

AD A 041 766

18 AFOSR TR- 77-0747 19

7
B.S.

6 Statistical Models for Time Series and Life Testing with Applications in Engineering Systems.

9 FINAL REPORT. 1 Apr 72 - 31 Mar 77

Air Force Office of Scientific Research

GRANT AFOSR 72-2363 ✓

11 31 Mar 77

1 April 1972 - 31 March 1977

12 28p.

10 G. E. P./Box,
Principal Investigator

G. K. Bhattacharyya,
Richard A. Johnson
Grace Mahba

Associate Investigators

DDC
JUL 19 1977

15 ✓ AF-AFOSR-2363-72

AD No. _____
DDC FILE COPY
Approved for public release
distribution unlimited.

ACCESSION for	
NTIS	White Section <input checked="" type="checkbox"/>
DDC	Buff Section <input type="checkbox"/>
UNANNOUNCED	<input type="checkbox"/>
JUSTIFICATION.....	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	Avail. and/or SPECIAL
A	

400243

STATISTICAL MODELS FOR TIME SERIES AND LIFE TESTING
WITH APPLICATIONS IN ENGINEERING SYSTEMS

OVERVIEW OF RESEARCH

Over the period of this grant 34 papers have been published, 5 papers have been accepted for publication and 36 technical reports have been written, one book has been written and another revised, all of these have received support from the grant. Many of the publications have reflected our continuing interest in the topics of stochastic systems (especially with discrete control) in model building, smoothing and curve estimation, methods of approximation with noisy data, inferences with censored data, and life testing and reliability with applications to systems. *are discussed.*

Time Series Models

Today much more efficient system control, surveillance, and prediction are possible because of more effective instrumentation, producing increased capability for continuous or rapid intermittent measurement. However, for the analysis of the resulting data, classical statistical methodology in which observations are assumed to be statistically independent are inappropriate. It has, therefore, been necessary to develop new techniques of stochastic model building for data analysis, forecasting, and stochastic control. Such methods have wide applicability to practical problems ranging from missile tracking to estimating the effect of a change in Air Force recruiting policy. Research in this contract period has led to deeper understanding and further development of appropriate model forms, to improved methods of model identification, more exact methods of estimation of parameters in stochastic models, and better methods of diagnostic checking of the fitted models.

New statistical methods for solving a number of other important practical problems in data analysis have been developed. A series of results in curve fitting and smoothing using splines provided a practical nonparametric method for smoothing noisy data, and provided a practical solution to the notoriously difficult problem of differentiating noisy data. Surface data can also be smoothed using these newly developed techniques. Recent work in regression has provided a technique whereby the user can choose between subset selection, ridge or principal components estimates for modelling a regression problem. This method can be used in some situations even where the number of variables exceeds the number of data points.

Remote sensing experiments typically lead to a data analysis problem requiring the approximate solution of a Fredholm integral equation of the first kind; where the data are noisy. It is notoriously difficult to obtain a good estimate to the true solution to these equations, and ad hoc methods are frequently employed. Research under this grant has led to the first practical, completely automatic method, with proven good properties, for obtaining an estimate of the solution.

Several other hitherto unsolved problems were solved with the support of this grant. Practical methods for estimating the optimum bandwidth parameter for spectral density and density estimates were obtained, thus allowing for the elimination of commonly used subjective methods for doing this. The problem of the optimum input for system identification of the stationary linear dynamic model, was solved. Advances in goodness-of-fit tests and the theory of optimal control were also made.

Life Testing and Reliability

Research supported by this grant during the past four years has produced major advancements in several challenging areas of system reliability modeling and statistical inferences with life testing experiments. The primary goal of our research program has been steadily directed to the needs of modern day technology where the progressive complexity of system structures demand a more sophisticated formulation of models. The models must adequately describe the chance failure mechanisms in order to reach implications regarding the system's successful completion of mission. Included in this broad perspective is our sustained effort to strengthen statistical theory and techniques for careful reliability analyses that are indispensable in preventing costly failures in complicated systems.

A dominant concentration of our research activity was aimed at improving models for system reliability so as to incorporate possible interaction between components and the uncertainties in the external stresses. Substantial progress was achieved in developing statistical analyses for reliability structures that extended far beyond the simplistic models treated in contemporary literature. Research contributions in this direction are of significant value because improved modeling is at the very root of success in implementing programs of maintenance and ascertaining system availability.

Another area of major achievement within our research program concerns optimal statistical procedures for dealing with studies of

equipment failures where practical constraints of cost and time necessitate an early termination of the experiments. Statistical inferences with censored data were intensively investigated during this period. Several remarkable results enriched this important theoretical area of statistics and at the same time had direct bearing on the analysis of censored life testing data.

