AN ABSTRACT OF THE THESIS OF

THOMAS DANFORTH BURNETT for the DOCTOR OF PHILOSOPHY
(Name) (Degree)

in STATISTICS presented on March 6, 1969
(Major) (Date)

Title: ESTIMATION OF STOCHASTICALLY VARYING REGRESSION PARAMETERS

Abstract approved: Redacted for Privacy
Donald Guthrie Jr.

The traditional assumption that in repeated regression experiments with observations $\underline{Y}_t = X_t' \underline{\beta}_t + \underline{\epsilon}_t$, $t = 1, 2, \ldots, T$, the parameter vector $\underline{\beta}_t$ is constant is replaced by the assumption that $\underline{\beta}_t$, $t = 1, \ldots, T$ is a sequence from a vector-valued wide-sense stationary stochastic process with covariance function $r(s)$. The existence of this structure interrelating the several experiments admits the existence of information relevant to the value of $\underline{\beta}_t$ in observations $\underline{Y}_{t-1}, \ldots, \underline{Y}_1$, and makes improvement over the usual Least Squares estimator possible.

An estimator is derived which utilizes information from previous time points based on the following definition: A Best Linear (BL) estimator of $\underline{\beta}_t$ is the linear combination

$$\underline{B}_t = \sum_{i=1}^{t} H_i^t \underline{Y}_i$$

which minimizes $E\|\underline{B}_t - \underline{\beta}_t\|^2$. It is shown that the BL estimator can be written as

$$\sum_{i=1}^{t} K_i^t \tilde{\underline{\beta}}_t$$

where

$$\tilde{\underline{\beta}}_t = X_t'^{+} \underline{Y}_t$$

is the unique Least Squares estimator of $\underline{\beta}_t$, with $X_t'^{+}$ the Moore-Penrose generalized inverse of X_t'.

The covariance matrix of the BL estimator is found and for the case where X_i' is of full rank, $i = 1, \ldots, t$, the BL estimator is compared with the usual Least Squares estimator and with the Bayes estimator with normal prior.

A consistent estimator of the covariance function $r(s)$ is obtained when $\underline{\beta}_i$, $i = 1, \ldots, t$ is either known or is estimated by $\tilde{\underline{\beta}}_i$ where X_i' is of full rank. Also given is an estimator of the process mean $\bar{\underline{\beta}}$.

Estimation of Stochastically Varying
Regression Parameters

by

Thomas Danforth Burnett

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

June 1969

APPROVED:

Redacted for Privacy

Associate Professor of Mathematics and Statistics

in charge of major

Redacted for Privacy

Chairman of Department of Statistics

Redacted for Privacy

Dean of Graduate School

Date thesis is presented March 6, 1969

Typed by Clover Redfern for Thomas Danforth Burnett

ACKNOWLEDGMENT

The author wishes to express his thanks to his wife, Margie, for her patience and encouragement, to his major professor, Dr. Donald Guthrie Jr., for his help and advice in the preparation of this thesis and to the National Science Foundation for support of this research under grant GA-379.

TABLE OF CONTENTS

LIST OF FIGURES

ESTIMATION OF STOCHASTICALLY VARYING REGRESSION PARAMETERS

I. INTRODUCTION

Statisticians have for many years concerned themselves with estimation of parameters in models constructed to represent structure of observed data. Traditionally, regression models have assumed that the parameters are temporally constant and that "observation error" is the source of variation in the observation.

However, when concerned with systems in the rapidly evolving economic and social structure of today's world it would seem unrealistic to assume that such parameters are unchanging with time. Similarly, in the physical domain, physical laws often dictate that macro-scale systems are statistically in a steady state, but this does not preclude the existence of variations in both the macro- and micro-scale. Indeed, it is through the observation and study of these variations that our knowledge of the physical laws are increased. It is therefore natural to examine models which admit stochastic variations among the parameters.

Once one accepts that the parameters associated with the system being studied cannot be assumed constant, he must make assumptions about the nature of their behavior. Clearly, a wide range of possibilities is present. The aim of this paper is not to investigate all possibilities, but only one specific behavioral pattern; that the parameters vary according to the probability law of a multi-dimensional stationary stochastic process. Whether such a model may profitably be used to represent an actual system under study depends naturally on the system, and one must justify any chosen model through his collective objective and subjective knowledge of the problem. However, it would seem an unlikely situation that temporal parameter variation has a random character; but conversely that there exists a non-zero serial correlation among the parameter values at different times.

The additional assumption of stationarity of parameter variation, while not applying to all situations, would be reasonable in many cases if one was careful in specification of the time domain within which to investigate the system.

This paper is concerned with the estimation of the present value of a vector of such parameters as they vary in time (or space) or with the prediction of the value of these parameters at some time (or place) in the future. It will be assumed that at each point t in time a regression structure exists in which the present value of the parameters plays the role of unknown constants called regression parameters.

To permit the application of conventional Least Squares (LS) regression analysis to the estimation of these parameters it is necessary that a full rank experiment is performed either instantaneously or during a period of time throughout which the regression parameters may be considered fixed.

In the following, however, it will be assumed in general that this condition is not met. In this light two problems are considered: the estimation and prediction of the value of the vector of regression parameters when in general sufficient experimentation may not be possible by utilizing the covariance structure of the stochastic processes; and the estimation of the mean and covariance function of the process when it is unknown.

It should be made clear that while we are concerned with the mathematical description of a system and primarily with the parameters associated with this description, the parameters are actually assumed to be random variables. We, however, shall refer to them as parameters as this is their primary role.

Such a dual role by the parameters is similar to the situation encountered in Bayesian estimation where the parameters to be estimated are assigned an apriori distribution and thus in a sense are assumed to be random variables. To carry the similarity a step further, we might think of estimation under the present assumption as a Bayesian procedure with the prior given in the form of a stochastic process. However, as no distributional form is associated with the stochastic process, the formal application of the Bayesian procedure cannot be performed. Nevertheless, in the following, it will be shown under special circumstances the Bayes estimator and the estimator derived in this paper are identical.

Throughout this paper we will consider the estimation of parameters in a linear regression function. By this we mean that the expected value of the observed dependent variable, the regressand is a linear combination of the observed values of the independent variables, the regressors, thus.

$$Y = X_1\beta^1 + X_2\beta^2 + \dots + X_p\beta^p + \epsilon$$

where Y is the regressand and $X_1, X_2, \dots, X_p$ are the regressors. In keeping with this terminology $\beta^1, \beta^2, \dots, \beta^p$ will be denoted regression parameters.

The question which obviously arises is how do we estimate the unknown $p \times 1$ vector of regression parameters $\underline{\beta}_t = (\beta_t^1 \beta_t^2 \dots \beta_t^p)'$ at time t. Using matrix notation we know from the classical theory of LS that if the rank of the $p \times n_t$ matrix X_t is p (denoted $\rho(X_t) = p$) where X_t is a matrix whose n_t columns are the vectors of regressors associated with n_t independent and identically distributed observations taken at time t, then the best (minimum variance), linear, unbiased estimator (BLUE) of $\underline{\beta}_t$ is

$$\tilde{\underline{\beta}} = (X_t X_t')^{-1} X_t \underline{Y}_t$$

where $\underline{Y}_t$ is the vector of observations (Scheffé, 1959). A restriction on the above is that the vector $\underline{\beta}_t$ must remain constant over the period of time the n_t observations are taken. This obviously would be the case if the observations were taken simultaneously. However, if the following not too unlikely conditions exist BLUE estimates cannot be obtained.

Suppose that the regression parameters are neither constant with time nor are sufficient observations obtained to yield a matrix X_t of rank p over the period of time when $\underline{\beta}_t$ may be considered constant, and hence that the assumptions underlying the optimality of Least Squares estimation are no longer valid. It will be shown in the following that an estimator with certain optimal properties can be obtained by assuming that the vector $\underline{\beta}_t$ behaves as a stationary stochastic process with known mean and covariance structure.

Before proceeding further, let us consider this problem historically. Kalman (1960) employed the Markovian model

$$\underline{\beta}_t = T(t, t-1)\underline{\beta}_{t-1} + \underline{\mu}_t$$

by which the value of the parameter vector at time t is a linear function of the value at time $t-1$ plus a random error. The estimate of $\underline{\beta}_t$ is then a weighted average of the predicted value of $\underline{\beta}_t$ based on the Markovian model and of the information obtained at time

t through a regression structure where the parameters to be estimated are the unknown parameters of this structure. Kalman assumed the transition matrix $T(t, t-1)$ was known but later others have extended this approach by estimating this matrix.

Jones (1966) simplified Kalman's model by letting the transition matrix be the identity matrix and by assuming the process is observed directly with error. This model leads to exponential smoothing with decreasing weight being given to observations further in the past.

Both of the approaches, because of their Markovian nature, do not allow for stochastic dependencies of lag greater than one if they exist. Thus, information through such dependencies is not utilized.

It is with this possibility in mind, that we investigate the estimation of $\underline{\beta}_t$ under the assumption that $\{\underline{\beta}_t\}_{t=1}^{\infty}$ is a stationary stochastic process. The existence of a non-zero covariance structure $r(s)$, $s = 0, 1, \ldots, k$ between $\underline{\beta}_t$ and $\underline{\beta}_{t-s}$ dictates by its very nature that there is information in all of the observations $\underline{Y}_{t-s}$, $s = 0, 1, \ldots, k$ relevant to the value of $\underline{\beta}_t$. On this basis an estimator of $\underline{\beta}_t$ is then formulated as that linear function of the last $k+1$ observation such that the expected Euclidian distance between the estimate and the true value of $\underline{\beta}_t$ is minimized.

The resulting estimator, called the Best Linear (BL) estimator, can also be shown to be a linear function of the LS estimators of the parameter vectors at the times of the last $k+1$ observations. In the case when the LS estimator is not unique at some or all of these times a particular LS estimator is chosen, one which has certain optimal properties. Through representation of the BL estimator in terms of the LS estimators, comparisons are made between its covariance matrix and that of the LS and Bayes estimators. A BL predictor is formulated in a similar manner.

Frequently the covariance structure may be unknown. We therefore present a method of constructing consistent estimates of the correlation function $r(s)$ when $\rho(X_t) = p$, $p \leq n_t$. However no method of estimating the covariance function has been found when $\rho(X_t) < p$. No study was made of the distribution of the estimator of $r(s)$. Both these problems seem to be suitable areas for further work, however, both appear to be quite difficult.

For the case when the process $\underline{\beta}_t$ has an unknown mean $\bar{\underline{\beta}}$, we derive a method of estimating $\bar{\underline{\beta}}$.

An Example

Let us consider the problem of studying the spatial and temporal variation of a field variable β such as the concentration of ozone in the atmosphere at 40 KM above sea level. If one is to use quantitative numerical techniques in such a study this is facilitated by the knowledge of the value of the field variable at orderly arranged locations throughout the field at successive points in time. However, direct observations of the field variable at these orderly arranged locations may be difficult or impossible to obtain.

If direct observations of the field at any given time are well spaced throughout the field and plentiful the grid field (the value of the field at the orderly spaced locations) may be estimated using ordinary smoothing and contouring techniques. Usually however, such quantities of observation are not available at any given time and other estimation techniques are necessary.

The BL estimation procedure which is developed in the following chapters may be used as such a technique if the field variable has certain well behaved properties. That is, if the field variable at any two positions in the field obeys the probability law of a bivariate stationary stochastic process. Thus associated with the grid field there is a non-zero spatial and temporal covariance structure which in effect says that there is information relevant to the present value of the grid field in observations taken in the past.

The BL estimator utilizes this covariance structure to optimally combine the information from past observations with the information in present observations to estimate the present value of the grid field.

To proceed mathematically let us consider a sub-grid of 16 points and a field with known mean. As we may always subtract the mean field from our observations we may without loss of generality assume the field has zero mean.

Labeling the sub-grid points $p_1, p_2, \ldots, p_{16}$, let $\beta_{t,i}$ be the value of the field variable at grid point p_i at time t. Let $C(s, p_i, p_j)$ be the time-space covariance function of lag s between the values of the field at grid points p_i and p_j. Thus the vector $\underline{\beta}_t = (\beta_{t,1}\beta_{t,2} \cdots \beta_{t,16})'$ of grid point field values obeys the probability law of a 16-dimensional stationary stochastic process with mean vector $\underline{0}$ and autocovariance function $r(s) = \{C(s, p_i, p_j)\}$, $i, j = 1, 2, \ldots, 16$, $s = 0, 1, 2, \ldots, t-1$.

We need now only to establish a relationship at time t between $Y_{t,j}$, the observed value of the field at location $l_{t,j}$ with the corresponding variables at the sub-grid points. Suppose, for example, that

$$Y_{t,j} = \sum_{i=1}^{16} W_{t,i,j}\beta_{t,i} + \epsilon_{t,j}$$

where the weights $W_{t,i,j}$ are inversely proportioned to the distance $\| l_{t,j} - p_i \|^2$ and

$$\sum_{i=1}^{16} W_{t,i,j} = 1.$$

The above example possesses the necessary elements for the employment of the BL estimator to estimate the field variable vector

$\underline{\beta}_t$ from data observed not only at time t but utilizing data from prior times. The advantage of such a procedure is apparent when one realizes that depending on the method of observation, the number of observation stations within range of the sub-grid may change with time and as such the amount of information about $\underline{\beta}_t$ available at time t may be minimal or non-existent.

Exactly how prior data can be utilized will be given in the following chapters.

