NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

SPACE TECHNOLOGY APPLICATIONS

Task 125
Final Report
April 23, 1971

JPL DOCUMENT 650-126

CRIME PREDICTION MODELING

R. G. Chamberlain
Task Team Leader

E. P. Froman, Manager
Public Safety Support

JET PROPULSION LABORATORY
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA

CONTENTS

CONTENTS (Contd)

FIGURES

TABLES

ABSTRACT

A study of techniques for the prediction of crime in the City of Los Angeles has been conducted by the Space Technology Applications Office of the Jet Propulsion Laboratory. This is a part of a continuing program of the application of technology to civil problems, funded by the Technology Applications Office of NASA. The motivation and immediate application for the study was the evaluation of the effectiveness of a new tactical system -- use of helicopters as police patrol vehicles.

Alternative approaches to crime prediction (causal, quasi-causal, associative, extrapolative, and pattern-recognition models) are discussed, as is the environment within which predictions were desired for the immediate application. The decision was made to use time-series (extrapolative) models to produce the desired predictions. The characteristics of the data and the procedure used to choose equations for the extrapolations are discussed. The usefulness of different functional forms (constant, quadratic, and exponential forms) and of different parameter estimation techniques (multiple regression and multiple exponential smoothing) are compared, and the quality of the resultant predictions is assessed.

Appendixes present a discussion of the different approaches to crime prediction that were considered, a technique for simultaneous consideration of arrests and offenses, and algorithms for analysis of time-series. Included is the development of a modification to the multiple exponential smoothing technique which eliminates the need for a priori estimates of the model parameters.

SECTION I

INTRODUCTION

A. BACKGROUND

The capability to predict crime could be of considerable value to the police in a number of ways. For example, strategic decisions relating to force structures and training emphases would be easier if reliable long-range forecasts of crime were available. Tactical decisions on the day-to-day deployment of forces would be eased by accurate short-range forecasts or by forecasts of specific crimes. Analysis of the effectiveness of new (and old) equipment can be performed more accurately if appropriate crime prediction capability exists. An understanding of the relationships between social, economic, political, and moral conditions and crime, expressed in a model, could be used to devise effective crime prevention campaigns. These potential uses have generated interest in the broad technical problem of predicting crime.

In 1968, the Jet Propulsion Laboratory (JPL) was called upon to assist the Los Angeles Police Department (LAPD) in evaluating the effectiveness of helicopters as police patrol vehicles. The requested assistance was provided by JPL's Space Technology Applications (STA) Office, as Task 86. (The results of the Task 86 evaluation are documented elsewhere*.) This was funded by the Technology Applications Office of NASA as a part of its continuing program in applying technology to civil problems. It was necessary to this evaluation that crime levels be predicted.

Investigation of the prediction of crime was undertaken separately, as ST Task 125, with the evaluation of the effectiveness of a new tactical system (that is, Task 86) to be used as a focus for the effort. In particular, a major part o the evaluation was to be a determination of the reduction in crime, if any, due

*Weaver, R. W., Effectiveness Analysis of Helicopter Patrols. JPL Documer 650-89. Jet Propulsion Laboratory, Pasadena, Calif., Jul. 27, 1970.

to the presence of the helicopters. In order to make the necessary comparisons, predictions of the crime levels that would have occurred without the helicopters were required.

B. OBJECTIVES

The objectives of Task 125 were as follows:

1) Assist Task 86, Effectiveness Analysis of Helicopter Patrols, in the evaluation of the effectiveness of helicopters as patrol vehicles by producing a set of crime predictions for use in comparisons with actual crime occurrences.
2) Determine current and potential applications of crime prediction to permit assessment of the usefulness of different types of crime prediction models. Determine input and output requirements and operational constraints for these models.
3) Review existing and potential crime prediction techniques to determine the most promising approaches for further development.

Thus, the task consisted of three interrelated parts which were carried on essentially in parallel: A determination of the potential applications of crime prediction, a study of the state of the art, and an exercise in the real world.

C. APPROACH

Crime prediction is but one application in the general problem of quantitative forecasting. Thus, in reviewing the state of the art, it was necessary to consider recent advances in forecasting procedures as well as the relatively scant literature on crime prediction.

Potential applications of crime prediction capability were determined by surveying the applicable literature and by considering operations within the LAPD. They were further explored in discussions with officials of the Los Angeles Police Department.

The most appropriate type of model for any particular application depends upon a number of factors (Appendix A). Primary among these is the type of application. Thus, in Task 86, helicopters were placed in 2 of 17 police divisions, making divisional boundaries important. The experiment was expected to show effects on numbers of offenses and arrests, so these quantities were selected. The helicopter experiment was to run for a year, so predictions were needed for the same length of time.

A second major factor in choosing the kind of model to use was the constraint on the resources available for the study. Thus, although one could speculate that the use of cause and effect relationships might provide the most reliable predictions, development of such relationships would have been a more extensive research project than possible within the limited resources of this task.

A third factor that had to be considered was the availability of data. Many data types exist in the files of various public agencies, but others would have to be collected in the field - often an expensive, time-consuming operation. Preiminary crime prediction work for Task 86 had revealed that division boundaries had been changed a number of times during the historical period of interest. Fortunately, the LAPD files included quarterly crime reports for each of about 600 reporting districts, which can be considered as building blocks out of which divisions are made. Thus, it was possible to re-combine the data from the reporting districts according to the geographical boundaries of the divisions during the test period.

As a result of the assessment of these factors, it was decided to use eight* years of quarterly** data from the LAPD files and to use extrapolative (i.e., time-series) models to predict for the ninth year, 1969.

Section II of this report discusses the generation of predictions in greater detail. In brief, the approach taken was as follows: Time-series for each of 27 crime types in 24 geographical areas were built from LAPD quarterly summaries for the basic geographical building blocks, the police reporting districts. Analytical techniques used for each time-series were chosen by comparing the predictive accuracy of each of a variety considered.

It was reasoned that arrests were related to offenses, and that a system that reduced offenses would tend to reduce arrests as a result. Further, an increase in arrest effectiveness could be expected to reduce offenses due to the detention of multiple offenders. This relationship was investigated, and a statistical test for evaluation of the experimental results, using arrests and offenses together, is discussed in Appendix B of this report.

---

* Eight years was chosen as the sample size to avoid a major jurisdictional boundary change, affecting five of the current divisions, which occurred in January, 1961.

** A smaller interval, such as a month, would, if the data were obtainable, provide more data points in the history. The main benefit provided by these additional points would be a more detailed determination of seasonal effects; they would provide little, if any, improvement in trend calculations. A larger interval, such as a year, would provide larger numbers to work with. (It is shown later that prediction uncertainties are usually smaller percentages where expected numbers are larger.) It was deemed impractical to use a yearly interval, however, because it could be anticipated that the effectiveness of the helicopter patrol might change after some break-in period.

SECTION II

PREDICTION OF CRIME

A. SUMMARY

In developing a prediction model, it is necessary to make technical decisions on three levels (see Table 1). The first decision level is the most comprehensive and has a major impact on the predictions eventually produced. At this level, the variables to be predicted and the type of model to be used must be chosen. The second decision level is the selection of independent variables and their functional relationships. The last decision level is the choice of specific parameter estimation techniques.

Selection of the variables to be predicted depends upon the use to which they will be put. Hence, for the evaluation of helicopter patrols, number of offenses - with an emphasis on property crimes - was chosen to allow investigation of anticipated crime reduction, and number of arrests was chosen to allow investigation of anticipated improvement in police operational effectiveness.

The quality and availability of data plays an important role in many of the modeling decisions which must be made. Since the immediate application called for accurate predictions of crime levels, it was decided to use extrapolative models to make quarterly predictions of 27 crime types (see Table 2) in 24 geographical areas (see Table 3 and Fig. 1).

The choices of functional relationships and of parameter estimation techniques were made on the basis of their predictive abilities. Standard econometric techniques were considered, but rejected because of the failure of key assumptions. Instead, 6 of the 8 years of data was used to predict the remaining 2 years for each of the 648* time-series, by each of the model-technique combinations. The resultant matrix of prediction results was then

*(27 crime types) x (24 geographical areas) = 648 time-series.

