Lucas. Generalized Asymptotic Regression and Non-Linear Path Analysis. Biometrics, Vol. 17 (1961), pages 120-143.

23. J. Mandel. Estimation of Weighting Factors in Linear Regression and Analysis of Variance. Technometrics, Vol. 6 (1964), pages 1-26.

24. L. G. Telser. Iterative Estimation of a Set of Linear Regression Equations. Journal of the American Statistical Association, Vol. 59 (1964), pages 845-863.

25. A. R. Dennis. An Adaptive Method for the Linear Estimation of Parameters in Discrete Nonstationary Time Series. TRW Systems Technical Memorandum 9990-6440-KU-000, TRW Inc., June 1963.

26. Stochastic Optimization of Ship's Inertial Navigation System. Autonetics Report C4-505/32, North American Rockwell Corporation, April 1964.

27. R. L. Anderson. The Problem of Autocorrelation in Regression Analysis. Journal of the American Statistical Association, Vol. 49 (1954), pages 113-129.

28. V. K. Murthy. On Fitting a Linear Trend and Testing Independence When the Residuals Form a Markov Process. Applied Mathematics and Statistics Laboratories Technical Report 6, Grant DA-ORD-43, Stanford University, November 1960.

29. N. R. Goodman. Measuring Signals in the Presence of Stationary Noise with the Spectral Density of the Noise Unknown. Rocketdyne Research Report 62-20, North American Rockwell Corporation, October 1962.

30. R. W. Parks. Efficient Estimation of a System of Regression Equations When Disturbances are Both Serially and Contemporaneously Correlated. Journal of the American Statistical Association, Vol. 62 (1967), pages 500-509.

31. H. Raiffa and R. Schlaifer. Applied Statistical Decision Theory. Harvard Business School, 1961.

32. H. Theil. On the Use of Incomplete Prior Information in Regression Analysis. Journal of the American Statistical Association, Vol. 58 (1963), pages 401-414.

33. T. L. Gunckel. The Estimation of Unknown Parameters Using Imperfect Information. Autonetics Inter-Office Letter 61-3243-4-PhA-58, North American Rockwell Corporation, September 1961.

34. A. N. Drucker. The Determination of Tracking System Requirements and the Design of an Optimized Test Regime for Evaluating Ballistic Missile Inertial Guidance Systems in Flight Test. Advisory Committee to the Air Force Systems Command Tracking and Data Analysis Panel, National Academy of Sciences, March 1964.

35. J. S. Chipman. On Least Squares with Insufficient Observations. Journal of the American Statistical Association, Vol. 59 (1964), pages 1078-1111.

36. G. G. Judge and T. Takayama. Inequality Restrictions in Regression Analysis. Journal of the American Statistical Association, Vol. 61 (1966), pages 166-181.

37. L. Villone, R. L. McCornack and K. R. Wood. Multiple Regression with Subsetting of Variables. System Development Corporation Field Note FN-6622, System Development Corporation, June 1962.

38. W. F. Massy. Principal Components Regression in Exploratory Statistical Research. Journal of the American Statistical Association, Vol. 60 (1965), pages 234-256.

39. J. J. Fortier and H. Solomon. Clustering Procedures. Proceedings of an International Symposium on Multivariate Analysis, edited by P. R. Krishnaiah. Academic Press, 1966.

40. B. King. Step-wise Clustering Procedures. Journal of the American Statistical Association, Vol. 62 (1967), pages 86-101.

41. A. F. Goodman. The Interface of Computer Science and Statistics. Western Division Paper 5204, McDonnell Douglas Astronautics Company, February 1969.

42. H. C. Rutemiller and D. A. Bowers. Estimation in a Heteroscedastic Regression Model. Journal of the American Statistical Association, Vol. 63 (1968), pages 552-557.