Techniques for the statistical control of manufacturing processes are invaluable for maintaining high quality production. In our study of the very popular technique, called a cusum test, we have alerted potential users of a serious shortcoming. When serial correlation is present, as is commonly the case, the technique may well fail to protect the designated level of quality. In addition, our approximations to the test and its properties provide the necessary alterations to surmount these difficulties.

Moreover, our fruitful method of analysis also leads to novel techniques for the sequential detection of time series model changes. These include statistical methods for continuously monitoring critical system parameters for time dependent models. These techniques and their extensions provide quantitative tools for maintaining complex systems to run at top performance.

Most importantly, our research activity was considerably broader in scope than the major goals stated above. It extended into several important general areas of statistical theory. Significant contributions were made in large sample theory, nonparametric methods, quality control,

multivariate analysis and inferences in stochastic processes. Aside from their role in enhancing knowledge in the domain of statistical theory, many of these results have potential applications in diverse areas of scientific investigation.

DESCRIPTION OF RESEARCH ACTIVITIES

Our research efforts have resulted in several innovative advances in modeling and inferences, and we now describe these in some detail.

Part 1 - Time Series Models

Further Development of Practical Stochastic Models

In [1] the question is addressed "How are stochastic difference equation time series models related to more traditional deterministic models-- how may imbedded deterministic components be recognized?"

In [2] the relation of linear stochastic difference equation models to exponential smoothing methods is elucidated. In [3] the effect of non-Normal innovations is considered. Research on the detection and tracking of parameter changes is discussed in [4] and [5] considers the appropriate choice of sampling interval.

In [6] the relation between parametric time series methods and the "state variable" approach are considered. In [7] the problem is addressed of how best to control a system when the allowable variation in the manipulated variable is constrained. In [8] the problem of optimum feed forward control is explored. Particularly after "linearization", models are liable to severe inhomogeneity of variance which if ignored can result in inefficient estimation. Frequently the difficulty can be corrected by an appropriate transformation. Methods for estimating the transformation are derived in [9].

Stochastic Model Building

Model Building is an iterative process which may be conveniently divided into three stages--(i) Identification (or specification) of the model form (autocorrelation and cross correlation analysis are of special importance at this stage); (ii) Estimation of system parameters for the tentatively identified model (for example, by maximum likelihood methods); (iii) Diagnostic checking of adequacy of the fitted model (by study of residuals). During this grant period advances have been made in all three areas. At the identification stage a deeper understanding of the nature of partial autocorrelation was obtained by considering their role in a Bayesian context [10]. In earlier work, system parameters were estimated using approximate maximum likelihood. Improved exact likelihood estimation was developed in [11]. In practice data often contains outliers or "bad" values. In [12] and [13] estimation of parameters in the presence of bad values is explored. Operating systems frequently employ feed back loops. Special methods needed in the analysis of the resulting data were studied in [14, 15, 16]. A portmanteau test statistic for detecting lack of model fit using the autocorrelation of residuals was earlier developed. A much closer and more useful approximation to the distribution of this statistic has now been devised [17]. It is important that major developments be brought to the attention of the audience who can put them to use. Therefore, summary accounts of recent developments have been prepared and published [18, 19]. Developments in the analysis of multiple time series and in intervention analysis were described in [20]. During the grant period a book on Bayesian Inference [21] has been published, research for which was partly sponsored by the grant. Also a revised edition of a very successful book in Time Series Forecasting and Control [22] has been published. Other results appear in [23-26].

Curve and Surface Smoothing, Numerical Differentiation of Noisy Data

The problem of nonparametric curve fitting when the model is $y(t_i) = f(t_i) + \epsilon_i$, $t_i \in [0,1]$, where the sequence of $\{\epsilon_i\}$ are white Gaussian noise and f is only assumed to be "smooth" has been solved: A smoothing spline is used to smooth the data, and the method of generalized cross-validation, developed under this grant, is used to estimate the optimum degree of smoothing from the data. Since the optimum degree of smoothing is nearly obtained, numerical differentiation can be implemented by differentiating the smoothing spline even for moderately noisy data. Theoretical properties, discussion of various applications, and numerical demonstrations of the amazingly good effectiveness of the method appear in [27,28,29,30]. A discussion of earlier methods appears in [31]. A new solution to the problem of smoothing irregularly spaced data on a surface, using an extension of the curve smoothing methods, appears in [32].

