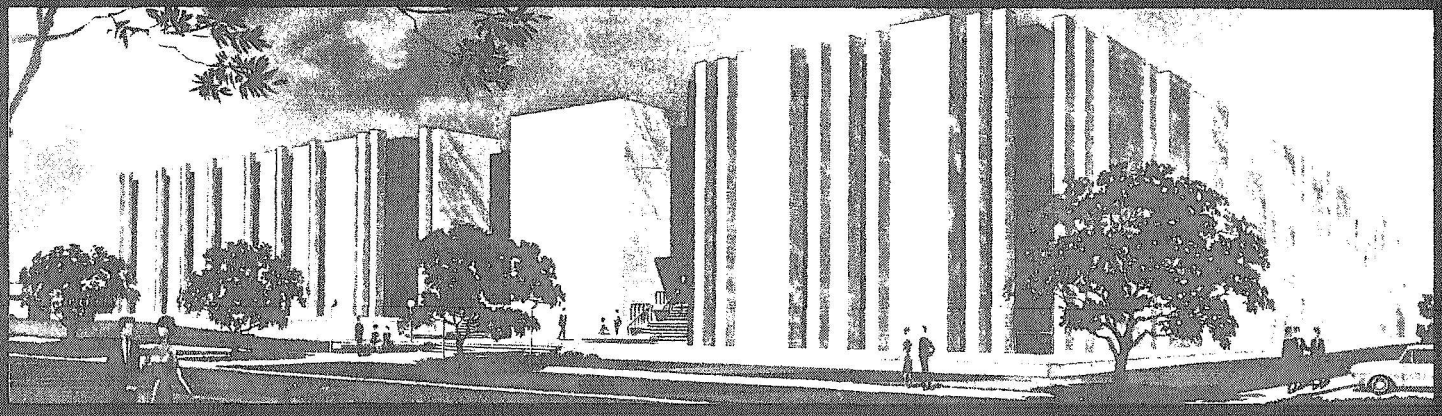
SEMIANNUAL STATUS REPORT

June 1, 1969, to November 30, 1969

TIME SERIES ANALYSIS WITH MISSING DATA

GRADUATE
INSTITUTE
OF
STATISTICS



TEXAS A&M UNIVERSITY · COLLEGE STATION

SEMIANNUAL STATUS REPORT

June 1, 1969, to November 30, 1969

TIME SERIES ANALYSIS WITH MISSING DATA

National Aeronautics and Space Administration

Grant #NGR 44-001-097

W. B. Smith, Associate Professor

Institute of Statistics

Texas A&M University

College Station, Texas 77843

SEMIANNUAL STATUS REPORT

TIME SERIES ANALYSIS WITH MISSING DATA

I. Introduction.

Hocking and Smith (1968) have presented a technique for estimating the parameters of a multivariate normal distribution from data vectors, some of which have missing elements. The purpose of this grant is to extend this technique to the estimation of the covariance function of a weakly stationary time series when the missing observations occur randomly. Then having estimated this covariance function one is able to estimate the spectral density function associated with the parent stochastic process, thus enabling an inference to be made on this parent process. An immediate application of this technique would be to a situation in which there is a random machine or telemetry failure on life data of an astronaut in space. This procedure would enable the scientist to estimate the missing observation, as well as, the general process which had been occurring as a function of time.

The estimation procedure of Hocking and Smith will also be applied to small sample estimation problems to aid in determining sample sizes, means, and variances. A mathematical formulation for the exact amount of information gleaned from the partial data vectors can be given.

The research along these lines is continuing in the following fashion:

(1) The procedure of Hocking and Smith is being extended to more general cases and more general problems, so its application to the problem of interest will be more easily attained.

(2) Other forms of stochastic processes are being considered; mainly the use of stochastic processes in compartmental analysis.