II. THE MATHEMATICAL MODEL

The Model

Consider, at time $\tau = t$, the regression structure

(2.1) $$\underline{Y}_t = X_t'\underline{\beta}_t + \underline{\epsilon}_t$$

where $\underline{Y}_t$ is a $n_t \times 1$ vector of independent and identically distributed observations, X_t' a $n_t \times p$ matrix of known coefficients, $\underline{\beta}_t$ a vector of unknown regression parameters and $\underline{\epsilon}_t$ a $n_t \times 1$ vector of normally distributed random errors with $E(\underline{\epsilon}_t) = 0$ and $E(\underline{\epsilon}_t\underline{\epsilon}_t') = \sigma^2 I_{n_t}$ where σ^2 may be known or unknown. It is also assumed that $\underline{\beta}_t$ and $\underline{\epsilon}_s$ are uncorrelated for all s and t.

Instead of the usual assumption in linear regression of the vector of regression parameters being fixed during the period of observation, assume that $\underline{\beta}_t$ is characterized in the following manner. Let $\{\mathscr{B}(\tau)\}$ be a real-valued, p-dimensional stationary stochastic process. Then the sequence $\{\underline{\beta}_t\}_{t=1}^{\infty}$, where $\underline{\beta}_t$ is the value of $\mathscr{B}(\tau)$ at $\tau = t$, is a p-dimensional discrete stationary stochastic process. For convenience, let us also assume that the mean of the process $\underline{\bar{\beta}}$ and the covariance function

$$r(s) = E(\underline{\beta}_t - \underline{\bar{\beta}})(\underline{\beta}_{t+s} - \underline{\bar{\beta}})', \quad s = 0, \pm 1, \pm 2, \ldots$$

of the discrete process are known reserving the problem of their estimation until later. We may then choose $\underline{\bar{\beta}} = 0$ without loss of generality and hence we find

$$r(s) = E(\underline{\beta}_t \underline{\beta}'_{t+s}).$$

Let us consider now the problem of estimating $\underline{\beta}_t$, the value of the process at time t.

Least Square Estimation

The assumption concerning the stochastic behavior of the $\{\mathscr{B}(\tau)\}$ process, however, does not negate the appropriateness of applying conventional LS regression analysis to obtain an estimate of $\underline{\beta}_t$ as long as we restrict ourselves in considering the regression structure (2.1) to a single point in time. This follows from the observation that the vector of regression parameters to be estimated is only required to be constant over the period of observation and the realized value of the random variable $\mathscr{B}(\tau)$ satisfies this at any given point in time.

Ignoring for the moment the stochastic character of $\underline{\beta}_t$, if we consider observations at time t only, it is well known that in the case of $\rho(X_t') = p$, the LS estimator

$$\tilde{\underline{\beta}}_t = (X_t X_t')^{-1} X_t \underline{Y}_t$$

is unique and minimum variance among linear unbiased estimates.

However, when $\rho(X_t) < p$, neither an unbiased nor unique LS estimator of $\underline{\beta}_t$ exists. In fact, an infinite number of LS estimators exist, none of which are unbiased. This naturally leads to further investigations of the estimation of $\underline{\beta}_t$ when $\rho(X_t') < p$.

Chipman (1964) considers the approach of specifying a complementary set of linear restrictions $X_t'^* \underline{\beta}_t = \underline{Y}_t^*$ such that the augmented design matrix

$$\begin{bmatrix} X_t' \\ X_t'^* \end{bmatrix}$$

is of full rank. Estimators obtained in this manner are conditionally unbiased, conditioned on the complementary restrictions. Chipman shows that there is a specific set of complementary restrictions which result in a procedure called "minimum bias estimation" which selects from the class of minimum bias estimators the one that has minimum variance. Furthermore, Chipman points out that "minimum bias estimation" is equivalent to an estimation procedure proposed by Penrose (1956). Before discussing Penrose's method it will be useful to discuss one of its principle ingredients, the generalized inverse; we shall restrict our interest to a particular generalized inverse defined by Moore (1935) and Penrose (1955) and discussed extensively by Greville (1959, 1960), Ben-Israel and Charles (1963) and Price (1964) among others. We shall designate the Moore-Penrose (M-P) inverse of a matrix U by U^+ and define it to be the unique matrix that satisfies the following relationships:

$$(2.2) \qquad \begin{aligned} &a)\ UU^+U = U \\ &b)\ U^+UU^+ = U^+ \\ &c)\ (U^+U)' = (U^+U) \\ &d)\ (UU^+)' = (UU^+). \end{aligned}$$

From these four defining properties additional features of the M-P inverse are obtained. Those which are of use in the following sections are given below:

$$(2.3) \qquad \begin{aligned} &a)\ (U^+)^+ = U \\ &b)\ U'^+ = U^{+'} \\ &c)\ \text{if } U \text{ is square and nonsingular } U^+ = U^{-1} \\ &d)\ (U'U)^+ = U^+U^{+'} \\ &e)\ (U^+U) \text{ is hermitian idempotent} \\ &f)\ UU^+ = I \text{ if } \underset{p \times q}{U} \text{ is of rank } p. \end{aligned}$$

Derivations are contained in Penrose (1955).

Return now to the estimation of $\underline{\beta}_t$ under the regression structure (2.1) where $\rho(X_t') < p$, the estimator suggested by Penrose which corresponds to Chipman's "minimum bias estimator" is

(2.4) $$\underline{\tilde{\tilde{\beta}}}_t = X_t'^{+} \underline{Y}_t.$$

$\underline{\tilde{\tilde{\beta}}}_t$ is unique since $X_t'^{+}$ is unique and it can be shown that

<u>Lemma 1.</u> $\underline{\tilde{\tilde{\beta}}}_t$ is a least square estimator of $\underline{\beta}_t$.

<u>Proof.</u> Following the notation of Scheffé (1959) a LS estimate of $\underline{\beta}_t$ is defined as any value of $\underline{b}_t$, where $\underline{b}_t$ is linear in the observations $\underline{Y}_t$, which minimized the function

$$L(\underline{Y}_t, \underline{b}_t) = \| \underline{Y}_t - X_t' \underline{b}_t \|^2$$

where $\|\underline{V}\|$ denotes the length of the vector $\underline{V}$. Let $M(X_t')$ be a linear manifold spanned by $\underline{\xi}_{t,1}, \underline{\xi}_{t,2}, \ldots, \underline{\xi}_{t,p}$, the columns of X_t'. Then

$$\begin{aligned} \underline{Z}_t &= X_t' \underline{b}_t \\ &= b_{t,1}\underline{\xi}_{t,1} + b_{t,2}\underline{\xi}_{t,2} + \ldots + b_{t,p}\underline{\xi}_{t,p} \end{aligned}$$

is an element of $M(X_t')$. With $\underline{Z}_t \in M(X_t')$ and letting $\underline{b}_t = D_t \underline{Y}_t$ for some D_t we see geometrically that $L(\underline{Y}_t, \underline{b}_t) = \|\underline{Y}_t - \underline{Z}_t\|^2$ attains a minimum when and only when $\underline{Z}_t = X_t' D_t \underline{Y}_t$ is a projection of $\underline{Y}_t$ onto $M(X_t')$. That is when $X_t' D_t$ is a projection operator onto $M(X_t')$. From (2.2c), (2.3e) and (2.4) we see that $X_t' D_t = X_t' X_t'^{+}$ is hermition idempotent and thus a projection operator onto $M(X_t')$.

Thus $\tilde{\underline{\beta}}_t$ is a LS estimate of $\underline{\beta}_t$. It should also be noted from (2.2), when $\rho(X_t') = p$

$$X_t'^{+} = (X_t X_t')^{-1} X_t.$$

Thus $\tilde{\underline{\beta}}_t = \tilde{\tilde{\underline{\beta}}}_t$. We then may in general define

$$\tilde{\underline{\beta}}_t = X_t'^{+} \underline{Y}_t$$

as the LS estimator of $\underline{\beta}_t$. In the following all references to a LS estimator of $\underline{\beta}_t$ refer to this estimator.

In the foregoing we consider the estimation of $\underline{\beta}_t$ under conditions equivalent to the assumption $r(s) = 0$, $s \neq 0$. If this is not the case but $r(s) \neq 0$ for $|s|$ less than some fixed number k, it seems apparent that there is information about $\underline{\beta}_t$ in the observations $\underline{Y}_t, \underline{Y}_{t-1}, \ldots, \underline{Y}_{t-k+1}$. This will be investigated in the following sections.

III. A BEST LINEAR ESTIMATOR OF $\underline{\beta}_t$

In this chapter we explore the "improvement" of the LS estimate of $\underline{\beta}_t$ utilizing the covariance structure of the $\{\mathcal{B}(\tau)\}$ process. When $\rho(X_t') = p$ this "improvement" takes the form of a reduction in the sum of the diagonal elements of the covariance matrix of the estimator of $\underline{\beta}_t$ at the expense of the property of (conditional) unbiasedness. When $\rho(X_t') < p$, the "improvement" takes the same form as above but since the original LS estimator was itself biased, the property of bias is changed only quantitatively not qualitatively.

Justification for the expectation that the LS estimator can be improved upon under the present model is embodied in the following notions. The existence of a non-zero covariance structure, by its very nature, dictates the presence in the observations $\underline{Y}_t, \underline{Y}_{t-1}, \ldots, \underline{Y}_1$ of information related to the value of $\underline{\beta}_t$. It thus seems clear that any estimator which does not use all this information is less than optimum. Conversely, an optimum weighting of the observations $\underline{Y}_t, \underline{Y}_{t-1}, \ldots, \underline{Y}_1$, which will be seen to be an optimum weighting of LS estimates $\tilde{\underline{\beta}}_t, \tilde{\underline{\beta}}_{t-1}, \ldots, \tilde{\underline{\beta}}_1$ would give zero weight to all but $\tilde{\underline{\beta}}_t$ if the LS estimate were indeed optimum. This we will find is not the case.

Before defining what shall be called a Best Linear estimator of $\underline{\beta}_t$ let us consider the joint distribution of the observations

$\underline{Y}_t, \underline{Y}_{t-1}, \ldots, \underline{Y}_1$.

Joint Distribution of the Observations

Under the following three previously mentioned assumptions:

i) $\underline{\beta}_t \sim$ p-variate normal with mean $\underline{0}$ and covariance matrix $r(0)$,

ii) $\underline{\epsilon}_t \sim n_t$-variate normal with mean $\underline{0}$ and covariance matrix $\sigma^2 I_{n_t}$,

iii) $E(\underline{\beta}_t \underline{\epsilon}_t') = 0$,

we see that

Lemma 1. $\underline{Y}_t$ is distributed n_t-variate normal with mean $\underline{0}$ and covariance matrix $[X_t' r(0) X_t + \sigma^2 I_{n_t}]$.

Under the additional assumptions:

iv) $\{\mathcal{B}(\tau)\}$ is a p-variate stationary gaussian process with mean $\underline{0}$ and covariance structure $E(\underline{\beta}_t \underline{\beta}'_{t+s}) = r(s)$,

v) $E(\underline{\beta}_t \underline{\epsilon}'_{t+s}) = 0, \quad s \neq 0$

vi) $E(\underline{\epsilon}_t \underline{\epsilon}'_{t+s}) = 0, \quad s \neq 0$

we see that,

Lemma 2. The vector of observations

(3.1) $$\underline{Y}(t) = (\underline{Y}_t' \underline{Y}_{t-1}' \ldots \underline{Y}_1')'$$

is distributed $N(t)$-variate normal with mean $\underline{0}$ and covariance matrix

$$\Phi_1(t) = (A'(t)R(t)A(t)+\sigma^2 I_{N(t)}) \tag{3.2}$$

where

$$A'(t) = \begin{bmatrix} X'_t & 0 & & 0 \\ 0 & X'_{t-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & X'_1 \end{bmatrix}, \tag{3.3}$$

$$R(t) = \begin{bmatrix} r(0) & r(1) & \cdots & r(t-1) \\ r'(1) & r(0) & \cdots & r(t-2) \\ \vdots & \vdots & \ddots & \vdots \\ r'(t-1) & r'(t-2) & \cdots & r(0) \end{bmatrix}$$

and

$$N(t) = \sum_{s=1}^{t} n_s .$$

<u>Proof.</u> We can write

$$\underline{Y}(t) = A'(t)\,\underline{\beta}(t) + \underline{\epsilon}(t)$$

where

$$\underline{\beta}(t) = (\underline{\beta}'_t\, \underline{\beta}'_{t-1} \cdots \underline{\beta}'_1)' \tag{3.4}$$

and

(3.5) $$\underline{\epsilon}(t) = (\underline{\epsilon}'_t \underline{\epsilon}'_{t-1} \cdots \underline{\epsilon}'_1)' .$$

Then we have

$$E(\underline{Y}(t)) = E(A'(t)\underline{\beta}(t)+\underline{\epsilon}(t)) = \underline{0}$$

and

$$\begin{aligned} E(\underline{Y}(t)\underline{Y}'(t)) &= E((A'(t)\,\underline{\beta}(t)+\underline{\epsilon}(t))(A'(t)\,\underline{\beta}(t)+\underline{\epsilon}(t))') \\ &= A'(t)E(\underline{\beta}(t)\,\underline{\beta}'(t))A(t) + E(\underline{\epsilon}(t)\,\underline{\epsilon}'(t)) \\ &= A'(t)R(t)A(t) + \sigma^2 I_{N(t)}. \end{aligned}$$

Let us now make the following notational conventions. A matrix $C(t)$, whose dimensions and/or properties may be a function of time will be denoted as the matrix C when no confusion will arise in doing so. In addition matrix dimensions will be omitted when their value is obvious.