Regression

It is well known that Gauss-Markov or minimum variance unbiased regression estimates can be very bad from the point of view of mean square error, and so the use of biased estimates (to reduce the mean square error) is becoming increasingly popular. A new technique has been developed (generalized cross-validation) for choosing the ridge parameter in a ridge estimate. This method can also be used to select a subset when subset selection regression methods are used and to choose the principal components when a principal components method is used. In fact, the technique can be used to choose between the

"best" ridge, subset selection or principal components estimate. The technique has the good properties of the popular Mallows C_p - C_L statistics but unlike C_p or C_L , can be used (in certain circumstances) when there are no degrees of freedom for estimating σ^2 , that is, when the number of candidate variables is as large or larger than the number of data points. These results appear in [33].

Approximate Solution of Linear Operator Equations Arising in Remote Sensing Experiments

Early work in this area focussed on development and evaluation of methods for obtaining approximate solutions of linear operator equations, including integral and differential equations, using projection methods in reproducing kernel Hilbert spaces [34,35,36,37,38]. The mathematics is formally similar to time series methods for continuous time, time series problems, although the applications are different. This theoretical work laid the foundations for practical methods for solving first kind integral equations when the data are noisy. It is very frequently necessary to solve these equations numerically in analyzing experimental data in physics, meteorology, biology, geophysics, etc. since this type of equation models the typical indirect sensing experiment. The method of regularization is one of the major techniques for solving first kind integral equations. In this method the user must choose the regularization parameter, which controls the bias-variance or stability-fidelity tradeoff of the solution. Numerous ad hoc methods have been proposed for choosing this parameter, but for many years it has been an open question how to choose this parameter from the data without prior information. This question has been answered

while the solver was supported by the grant, the result will be published shortly [39].

Density and Spectral Density Estimation

Density estimates are commonly used as descriptive procedures for data. Also questions arising in a variety of applied problems reduce to questions concerning properties of a density, and, ultimately, multivariate density or likelihood function estimates will provide the answer to certain particularly difficult classification problems. Several new density estimates have been proposed and their properties obtained [40, 41, 42, 43]. A fundamental theoretical result was obtained in this area. It is the establishment of the fact that all the good estimates have, asymptotically, the same mean square error convergence rates under comparable circumstances if their control parameter, which controls the squared bias-variance tradeoff is chosen correctly [41, 42, 43, 44]. The fundamental practical culmination of this work is the development of a density estimation technique for which the value of the control parameter which minimizes integrated mean square error, can be estimated from the data [43]. This method can also be used to choose the optimal (integrated mean square error) bandwidth parameter in a window-type, spectral density estimate [28], thus answering another long-open question.

Optimal Experimental Design for System Identification

The problem of selecting the optimal input for estimating the parameters of a stationary linear dynamic model was solved. [45, 46 47] This work is frequently referenced.

k-Spacings and Goodness of Fit Tests

The k-spacings are defined as the distances between every k-th order statistics.

In [48] the asymptotic distribution of certain goodness of fit tests based on k-spacings is obtained. It is shown that tests with $k > 1$ are more powerful asymptotically than the usual spacings tests based on $k=1$.

In [49] a relationship is established between the empirical distribution functions (e.d.f.) of distributions with unknown scale parameters and the e.d.f. of independent random variables subject to scale perturbations. The relation is exploited to obtain significance points for some tests based on the e.d.f. of the k-spacings.

Reports [50] and [51] are the first two of a series of four reports dealing with the theory of k-spacings and their applications to goodness of fit tests. Report [50] contains an introduction to the series and shows the main properties and applications of the Dirichlet distribution. Report [51] reviews several methods used to study the distribution of the k-spacings for $k=1$ and extends them to the case $k > 1$. In particular, an important theorem of LeCam (1958) is extended.

Stiff Differential Equations

These equations may have highly oscillatory solutions which are exceedingly difficult to compute numerically. A completely new approach to their numerical solution has been proposed. Instead of trying to compute the exact solution, which may be impossible, a method for computing the low-pass-filtered solution was developed [52].

Control Theory

It was shown that the general linear plant problem can be solved by reproducing kernel Hilbert space techniques when arbitrary linear constraints which can be expressed as continuous linear functionals are imposed [53]. A convergent numerical method was obtained for minimizing a quadratic functional (on a reproducing kernel Hilbert space), when a continuous family of linear inequality constraints are imposed [54].

Part 2 - Life Testing and Reliability

Inferences from Censored Data

Experiments in life testing often must be terminated before all units fail. Consequently, statistical procedures based on censored data play an important role in reliability studies, especially of highly reliable components or systems. Advances were made on several aspects of this primary problem area of censored life tests.