Table 1. The three levels of modeling decisions

Level 1. Type of Model and Dependent Variables to be Predicted
a. Types of Models (see Appendix A): Extrapolative, associative, quasi-causal, causal, pattern recognition.
b. Dependent Variables to be Predicted: Number of offenses and/or arrests, probability of a crime event, specific crimes, calls for police service, and others.
c. Geographical Boundaries: Block, census tract or reporting district, division, city, county or region, state, nation.
d. Frequency and Length of Prediction: Hourly, daily, weekly, monthly, quarterly, annually.
Level 2. Independent ("Explanatory") Variables and Functional Relationships
a. Independent Variables: Proxy variables such as time, correlated variables such as population, variables resulting from theory such as measures of social pressure.
b. Functional Relationships: Linear or nonlinear, feedback control systems, and others.
Level 3. Parameter Estimation Techniques
a. Mathematical: Regression (least-squares), exponential smoothing, Fourier analysis, spectral analysis, moving averages, and others.
b. Nonmathematical: Trial and error, "eyeballing", simulation, physical modeling, and others.
c. Treatment of outlying points resulting from unusual occurrences and/or clerical errors.

Table 2. Crime types predicted for Task 86

1.	Murder, rape, and aggravated assault offenses
2.	Street robbery offenses
3.	Other robbery offenses
4.	Total robbery offenses (2 + 3)
5.	Residence burglary offenses
6.	Business burglary offenses
7.	Phone booth and other burglary offenses
8.	Total burglary offenses (5 + 6 + 7)
9.	Theft from person offenses
10.	Theft and burglary from auto offenses
11.	Bicycle and other theft offenses
12.	Total theft offenses (9 + 10 + 11)
13.	Auto theft offenses
14.	Total property offenses (4 + 8 + 12 + 13)
15.	Total property arrests (19 + 20 + 23 + 24)
16.	Total "Part I" offenses* (14 + 1)
17.	Other arrests (murder, rape, drunk, etc.)
18.	Aggravated assault arrests
19.	Robbery arrests
20.	Burglary arrests
21.	Felony theft arrests
22.	Misdemeanor theft arrests
23.	Total theft arrests (21 + 22)
24.	Auto theft arrests
25.	Narcotics arrests
26.	Total arrests other than traffic and forgery (17 + 18 + 19 + 20 + 23 + 24 + 25)
27.	Total arrests (26 + traffic + forgery)

*The FBI publishes the Uniform Crime Report annually. The "seven major" or "Part I" crimes, as defined in that publication, are murder and non-negligent manslaughter, forcible rape, aggravated assault, robbery, burglary, theft, and auto theft.

Table 3. Geographic areas used for Task 86 predictions

1. Central Division	13. Newton Street Division
2. Rampart Division	14. Venice Division
3. University Division	15. North Hollywood Division
4. Hollenbeck Division	16. Foothill Division
5. Harbor Division	17. Devonshire Division
6. Hollywood Division	18. Area 2 (Divisions 3, 7, 12, 13)
7. Wilshire Division	19. Area 3 (Divisions 1, 2, 4, 6, 11)
8. West Los Angeles Division	20. Area 4 (Divisions 9, 10, 15, 16, 17)
9. Van Nuys Division	21. Area 5 (Divisions 5, 8, 14)
10. West Valley Division	22. Area 2 less University Division
11. Highland Park Division	23. Area 4 less West Valley Division
12. 77th Street Division	24. Los Angeles City (Sum of 1 thru 17)

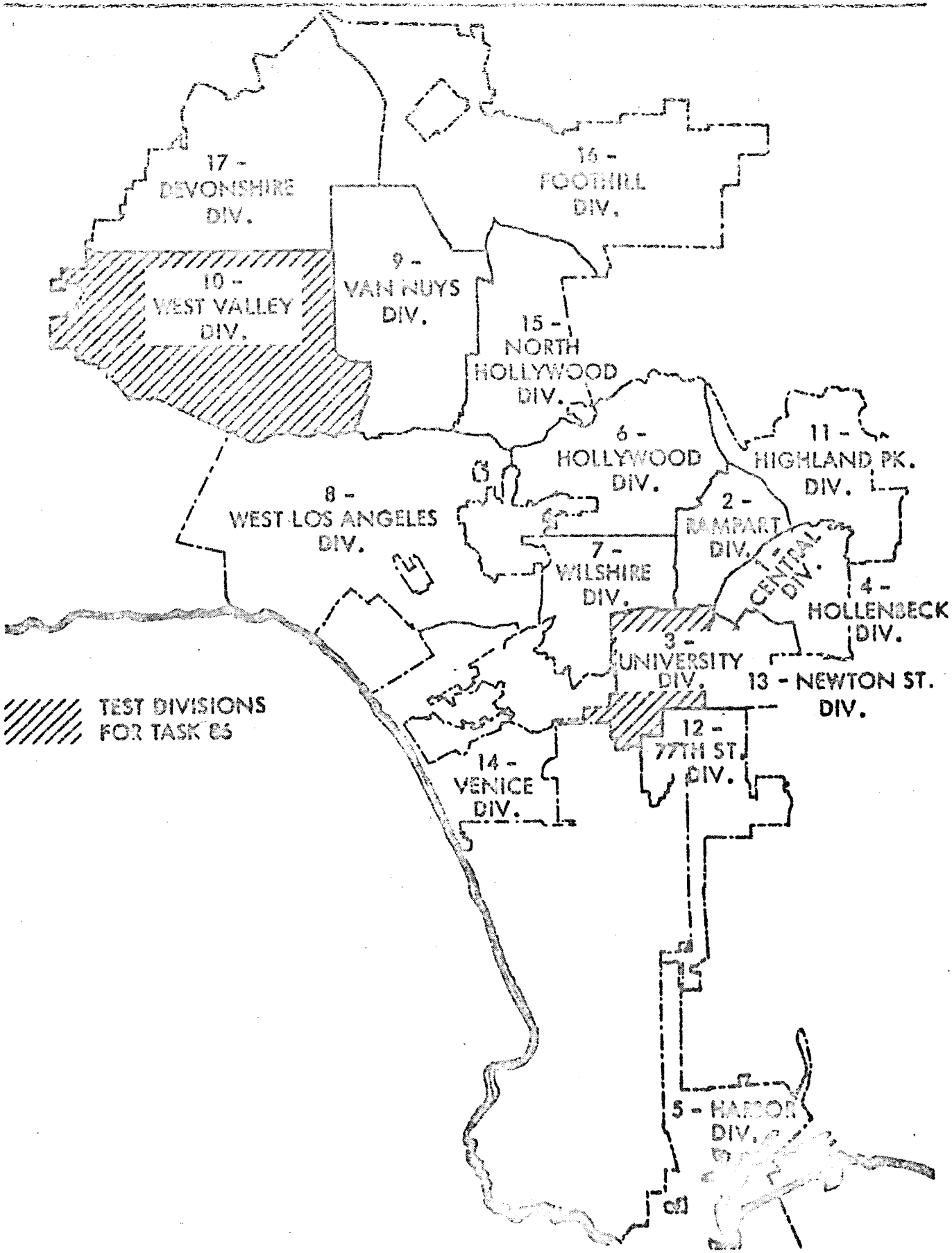

Fig. 1. Police division boundaries

compared with plots of the time-series to select model-technique combinations for the final predictions.

Figures 2 and 3 show the steps involved in proceeding from the raw data (LAPD quarterly reports, internal memoranda, jurisdictional boundary maps, and discussions with personnel within the department) to the final predictions.

First, a "dictionary" was constructed. By entering this dictionary with a reporting district number and a date, it was possible to determine, among other things, which division contained that reporting district at any other specified date. Program 1 used this information to recombine the historical data from the LAPD reporting districts given in the quarterly crime reports into time-series data for the LAPD divisions as they were during the study.

This data was then used by Program 2 to evaluate a set of models. There were 54 models in all that were compared, these being formed through combinations of the following (see Fig. 4):

1) Constant, quadratic, and exponential dependence on time. (Time was used as a proxy variable for the underlying causal factors.)
2) Outlier rejection criteria of 2, 4, ∞ standard deviations of the (uncensored) trial fit. (The data contains some "outlying" points, that is, points that might decrease prediction accuracy. Such points could be the result of clerical error or unusual events, such as riots.)
3) Multiple regression (least squares) and modified multiple exponential smoothing algorithms for determination of model parameters. Five smoothing constants (0.01, 0.03, 0.1, 0.3, 0.5) were tried with the exponential smoothing algorithm.

Program 2 applied each of the model-technique combinations to the first 6 years of each time-series and determined prediction errors during the seventh and eighth years of data.

These results were then compared with plots of the time-series, produced by Program 3, and models were chosen for the extrapolation into the test period.