The Best Linear Estimator

Restriction of our investigation to estimates in the class of estimates linear in the observations we shall make the following

Definition 3.1. A Best Linear estimator of $\underline{\beta}_t$ is the value of the linear combination

(3.6) $$\underline{B}_t = \sum_{i=0}^{t-1} H^t_{t-i}\underline{Y}_{t-i}$$

which minimizes the loss function

$$L_1 = E\|\underline{B}_t - \underline{\beta}_t\|^2. \tag{3.7}$$

Knowing the joint distribution of the observations reduces the problem of estimation to choosing $H_t^t, \ldots, H_1^t$ to minimize (3.7).

The correct choice is given by

Theorem 3.1. The matrices $H_t^t, \ldots, H_1^t$ which minimize (3.7) when $\underline{B}_t$ is given by (3.6) satisfy the matrix equation

$$[H_t^t H_{t-1}^t \ldots H_1^t]\,\Phi_1(t) = \psi'(t)A(t)$$

where

$$\psi'(t) = [r(0)r(1)\ldots r(t-1)].$$

Proof. We may express the loss function as

(3.8)

$$\begin{aligned} L_1 &= E\|\underline{B}_t - \underline{\beta}_t\|^2 \\ &= E(\underline{B}_t - \underline{\beta}_t)'(\underline{B}_t - \underline{\beta}_t) \\ &= \text{Trace } E(\underline{B}_t - \underline{\beta}_t)(\underline{B}_t - \underline{\beta}_t)' \\ &= \text{Trace } E\left(\sum_{i=0}^{t-1} H_{t-i}^t \underline{Y}_{t-i} - \underline{\beta}_t\right)\left(\sum_{i=0}^{t-1} H_{t-i}^t \underline{Y}_{t-i} - \underline{\beta}_t\right)' = \end{aligned}$$

$$= \text{Trace } [-I H^t_t \ldots H^t_1] \begin{bmatrix} r(0) & \psi'(t)A(t) \\ A'(t)\psi(t) & \Phi_1(t) \end{bmatrix} \begin{bmatrix} -I \\ H^{t'}_t \\ \vdots \\ H^{t'}_1 \end{bmatrix}$$

$$= \text{Trace } \left(r(0) - [H^t_t \ldots H^t_1] A'(t)\psi(t) - \psi'(t)A(t) \begin{bmatrix} H^{t'}_t \\ \vdots \\ H^{t'}_1 \end{bmatrix} \right.$$

$$\left. + [H^t_t \ldots H^t_1]\Phi_1(t) \begin{bmatrix} H^{t'}_t \\ \vdots \\ H^{t'}_1 \end{bmatrix} \right),$$

from (3. 2) and the relation

$$E(\underline{\beta}_t \underline{Y}'(t)) = E(\underline{\beta}_t (A'(t)\, \underline{\beta}(t) + \underline{\epsilon}(t))')$$

$$= [r(0) r(1) \ldots r(t-1)] A(t).$$

We also find that

$$(3.9) \quad L_1 = \text{Trace } r(0) - 2 \text{ trace } \sum_{i=0}^{t-1} H^t_{t-i} \theta_{t-i}$$

$$+ \text{ trace } \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} H^t_{t-i} \Phi_{ij}(t) H^{t'}_{t-j}$$

where

$$\theta_{t-i} = X'_{t-i} r'(i), \quad i = 0, \ldots, t-1$$

and

$$\Phi_{ij} = X'_{t-i} r(i-j) X_{t-j} + \delta(i-j) \sigma^2 I_{n_i}$$

where

$$\delta(i-j) = \begin{cases} 1 : i = j \\ 0 : i \neq j \end{cases} .$$

Letting $C_{i,\ell}$ denote the ℓ-th column and $C_{i,\ell'}$ the ℓ-th row of a matrix C, (3. 9) becomes

$$\text{(3. 10)} \quad L_1 = \text{trace } r(0) - 2 \sum_{i=0}^{t-1} \sum_{\ell=1}^{p} H^t_{t-i,\ell'} \theta_{t-i,\ell}$$

$$+ \sum_{i=0}^{t-1} \sum_{j=0}^{t-1} \sum_{\ell=1}^{p} H^t_{t-i,\ell'} \Phi_{ij} H^{t'}_{t-j,\ell} .$$

Differentiating (3. 10) with respect to each row $H^t_{t-i,\ell'}$ and setting the derivatives to zero, we obtain tp "normal" equations

$$-2\theta_{t-i,\ell} + 2 \sum_{j=0}^{t-1} \Phi_{ij} H^{t'}_{t-j,\ell} = 0, \quad j = 0, \ldots, t-1; \ \ell = 1, \ldots, p,$$

which can be written as a single matrix equation

$$\text{(3. 11)} \qquad [H^t_t \ldots H^t_1] \Phi_1(t) = \psi'(t) A(t).$$

and the theorem is proved.

From the relation "If $A = B + C$, B is positive definite, and C is skew symmetric, then $|A| \geq |B|$" (Rao, 1965a) we see that $|\Phi_1(t)| > 0$ and thus $\Phi_1^{-1}(t)$ exists. Therefore

$$[H_t^t \ldots H_1^t] = \psi'(t)A(t)\Phi_1^{-1}(t) \tag{3.12}$$

and

$$\underline{\hat{\beta}}_t = \psi'(t)A(t)\Phi_1^{-1}(t)\underline{Y}(t) \tag{3.13}$$

is that value of $\underline{B}_t$ such that

$$E\|\underline{\hat{\beta}}_t - \underline{\beta}_t\|^2 = \underset{\underline{B}_t}{\mathrm{Min}}\, E\|\underline{B}_t - \underline{\beta}_t\|^2.$$

This result can be seen to be an extension of the following result obtained by Rao (1965b). Rao considered the following problem which, in the notion of the present paper, assumes $\underline{Y}$ is a vector random variable with the structure

$$\underline{Y} = X'\underline{\beta} + \underline{\epsilon}$$

where $\underline{\beta}$ and $\underline{\epsilon}$ are unobservable random variables and X' is a matrix of rank p. Further Rao assumes

$$E(\underline{\beta}) = \underline{0}, \quad E(\underline{\epsilon}) = \underline{0},$$

$$E(\underline{\beta} - \underline{\bar{\beta}})(\underline{\beta} - \underline{\bar{\beta}})' = r(0),$$

$$E(\underline{\epsilon}\,\underline{\epsilon}') = \sigma^2 I,$$

and

$$E(\underline{\beta}, \underline{\epsilon}') = 0.$$

Under these assumptions Rao found the linear estimator of $\underline{\beta}$ for which $E\|\check{\underline{\beta}} - \underline{\beta}\|^2$ is a minimum is

$$\check{\underline{\beta}} = r(0)X(X'r(0)X + \sigma^2 I_N)^{-1}\underline{Y}$$

which can be seen to be identical to the present estimator.

$$\hat{\underline{\beta}}_t = \psi'(t)A(t)\Phi_1^{-1}(t)\underline{Y}(t)$$

if only data taken at time t is considered.

From (3.8) L_{min} obtained under (3.13) is

$$L_{min} = \text{trace}\,[r(0) - \psi'(t)A(t)\Phi_1^{-1}(t)A'(t)\psi(t)]$$

and the covariance matrix of $\hat{\underline{\beta}}_t$ is

$$E(\hat{\underline{\beta}}_t - \underline{\beta}_t)(\hat{\underline{\beta}}_t - \underline{\beta}_t)' = [r(0) - \psi'(t)A(t)\Phi_1^{-1}(t)A'(t)\psi(t)].$$

The form of the BL estimator of $\underline{\beta}_t$ as a linear combination of the observation does not lend itself readily to the investigation of the property of conditional unbiasedness nor to the evaluation of the covariance matrix of this estimator or comparison with the covariance matrix of $\tilde{\underline{\beta}}_t$. Consequently an alternative expression for the

BL estimator will be given in the following

<u>Theorem 3. 2.</u> Let $\tilde{\underline{\beta}}_t, \ldots, \tilde{\underline{\beta}}_1$ be LS estimators of $\underline{\beta}_t, \ldots, \underline{\beta}_1$ as defined by (2. 4). There exists $K_t^t, \ldots, K_1^t$, where K_i^t is $p \times p$, such that

$$\hat{\hat{\underline{\beta}}}_t = \sum_{i=0}^{t-1} K_{t-i}^t \tilde{\underline{\beta}}_{t-i} = \hat{\underline{\beta}}_t$$

and

$$E(\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t)(\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t)' = E(\hat{\underline{\beta}}_t - \underline{\beta}_t)(\hat{\underline{\beta}}_t - \underline{\beta}_t)' .$$

<u>Proof.</u> Let

(3. 14) $$[K_t^t \ldots K_1^t] = [H_t^t \ldots H_1^t]A'(t) .$$

Then

$$\hat{\hat{\underline{\beta}}}_t = [H_t^t \ldots H_1^t]A'(t)\, \tilde{\underline{\beta}}(t)$$

$$= \psi' A(A'RA+\sigma^2 I_N)^{-1} A'A^{+'} \underline{Y}$$

where $\tilde{\underline{\beta}}(t)$ is analogus to (3. 4). But

$$A(A'RA+\sigma^2 I_N)^{-1} A'A^{+'}$$

$$= A(A'RA+\sigma^2 I_N)^{-1} A'A^{+'} (A'RA+\sigma^2 I_N)(A'RA+\sigma^2 I_N)^{-1}$$
$$= A(A'RA+\sigma^2 I_N)^{-1} (A'RA+\sigma^2 I_N)A'A^{+'} (A'RA+\sigma^2 I_N)^{-1}$$
$$= AA'A^{+'}(A'RA+\sigma^2 I_N)^{-1}$$
$$= A(A'RA+\sigma^2 I_N)^{-1} .$$

Thus

$$\hat{\hat{\underline{\beta}}}_t = \psi' A(A'RA+\sigma^2 I_N)^{-1} Y = \hat{\underline{\beta}}_t .$$

Also

(3.15)
$$E(\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t)(\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t)' = E\Big(\sum_{i=0}^{t-1} K^t_{t-i}\tilde{\underline{\beta}}_{t-i} - \underline{\beta}_t\Big)\Big(\sum_{i=0}^{t-1} K^t_{t-i}\tilde{\underline{\beta}}_{t-i} - \underline{\beta}_t\Big)'$$

$$= [-I K^t_t \dots K^t_1] \begin{bmatrix} r(0) & \psi' AA^+ \\ AA^+\psi & \Phi_2 \end{bmatrix} \begin{bmatrix} -I \\ K^{t'}_t \\ \vdots \\ K^{t'}_1 \end{bmatrix}$$

where

$$\Phi_2(t) = A(t)A^+(t)(R(t)+\sigma^2(A(t)A'(t))^+)A(t)A(t)^+.$$

This follows from the relation

$$E(\underline{\beta}_t \tilde{\underline{\beta}}'(t) = E(\underline{\beta}_t (A^{+'}\underline{Y}(t))')$$

$$= E(\underline{\beta}_t (A^{+'}(A^+\underline{\beta}(t)+\underline{\epsilon}(t)))')$$

$$= E(\underline{\beta}_t \, \underline{\beta}'(t)AA^+)$$

$$= [r(0)r(1)\dots r(t-1)]AA^+$$

and

$$E(\tilde{\underline{\beta}}(t)\tilde{\underline{\beta}}'(t)) = E((A^{+'}(A^+\underline{\beta}(t)+\underline{\epsilon}(t)))(A^{+'}(A^+\underline{\beta}(t)+\underline{\epsilon}(t)))')$$

$$= (A^{+'}A'E(\underline{\beta}(t)\underline{\beta}'(t))A'A^+) + (A^{+'}E(\underline{\epsilon}(t)\underline{\epsilon}'(t))A^+)$$

$$= A^{+'}A'R(t)AA^+ + \sigma^2 A^{+'}A^+$$

$$= AA^+(R(t)+\sigma^2(AA')^+)AA^+ .$$

Using the identity

$$A'AA^{+}(R+\sigma^{2}(AA')^{+})AA^{+}A = (A'RA+\sigma^{2}I_N)A^{+}A$$

it follows from (3.12), (3.14) and (3.15) that

(3.16)

$$\begin{aligned} E(\hat{\hat{\underline{\beta}}}_t-\underline{\beta}_t)(\hat{\hat{\underline{\beta}}}_t-\underline{\beta}_t)' &= r(0) - [K_t^t \ldots K_1^t]AA^{+}\psi - \psi'AA^{+}\begin{bmatrix} K_t^{t'} \\ \vdots \\ K_1^{t'} \end{bmatrix} \\ &\quad + [K_t^t \ldots K_1^t]\Phi_2 \begin{bmatrix} K_t^{t'} \\ \vdots \\ K_1^{t'} \end{bmatrix} \\ &= r(0) - \psi'A(A'RA+\sigma^{2}I_N)^{-1}A'AA^{+}\psi \\ &\quad - \psi'AA^{+}A(A'RA+\sigma^{2}I_N)^{-1}A'\psi \\ &\quad + \psi'A(A'RA+\sigma^{2}I_N)^{-1}A'AA^{+}(R+\sigma^{2}(AA')^{+}) \\ &\quad \times AA^{+}A(A'RA+\sigma^{2}I_N)^{-1}A'\psi \\ &= r(0) - \psi'A(A'RA+\sigma^{2}I_N)^{-1}A'\psi \\ &= E(\hat{\underline{\beta}}_t-\underline{\beta}_t)(\hat{\underline{\beta}}_t-\underline{\beta}_t)'. \end{aligned}$$

This completes the proof of Theorem 2.

<u>Corollary 3.1.</u> $\hat{\hat{\underline{\beta}}}_t$ is a BL estimator of $\underline{\beta}_t$.