Locally most powerful rank tests for a general parameter in a two-sample single-censored situation are derived in [1] and the work includes an investigation of asymptotic power and efficiency. Asymptotically optimal inference procedures are developed in [2] for inferences with censored data using only very weak regularity conditions. Our investigations with censored data were extended to multiple-censored situations where blocks of order statistics are censored. In [3], several important relationships have been established between the moments of the hazard rate, or its multiple censoring extensions, and the terms in the expressions for the uncensored situation. This study leads to an exact expression for the finite sample Fisher information, whereas only the asymptotic expressions were previously available. The ideas developed in [1] are extended to derive locally most powerful tests in multiple-censored data [6]. The completeness properties of both parametric and non-parametric families of life distributions under censoring, and implications regarding inferences on reliability are investigated in [4]. The asymptotic sufficiency of the ranks is established in [5] for the two sample situation when the observations are censored at the

rth order statistic. Several important conclusions are drawn concerning optimal tests.

Statistical Modeling and Inference for Complex Systems

A major thrust of our research was directed towards reliability studies of multicomponent systems by monitoring the failures of components and modeling their interactive behavior. Inferences with the bivariate reliability model of Marshall and Olkin are investigated in [7], [8] where existence, uniqueness and asymptotic properties of maximum likelihood estimator are studied and a uniformly most powerful test of independence is derived. Important advances are made in [9], [10] in terms of constructing realistic and tractable models for the reliability of a system whose components have variable strengths and which are subjected to random stresses emanating from the operating environment. In a parametric setting, the UMVU and maximum likelihood estimators as well as confidence bounds are derived in [9] for the reliability of an s out of k system. Optimal nonparametric estimates, their large sample distribution and efficiency are studied in [11]. Important generalizations of the stress-strength models to more complex systems and associated inference problems are treated in [10] which also includes a Bayesian technique for certain parametric models. In addition, optimal estimation procedures for system reliability and large sample confidence bounds are developed in [12] for the experimental situation where groups of components are tested under common stresses and, instead of individual strength and stress measurements, only the survivor counts are recorded.

The study includes an evaluation of the loss of efficiency in using the count data.

Statistical Techniques for the Control of Quality

Our investigation started by developing large sample approximations to the distribution of Page's cumulative sum (cusum) test. Considerable effort was expended in order to study the complete run length distribution rather than just the mean time to signal 'out-of-control'. The basic results were derived by relating the statistic to a functional of a Wiener process. Next, identifying the problem with that of a continuous random walk with one reflecting and one absorbing barrier, we obtained the distributional results in [13].

Some practical consequences of high import then followed directly. The same mathematical derivation also applies when the observations exhibit serial correlation. Thus, in [15], [16] we were able to study the change in distribution, of the time to signal, under auto-regressive and other dependent models. Our primary conclusion was that the cusum tests are not robust with respect to departures from independence. The use of cusum tests is now widespread and the presence of serial correlation so common, that attention must be drawn to the seriousness of this lack of robustness.

Our investigation continued in [14] where, for the first time, an analytic derivation is given for the optimal choice of a reference value based on the Wiener approximation. This value agrees with the rule suggested by others on the basis of Monte Carlo studies. Secondly,

we determined the effect of using an estimate for variance and the manner in which the use of an estimate interacts with the choice of a reference value.

Pursuing this line of attack, we obtained potentially the most important breakthrough concerning the development of statistical techniques for monitoring the parameters of integrated autoregressive moving average time series [17]. These time series models are among the most widely used statistical models in both forecasting situations and basic modeling of critical system performance indicators. In either situation, environmental shocks can cause disruption and result in model changes which must be detected quickly. Our techniques form the first steps of an extensive search for new statistical methods designed to meet this purpose.

Other advances in research

In [18], our goal was to study classification procedures from the point of view of finding low dimensional hyperplanes which in some sense best represent the p -variate population distributions and their samples. Our criterion of weighted loss of distance between observations and sample centroids shows explicitly the manner in which the prior probabilities enter. We establish two new optimality properties for Fisher's solution to finding a lower dimensional representation. Moreover, our class of choices including Fisher's, can be viewed as an orthogonal transformation of the standardized principal components from the common covariance matrix.

In [19], a uniformly most powerful unbiased test is derived for

testing the equality of inter-transition time distributions for a two-state semi-Markov process under an inverse sampling scheme. Tests with regular sampling, their exact power and asymptotic properties are also investigated for semi-Markov processes. These processes are useful in studying system failure and preventive maintenance policies, consumer brand switching patterns in marketing and also in modeling various stochastic phenomena in other fields.