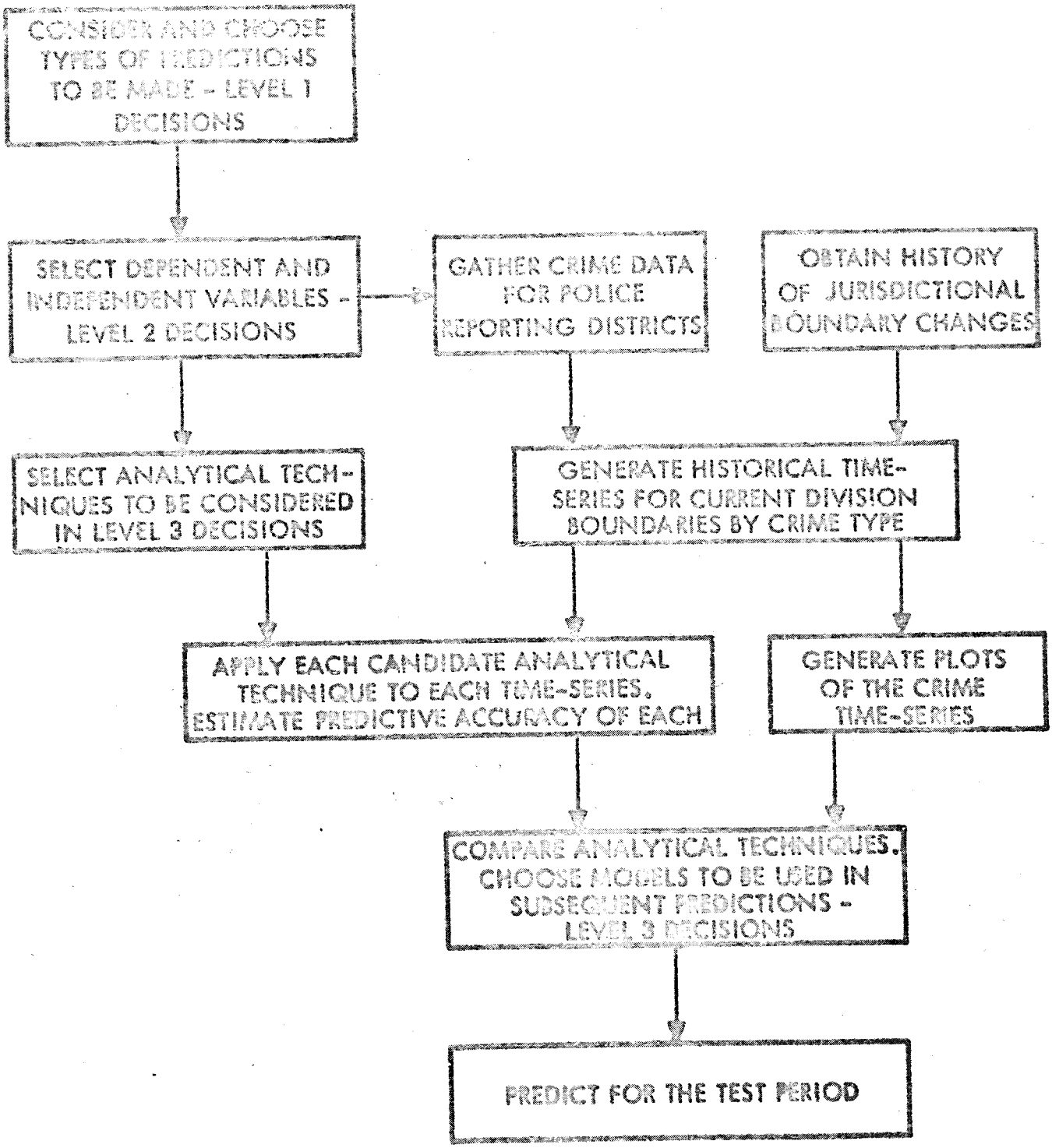

Fig. 2. Steps in the prediction of crime

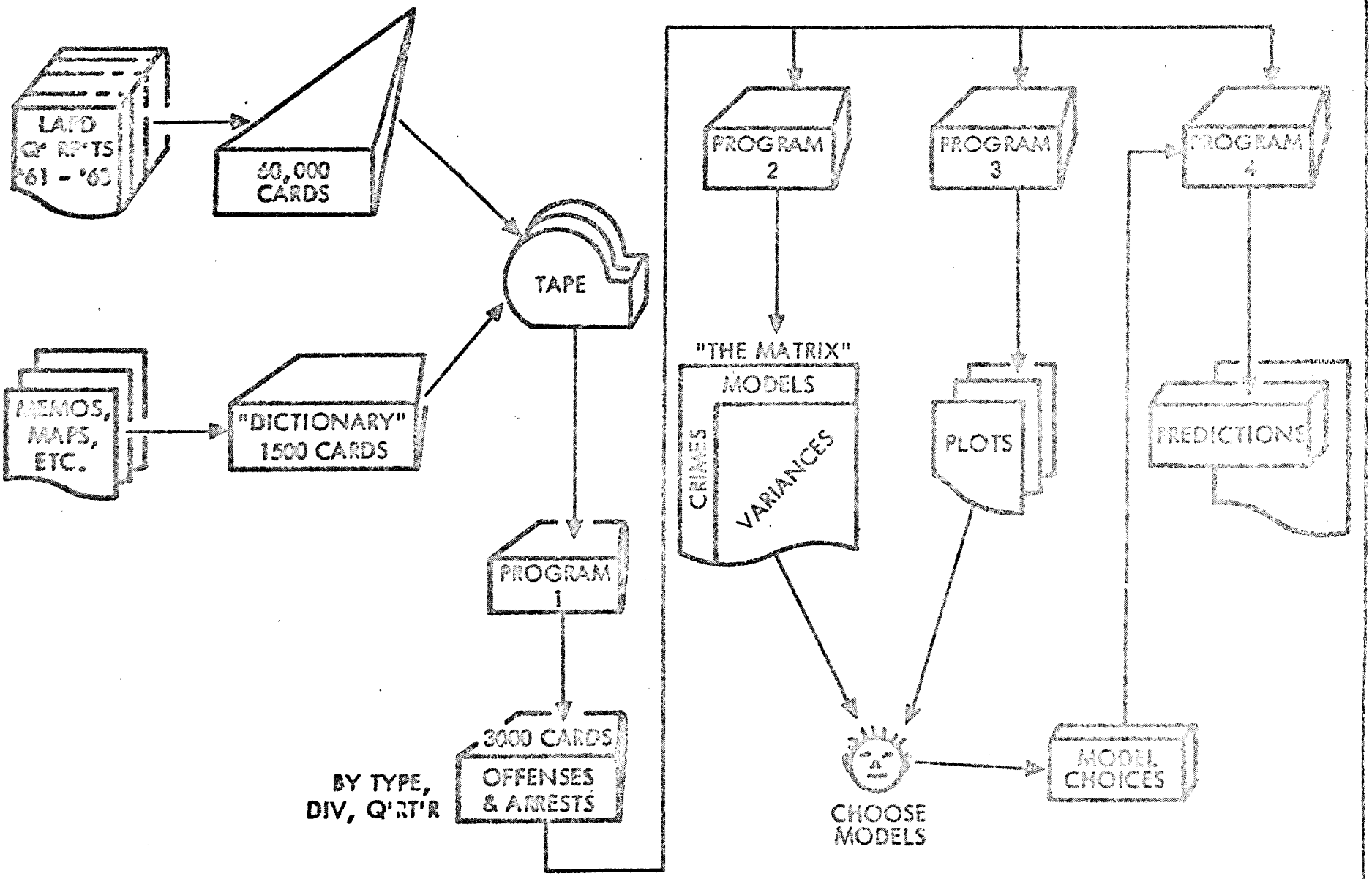

Fig. 3. Flow of information from raw data to prediction

JPL-STA 650-126

Functional Form of Prediction Model		
$\hat{y}_t = b_o$ + seasonals	$\hat{y}_t = b_o + b_1 t + b_2 t^2$ + seasonals	$\hat{y}_t = \exp(b_o + b_1 t)$ x seasonals
where $\hat{y}_t$ is the predicted amount of crime during period t t is the time proxy b_o, b_1, b_2, seasonals } are the parameters of the models.		

X

Criterion for Rejection of Outlying Data Points		
2σ	4σ	$\infty\sigma$ (No rejection)
where σ is the standard deviation of a fit to all the data.		