Proof.

$$\|\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t\|^2 = \text{Tr } E((\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t)(\hat{\hat{\underline{\beta}}}_t - \underline{\beta}_t)')$$
$$= \text{Tr } E((\hat{\underline{\beta}}_t - \underline{\beta}_t)(\hat{\underline{\beta}}_t - \underline{\beta}_t)')$$
$$= E\|\hat{\underline{\beta}}_t - \underline{\beta}_t\|^2.$$

As $\hat{\hat{\underline{\beta}}}_t = \hat{\underline{\beta}}_t$ we shall dispense with the duplicate notation and let $\hat{\underline{\beta}}_t$ denote the BL estimator whether it is a linear combination of the LS estimators or of the original observations.

It will be useful later to note that

$$A\Phi_1^{-1}A' = \Phi_2^+ . \tag{3.17}$$

This follows by writing

$$\Phi_2 = AA^+(R+\sigma^2(AA')^+)AA^+ = (A^{+'}(A'RA+\sigma^2I_N)A^+) \quad .$$

and noting that

$$A\Phi_1^{-1}A' = A(A'RA+\sigma^2I_N)^{-1}A' = (A^{+'}(A'RA+\sigma^2I_N)A^+)^+$$

That the latter is true follows from the definition of the M-P inverse.

The Best Linear Predictor

Consider the parallel problem to estimation, that of prediction, under the present model. A non-zero covariance function $r(s)$

again dictates the existence of information relevant to the expected value of $\underline{\beta}_{t'}$ in observations $\underline{Y}_{t'}, \ldots, \underline{Y}_1$ for $t' > t$. The question then arises, how can this information be used to predict $\underline{\beta}_{t'}$ or a linear function of $\underline{\beta}_{t'}$.

Analogous to the BL estimator, the BL predictor is given by the following

Definition 3.2. The Best Linear (BL) predictor $\underline{\beta}^*_{t'}$ of $\underline{\beta}_{t'}$ is that value of

$$B^*_{t'} = \sum_{i=0}^{t-1} P^t_{t-i} \underline{Y}_{t-i} \tag{3.18}$$

which minimizes the loss function

$$L_2 = E \| \underline{B}^*_{t'} - \underline{\beta}_{t'} \|^2 .$$

Analogous to (3.8) we may write

$$L_2 = \text{trace } [-I \, P^t_t \ldots P^t_1] \begin{bmatrix} r(0) & \psi'_{t'}(t)A(t) \\ A'(t)\psi_{t'}(t) & \Phi_1(t) \end{bmatrix} \begin{bmatrix} -I \\ P^{t'}_t \\ \vdots \\ P^{t'}_1 \end{bmatrix}$$

where

$$\psi'_{t'}(t) = [r(t'-t) r(t'-t+1) \ldots r(t'-1)] .$$

Consequently we find

<u>Theorem 3.3.</u> The values of the matrices $P^t_t, \ldots, P^t_1$ which minimize L_2 where $\underline{B}^*_t$ takes the form of (3.18) satisfy the matrix equation

$$[P^t_t P^t_{t-1} \ldots P^t_1]\Phi_1(t) = \psi'_t(t)A(t).$$

<u>Proof.</u> The proof follows directly from Theorem 1.

Due to the parallel formulation it can be seen that all results obtained in the following chapters pertaining to the BL estimator equally pertain to the BL predictor.

IV. ESTIMATION UNDER FULL RANK EXPERIMENTS

In this chapter further results are given under the constraint $\rho(X_i') = p$, $i = 1, 2, \ldots, t$. These results take the form of:

i) A computational scheme for obtaining the coefficient matrices $K_t^t \ldots K_1^t$ with matrix inversions of order p.

ii) Evaluation of the expression of the covariance matrix of $\underline{\beta}_t$.

iii) Comparison of the BL estimator with the LS estimator and Bayesian estimator with normal prior.

iv) Investigation of bias in connection with the BL estimator.

Before proceeding let the BL estimator be generalized by the following

Definition 4.1. An order - ℓ BL estimator of $\underline{\beta}_t$ is given by the value of

$$\underline{B}_{t,\ell} = \sum_{i=0}^{\ell-1} K_{t-i}^{t,\ell} \underline{\tilde{\beta}}_{t-i}$$

which minimizes

$$E\|\underline{B}_{t,\ell} - \underline{\beta}_t\|^2. \tag{4.1}$$

Denoting this estimator by $\underline{\hat{\beta}}_{t,\ell}$ we see that $\underline{\hat{\beta}}_{t,t} = \underline{\hat{\beta}}_t$. It can be seen that the results of previous chapters relevant to the estimator $\underline{\hat{\beta}}_t$ apply directly to the order - ℓ BL estimator, $\ell \leq t$. This follows from the observation that the problem of obtaining an order-ℓ BL estimator is of the same dimension and formulated in the same manner as that of obtaining the BL estimator $\underline{\beta}_t$ when $t = \ell$.

A generalization of notation is required to deal with the order-ℓ BL estimator. From Definition 4.1 we see

$$\underline{\beta}_{t,\ell} = \sum_{i=0}^{\ell-1} H_{t-i}^{t,\ell} \underline{Y}_{t-i}$$

$$= \sum_{i=0}^{\ell-1} K_{t-i}^{t,\ell} \underline{\tilde{\beta}}_{t-i}$$

where $[H_t^{t,\ell} \ldots H_{t-\ell+1}^{t,\ell}]$ and $[K_t^{t,\ell} \ldots K_{t-\ell+1}^{t,\ell}]$ are such that (4.1) is minimized. Analogus to (3.12) and (3.14) respectively we see that

$$[H_t^{t,\ell} \ldots H_{t-\ell+1}^{t,\ell}] = \psi'(t,\ell)A(t,\ell)\Phi_1^{-1}(t,\ell)$$

and

$$[K_t^{t,\ell} \ldots K_{t-\ell+1}^{t,\ell}] = (H_t^{t,\ell} \ldots H_{t-\ell+1}^{t,\ell})A'(t,\ell)$$

where

$$\psi'(t,\ell) = [r(0)r(1)\ldots r(\ell-1)]$$

and $A'(t,\ell)$ and $\Phi_1(t,\ell)$ are the upper left $N(t,\ell) \times \ell p$ and $\ell p \times \ell p$ matrices of $A'(t)$ and $\Phi_1(t)$ respectively. It follows that

$$[K_t^{t,\ell} \ldots K_{t-\ell+1}^{t,\ell}] = \psi'(t,\ell)A(t,\ell)(A'(t,\ell)R(t,\ell)A(t,\ell)+\sigma^2 I_{N(t,\ell)})^{-1}A'(t,\ell)$$

where $R(t, \ell) = R(\ell)$.

Let us turn now to the investigation of the order-ℓ BL estimator under the requirement that $\rho(X_j') = p$, $j = t-\ell+1, \ldots, t$.

From (2. 3f) and (3. 3) it follows that for $\rho(X_i') = p$, $i = t-\ell+1, \ldots, t$

$$A(t, \ell)A^{+}(t, \ell) = I_{N(t, \ell)} .$$

It then follows from (3. 17) that

$$(4.2) \qquad [K_t^{t, \ell} \ldots K_{t-\ell+1}^{t, \ell}] = \psi'(t, \ell)(R(t, \ell)+\sigma^2(A(t, \ell)A'(t, \ell))^{-1})^{-1}$$

where $(R(t, \ell)+\sigma^2(A(t, \ell)A'(T, \ell))^{-1})^{-1}$ is seen to exist from the discussion following (3. 11). Under these conditions it appears the inversion of a matrix of order ℓp is required in order to obtain $\hat{\underline{\beta}}_{t, \ell}$. It will be seen by the following that this is not the case.

<u>Calculation of $K_t^{t, \ell}, \ldots, K_{t-\ell+1}^{t, \ell}$</u>

In this section the computational method will be given for calculating the coefficient matrices $K_t^{t, \ell}, \ldots, K_{t-\ell+1}^{t, \ell}$. This will initially be done for $\ell = t$ and then adaptations will be made to satisfy the case, $\ell < t$. Let

$$(4.3) \qquad \phi_t = (R(t)+\sigma^2(A(t)A'(t))^{-1})$$

and let

$$\phi_{t+1} = \begin{bmatrix} \phi_{11} & \phi_{1t} \\ \phi_{t1} & \phi_t \end{bmatrix}$$

where

$$\phi_{11} = (r(0) + \sigma^2 (X_{t+1} X'_{t+1})^{-1})$$

and

$$\phi_{1t} = \phi'_{t1} = [r(1) \ldots r(t)].$$

Let us also assume ϕ_t^{-1} is known. Choosing A_{ij} and B_{ij}, $i, j = 1, 2,$ so that

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}$$

we find that

$$\text{(4. 5)} \qquad B_{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1},$$

$$B'_{12} = B_{21} = -A_{22}^{-1} A_{21} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1},$$

$$B_{22} = A_{22}^{-1} + A_{22}^{-1} A_{21} (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1} A_{12} A_{22}^{-1}.$$

From this and (4. 4) it follows that

$$\phi_{t+1}^{-1} = \begin{bmatrix} (\phi_{11} - \phi_{1t}\phi_t^{-1}\phi_{t1})^{-1} & -(\phi_t^{-1}\phi_{t1}(\phi_{11} - \phi_{1t}\phi_t^{-1}\phi_{t1})^{-1})' \\ -\phi^{-1}\phi_{t1}(\phi_{11} - \phi_{1t}\phi_t^{-1}\phi_{t1})^{-1} & \phi_t^{-1} + \phi_t^{-1}\phi_{t1}(\phi_{11} - \phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\phi_{1t}\phi_t^{-1} \end{bmatrix}$$

is obtainable with matrix inversions of order p. From (4. 2) we find

$$K_{t+1}^{t+1} = \psi(t+1) \begin{bmatrix} (\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1} \\ -\phi_t^{-1}\phi_{t1}(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1} \end{bmatrix}$$

$$= r(0)(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1} - \phi_{1t}\phi_t^{-1}\phi_{t1}(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}$$

$$= I - \sigma^2(X_{t+1}X_{t+1}')^{-1}(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}.$$

In addition

$$[K_t^{t+1} \ldots K_1^{t+1}] = \psi(t+1) \begin{bmatrix} -(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\phi_{1t}\phi_t^{-1} \\ \phi_t^{-1} + \phi_t^{-1}\phi_{t1}(\phi_{1t}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\phi_{1t}\phi_t^{-1} \end{bmatrix}$$

$$- r(0)(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\phi_{1t}\phi_t^{-1} + \phi_{1t}\phi_t^{-1}$$

$$+ \phi_{1t}\phi_t^{-1}\phi_{t1}(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\phi_{1t}\phi_t^{-1}$$

$$= \sigma^2(X_{t+1}X_{t+1}')^{-1}(\phi_{11}-\phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\phi_{1t}\phi_t^{-1}$$

$$= (I-K_{t+1}^{t+1})\phi_{1t}\phi_t^{-1}.$$

Since ϕ_1^{-1} is obtained from ϕ_1 by an inversion of order p the derivation is complete.

Let us now turn our attention to the coefficient matrices associated with an order - ℓ BL estimate. Let $\phi^{-1}(t, \ell) = \phi_{t,\ell}^{-1}$ be known. By the above scheme, $\phi^{-1}_{t+1,\ell+1}$ can be determined. Let

$$\text{(4.6)} \qquad \phi^{-1}_{t+1,\ell+1} = \begin{bmatrix} \phi_{t+1,\ell} & \theta'_{t+1,\ell} \\ \theta_{t+1,\ell} & \theta_{\ell,\ell} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

where

$$\theta_{t+1,\ell} = [r'(\ell)r'(\ell-1)\ldots r'(1)]$$

and

$$\theta_{\ell,\ell} = r(0) + \sigma^2 (X_{t-\ell}X'_{t-\ell})^{-1}.$$

Then

$$\phi_{t+1,\ell+1} = \begin{bmatrix} \phi_{t+1,\ell} & \theta'_{t+1,\ell} \\ \theta_{t+1,\ell} & \theta_{\ell,\ell} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$$

and $\phi_{t+1,\ell} = B_{11}$ if and only if $\phi^{-1}_{t+1,\ell} = B^{-1}_{11}$ and from (4.5)

$$\phi^{-1}_{t+1,\ell} = B^{-1}_{11} = (A_{11} - A_{12}A^{-1}_{22}A_{21})$$

which is obtained directly from (4.6). Thus

$$[K^{t+1}_{t+1}\ldots K^{t+1}_{t-\ell+2}] = \psi(t+1,\ell)\phi^{-1}_{t+1,\ell}\,.$$

Let us now consider the covariance matrix of $\hat{\underline{\beta}}_t$. From (4.2) we see

$$E(\underline{\hat{\beta}}_t - \underline{\beta}_t)(\underline{\hat{\beta}}_t - \underline{\beta}_t)' = r(0) - [K_t^t \ldots K_1^t]\phi_t \begin{bmatrix} K_t^{t'} \\ \vdots \\ K_1^{t'} \end{bmatrix}$$

$$= r(0) - \psi'(t)\phi_t^{-1}\psi(t)$$

$$= r(0) - [K_t^t \ldots K_1^t]\psi(t).$$

From the relation

$$[K_t^t \ldots K_1^t]\phi_t = \psi'(t)$$

$$= \psi'(t) + [\sigma^2(X_tX_t')^{-1} 0 \ldots 0] - [\sigma^2(X_tX_t')^{-1} 0 \ldots 0]$$

it follows that

$$[K_t^t \ldots K_1^t] = [\psi'(t) + \sigma^2(X_tX_t')^{-1} 0 \ldots 0)]\phi_t^{-1} - [\sigma^2(X_tX_t')^{-1} 0 \ldots 0]\phi_t^{-1} .$$

Thus

$$[K_t^t \ldots K_1^t]\psi(t) = [I\, 0 \ldots 0]\psi(t) - [\sigma^2(X_tX_t')^{-1} 0 \ldots 0]\phi_t^{-1}\psi(t)$$

$$= r(0) - [\sigma^2(X_tX_t')^{-1} 0 \ldots 0] \begin{bmatrix} K_t^{t'} \\ \vdots \\ K_1^{t'} \end{bmatrix}$$

$$= r(0) - \sigma^2(X_tX_t')^{-1}K_t^{t'} ,$$

and

(4.7)

$$E(\underline{\hat{\beta}}_t - \underline{\beta}_t)(\underline{\hat{\beta}}_t - \underline{\beta}_t)' = \sigma^2(X_tX_t')^{-1}K_t^{t'}$$

$$= \sigma^2(X_tX_t')^{-1} - \sigma^2(X_tX_t')^{-1}(\phi_{11} - \phi_{1t}\phi_t^{-1}\phi_{t1})^{-1}\sigma^2(X_tX_t')^{-1}.$$

Comparing this with the covariance matrix of the LS estimator,

$$E(\tilde{\underline{\beta}}_t - \underline{\beta}_t)(\tilde{\underline{\beta}}_t - \underline{\beta}_t)' = \sigma^2 (X_t X_t')^{-1}$$

we see that the former is smaller by a factor of

$$\sigma^2 (X_t X_t')^{-1} (\phi_{11} - \phi_{1t}\phi_t^{-1}\phi_{t1})^{-1} \sigma^2 (X_t X_t')^{-1}.$$

Thus, the greater the information contained in the past data, as reflected by the covariance function $r(s)$, the smaller the value of K_t^t and the greater the improvement over the LS estimator.