A purchase incidence model is introduced in [20] where the interpurchase times are described by a two-parameter inverse Gaussian distribution and the population heterogeneity is modeled by the natural conjugate family. This model is more flexible than the exponential and one-parameter gamma models which were previously used for purchase incidence.

Significant advances were made to the understanding of the large sample behavior of posterior distributions in [21] and [22]. These results have importance in almost all situations where Bayesian methods of inference are employed, including applications to life testing. The posterior distribution is considered in the general multiparameter situation when the population belongs to some exponential family. First an asymptotic expansion in powers of $n^{-1/2}$ is obtained, for the posterior distribution, having the limiting normal as a leading term. The following terms can be used to correct the limiting normal approximation.

Besides the basic result on the expansion of the posterior distribution we obtain a similar result for risk. For the first time, expansions are also given for the marginal distributions and conditional distributions for the parameters. These motivate an interesting discussion of asymptotic independence. We conclude with a novel study of an approximation to the regions of highest posterior density obtained by modifying the normal ellipsoidal regions by using a correction term.

Families of discrete distributions were characterized by probability generating functions involving hypergeometric or confluent hypergeometric functions. Estimators of the parameters were obtained and their behavior examined on the basis of asymptotic relative efficiency [23].

The distortion of the t-distribution has been previously examined when the parent population is a mixture of normals. During the present research period some equal probability contours were computed showing precisely the amount of distortion from a pre-specified level of significance [25].

A solution to the Behrens-Fisher problem was previously developed. During the present research period the results were revised and re-written in a form for suitable publication [24].

PUBLICATIONS AND REPORTS

Part 1

- [1] Box, G. E. P. and Abraham, Bovas, "Stochastic Difference Equation Models", Technical Report #438, (1975).
- [2] Ledolter, Johannes and Box, G.E.P., "When are Exponential Smoothing Forecast Procedures Optimal?", Technical Report #447, (1975).
- [3] Ledolter, Johannes and Box, G.E.P., "ARIMA Time Series Models With Non-Normal Shocks", Technical Report #448, (1976).
- [4] Ledolter, Johannes and Box, G.E.P., "Topics in Time Series Analysis IV. Various Aspects of Parameter Changes in ARIMA Models", Technical Report #449, (1976).
- [5] Box, G.E.P. and Abraham, Bovas, "Linear Models, Time Series and Outliers - 4: Sampling Interval", Technical Report #439, (1975).
- [6] MacGregor, John F., "Topics in Control 1. State Variable Approach to Time Series Representation and Forecasting", Technical Report #306, (1972).
- [7] MacGregor, John F., "Topics in Control 2. Constrained Feedback Control of Linear Dynamic-Stochastic Systems", Technical Report #307, (1972).
- [8] Box, G.E.P. and MacGregor, John F., "Topics in Control 3. Feed-forward Control", Technical Report #308, (1972).
- [9] Box, G.E.P. and Hill, W.J., "Correcting Inhomogeneity of Variance with Power Transformation Weighting", Technometrics, Vol. 16, No. 3, pp.385-389, (1974).
- [10] Box, G.E.P., Jenkins, G.M. and Guttman, I., "Partial Autocorrelations from a Bayesian Viewpoint and Orthogonal Parameterization", Metron, Vol. XXX, N. 1-4, 31 - XII, (1972).
- [11] Ljung, G. and Box, G.E.P., "Studies in the Modelling of Discrete Time Series 2. Maximum Likelihood Estimation in the Autoregressive Moving Average Model", Technical Report #476, (1976).
- [12] Box, G.E.P. and Abraham, Bovas, "Outliers in Linear Models", Technical Report #437, (1975).
- [13] Box, G.E.P. and Abraham, Bovas, "Outliers in Time Series", Technical Report #440, (1975).
- [14] Box, G.E.P. and MacGregor, John F., "Topics in Control 4. The Analysis of Closed-Loop Dynamic-Stochastic Systems", Technical Report #309, (1972).
- [15] Box, G.E.P. and MacGregor, John F., "The Analysis of Closed-Loop Dynamic-Stochastic Systems", Technometrics, Vol. 16, No. 3, p.391-398, (1974).