X

Parameter Estimation Technique					
Regression (Least Squares)	Exponential Smoothing Constant				
	0.01	0.03	0.1	0.3	0.5

$$\begin{pmatrix}\text{3 Functional}\\ \text{Forms}\end{pmatrix} \text{X} \begin{pmatrix}\text{3 Outlier Rejec-}\\ \text{tion Criteria}\end{pmatrix} \text{X} \begin{pmatrix}\text{6 Parameter Esti-}\\ \text{mation Techniques}\end{pmatrix} = \begin{pmatrix}\text{54 Model-technique}\\ \text{Combinations}\end{pmatrix}$$

Fig. 4. Model-technique combinations

Finally, the data from Program 1 and the chosen model-technique combinations were used by Program 4 to produce predictions for the test period.

B. FUNCTIONAL FORMS

Once it has been decided to produce predictions by extrapolation of historical time-series, it is necessary to select candidates for the equations to be used for this extrapolation. In the present case, this problem has three parts: treatment of trends, incorporation of seasonal variation, and the handling of cyclical variations.

Time-series prediction models use time as a proxy for the unknown variables that cause crime levels to change. If it is a poor proxy, or if these variables do not change with time, then no variation with time would be expected. Hence, a constant model was included among the candidates. A linear model accounts for a steady change in crime, while a quadratic model allows a steady change in that rate of change. Since a quadratic model can also show a steady change in crime, and since it was desired to keep the number of functional forms investigated to a tractable number, the linear model was not included as a candidate, while the quadratic model was. The rate of change of many processes is dependent upon the current state of that process, thus leading to exponential behavior with time. To account for such growth processes among the factors causing changes in crime levels, an exponential model was included among the candidates. The growth rate was arbitrarily allowed to change linearly with time to provide slightly greater flexibility. Since the model was not based on an understanding of the basic phenomena involved, it was felt that additional candidates would be superfluous.

Seasonal variation can be incorporated in a number of ways, the simplest of which - the use of additive* seasonal constants - was used here. If a large

*The exponential model is analyzed as linear with time after the dependent variable is replaced by its logarithm. Additive seasonal constants become multiplicative when the antilog is taken.

number of "seasons" were used (such as months or weeks, instead of quarters), this procedure could require too many degrees of freedom (that is, "use up" more than its "share" of the data), and more sophisticated techniques, such as Fourier analysis, might be required.

Possible cyclical variations were ignored, with the expectation that the provisions for modeling changes in trends would account for the most prominent cyclical changes, if any. Further, the data base was not long enough to anticipate much success in determining cyclic variations if present.

C. DATA PROBLEMS

A history of the variable to be predicted is the only data required for prediction by pure extrapolation. This history, often available for some period of time, can be quite extensive. The 27 crime types (see Table 2) for each of 24 geographic areas (see Table 3 and Fig. 1) constitute 648 different time-series. Quarterly data for 8 years provides 32 data points in each series. Whether 32 data points is sufficient for development of a satisfactory prediction model depends not only on the use of the model, but on the dispersion of the data itself. The minimum amount of data needed cannot be predicted ahead of time, but the success of this task indicates that, for most of the crime types treated, the data obtained was sufficient.

Available crime data suffers from a number of problems. One of these is that reporting criteria may vary considerably from place to place and time to time. As an extreme example, when New York City introduced centralized record keeping in 1950, the reported number of burglaries jumped 1300% over the preceding year.

Another problem is that the jurisdictional boundaries of various police organizational units within a city are generally determined by operational considerations. As a result, recorded data for these units suffers from discontinuous changes. From 1958 through 1968, for example, the LAPD underwent six major and about 30 minor boundary shifts, going from 13 divisions at the beginning of 1958 to 17 at the start of 1969 (see Fig. 5, which shows only the

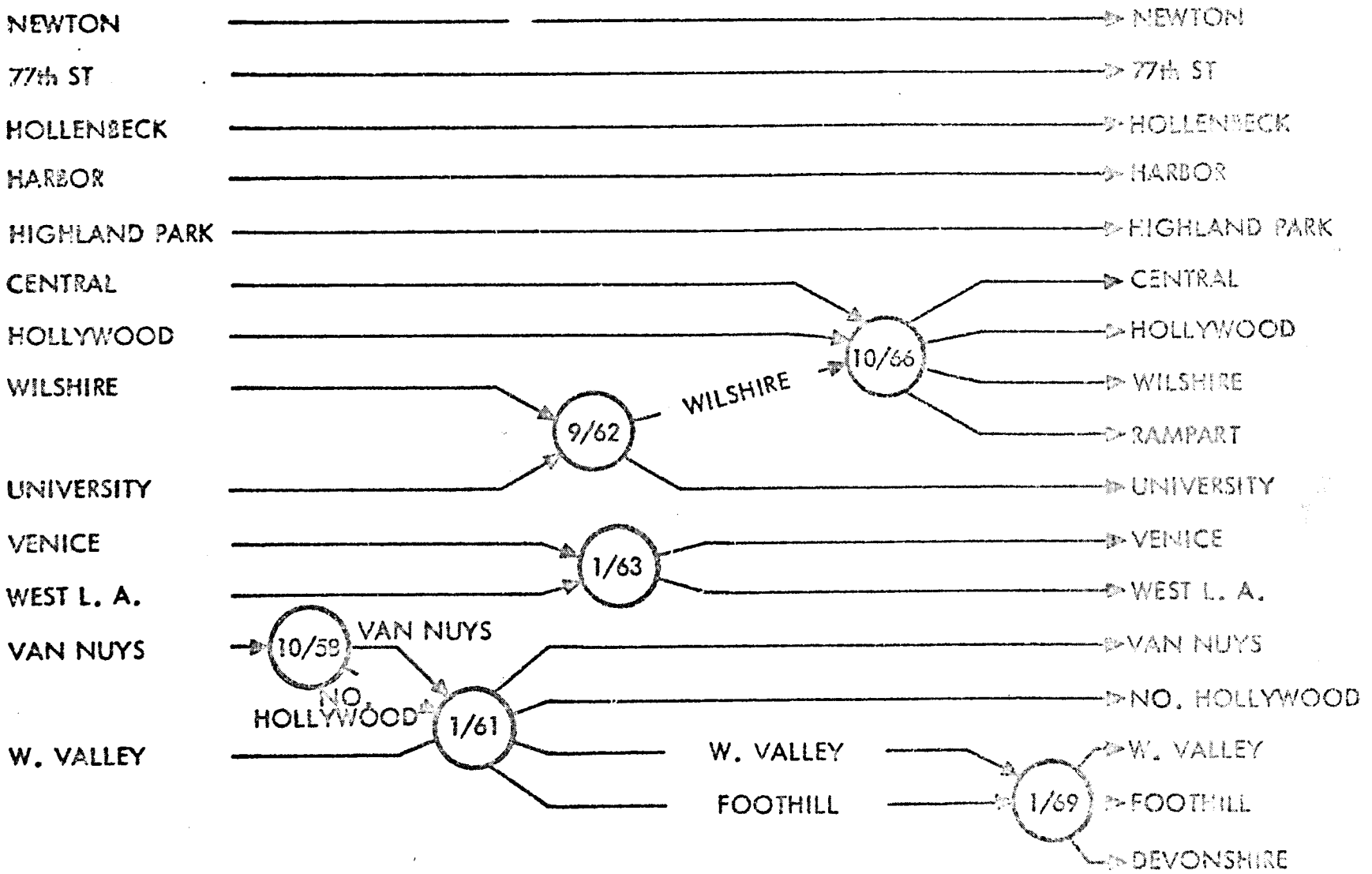

Fig. 5. Division genealogy - history of major boundary changes

major changes). Fortunately, many agencies gathering data about people, including LAPD, are now using census tracts as basic units, so that it was possible to obtain consistent histories for geographical areas.

A further problem is that unusual events are commonly recorded in the standard categories. For example, burglary arrests in the University Division in August 1965 were 10 times the normal average for August because of looting during the Watts riot. Such points must be identified and properly handled.*

A slightly different kind of problem is the appropriateness of the data type used. Specifically, in order to detect whether helicopter patrol has a repressive effect on crime, it is desirable to deal with numbers of offenses of various types of crime.. However, a large fraction of crimes committed are never reported to the police. Yet the files of "offenses known to police" are the primary (or only available) source of historical data. Changes in any of several factors, such as police-community relations, could conceivably change the fraction significantly. To validate the prediction techniques and to detect whether such changes were occurring, predictions were made for the divisions that did not use helicopter patrols. It may be noted that the divisions without helicopters did indeed perform as predicted**.

To determine whether the helicopter patrols contribute directly to the effectiveness of police operations, it would be desirable to deal with the number of offenders caught and the number of crimes solved. Conceivably, court conviction records could be used, but this data would be difficult to obtain and

*Since the objective in this task was to predict under "normal" conditions, such extremely unusual data points were simply rejected as inappropriate.