It should be noted, however, that this comparison is in a sense inappropriate as the LS estimator is not dependent upon assumptions concerning the mean and variance of the parameters while the BL estimator is. A more relevent comparison can be made between the BL estimator and a closer relative, the Bayes estimator under a normal prior with $E(\underline{\beta}_t) = 0$, $E(\underline{\beta}_t \underline{\beta}_t') = r(0)$.

Here the Bayes estimator is (Raiffa and Schlaifer, 1961)

$$\underline{\beta}_t^* = r(0)(r(0) + \sigma^2 (X_t X_t')^{-1})^{-1} \tilde{\underline{\beta}}_t$$

and has covariance matrix

$$\begin{aligned} & E(\underline{\beta}_t^* - \underline{\beta}_t)(\underline{\beta}_t^* - \underline{\beta}_t)' \\ & = \sigma^2 (X_t X_t')^{-1} - \sigma^2 (X_t X_t')^{-1} (r(0) + \sigma^2 (X_t X_t')^{-1})^{-1} \sigma^2 (X_t X_t')^{-1}. \end{aligned}$$

A qualitative comparison of these two covariance matrices can be made using the following

Definition 4.2. Let A, B be square matrices of the same order where $A > B$ means $A - B$ is positive definite and $A \geq B$ means $A - B$ is non-negative definite.

We now have the following

Lemma 4.1. Let $A, B, A-B$ be positive definite. Then $(A-B)^{-1} > A^{-1}$.

Proof. $(A-B)^{-1} - A^{-1} = -A^{-1}(A^{-1}-B^{-1})^{-1}A^{-1}$.

If one of the following are true all are true.

$$-A^{-1}(A^{-1}-B^{-1})^{-1}A^{-1} \geq 0,$$

$$A^{-1}(B^{-1}-A^{-1})^{-1}A^{-1} \geq 0,$$

$$(B^{-1}-A^{-1})^{-1} \geq 0,$$

$$(B^{-1}-A^{-1}) \geq 0.$$

That the latter is true is found in Rao (1965), thus $(A-B)^{-1} \geq A^{-1}$ but only if $B = 0$ can the equality hold. Thus $(A-B)^{-1} > A^{-1}$.

Theorem 4.1. $E(\underline{\beta}_t^*-\underline{\beta}_t)(\underline{\beta}_t^*-\underline{\beta}_t)' > E(\hat{\underline{\beta}}_t-\underline{\beta}_t)(\hat{\underline{\beta}}_t-\underline{\beta}_t)'$.

Proof. This theorem follows directly from Lemma 4.1 with

$A = (r(0)+\sigma^2(X_t X_t')^{-1}) = \phi_{11}$ and $B = \phi_{1t}\phi_t^{-1}\phi_{t1}$.

Here again the greater the information contained in past data the greater the improvement of the BL estimator over the Bayes estimator.

It should also be noted that since $\phi_{11} = (r(0)+\sigma^2(X_t X_t')^{-1})$ if $\phi_{1t}\phi_t^{-1}\phi_{t1} = 0$, that is if $r(s) = 0,\ s > 0$, the BL estimator reduces to the Bayes estimator with normal prior with $E(\underline{\beta}_t) = 0$ and $E(\underline{\beta}_t\ \underline{\beta}_t') = r(0)$. This is equivalent to saying that the order-1 BL estimator is the Bayes estimator. From this point of view the multiple lag BL estimator might be thought of as a Bayesian estimator with a prior given in the form of a stochastic process.

Bias

Consideration of the property of bias under the assumption that the mean of the process is zero leads to the observation that all estimators linear in the observations are unconditionally unbiased if the observational errors have expectation zero. However, if one wishes to calculate estimator bias for the process at time t, the property which is relevant is that of conditional bias, conditional on time t. Considering this we find that in general the BL estimator is not conditionally unbiased. This follows from the observation that the order-1 BL estimator is identical to the conditionally biased Bayes estimator with normal prior. In this particular case the amount of bias can be determined if so desired. Such is not the case if we are considering BL estimators of order greater than one. This follows from the observation that such an estimator involves information related to the parameters $\underline{\beta}_i$, $i < t$, and if $\underline{\beta}_i \neq \underline{\beta}_t$ no simple expectation relation exists between past observations and $\underline{\beta}_t$. Thus no meaningful quantitative evaluation of bias can be made in this case.

V. ESTIMATION OF $r(s)$ AND $\overline{\underline{\beta}}$

Estimation of $r(s)$

In this chapter a consistent estimator of the autocovariance function $r(s)$ for $s = 0, \pm 1, \pm 2, \ldots, \pm t-1$ is given when either $\underline{\beta}_i$ or its unbiased LS estimate $\tilde{\underline{\beta}}_i$ is known for $i = 1, 2, \ldots, t$. Obviously, this last condition requires $\rho(X_i') = p$ for $i = 1, 2, \ldots, t$. Also a minimum variance linear unbiased estimator of $\overline{\underline{\beta}}$ is given.

In the development of the estimator of $r(s)$ we shall adapt the method of Lomnicki and Zaremba (1957) dealing with the estimation of autocorrelation in a one-dimensional time series. It will be shown that the estimator of the autocovariance function in the present case is mathematically analogous to the estimator given by Lomnicki and Zaremba in the one-dimensional case. Consequently, their results may be adapted directly.

Let $\{\underline{\eta}_t\}$ be a p-dimensional discrete stationary stochastic process with

$$E(\underline{\eta}_t) = 0, \quad E(\underline{\eta}_t \underline{\eta}_t') = \xi^2 I_p, \quad E(\underline{\eta}_t \underline{\eta}_s') = 0, \quad s \neq t.$$

Let $\sum_{s=0}^{\infty} m_s^{ij}$ be absolutely convergent for each i, j (where $M_s = [m_s^{ij}]$, $i, j = 1, 2, \ldots, p$). Then the sequence $\{\underline{Z}_t\}$ defined by the moving average

(5.1) $$\underline{Z}_t = \sum_{s=0}^{\infty} M_s \underline{\eta}_{t-s}$$

is a stationary stochastic process (Whittle, 1963). Assuming that this process differs from the stationary process $\{\underline{\beta}_t\}$ only by a constant $\overline{\underline{\beta}}$, the mean value of the $\{\underline{\beta}_t\}$ process, which will be assumed non-zero and unknown, we have

(5.2) $$\underline{\beta}_t = \overline{\underline{\beta}} + \underline{Z}_t \,.$$

In addition, the following assumptions on the series $\{\underline{\eta}_t\}$ are made:

(5.3) i) All finite moments of $\underline{\eta}_t^i$ exist and are finite where

$$\underline{\eta}_t = \begin{bmatrix} \eta_t^1 \\ \vdots \\ \eta_t^p \end{bmatrix}$$

ii) $$E(\underline{\eta}_{t_1}^{i_1} \underline{\eta}_{t_2}^{i_2} \cdots \underline{\eta}_{t_m}^{i_m}) = E(\underline{\eta}_{t_1+r}^{i_1} \underline{\eta}_{t_2+r}^{i_2} \cdots \underline{\eta}_{t_m+r}^{i_m})$$

where

$t_1, t_2, \ldots, t_m, r$ and $0 \leq i_1, i_2, \ldots, i_m \leq p$ are integers.

iii) The moments of the process $\{\underline{\eta}_t\}$ behave as if they were independent. Namely:

$$E((\underline{\eta}_{t_1}^{i_1})^{\lambda_1} (\underline{\eta}_{t_2}^{i_2})^{\lambda_2} \cdots (\underline{\eta}_{t_m}^{i_m})^{\lambda_m}) = E(\underline{\eta}_{t_1}^{i_1})^{\lambda_1} E(\underline{\eta}_{t_2}^{i_2})^{\lambda_2} \cdots E(\underline{\eta}_{t_m}^{i_m})^{\lambda_m}$$

with $t_1, \ldots, t_m$, $i_1, \ldots, i_m$ as in (ii) and $\lambda_1, \ldots, \lambda_m$ any set of non-negative integers.

It follows from (5. 1) that

$$r(\ell) = \mathrm{Cov}(\underline{\beta}_t, \underline{\beta}_{t+\ell}) = \mathrm{Cov}(\underline{Z}_t, \underline{Z}_{t+\ell}) = \xi^2 \sum_{s=0}^{\infty} M_s M'_{s+\ell}$$

where $\mathrm{Cov}(\underline{\beta}_t, \underline{\beta}_{t+\ell})$ represents the matrix of covariances of the individual elements of the vectors $\underline{\beta}_t$ and $\underline{\beta}_{t+\ell}$.

Letting

$$\underline{\beta}_t = \begin{bmatrix} \beta_t^1 \\ \beta_t^2 \\ \vdots \\ \beta_t^p \end{bmatrix}, \quad M_s = \begin{bmatrix} M_s^1 \\ \vdots \\ M_s^p \end{bmatrix}, \quad \underline{Z}_t = \begin{bmatrix} Z_t^1 \\ \vdots \\ Z_t^p \end{bmatrix}$$

and

$$r(\ell) = [r^{ij}(\ell)], \quad i, j = 1, \ldots, p,$$

we see that

$$Z_t^i = \sum_{s=0}^{\infty} M_s^i \underline{\eta}_{t-s} \tag{5. 4}$$

and

$$r^{ij}(\ell) = \xi^2 \sum_{s=0}^{\infty} M_s^i M_{s+\ell}^{j'} \, . \tag{5. 5}$$

Let us define

$$(5.6) \qquad r^{ij}(\ell, N) = \frac{1}{N-\ell} \sum_{s=1}^{N-\ell} Z_s^i Z_{s+\ell}^j$$

and

$$(5.7) \qquad C^{ij}(\ell, N) = \frac{1}{N-\ell} \sum_{s=1}^{N-\ell} (\beta_s^i - \overline{\overline{\beta}}{}_s^i)(\beta_{s+\ell}^j - \overline{\overline{\beta}}{}_{s+\ell}^j)$$

where

$$(5.8) \qquad \overline{\overline{\beta}}{}_{s+q}^i = \frac{1}{N-\ell} \sum_{j=1+q}^{N-\ell+q} \overline{\beta}_j^i \qquad \text{for} \quad q = 0, \ell.$$

The above assumptions and definitions are analogous to those presented by Lomnicki and Zaremba with the exception of (5.4) and (5.5) which are expressed in matrix notation as opposed to scaler. However, by reindexing it can be seen that

$$(5.9) \qquad Z_t^i = \sum_{s=0}^{\infty} M_s^i \underline{\eta}_{t-s} = \sum_{s=0}^{\infty} \sum_{j=1}^{p} m_s^{ij} \eta_{t-s}^i = \sum_{s'=0}^{\infty} m_{s'}^i \eta_{t-s'}$$

where the elements of the set $\{s': 0, 1, 2, \ldots\}$ correspond one-to-one to those of the set $\{(s, j) : (0,1), \ldots, (0,p), (1,1), \ldots, (1,p), (2,1), \ldots\}$. It follows that

$$r^{ij}(\ell) = \xi^2 \sum_{s'=0}^{\infty} m_{s'}^i m_{s'+p\ell}^{j'} \,.$$

Moreover in view of the absolute convergence of

$\sum_{s=0}^{\infty} m_s^{ij}$, $i, j = 1,\ldots,p$, $\sum_{s'=0}^{\infty} m_{s'}^{i}$ and consequently $\sum_{s=-\infty}^{\infty} r^{ij}(\ell)$ is absolutely convergent. Using the above in addition to the obvious assumptions on the one-dimensional $\{\eta_{s'}\}$ process as derived from those on the p-dimensional $\{\underline{\eta}_s\}$ process it is seen that the results due to Lomicki and Zaremba can be adapted to the present situation.