- [16] Box, G.E.P. and MacGregor, John F., "Parameter Estimation for Dynamic-Stochastic Models Using Closed-Loop Operating Data", Proceedings of the Sixth Triennial World Congress of the International Federation of Automatic Control, Boston, Technometrics, Vol. 18, No. 4, pp.371-380, (1976).
- [17] Ljung, G. and Box, G.E.P., "Studies in the Modelling of Discrete Time Series 3. A Modification of the Overall χ^2 Test for Lack of Fit in Time Series Models", Technical Report #477 (1976).
- [18] Box, G.E.P., Jenkins, G.M. and MacGregor, John F., "Some Recent Advances in Forecasting and Control". Pt. 11, Technical Report #310, (1972).
- [19] Box, G.E.P., Jenkins, G.M. and MacGregor, John F., "Some Recent Advances in Forecasting and Control, Part II", J. Royal Statistical Society, Series C (Applied Statistics) Vol. 23, No. 2, pp.158-179, (1974).
- [20] Box, G.E.P. and Tiao, G.C., "Applications of Time Series Analysis", Contributions to Festschrift for H. O. Hartley to be published. Technical Report #479
- [21] Box, G.E.P. and Tiao, G.C., "Bayesian Inference in Statistical Analysis", 588pp. Addison Wesley (1973).
- [22] Box, G.E.P. and Jenkins, G.M., "Time Series Analysis Forecasting and Control", 2nd Ed., Holden-Day (1976).
- [23] Haugh, Larry D. and Box, G.E.P., "Identification of Dynamic Regressive (distributed lag) Models Connecting Two Time Series", Journal of the American Statistical Association, 72, No. 357, pp.121-130, (1977).
- [24] Haugh, Larry D., "Checking the Independence of Two Covariance-Stationary Normal Time Series", Technical Report #371, (1974).
- [25] Box, G.E.P. and MacGregor, John F., "Parameter Estimation for Dynamic Models Using Closed-Loop Operating Data", Technical Report #401, (1974).
- [26] Box, G.E.P. and Kanemasu, H., "Topics in Model Building, Part II: On Nonlinear Least Squares", Technical Report #321, (1972).
- [27] Wahba, Grace and Wold, Svante, "A Completely Automatic French Curve: Fitting Spline Functions by Cross-Validation", Communications in Statistics, 4, 1 (1975).
- [28] Wahba, Grace and Wold, Svante, "Periodic Splines for Spectral Density Estimation: The Use of Cross-Validation for Determining the Degree of Smoothing", Communications in Statistics, 4, 2 (1975).

- [29] Wahba, Grace and Craven, Peter, "Smoothing Noisy Data with Spline Functions: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation", Technical Report #445, (1977).
- [30] Wahba, Grace, Discussion to the paper "Non-Parametric Regression" by Charles Stone, to appear, Ann. Statist. (1977).
- [31] Wahba, Grace, "Smoothing Noisy Data by Spline Functions", Num. Math. 24 (1975).
- [32] Wahba, Grace, "A Canonical Form for the Problem of Estimating Smooth Surfaces", Technical Report #420, (1975).
- [33] Wahba, Grace, Golub, Gene and Heath, Michael, "Cross-Validation and Optimum Ridge Regression", Technical Report #491 (1977).
- [34] Wahba, Grace, "Convergence Rates for Certain Approximate Solutions to First Kind Integral Equations", J. of Approximation Theory, 7, 2 (1973).
- [35] Wahba, Grace, "A Class of Approximate Solutions to Linear Operator Equations", J. of Approximation Theory, 9, 1 (1973).
- [36] Wahba, Grace and Nashed, M.Z., "Convergence Rates of Approximate Least Squares Solutions of Linear Integral and Operator Equations of the First Kind", Math. Comp. 28, 125 (1974).
- [37] Wahba, Grace and Nashed, M.Z., "Regularization and Approximation of Linear Operator Equations in Reproducing Kernel Spaces", Bull. Am. Math. Soc., 80, 6 (1974).
- [38] Wahba, Grace and Nashed, M.Z., "Some exponentially Decreasing Error Bounds for a Numerical Inversion of the Laplace Transform", J. Math. Anal. Applic., 52, 3 (1975).
- [39] Wahba, Grace, "Practical Approximate Solutions to Linear Operator Equations When the Data are Noisy", Technical Report #430, (1975). To appear, SIAM J. Numerical Analysis, Sept. 1977.
- [40] Wahba, Grace, "A Polynomial Algorithm for Density Estimation", Ann. Math. Statist., 42, 6 (1971).
- [41] Wahba, Grace, "Interpolating Spline Methods for Density Estimation. I. Fixed Knots", Ann. Statist., 3, 1 (1975).
- [42] Wahba, Grace, "Histosplines with Knots Which are Order Statistics", J. Roy. Stat. Soc., Series B., 38, 2 (1976).
- [43] Wahba, Grace, "Optimal Smoothing of Density Estimates", Technical Report #469, (1976). To appear, Classification and Clustering, John Van Ryzin, Ed. Academic Press (1977).