**There is one chance in seven that the non-test divisions would have looked more effective than they did, but only three chances in 100,000 that the test divisions would have made a better showing, as reported in the Task 86 final report (Weaver, R. W., Effectiveness Analysis of Helicopter Patrols, JPL Document 650-89. Jet Propulsion Laboratory, Pasadena, Calif., July 27, 1970.)

of questionable value*. Arrest statistics, on the other hand, are readily available. If it is recognized that arrests are made within constraints set by law and by police policy, and if it is assumed that those constraints do not change just before or during the prediction time period, the number of arrests may serve as a satisfactory measure of police performance in the apprehension of offenders.

D. DATA BASE

Each police division in Los Angeles contains about 30 or 40 police reporting districts (roughly equivalent to census tracts), which have almost always remained intact during jurisdictional boundary changes. In addition, the LAPD has a file of quarterly crime summaries by reporting district.

A detailed history of the 600 or so police reporting districts was obtained from files of LAPD internal memoranda. A computer program to generate time-series for police divisions as constituted in 1969 was applied to the 60,000 IBM cards containing all crime data quarterly reports.

The quarterly reports used for this data base were compiled from reports filed on each incident and contained some clerical mistakes, thus placing a limit on the possible accuracy for subsequent predictions. By way of illustration, it may be noted that about 2 1/2% of the reported offenses and arrests were attributed to nonexistent** reporting districts.

---

*In addition to the desired information, conviction data may also reflect the police department's effectiveness in gathering and processing evidence and the capabilities of the District Attorney's office in prosecuting cases.

**Reporting districts are identified by four-digit numbers, the first two of which correspond to the division number. Thus, about 60% of the possible numbers are not assigned. The first two digits were assumed to correctly identify the division. Consequently, most of the incorrectly numbered reporting districts were probably placed in the correct divisions by the aggregation process.

E. MODELING TECHNIQUES

Time-series prediction models can be expressed as equations containing parameters chosen to provide a best fit, in some sense, of the model to the historical data.

The traditional statistical technique, multiple regression, finds those values that minimize the sum of the squares of the differences between the fit and the data. The validity of this criterion rests on three assumptions:

1) The variables used in the model (in this case, time and season) are adequate proxies for the factors that control the underlying process and are modeled in a functionally correct way.
2) The process remains stable, in the sense that the "true" values of the parameters do not change. If, for example, crime of a particular type is constant, it must be assumed that that constant does not change.
3) Variations about the general trend result from a multitude of small, unmodeled causes, independent from one time period to the next and independent of each other.

There is no reason to believe, however, that any of these assumptions are fully satisfied in the case of crime prediction. Hence, it seems likely that placing greater emphasis on more recent data may lead to a better fit during the period for which predictions are desired. Fortunately, a relatively new technique, exponential smoothing, does this by giving exponentially decreasing weight to past data. This technique has shown promise in marketing and inventory control applications.

The published procedure for exponential smoothing requires initial estimates of the parameters. To avoid estimating thousands of initial values, a modified procedure was developed. Algorithms for this procedure and for multiple regression are given in Appendix C of this report.

It was not known beforehand whether the parameters derived through exponential smoothing or those from multiple regression would produce the better

predictions. Nor was it clear how far out an outlying point had to be before it would unduly influence a prediction. Further, when exponential smoothing was used, a wide range of possible weight decay rates were potentially competitive. Standard econometric techniques to compare the predictive capabilities of different model-technique combinations were investigated, but were rejected due to failure of key assumptions.

It was decided to use a direct approach: applying each model-technique combination to the first 6 years of data of each time-series to generate predictions for the seventh and eighth years*. The variances of these predictions were then compared, and plots of the time-series were inspected to select model-technique comparisons for extrapolation to the ninth year, 1969.

F. STATISTICAL EVALUATION TECHNIQUES

For some purposes, predictions can be used directly. Other applications, however, require that the predictions be processed in some way. When evaluation of the effectiveness of a new tactical system is the goal, the predictions must be used as a basis of comparison with the crime levels that actually occurred.

To be meaningful, these comparisons must be made statistically. That is, the predictions cannot be expected to match the actuals exactly. Statistical evaluation is required to determine whether the differences should be attributed to chance or to the new tactical system.

The differences to be expected as a result of chance are described by the standard deviation of the prediction error. Estimates of the standard deviations were also determined directly**: Once the models were chosen, 4 years of data were used to predict the fifth year, 5 years to predict the sixth, 6 years to predict the seventh, and 7 years to predict the eighth. These 16 quarters

*Two years were used for the comparison period in order to have a larger number of data points for the estimation of the prediction variance.

**That is, they were determined by consideration of the models' predictive performance, rather than by the usual econometric procedure of dealing solely with errors in fitting historical data.

of prediction errors were then used to estimate the quarterly error* for each time-series.

The validity of the models and estimated standard deviations was determined by applying the same techniques to the divisions without the new tactical system (that is, without the helicopters). It was expected that statistical evaluation would indicate that prediction errors in these cases would be attributable to chance.

Results during the test period were then compared to confidence limits based on these estimated standard deviations. In particular, the probability is 0.10 that offenses, for example, of a particular type in a particular division during a given quarter would be lower than the 90% lower confidence limit.

Treatment of offenses and arrests separately is not necessarily sufficient. A reduction in offenses could be expected to reduce the opportunities for arrests. (Consider, for example, the extreme of no offenses.) Thus, effectiveness in reducing offenses could mask an increase in the fraction of offenders apprehended if these related variables are only considered separately. Scattergrams of arrests versus offenses were plotted for several crime types in several divisions. The resultant clouds of points showed the anticipated correlation and suggested that the relationship may be linear. This correlation was used to devise a test for determining the statistical significance of offense-arrest vectors (that is, pairs of values).

Results from several quarters or several crime types can be compared by using a test that considers the directions and/or the amounts by which offense-arrest vectors differ from predicted offense-arrest vectors. This test can be described by the following analogy: If arrows shot at a target with no crosswind present cluster about the center of the target, then a cluster of

*Since a longer historical period can usually be expected to produce better predictions, these estimates are, in the main, conservative (i.e., slightly too large).

arrows to one side of the target is evidence of the presence of a crosswind. It is important to note that the conclusion does not depend on the size of the cluster. The techniques involved in performing these statistical tests are discussed in Appendix B of this document.

Unfortunately, the usefulness of this test relies upon the assumption that prediction errors are not correlated. As there was some evidence of correlation, the results of this test, though supporting the conclusions* of other analyses, were inconclusive by themselves.

*See the Task 86 final report (Weaver, R. W., Effectiveness Analysis of Helicopter Patrols. JPL Document 650-89. Jet Propulsion Laboratory, Pasadena, Calif. July 27, 1970.)

SECTION III

ANALYSIS OF THE SELECTION PROCESS AND RESULTS

A. THE MATRIX

It was noted in parts A and E of Section II that 54 model-technique combinations were applied to the first 6 years of data from each of the 648 time-series to produce prediction errors for the remaining 2 years for which data was available. The resultant 35 thousand comparisons are too voluminous to publish in this report. Table 4, restricted to the test divisions and suppressing the variation with different outlier rejection criteria* and different exponential smoothing constants**, suggests the nature of the information used. The body of this table contains the ratio of the standard deviation of the prediction errors during the two years 1967 and 1968 produced by the model-technique combination at the top of the column to the standard deviation about a horizontal line fit by least squares through the historical data. For each crime type, the denominator of the ratio is the same for all model-technique combinations, so differences result entirely from differences in the prediction accuracy of the model-technique combinations. The purpose of forming the ratios was to provide "normalized" numbers describing the prediction accuracy which could be compared among crime types and among divisions more easily than the "unnormalized" estimated standard deviations of the prediction errors. The reader is referred back to Table 2 for a list of crime types.

To illustrate how this table can be used, consider total burglary offenses (crime type 8) in these two divisions. The data for these two time-series is plotted in Figs. 6 and 7. The strong trend noticeable in both divisions explains

---

* Figures are shown for the best rejection of outlying points on the regression fits and no rejection when exponential smoothing was used.

** Figures are shown for those smoothing constants that indicated the best performance (smallest ratios). Results for exponential smoothing with the quadratic model are not shown because a programming error existed when the matrix was prepared.