Some results of immediate interest are:

a) $r^{ij}(\ell, N)$ is unbiased for $r^{ij}(\ell)$.

b) $r^{ij}(\ell, N)$ is a consistent estimator of $r^{ij}(\ell)$ if $M_s = 0$, $s < 0$.

c) $C^{ij}(\ell, N)$ is an asymptotically unbiased estimator of $r^{ij}(\ell)$ and the bias is of order N^{-1}.

d) $C^{ij}(\ell, N)$ is a consistent estimator of $r^{ij}(\ell)$.

All these results then also apply for the matrices

$$r(\ell, N) = [r^{ij}(\ell, N)] \quad \text{and} \quad C(\ell, N) = [C^{ij}(\ell, N)],$$
$$i, j = 1, \ldots, p.$$

Above we have given an estimator of the covariance function when the value of the process is known at times $s = 1, 2, \ldots, t$. However, let us consider the estimation of $r(\ell)$ when we only have estimates of $\underline{\beta}_s$, $s = 1, 2, \ldots, t$. Suppose $\rho(X_s') = p$, $s = 1, 2, \ldots, t$. Then $\tilde{\underline{\beta}}_s = (X_s X_s')^{-1} X_s \underline{Y}_s$, the LS estimate of $\underline{\beta}_s$ exists. Setting

$$(5.10) \qquad \underline{\tilde{\beta}}_t = \underline{\beta} + \underline{\tilde{Z}}_t$$

we find

$$(5.11) \qquad \underline{\tilde{Z}}_t = \sum_{s=0}^{\infty} M_s \underline{\eta}_{t-s} + \underline{\Delta}_t$$

where

$$(5.12) \qquad \underline{\Delta}_t = \underline{\tilde{\beta}}_t - \underline{\beta}_t.$$

Let us define

$$(5.13) \qquad \underline{\eta}_t^* = \frac{\xi}{\sigma} U_t^{-1/2} \underline{\Delta}_t$$

and

$$(5.14) \qquad M_t^* = \frac{\sigma}{\xi} U_t^{1/2}$$

where

$$(5.15) \qquad E(\underline{\Delta}_t \underline{\Delta}_t') = \sigma^2 U_t, \quad E(\underline{\Delta}_t \underline{\Delta}_s') = 0, \quad s \neq t$$

from the assumptions of LS regression analysis.

It is assumed that assumptions (5.3i), (5.3ii) and (5.3iii) pertaining to the $\{\underline{\eta}_t\}$ process are also valid assumptions on the $\{\{\underline{\eta}_t\}, \underline{\eta}_t^*\}$ process. Also it can be seen that $\underline{\eta}_t^*$ is independent from all $\underline{\eta}_s$. From (5.11), (5.13) and (5.14) it follows that

$$\underline{\tilde{Z}}_t = \sum_{s=0}^{\infty} M_s \underline{\eta}_{t-s} + M_t^* \underline{\eta}_t^* .$$

Corresponding to the reindexing of (5.9)

$$\underline{\tilde{Z}}_t^i = \sum_{s'=0}^{\infty} M_{s'}^i \eta_{t-s'} + \sum_{j=1}^{p} m_t^{ij} \eta_t^{*j} = \sum_{s'=-p}^{\infty} m_{s'}^i \eta_{t-s'}$$

where $m_{-p}^i, m_{-p+1}^i, \ldots, m_{-1}^i, \eta_{t+p}, \eta_{t+p}, \ldots, \eta_{t+1}$ are understood to be $m_t^{*i1}, m_t^{*i2}, \ldots, m_t^{*ip}, \eta_t^{*1}, \eta_t^{*2}, \ldots, \eta_t^{*p}$ respectively for fixed t.

It then follows from (5.10-5.15) that

(5.16)
$$\mathrm{Cov}(\underline{\tilde{\beta}}_t, \underline{\tilde{\beta}}_{t+\ell}) = \mathrm{Cov}(\underline{\tilde{Z}}_t, \underline{\tilde{Z}}_{t+\ell})$$

$$= \xi^2 \sum_{s'=0}^{\infty} M_{s'} M'_{s'+\ell} + \delta_\ell \sigma^2 U_t$$

$$= r(\ell) + \delta_\ell \sigma^2 U_t,$$

where

$$\delta_\ell = \begin{cases} 1, & \text{if } \ell = 0 \\ 0, & \text{if } \ell \neq 0 \end{cases},$$

and

$$U_t = (X_t X_t')^{-1}.$$

Define analogous to (5.6-5.8)

(5.17)
$$\tilde{r}^{ij}(\ell, N) = \frac{1}{N-\ell} \sum_{t=1}^{N-\ell} (\tilde{Z}_t^i \tilde{Z}_{t+\ell}^j - \delta_\ell \sigma^2 U_t^{ij})$$

and

$$C^{ij}(\ell, N) = \frac{1}{N-\ell} \sum_{t=1}^{N-\ell} (\tilde{\beta}_t^i - \bar{\tilde{\beta}}_t^i)(\tilde{\beta}_t^j - \bar{\tilde{\beta}}_t^j)' - S_N^2 U_t^{ij}$$

where

$$\bar{\tilde{\beta}}^i_{t+s} = \frac{1}{N-R} \sum_{j=1+s}^{N-\ell+s} \tilde{\beta}^i_j \qquad \text{for} \quad s = 0, \ell$$

and where

$$S^2_N = \frac{1}{\sum_{t=1}^{N} (n_t - p)} \sum_{t=1}^{N} (n_t - p) s^2_t$$

with

$$s^2_t = \frac{1}{n_t - p} (\underline{Y}_t - X'_t \underline{\tilde{\beta}}_t)'(\underline{Y}_t - X'_t \underline{\tilde{\beta}}_t).$$

Theorem 5. 1.

a) $\tilde{r}^{ij}(\ell, N)$ is unbiased for $r^{ij}(\ell)$.

b) $\tilde{r}^{ij}(\ell, N)$ is a consistent estimator of $r^{ij}(\ell)$.

c) $\tilde{C}^{ij}(\ell, N)$ is an asymptotically unbiased estimator of $r^{ij}(\ell)$ and the bias is of order N^{-1}.

d) $\tilde{C}^{ij}(\ell, N)$ is a consistent estimator of $r^{ij}(\ell)$.

Proof. For $\ell \neq 0$: $\tilde{r}^{ij}(\ell, N)$ and $\tilde{C}^{ij}(\ell, N)$ are exactly analogous to $r^{ij}(\ell, N)$ and $C^{ij}(\ell, N)$ with Z^i_t replaced by $\tilde{Z}^i_t$ and the theorem follows from the results of Lomnicki and Zaremba.

For $\ell = 0$: S^2_N is an unbiased and consistent estimator of σ^2. Thus $r^{ij}(0, N)$ has the same asymptotic distribution as

$$\tilde{\tilde{r}}^{ij}(0,N) = \frac{1}{N}\sum_{t=1}^{N}(\tilde{Z}_t^i\tilde{Z}_t^j - \sigma^2 U_t^{ij}).$$

Also it is apparent that

$$\mathrm{Var}(\tilde{\tilde{r}}^{ij}(0,N)) = \mathrm{Var}\left(\frac{1}{N}\sum_{t=1}^{N}(\tilde{Z}_t^i\tilde{Z}_t^j)\right).$$

Consider

$$\lim_{N\to\infty} N\,\mathrm{Cov}\left(\frac{1}{N}\sum_{t=1}^{N}\tilde{Z}_t^i\tilde{Z}_t^j,\ \frac{1}{N}\sum_{s=1}^{N}\tilde{Z}_{s+\ell}^i\tilde{Z}_{s+\ell}^j\right)$$

$$= \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}\sum_{t=1}^{N}\mathrm{Cov}(\tilde{Z}_t^i\tilde{Z}_t^j,\ \tilde{Z}_{t+\ell}^i\tilde{Z}_{t+\ell}^j)$$

where

(5.18) $$\mathrm{Cov}(\tilde{Z}_t^i\tilde{Z}_t^j, \tilde{Z}_{t+\ell}^i\tilde{Z}_{t+\ell}^i)$$

$$= \sum_{a'=-\infty}^{\infty}\sum_{b'=-\infty}^{\infty}\sum_{c'=-\infty}^{\infty}\sum_{d'=-\infty}^{\infty} m_{a'}^i m_{b'}^j m_{c'}^i m_{d'}^j E(\eta_{t-a'}\eta_{t-b'}\eta_{t-c'}\eta_{t-d'})$$

$$- (r^{ij}(0) + \sigma^2 U_t^{ij})(r^{ij}(0) + \sigma^2 U_{t+\ell}^{ij})$$

$$= \sum_{r=-\infty}^{\infty} m_{t-r}^i m_{t-r}^i m_{t+\ell p-r}^i m_{t+\ell p-r}^i E(\eta_r^4)$$

$$+ \sum_{r=-\infty}^{\infty}\ \sum_{\substack{q=-\infty \\ r\neq q}}^{\infty} m_{t-r}^i m_{t-r}^j m_{t+\ell p-q}^i m_{t+\ell p-q}^j E(\eta_r^2)E(\eta_q^2) \ +$$

$$+ \sum_{\substack{r=-\infty \\ r \neq q}}^{\infty} \sum_{q=-\infty}^{\infty} m^i_{t-r} m^j_{t-q} m^i_{t+\ell p-r} m^j_{t+\ell p-q} E(\eta_r^2) E(\eta_q^2)$$

$$+ \sum_{\substack{r=-\infty \\ r \neq q}}^{\infty} \sum_{q=-\infty}^{\infty} m^i_{t-r} m^j_{t-q} m^i_{t+\ell p-q} m^j_{t+\ell p-r} E(\eta_r^2) E(\eta_q^2)$$

$$- (r^{ij}(0) + \sigma^2 U^{ij}_t)(r^{ij}(0) + \sigma^2 U^{ij}_{t+\ell}).$$

Recalling that $E(\eta_{a'}\eta_{b'}) = 0$ for $t+1 < a' < t+p$ and $-\infty \leq b' \leq t$ and by adding and subtracting

$$\xi^4 \sum_{r=-\infty}^{\infty} m^i_{t-r} m^j_{t-r} m^i_{t+\ell p-r} m^j_{t+\ell p-r}$$

to the second, third and fourth terms of (5. 18) we see that

$$\mathrm{Cov}\,(\tilde{Z}^i_t \tilde{Z}^j_t, \tilde{Z}^i_{t+\ell} \tilde{Z}^j_{t+\ell})$$

$$= \sum_{r=-\infty}^{t} m^i_{t-r} m^j_{t-r} m^i_{t+\ell p-r} m^j_{t+\ell p-r} E(\eta_r^4) + \delta_\ell \sum_{r=t+1}^{t+p} m^i_{t-r} m^j_{t-r} m^i_{t-r} m^j_{t-r} E(\eta_r^4)$$

$$+ (r^{ij}(0) + \sigma^2 U^{ij}_t)(r^{ij}(0) + \sigma^2 U^{ij}_{t+\ell}) - \xi^4 \sum_{r=-\infty}^{\infty} m^i_{t-r} m^j_{t-r} m^i_{t+\ell p-r} m^j_{t+\ell p-r}$$

$$+ (r^{ij}(\ell) + \delta_\ell \sigma^2 U^{ij}_t)(r^{ij}(\ell) \_ \delta_\ell \sigma^2 U^{ij}_t) - \xi^4 \sum_{r=-\infty}^{\infty} m^i_{t-r} m^j_{t-r} m^i_{t+\ell p-r} m^j_{t+\ell p-r} \div$$

$$+ (r^{ij}(\ell)+\delta_\ell \sigma^2 U_t^{ij})(r^{ij}(\ell)+\delta_\ell \sigma^2 U_t^{ij}) - \xi^4 \sum_{r=-\infty}^{\infty} m_{t-r}^i m_{t-r}^j m_{t+\ell p-r}^i m_{t+\ell p-r}^j$$

$$- (r^{ij}(0)+\sigma^2 U_t^{ij})(r^{ij}(0)+\sigma^2 U_{t+\ell}^{ij})$$

$$= \delta^4 \sum_{r=-\infty}^{\infty} m_{t-r}^i m_{t-r}^j m_{t+\ell p-r}^i m_{t+\ell p-r}^j + \delta_\ell \sum_{r=t+1}^{t+p} m_{t-r}^i m_{t-r}^j m_{t-r}^i m_{t-r}^j E(\eta_t^4)$$

$$+ (r^{ij}(\ell)+\delta_\ell \sigma^2 U_t^{ij})^2 + (r^{ij}(\ell)+\delta_\ell \sigma^2 U_t^{ij})(r^{ij}(\ell)+\delta_\ell \sigma^2 U_t^{ij})$$

where

$$E(\eta_r^4) = \kappa_4 + 3\xi^4.$$

Thus from (28) of Lomnicki and Zaremba

$$\lim_{N\to\infty} N \operatorname{Var}(\tilde{\tilde{r}}^{ij}(0, N))$$

$$= \frac{\kappa_4}{\xi^4}[r^{ij}(0)]^2 + 2 \sum_{q=-\infty}^{\infty} [r^{ij}(q)]^2$$

$$+ \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} \sum_{r=t-1}^{t+p} m_{t-r}^i m_{t-r}^j m_{t-r}^i m_{t-r}^j E(\eta_t^4) + (\sigma^2 U_t^{ij})^2$$

$$+ \sigma^4 U_t^{ii} U_t^{jj} + 2\sigma^2 r^{ij}(0) U_t^{ij} + \sigma^2 r^{ii}(0) U_t^{jj} + \sigma^2 r^{jj}(0) U_t^{ii} \quad .$$

Since $\rho(X_t') = p$, U_t^{ii} is finite and thus so is $m_{s'}^i$, $-p < s' < -1$. Let U_t^{ii} and $m_{s'}^i \le M$. Then

$$\lim_{N \to \infty} N \operatorname{Var} (\tilde{\tilde{r}}^{ij}(0, N))$$

$$\leq \frac{\kappa_4}{\xi^4} [r^{ij}(0)]^2 + 2 \sum_{q=-\infty}^{\infty} (r^{ij}(q))^2$$

$$+ \lim \frac{1}{N} \sum_{t=1}^{N} (pM^4 E(\eta_t^4) + 2\sigma^4 M^2 + \sigma^2 M(r^{ij}(0) + 2r^{ij}(0) + r^{jj}(0)))$$

which is finite.