- [44] Wahba, Grace, "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation", Ann. Statist., 3, 1 (1975).
- [45] Viort, Bernard, "On the Dispersion Matrix of the Maximum Likelihood Estimators of the Parameters of a Dynamic Model", Technical Report #313, (1972).
- [46] Viort, Bernard, "D-Optimal Designs for Dynamic Models, Pt. I. Theory", Technical Report #314, (1972).
- [47] Viort, Bernard, "D-Optimal Designs for Dynamic Models, Pt. II. Applications", Technical Report #316, (1972).
- [48] del Pino, Guido E., "On the Asymptotic Distribution of Some Goodness of Fit Tests Based on Spacings", Technical Report #405, (1975).
- [49] del Pino, Guido E. and Stephens, Michael A., "Two Related Goodness-of-Fit Problems with Applications to Spacings", Technical Report #436, (1975).
- [50] del Pino, Guido E., "K-Spacings and Goodness of Fit Tests, Part I: Introduction and the Dirichlet Distribution", Technical Report #483, (1977).
- [51] del Pino, Guido E., "K-Spacings and Goodness of Fit Tests, Part II: Distribution Theory for Uniform Spacings and Some Functions Based on Them", Technical Report #484, (1977).
- [52] Wahba, Grace and Miranker, W.L., "An Averaging Method for the Stiff Highly Oscillatory Problem", Mathematics of Computation, 30, 135 (1976).
- [53] Kuhn, Robert M., "Reproducing Kernel Spaces and Stochastic Multivariate Differential Equations with a Deterministic Control Theory Application", Technical Report #336, (1973).
- [54] Wahba, Grace, "On the Minimization of a Quadratic Functional Subject to a Continuous Family of Linear Inequality Constraints", SIAM J. Control, 11, 1 (1973).

Part 2

- [1] Johnson, R. A., "Some optimality results for one and two sample procedures based on the smallest r order statistics." Proceedings of Symposium on Reliability and Biometry, SIAM. (1974) 459-493.
- [2] Johnson, R. A., "Asymptotic results for inference procedures based on the r smallest observations." Annals of Statistics 2, 1138-1151 (1974).
- [3] Bhattacharyya, G. K., Johnson, R. A. and Mehrotra, K., "Extended Hazard Rate Functions: Properties and Applications" TR#418.
- [4] Bhattacharyya, G. K., Johnson, R. A. and Mehrotra, K., "On the Completeness of Minimal Sufficient Statistics With Censored Observations", Ann. Statist. 5, 547-553, (1977).
- [5] Johnson, R. A. and Mehrotra, K., "Asymptotic sufficiency and asymptotically most powerful tests for the two sample censored situation". Annals of Statistics 4, 589-596, (1976).
- [6] Mehrotra, K. G., Johnson, R. A., Bhattacharyya, G. K. "Locally most powerful rank test for multiple-censored data". Communications in Statistics - Theory & Methods A6(5),459-469 (1977)
- [7] Bhattacharyya, G. K., and Johnson, R. A., "On a test of independence in a bivariate exponential distribution". Jour. Amer. Statist. Assoc., 68, 704 (1973).
- [8] Bhattacharyya, G. K. and Johnson, R. A., "Maximum likelihood estimation and hypothesis testing in the bivariate exponential model of Marshall and Olkin". Technical Report No. 276, University of Wisconsin (1973).
- [9] Bhattacharyya, G. K. and Johnson, R. A., "Estimation of Reliability in a Multicomponent Stress-Strength Model". Jour. Amer. Statist. Assoc., 69, 966-970 (1974).
- [10] Bhattacharyya, G. K. and Johnson, R. A., "Stress-Strength models for system reliability". Proceedings of the Conference on Reliability and Fault Tree Analysis, 509-532 (1975) SIAM.
- [11] Bhattacharyya, G. K. and Johnson, R. A., "Nonparametric estimation of reliability for an s out of k stress-strength system". Technical Report No. 341, University of Wisconsin (1973).
- [12] Bhattacharyya, G. K. "Reliability estimation from survivor count data in a stress-strength setting". Technical Report No. 466, University of Wisconsin (1976).