Table 4. A portion of the matrix

(a) University Division

Model	Constant		Quadratic		Exponential	
Algorithm	Regression	Exponential Smoothing	Regression	Exponential Smoothing	Regression	Exponential Smoothing
Crime Type						
1	3.6	2.2	1.4		1.9	1.3
2	2.6	2.4	4.6		1.3	0.7
3	2.6	1.8	1.6		2.2	1.6
4	3.7	2.5	2.0		2.6	1.4
5	1.5	1.2	1.2		1.4	1.3
6	1.0	0.8	3.7	Not	0.8	0.6
7	0.5	0.7	5.3	available	2.4	1.0
8	1.7	0.9	0.9	because of	1.1	0.7
9	1.6	1.8	1.9	program-	2.7	1.0
10	1.9	1.1	1.2	ming error	0.9	0.8
11	1.7	1.5	2.3		1.4	1.5
12	2.3	1.6	1.1		1.3	1.1
13	3.6	2.8	2.9		1.5	1.3
14	2.8	1.6	1.0		1.0	0.9
15	1.9	2.0	1.5		2.2	1.3
16	3.0	1.7	1.0		1.0	0.9
17	1.2	0.3	1.2		7.6	0.4
18	1.8	0.8	0.6		0.9	0.6
19	0.9	0.8	0.9		1.8	0.8
20	0.7	0.7	0.4		0.6	0.6
21	0.8	0.9	2.1		1.5	1.6
22	0.8	1.0	0.8		1.7	0.8
23	0.9	1.0	1.0		1.9	0.7
24	2.9	2.7	4.1		2.2	1.6
25	3.6	1.4	1.4		2.6	1.6
26	1.7	1.0	2.0		1.3	1.0
27	1.9	1.0	1.8		1.2	1.0

Table 4 (Contd)

(b) West Valley Division

Model	Constant		Quadratic		Exponential	
Algorithm	Regression	Exponential Smoothing	Regression	Exponential Smoothing	Regression	Exponential Smoothing
Crime Type						
1	3.0	1.5	1.4		1.1	1.0
2	3.0	2.9	5.6		1.9	1.7
3	2.9	2.6	5.0		2.0	1.7
4	3.1	2.2	1.3		1.3	1.4
5	1.7	0.9	0.4		2.1	0.7
6	0.9	0.8	0.8		2.3	2.3
7	1.5	1.4	2.1	Not	2.2	1.8
8	2.0	0.4	1.8	available	0.9	0.6
9	2.6	1.8	2.0	because of	1.9	2.0
10	2.6	1.0	1.2	program-	0.7	0.4
11	2.2	0.7	1.4	ming error	0.8	0.6
12	2.5	0.8	1.3		0.7	0.4
13	2.6	1.0	1.1		0.7	0.7
14	2.4	0.6	1.4		0.8	0.3
15	1.9	0.8	1.5		1.1	0.6
16	2.4	0.7	1.4		0.7	0.3
17	1.1	0.2	2.0		18.7	0.9
18	1.3	0.6	0.8		0.9	0.8
19	4.8	4.5	4.1		4.2	4.2
20	3.2	1.7	1.5		1.0	0.6
21	1.2	0.7	2.1		1.7	1.5
22	1.3	0.7	1.4		2.2	0.7
23	1.3	0.7	1.7		2.2	0.7
24	1.5	1.6	1.9		2.0	2.2
25	22.4	20.4	19.3		21.0	14.8
26	2.7	1.4	1.0		0.9	0.7
27	1.7	1.0	0.6		0.5	0.7