II. Research Reports.

Technical Report #1 submitted in September, 1969, entitled "Selection Index Estimation from Partial Multivariate Normal Data" represented an extension of the Hocking and Smith technique to the estimation of the well-known selection index. The selection index is based upon multivariate time-dependent data vectors, so it constitutes a legitimate extension of the partial data technique developed by Hocking and Smith. The procedure is spelled out in considerable detail in the technical report but can be summarized as follows:

The technique is applied to the estimation of the selection index, when incomplete multivariate normal data vectors are available. The procedure utilizes all data available, both full and partial vectors, and represents an improvement in the precision of the estimator over that found by using the full data vectors only. Estimates of the phenotypic means and the variance-covariance matrix are also found from the procedure. Individuals with partial data vectors are indexed

by an extension of the technique and this index is compared to that proposed by Henderson. The results of computer simulations are tabulated.

That is defining $I_j = b'x_j$, where I_j is the composite index value associated with the J^{th} member of the population, b is an n x 1 vector of unknown coefficients, and x_j is an n x 1 vector of phenotypic values (observations) on the J^{th} member of the population, Smith (1936) shows that $b = P^{-1}G\alpha$. In this context P is the n x n variance-covariance matrix of phenotypic values, G is the n x n matrix of genotypic values, and α is an n x 1 vector of economic weights associated with the n traits. The problem of estimating b is attacked in Technical Report #1 by estimating P the matrix of phenotypic values. This estimation from both full and incomplete data vectors represents an extension directly of the Hocking and Smith technique. For further details on the procedure and mathematical formulation of it see Technical Report #1.

Another area of research currently being carried out is the contents of Technical Report #2, "Stochastic Compartmental Analysis: Model and Least Squares Estimation from Time Series Data". This report, by J. H. Matis and H. O. Hartley, submitted November, 1969, can be summarized as follows:

Efficient estimation of the transition rate parameters of a stochastic m compartmental system requires the associated distribution. This report advances the distribution theory considerably by providing for the first

time a compact analytic solution. Often in practice individual compartments are inaccessible for observation and instead of time series, data is available only on the passage of material to the system exterior. The covariance kernel of such observations is derived in this paper and utilized for efficient parameter estimation. Finally, the estimation procedure is demonstrated from simulated data.

This research is into an area which represents a new branch of biomathematical modeling. This branch, called compartmental analysis, assumes a system may be divided into homogeneous components, or compartments. Various characteristics of the system are determined by observing the movement of tracer elements through these compartments. Mathematical models for such movements have been under the implicit assumption of deterministic flow patterns. In practice such patterns are entirely unrealistic. This report represents a first attempt at using stochastic flow patterns.

III. Future Research.

The lines of future research under this grant are three-fold: First, the extension of the Hocking-Smith technique to time series data in which there are a combination of types of data, that is, discrete and continuous random variables. An extension of the procedure to this combination will be significant and is being carried on at the present time. Secondly, further research on compartmental analysis and results of actual experiments on the same will be reported on. Finally,

the basic problem of estimating the spectral density function from the covariance kernel of a stochastic process in the presence of partial data will be attacked using the advanced methods developed by earlier papers and by earlier work in this contract.

IV. Technical Reports.

Technical Report #1, "Selection Index Estimation from Partial Mul Multivariate Normal Data", by W. B. Smith and R. C. Pfaffenberger, September, 1969. (Submitted to Biometrics).

Technical Report #2, "Stochastic Compartmental Analysis: Model and Least Squares Estimation from Time Series Data", by J. H. Matis and H. O. Hartley, November, 1969. (To be submitted for publication).

V. References

1. Hocking, R. R., and W. B. Smith, (1968). "Estimation of Parameters in the Multivariate Normal Distribution with Missing Observations", Journal of the American Statistical Association, 63, 159-173.

2. Henderson, C. R., (1963). "Selection Index and the Expected Genetic Advance", NAS-NRC Publication 982, 141-163.

3. Smith, H. F., (1936). "A Discriminant Function for Plant Selection", Ann. Eugen. Lond., 7, 240-250.