- [13] Bagshaw, M. and Johnson, R. A., "Sequential detection of a drift change in a Wiener Process". Communications in Statistics 4(8), 787-796 (1975).
- [14] Bagshaw M. and Johnson, R. A., "The influence of reference values and estimated variance of the ARL of CUSUM tests". Journal of the Royal Stat. Soc., B 37, 413-420 (1975).
- [15] Bagshaw, M. and Johnson R. A., "The effect of serial correlation on the performance of cusum tests." Technometrics, 16, 103-112 (1974).
- [16] Bagshaw, M. and Johnson, R. A., "The effect of serial correlation on the performance of CUSUM Tests II." Technometrics, 17, 73-80 (1975).
- [17] Bagshaw, M. and Johnson, R. A., "Sequential Procedures For Detecting Parameter Changes in a Time Series Model", University of Wisconsin Technical Report No. 422 to appear J. American Statistical Association.
- [18] Hudlet, R. and Johnson, R. A. "Linear discrimination and some further results on best lower dimensional representations." Proceedings of Symposium on Classification and Clustering. (To appear).
- [19] Banerjee, A. K. and Bhattacharyya, G. K., "Testing hypotheses in a two-state semi-Markov process." To appear in Sankhya, Series A. (also University of Wisconsin Technical Report No. 373, July 1974).
- [20] Banerjee, A. K. and Bhattacharyya, G. K. "A purchase incidence model with inverse Gaussian interpurchase times". Journal of the American Statistical Association. (1976) 823-829.
- [21] Johnson, R.A. and Ladalla, J., "The large sample behavior of posterior distributions when sampling from multiparameter exponential family models and allied results. Part I. An asymptotic expansion of the posterior distribution", Technical Report #463, (1976).
- [22] Johnson, R.A. and Ladalla, J., "The large sample behavior of posterior distributions when sampling from multiparameter exponential family models and allied results. Part II. Applications and expansions for risk, marginal and conditional distributions, HPD regions", Technical Report #464, (1976).

- [23] Gurland, J. and Tripathi, R. C., "Estimation of parameters in some extensions of the Katz family of discrete distributions involving hypergeometric functions." Statistical Distributions in Scientific Work, Vol. 1, Models and Structures. International Conference on Characterization of Statistical Distributions. Calgary. (1975) 59-82.
- [24] Lee, A. F. S. and Gurland, J. "On testing equality of means of two normal populations with unequal variances." Journal of American Statistical Association, 70, (1975) 933-941.
- [25] Lee, A. F. S. and Gurland, J. "One sample t-test when sampling from a non-normal population." To appear in Annals of Statistics.

Personnel Supported

Principal Investigators	G.E.P. Box	2 months summer, 1972-76
	John Gurland	2 months summer, 1974
Associate Investigators	G.K. Bhattacharyya	1 month summer 1973, 2 months summer, 1974-76
	R.A. Johnson	1 month summer 1973, 2 months summer, 1974-76
	Grace Wahba	2 months summer, 1972-76

Graduate Student Research Assistants

John MacGregor	July, August, 1972	
P. C. Kartha	July, August, 1972	
Johannes Ledolter	Sem. 1, 1972-73	1/6 time
Greta Ljung	Sem. 2, 1972-73	1/2 time
Joseph Ladalla	Sem. 2, 1972-73	1/6 time
Reba Chakrabarti	Sem. 1, 1973-74	1/4 time
William Davis	" "	1/8 time
R. Hudlet-Yanez	" "	1/6 time
H. Liao	" "	1/6 time
S. Tam	" "	1/6 time
L. J. Wei	" "	1/8 time
C. Herrmann	Sem. 2, 1973-74	1/4 time
J. Ledolter	" "	1/4 time
L. Pallesen	" "	1/6 time
B. Singh	" "	1/4 time
W. S. Wei	" "	1/4 time
B. Abraham	July, August, 1974	1/2 time
S. H. Li	July, August, 1974	1/2 time
K. Rai	July, August, 1974	1/2 time
Fung Chao	Academic year 1974-75	44% time
J. Ledolter	Academic year 1974-75	1/2 time
G. Ljung	Academic year 1974-75	1/2 time
G. del Pino	Sem. 1, Academic Year 1975-76	1/2 time
R. Hudlet-Yanez	Academic Year 1975-76	1/2 time
G. Ljung	Sem. 1, Academic Year 1975-76	1/6 time
A. Krug	Sem. 2, Academic Year 1975-76	28% time
M. Jaeger	Sem. 2, Academic Year 1975-76	18% time
W. Fortney	Sem. 2, Academic Year 1975-76	18% time

M. Akritas	July, August, 1976 Sem. 1, Academic Year 1976-77	1/2 time
G. Chen	July, August, 1976 Sem. 2, Academic Year 1976-77	1/2 time
S. T. Liu	July, August, 1976	1/2 time
S. C. Wu	July, August, 1976	1/2 time
F. Hernandez	Academic Year 1976-77	1/2 time
F. T. Hsieh	February and March 1977	1/2 time
S. Graves	February and March 1977	1/2 time

Technical Typist

Candy Smith

December 10, 1972-February 17, 1973