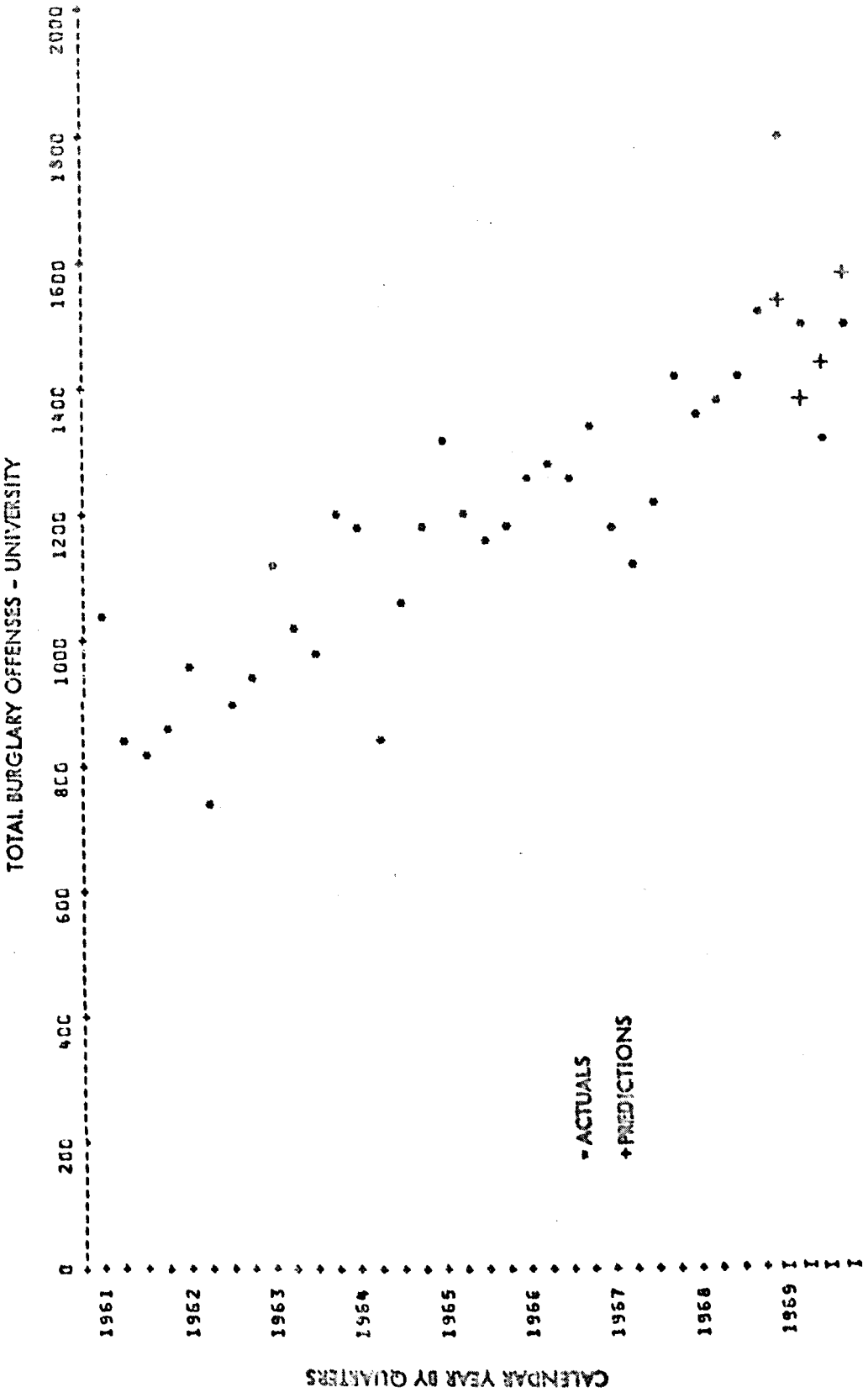

Fig. 6. Total burglary offenses - University Division

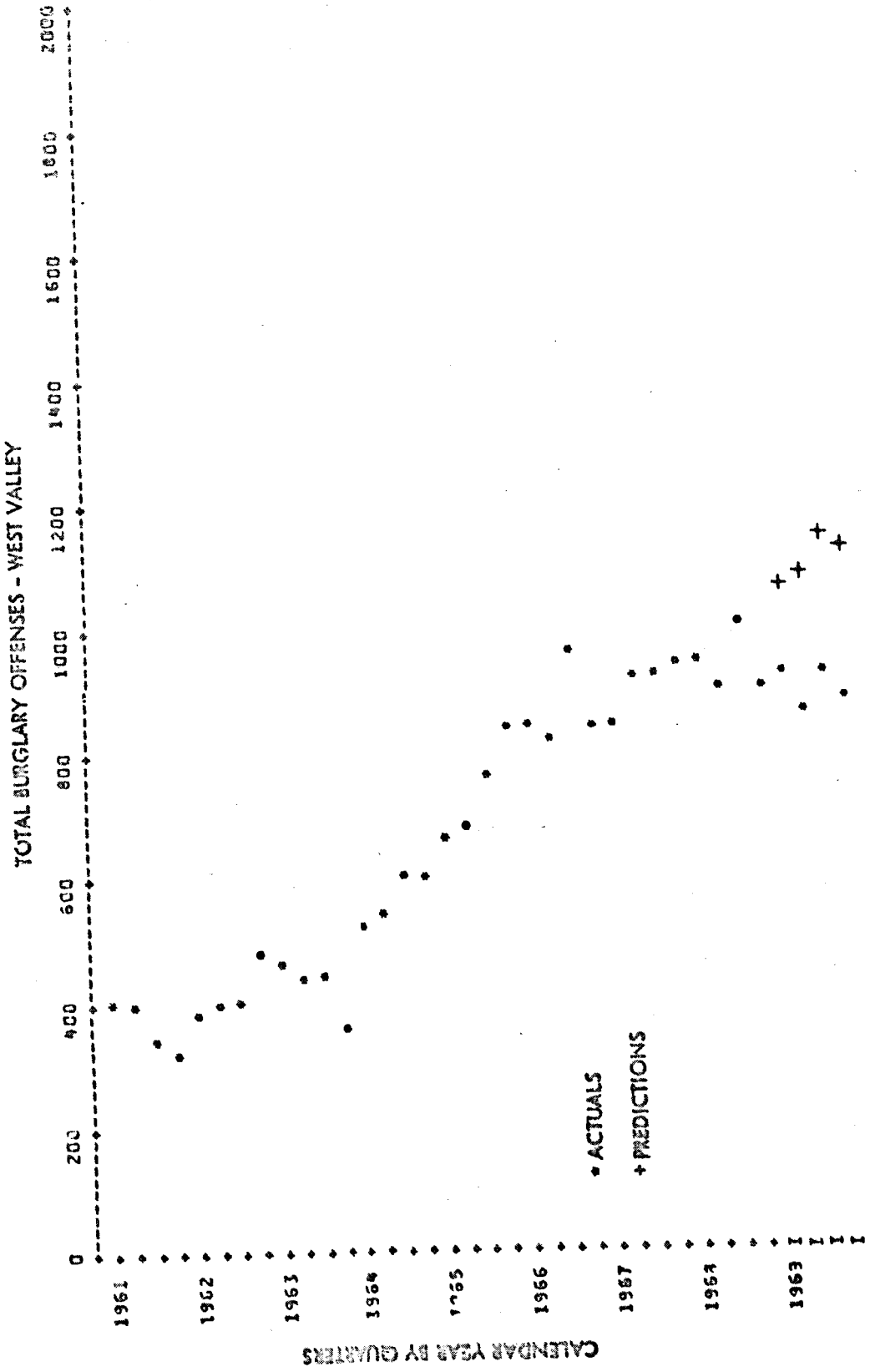

Fig. 7. Total burglary offenses - West Valley Division

the poor predictions obtained by the regression fit of the constant model, while the robustness of the exponential smoothing algorithm is shown by its relatively good performance even with the constant model. (The highest value of the smoothing constant was required to obtain this performance.) The S-shape of West Valley's curve explains the poor performance of the quadratic model. In both cases, the exponential model with the exponential smoothing algorithm was eventually chosen. It may be noted that the prediction error in the West Valley Division would be expected to be smaller than in the University Division, because there is noticeably less variability in the data. (Close inspection of the plot for University, however, will show that much of the apparent variability is seasonal.)

B. MODELS SELECTED

Model-technique combinations* cannot be chosen by simply finding the smallest numbers in the matrix. Many other factors must be considered. Among these are the following (not necessarily in order of importance):

1) For each crime type in the several divisions, and in each division for the several crime types, somewhat similar processes can be expected to be at work. Hence, models chosen should show consistency in both directions unless there are reasons to choose otherwise.

2) A poor fit by the regression algorithm (for a particular model) suggests that the model shape is inappropriate, so that a good fit with exponential smoothing should be viewed with skepticism**. Similarly, a fit that improves drastically as older data is discounted more and more rapidly (that is, as the expenential smoothing constant increases), also suggests an inappropriate model.

*That is, a model (constant, quadratic, exponential), an algorithm (regression, exponential smoothing), a smoothing constant (0.01, 0.03, 0.1, 0.3, 0.5), and an outlier rejection criterion (no rejection, rejection of points outside 2 or 4 standard deviations).

**For example, crime type 17 (other arrests), with the exponential model, in the West Valley Division.

3) Poor fits by all models* suggest that data in the two years used for evaluation of the prediction errors were unusual in some way and that model-technique combinations should be chosen some other way.

4) Visual inspection of plots of the time-series can be very helpful in deciding upon the appropriateness of various models and the usefulness of results in the matrix.

5) Generally speaking, when the exponential smoothing algorithm is to be used, small values of the smoothing constant are to be preferred, because large values essentially ignore a great deal of the available data.

6) Occasionally, trend parameter estimates, seasonal parameter estimates, and data during the final 2 years combine in peculiar ways to suggest inappropriate models.

Table 5 shows the final choices of models for each of the 648 time-series, and Table 6 gives a summary of the frequencies of choice. Refer to Table 2 for the list of crime types and Table 3 for the list of divisions.

It may be observed from Table 6 that the exponential model was chosen more than half the time, suggesting that much of the crime in the city of Los Angeles is growing exponentially and that better predictions might have been obtained if population had been used as a proxy variable instead of or in addition to time. It may also be noted (and this was also apparent in Table 4, showing a portion of the matrix) that the exponential smoothing algorithm was chosen more than 4-1/2 times as often as the regression algorithm, when both algorithms were equally available.

Several general conclusions were drawn about the selection of model-technique combinations for predicting crime:

1) Automatic rejection of outliers is very rarely useful with the modified exponential smoothing technique.

2) Automatic rejection of outliers is rarely useful with multiple regression; when it is useful, 4σ and 2σ seem about equally effective.

*For example, crime type 25 (narcotics arrests) in the West Valley Division.

Table 5. Crime prediction models chosen

Crime Type \ Division	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1212	3215	2110	3211	3214	3213	3212	3211	3211	3212	2110	2110	2211*	3214	3214	3211	3212	3211	3211	3215	3213	3215	3215	3215
2	1214	3215	3212	3213	1211	2110	3211	2110	3214	3211	3211	3213	2110	2110	1214	2110	2110	3211	3110	2110	3211	3211	2110	3214
3	3215	1211	2110	1215	3211	3213	3211	3212	3211	3211	3213	3214	1215	3211	3215	2110	3211	1110	3212	3211	3213	1211	3110	3211
4	1214	2110	2110	3213	2110	3211	3211	3211	3211	2110	3214	2110	3211	3211	3211	2110	2110	3212	3211	2110	3211	2110	2110	2110
5	1211	1213	2110	2110	2110	3110	3211	1215	1215	2110	2110	1215	3213	3110	1215	2110	1213	2110	2110	2110	3215	3211	2110	2110
6	1211	3215	3213	1215	3211	3211	1213	1213	1215	1212	3211	3211	1110	3110	1215	2110	1213	3212	3211	1214	3211	3211	1215	3213
7	1213	2110	3212	3110	2130	1110	1211	3130	1213	2130	3212	1213	3212	1215	1110	1214	1213	1213	1212	1212	1214	1214	1211	1213
8	1130	3211	3213	2110	3110	2130	1215	3110	1215	3213	2110	3211	2110	2110	1215	3211	3211	3130	2110	1215	3110	3130	1215	3211
9	3214	2110	2212*	2110	1110	3213	3213	3211	1211	1215	3211	3213	2213*	2110	3110	3110	1110	3213	1110	3213	2110	3212	3213	2213*
10	2212*	3213	2110	3110	3213	2130	3212	3110	3213	3212	2110	3214	1215	3110	3214	2110	3212	3211	3110	3110	3110	3211	3110	3110
11	3130	1212	3214	3213	1213	3212	3214	3130	3211	3211	2110	1214	3211	1215	1213	2110	3211	3213	3213	3211	3211	3213	3211	2110
12	2214*	3211	2110	3110	3110	2130	2110	2110	3211	3211	3110	2130	3211	3212	2110	2110	3212	3213	3212	3211	3110	2130	3212	3212
13	3214	2110	3211	3212	2110	3211	3211	3110	3110	3211	2110	2110	2110	3213	3211	2110	1215	3211	3212	3130	3214	3110	3110	2110
14	3214	3211	2110	2110	3214	2130	3110	3110	3211	3211	2110	2110	3215	3213	2110	3214	3211	2110	2130	3211	3213	2110	3211	3110
15	2212*	3212	2130	2211*	2110	2110	3211	3211	3213	3211	3110	3213	2130	2110	1215	3213	3130	3212	2110	2110	3214	2130	3110	3214
16	2130	3211	2110	2110	3214	2130	3110	3110	3211	3211	2130	2110	3214	3110	3211	2110	3211	2110	3212	3211	3213	2110	3214	3110
17	1211	2110	2110	1211	3213	3211	3211	3211	3214	1215	1211	3211	1215	3215	3211	3212	3211	1211	1211	3213	3213	3211	3215	2110
18	2110	3213	3213	1110	2212*	2110	3213	2211	3212	1215	2110	3213	3215	3213	2110	3211	2110	2110	3214	3213	2110	3214	3213	2110
19	2110	1110	1110	2211*	2212*	2211*	3212	2110	3212	1215	1214	3211	1130	1110	1214	3110	2110	1215	2110	2110	2211*	1215	3211	2110
20	3110	2110	2130	2130	2222*	2130	1213	3215	3213	3215	2110	2130	3224	2110	3211	3211	2110	2130	2130	2110	2211*	2130	3213	2130
21	3212	1110	2110	2110	1130	1110	3212	3212	2110	1213	1211	1110	1211	3211	3211	2130	1214	1110	3110	3110	1215	1110	1215	1130
22	3213	2110	2110	3213	1110	1213	3211	1212	3214	1213	1213	3213	3213	3110	3213	3215	3213	3211	3212	2110	3213	3110	1215	3213
23	3213	2110	2110	3213	1215	1212	3211	1213	3214	1214	2211*	3213	3211	3110	1130	3215	1215	3211	3211	2110	3211	3110	1215	3212
24	2212*	3213	3211	3110	3211	3213	3110	1211	3130	1110	3214	3211	3213	3110	3211	3211	1211	3211	3213	3214	3223	3211	3212	3211
25	1212	3214	3213	3213	3215	2110	2110	3213	3213	3213	2110	3211	3211	2110	3214	3215	3214	2110	3215	3211	3215	2110	3213	3214
26	1212	2110	3211	3213	3211	3211	2130	3211	3110	3211	2110	2110	1215	2110	3211	3214	3110	3110	3211	2110	2140	3130	3214	3211
27	1212	2110	3211	3211	3212	3211	3214	3211	3211	3130	2110	3110	2130	3215	2110	3214	3213	3211	3211	3130	2110	3211	3215	3211