Thus $\tilde{\tilde{r}}^{ij}(0, N)$ is consistent which in turn implies $\tilde{r}^{ij}(0, N)$ is consistent. Also from (5. 16) and (5. 17) $\tilde{\tilde{r}}^{ij}(0, N)$ is unbiased. In addition $\tilde{C}^{ij}(0, N)$ differs from $\tilde{r}^{ij}(0, N)$ by exactly the same algebraic amount as $C^{ij}(0, N)$ differs from $r^{ij}(0, N)$ and thus $\tilde{C}^{ij}(0, N)$ is asymptotically unbiased and consistent for $r^{ij}(0)$.

Let us turn now to the estimation of $\underline{\bar{\beta}}$.

Estimation of $\underline{\bar{\beta}}$

Previously we have assumed that the mean value $\underline{\bar{\beta}}$, of the stochastic process $\{\mathcal{B}(\tau)\}$ was known and thus without loss of generality could assume that $\underline{\bar{\beta}} = 0$. However, as this will not always be the case, let us consider the estimation of $\underline{\bar{\beta}}$.

At time t the regression model may be written

$$\underline{Y}_t = X_t' \underline{\bar{\beta}}_t + X_t'(\underline{\beta}_t - \underline{\bar{\beta}}) + \underline{\epsilon}_t .$$

Letting $\underline{\bar{\beta}}(t) = (\underline{\bar{\beta}}' \ldots \underline{\bar{\beta}}')'$ be a $tp \times 1$ vector and $\underline{Y}(t)$, $A'(t)$, $\underline{\beta}(t)$ and $\underline{\epsilon}(t)$ be as previously defined we may write a more general regression model as

(5.19) $$\underline{Y}(t) = A'(t)\underline{\bar{\beta}}(t) + A'(t)(\underline{\beta}(t) - \underline{\bar{\beta}}(t)) + \underline{\epsilon}(t)$$

Let us consider the following theorem:

Theorem 5. . Under model (5.1) with

(5.20) $$E(\underline{\epsilon}(t)) = \underline{0}, \quad E(\underline{\epsilon}(t)\underline{\epsilon}'(t)) = \sigma^2 I$$

(5.21) $$E(\underline{\epsilon}(t)\beta'(t)) = 0$$

and

(5.22) $$E(\underline{\bar{\beta}}(t)(\underline{\beta}(t) - \underline{\bar{\beta}}(t))') = 0$$

the minimum variance linear unbiased estimate of $\underline{\bar{\beta}}$ is given by

$$\underline{\bar{\beta}} = (X(t)X'(t))^{-1}X(t)\underline{Y}(t)$$

when $X(t) = [X_t, X_{t-1}, \ldots, X_1]$ is of rank p.

Proof. Let $\underline{\bar{\beta}}^* = C\underline{Y}(t)$ be a general linear function of $\underline{Y}(t)$ where C is a $p \times N(t)$ constant matrix. For $\underline{\bar{\beta}}^*$ to be unbiased we must have $C = (X(t)X'(t))^{-1}X(t) + D$ where $D = [D_t, D_{t-1}, \ldots, D_1]$ and D_i is a $p \times n_i$ constant matrix and

$$\sum_{i=1}^{t} D_i X_i' = 0.$$

This follows from

$$E(\underline{\overline{\beta}}^*) = E(C\underline{Y}(t))$$

$$= E[((X(t)X(t))^{-1}X(t) + D)(A'(t)\underline{\overline{\beta}}(t) + A'(t)(\underline{\beta}(t) - \underline{\overline{\beta}}(t)) + \underline{\epsilon}(t))]$$

$$= (\sum_{i=1}^{t} X_i X_i')^{-1} (\sum_{i=1}^{t} X_i X_i')\underline{\overline{\beta}} + \sum_{i=1}^{t} D_i X_i' \underline{\overline{\beta}}$$

$$= \underline{\overline{\beta}} .$$

We now must find D such that the resulting estimator has minimum variance in that the diagonal elements of $E(\underline{\overline{\beta}}^* - \underline{\overline{\beta}})(\underline{\overline{\beta}}^* - \underline{\overline{\beta}})'$ are minimized. Using (5. 20- 5. 22) and

$$\sum_{i=1}^{t} D_i X_i' = 0$$

and letting $(X(t)X'(t)) = S$ we find

$$E(\underline{\overline{\beta}}^* - \underline{\overline{\beta}})(\underline{\overline{\beta}}^* - \underline{\overline{\beta}})' \tag{5. 23}$$

$$= E(S^{-1}X(t)A'(t)\underline{\beta}(t) + S^{-1}X(t)\underline{\epsilon}(t) + D\underline{\epsilon}(t))$$

$$\times (S^{-1}X(t)A'(t)\underline{\beta}(t) + S^{-1}X(t)\underline{\epsilon}(t) + D\underline{\epsilon}(t))' =$$

$$= E(S^{-1}X(t)A'(t)\underline{\beta}(t)\underline{\beta}'(t)A(t)X'(t)S^{-1} + S^{-1}X(t)\underline{\epsilon}(t)\underline{\epsilon}'(t)X'(t)S^{-1}$$

$$+ D\underline{\epsilon}(t)\underline{\epsilon}'(t)D')$$

$$= S^{-1}X(t)A'(t)E(\underline{\beta}(t)\underline{\beta}'(t))A(t)X'(t)S^{-1} + \sigma^2(S^{-1} + DD').$$

As $G = DD'$ is positive semi-definite, its diagonal elements $g_{ii} \geq 0$. Thus the diagonal elements of (5.23) are individually minimized when $D = 0$. Thus $C = (X(t)X'(t))^{-1}X(t)$ and the theorem is proved.

VI. A NUMERICAL EXAMPLE

In order to study the effectiveness of the BL estimation procedure the following example was simulated on the CDC 3300 computer. A discrete three-dimensional wide-sense stationary stochastic process was generated with the following distributional characteristics.

$$E(\underline{\beta}_t) = 0$$

and

$$E(\underline{\beta}_t \underline{\beta}'_{t+s}) = r(s), \quad s = 0, 1, \ldots, k-1.$$

This was accomplished by generating a sequence $\{\underline{e}_i\}_{i=1}^{\infty}$ of three-dimensional variates with

$$E(\underline{e}_i) = 0$$

and

$$E(\underline{e}_i \underline{e}'_{i+j}) = \delta_j I.$$

Then the elements of the sequence $\{\underline{\gamma}_i\}_{i=1}^{\infty}$ with $\underline{\gamma}_i = (r(0))^{1/2}\underline{e}_i$ have

$$E(\underline{\gamma}_i) = \underline{0}$$

and

$$E(\underline{\gamma}_i \underline{\gamma}'_{i+j}) = \delta_j r(0).$$

Letting

$$\underline{\beta}_t = \sum_{i=1}^{k} a_i \underline{Y}_{t+i-1}$$

with

$$\sum_{i=1}^{k} a_i^2 = 1$$

it follows that the elements of the sequence $\{\underline{\beta}_t\}_{t=1}^{\infty}$ have

$$E(\underline{\beta}_t) = 0$$

and

$$(6.1) \qquad E(\underline{\beta}_t \, \underline{\beta}'_{t+s}) = (\sum_{i=1}^{k-s} a_i a_{i+s}) r(0) = r(s).$$

For the purposes of this example the parameters used were $k = 16$, $a_i = .25$ for $i = 1, \ldots, 16$ and

$$r(0) = \begin{bmatrix} .010 & .008 & .008 \\ .008 & .010 & .008 \\ .008 & .008 & .010 \end{bmatrix} .$$

From (6.1)

$$r(s) = (\frac{16-s}{16}) r(0).$$

These values were selected to yield a process $\{\underline{\beta}_t\}_{t=1}^{\infty}$ which was reasonably smooth (see Figures 1-5).

Using the generated sequence $\{\underline{\beta}_t\}_{t=1}^{\infty}$, for each value t a $n_t \times 1$ vector of observation

$$\underline{Y}_t = X_t' \underline{\beta}_t + \underline{\epsilon}_t$$

was generated where

$$E(\underline{\epsilon}_t) = \underline{0}$$

and

$$E(\underline{\epsilon}_t \underline{\epsilon}_{t+s}') = \delta_s(\sigma^2 I).$$

The third step in the simulation was to use these data to estimate the unknown parameter vector $\underline{\beta}_t$ utilizing the following estimators:

(6. 2) i) Least Squares, $n_t = 4$

ii) Bayes with normal prior, $n_t = 4$

iii) Best Linear estimator

a. order-3, $n_t = 4$

b. order-3, $n_t = 2$

c. order-5, $n_t = 1$.

$n_t = 4$ was chosen as the basic size of the individual experiment to permit the estimation of $\underline{\beta}_t$ by LS but yet allow for the evaluation and comparison of the effectiveness of the BL estimator when the quantity of data was a minimum. The design matrix X_t'

was the same for each experiment so that the quantity $E\|\underline{\beta}_t(\text{est})-\underline{\beta}_t\|^2$ associated with a given estimator was the same at each time point and thus could be estimated from the corresponding finite sequence of estimates $\{\underline{\beta}_t(\text{est})\}_{t=1}^{n}$. The particular design matrix used was chosen so that the absolute value of any element was less than ten and such that the inner product of its columns were small. This latter restriction contributes to the decrease of the covariance matrix of the LS estimator (Rao, 1965a). With regard to the former restriction, it is observed that the covariance matrix of the LS, Bayesian and BL estimators are a function of the design matrix and σ^2 only in the form $\sigma^2(X_tX_t')^{-1}$. Thus it is not the absolute magnitude of the estimates but their value relative to σ^2. Hence, for a given covariance structure, $r(s)$, the comparitive effectiveness of the various estimators under different ratios of X_t' and σ^2 can be studied by holding X_t' fixed and varying the value of σ^2.

Three independent sequences were generated as were corresponding sequences of experiments, each with one of the following values of σ^2: 0.09, 0.25, and 0.64. Label these sequences of experiments respectively A, B and C. The following parameters were estimated: the value of $\underline{\beta}_t$ for the first 25 time points in sequences A, B and C by the estimators indicated in (6.2i), (6.2ii) and (6.2iiia) and the additional estimators indicated in (6.2iiib) and (6.2iiic) for sequence B only. For each of these cases the quantity

Ave $\|\underline{\beta}_t(\text{est}) - \underline{\beta}_t\|^2$ was calculated to compare the effectiveness of the several estimators under varying conditions. The results are found in Table 1.

Table 1. Ave $\|\underline{\beta}_t(\text{est}) - \underline{\beta}_t\|^2$.

Sequence	Estimators				
	Least Squares $n_t = 4$	Bayes $n_t = 4$	Best Linear $n_t = 4$ order-3	Best Linear $n_t = 2$ order-3	Best Linear $n_t = 1$ order-5
A ($\sigma^2 = .09$)	.0071	.0070	.0036	-----	-----
B ($\sigma^2 = .25$)	.0144	.0076	.0044	.0077	.0091
C ($\sigma^2 = .64$)	.0405	.0121	.0059	-----	-----

As seen in Table 1 in all three sequences the BL estimator of order-3 yields substantial improvement over both the LS and Bayes estimators based on the same data with the degree of improvement in the cases given here depending on the value of σ^2. The fact that these sequences are independent gives support to the contention that this improvement is real and not a function of the particular sequence chosen. It is also seen that the BL estimators with $n_t = 2$ and $n_t = 1$ are improvements over the LS estimator in sequence B. This result, while not unexpected, is pleasing as the initial purpose of the investigation was to derive an estimator of the present value of a parameter vector varying according to a stationary stochastic process when data were minimal. Table 1 indicates the BL estimator is such an estimator.

To illustrate the comparative accuracy of the various estimators the sequences of estimated parameters are plotted in Figures 1-5 against the true values for the five different estimators given in (6. 2). Sequence B was used in each case and only the first element in the parameter vector was considered as it is representative of all three.

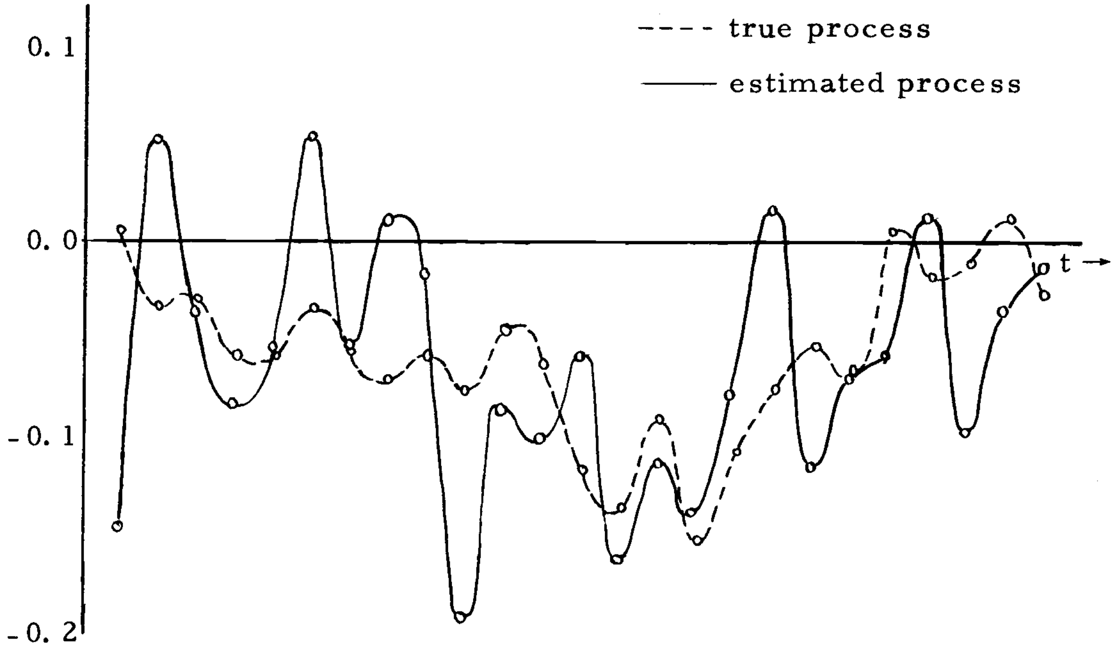

Figure 1. $\{\underline{\beta}_{t,1}\}_{t=1}^{25}$ estimated by Least Squares estimator, $n_t = 4$.

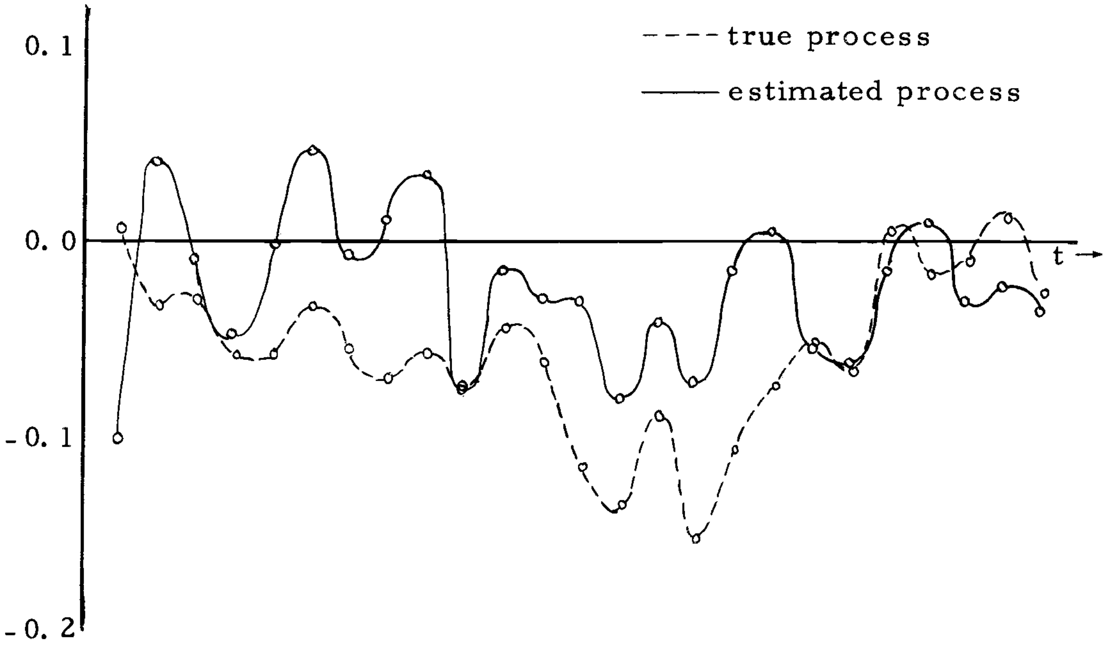

Figure 2. $\{\underline{\beta}_{t,1}\}_{t=1}^{25}$ estimated by Bayes estimator, $n_t = 4$.

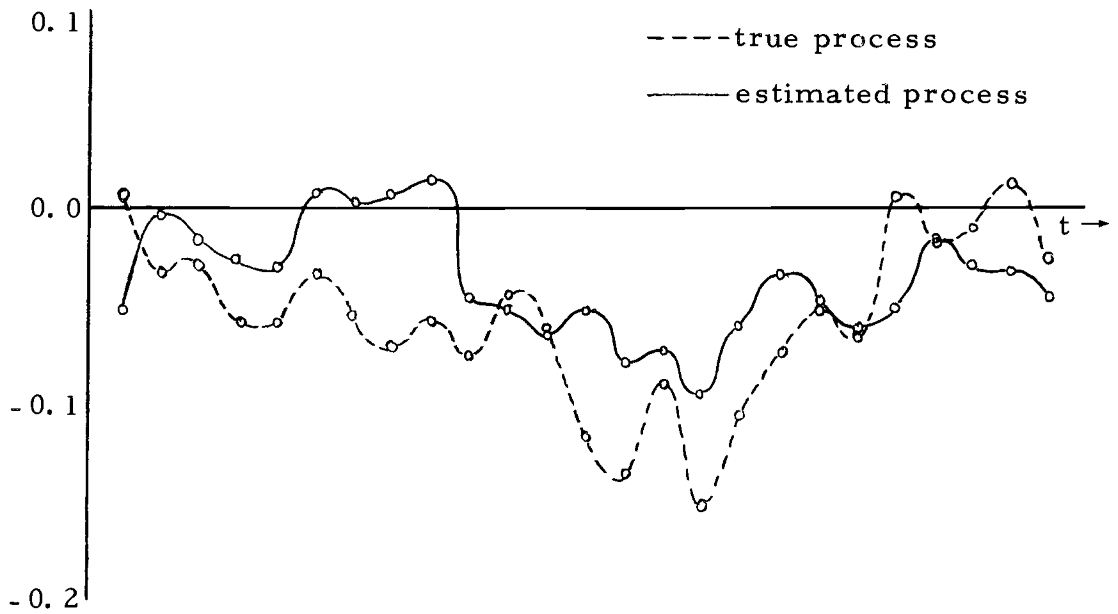

Figure 3. $\{\underline{\beta}_{t,1}\}_{t=1}^{25}$ estimated by order-3 Best Linear estimator, $n_t = 4$.

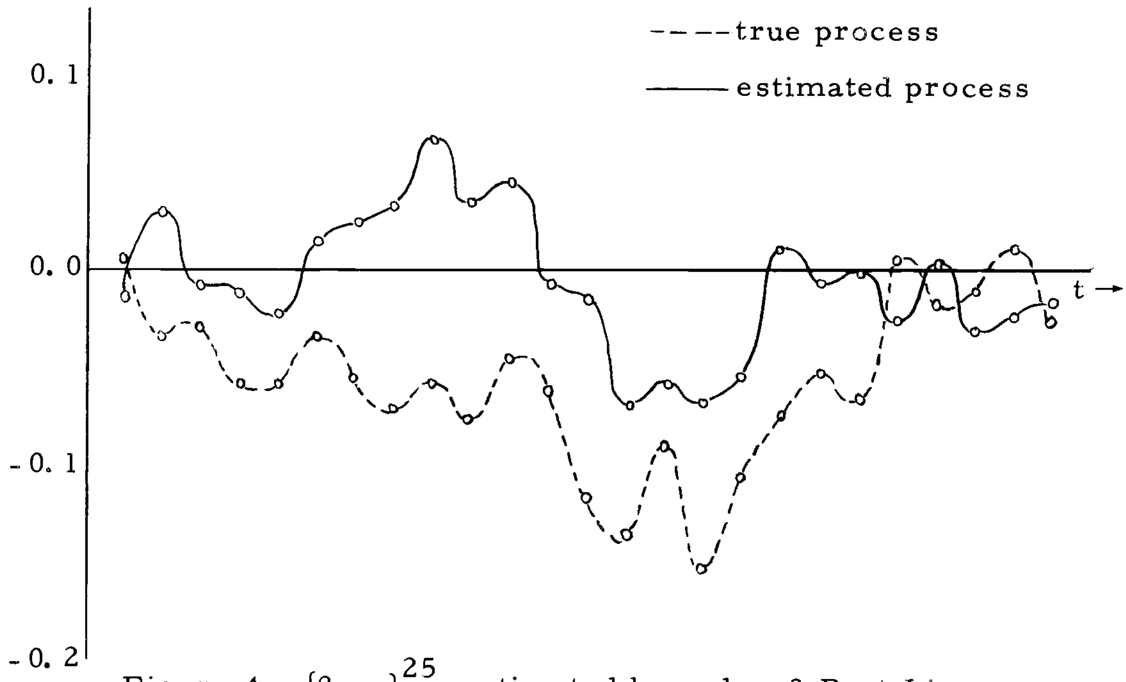

Figure 4. $\{\underline{\beta}_{t,1}\}_{t=1}^{25}$ estimated by order-3 Best Linear estimator, $n_t = 2$.

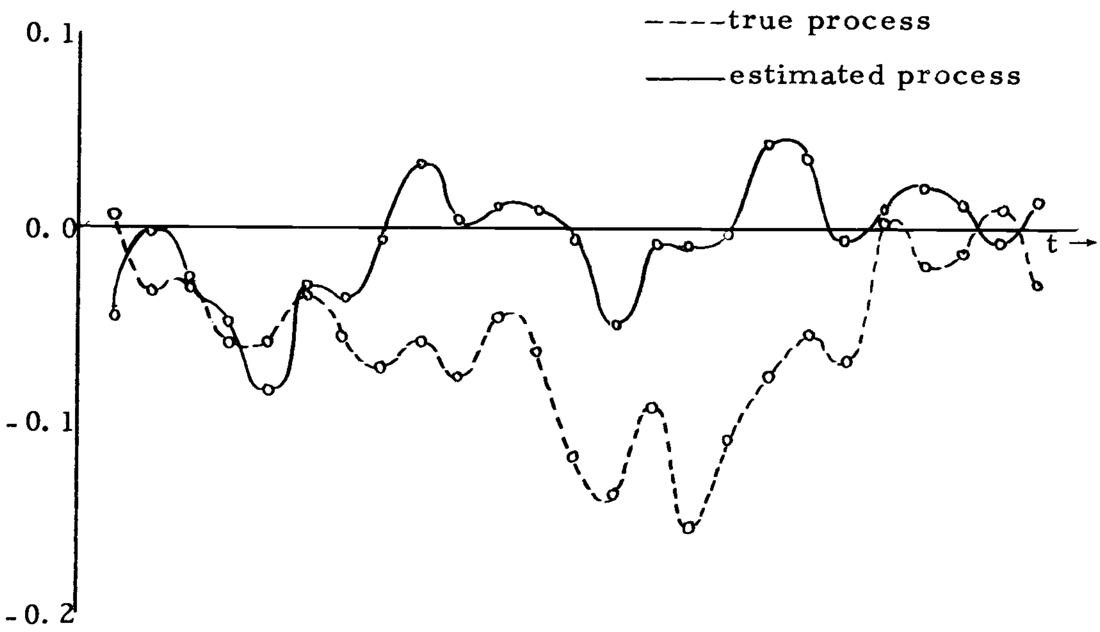

Figure 5. $\{\underline{\beta}_{t,1}\}_{t=1}^{25}$ estimated by order-5 Best Linear estimator, $n_t = 1$.

BIBLIOGRAPHY

Ben-Israel, A. and A. Charnes. 1963. Contributions to the theory of generalized inverses. Journal of the Society for Industrial and Applied Mathematics 2:667-699.

Chipman, John S. 1964. On least squares with insufficient observations. Journal of the American Statistical Association 59:1098-1111.

Greville, T. N. E. 1959. The pseudoinverse of a rectangular or singular matrix and its applications to the solution of systems of linear equations. SIAM Review 1:38-43.

Greville, T. N. E. 1960. Some applications of the pseudoinverse of a matrix. SIAM Review 2:15-22.

Jones, Richard H. 1966. Exponential smoothing for multivariate time series. Journal of the Royal Statistical Society, ser. B, 28:241-251.

Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. Transactions of the American Society of Mechanical Engineers, ser. D, Journal of Basic Engineering 82:35-45.

Lomnicki, Z. A. and S. K. Zaremba. 1957. On the estimation of auto-correlation in time series. The Annals of Mathematical Statistics 28:140-158.

Moore, Eliakim Hastings. 1935. General analysis. Philadelphia. 231 p. American Philosophical Society. Memoirs, vol. 1, part I.

Penrose, R. 1955. A generalized inverse for matrices. Proceedings of the Cambridge Philosophical Society 51:406-413.

Penrose, R. 1956. On best approximate solutions of linear matrix equations. Proceedings of the Cambridge Philosophical Society 52:17-19.

Price, Charles M. 1964. The matrix inverse and minimal variance estimates. SIAM Review 6:115-120.

Raiffa, Howard and Robert Schlaifer. 1961. Studies in management economics. Boston, Harvard University. 356 p.

Rao, C. Radhakrishna. 1965a. Statistical inference and its applications. New York, John Wiley and Sons. 522 p.

Rao, C. Radhakrishna. 1965b. The theory of least squares when the parameters are stochastic and its applications to the analysis of growth curves. Biometrika 52:447-458.

Scheffé, Henry. 1959. The analysis of variance. New York, John Wiley and Sons. 477 p.

Whittle, P. 1963. Regulation and prediction. London, English University Press. 147 p.