*The computer implementation of the exponential smoothing algorithm with the quadratic model contained a programming error. However, plots of all predictions, along with their historical time-series, were prepared and inspected visually. In all cases, including these, predictions were consistent with "eye-ball" extrapolations.

The model-technique combinations are identified by the following 4-digit code:

First Digit - Functional Form

1 = constant model

$$y_t = a_o + b_{quarter} + \text{residual}$$

2 = quadratic model

$$y_t = a_o + a_1 t + a_2 t^2 + b_{quarter} + \text{residual}$$

3 = exponential model

$$\log y_t = a_o + a_1 t + b_{quarter} + \text{residual}$$

Second Digit - Algorithm

1 = multiple regression

2 = modified multiple exponential smoothing

Third Digit - Outlier Rejection Criterion

1 = outliers beyond 2 sigma are rejected

2 = outliers beyond 4 sigma are rejected

3 = no outliers are rejected.

Fourth Digit - Exponential Smoothing Constant α

0 : used with multiple regression; α not relevant

1 : $\alpha = 0.01$

2 : $\alpha = 0.03$

3 : $\alpha = 0.1$

4 : $\alpha = 0.3$

5 : $\alpha = 0.5$

Table 6. Frequencies of choice of models

Functional Form	Algorithm		Total
	Multiple Regression	Exponential Smoothing	
Constant + Seasonals	23	92	115
Quadratic + Seasonals	155	18*	173
Exponential + Seasonals	62	298	360
TOTAL	240	408	648

*The computer implementation of the exponential smoothing algorithm for the quadratic model contained a programming error. Consequently, this model-technique combination was chosen much less often than it would have been otherwise. It is probable that most of the quadratic model-multiple regression technique choices would have been displaced if the programming error had not been present.

It is significant to note that the exponential smoothing technique was chosen 4-1/2 times as often as the regression technique with the other two models.

3) Rejection of outliers, if done, should be a manual operation, based on additional knowledge that the data to be censored is atypical.

4) Multiple exponential smoothing models with large smoothing constants are prone to overrespond in a fashion similar to that of multiple regression models with too few degrees of freedom.

C. RESULTS

Quarterly predictions and associated prediction uncertainties for 27 crime types for 24 geographical areas for 1969 were obtained. Crime in the divisions without helicopters matched the predictions well.* The uncertainties (magnitudes of one standard deviation) varied with the crime type and sample size, but were usually about 15% (see Fig. 8). As would be expected, uncertainties are, in general, larger when the expected number is smaller and smaller when the expected number is larger.

*That is, a chi-square test on the distribution of residuals, measured in standard deviations, produced values well within those that would be expected to occur at random if drawn from a Gaussian distribution.

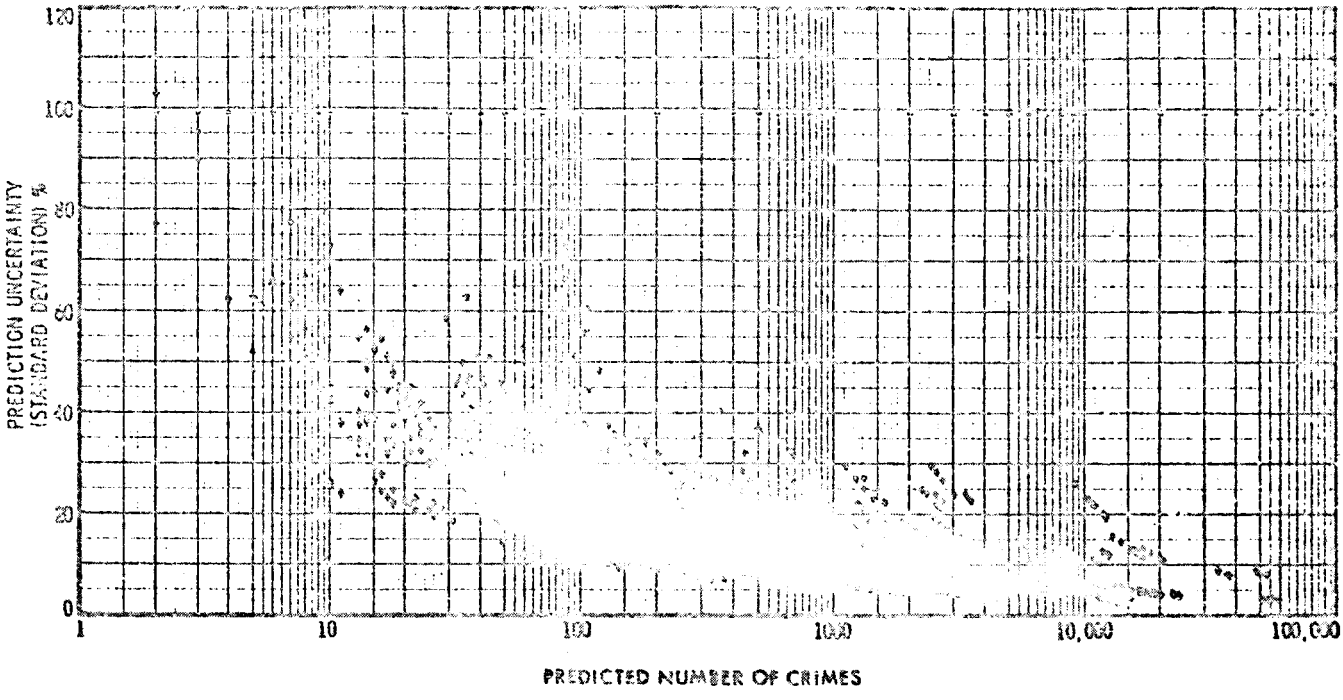

Fig. 8. Prediction uncertainties (magnitude of one standard deviation in %) for all crime types considered and all geographical areas versus magnitude of prediction

SECTION IV

CRIME PREDICTION APPLICATIONS

A. POTENTIAL APPLICATIONS

Crime prediction methodology is relatively new to the police system. As a result, its full potential is largely a matter of speculation. Some conjectures are presented in this section.

Long range (5 or 10 year) forecasts could be used by police policy-making echelons in planning recruitment campaigns, police academy curricula, force structure, and equipment acquisition. Forecasts would be particularly useful in this area if they could be relied upon to reflect changing trends.

Policy planners and operational commanders must know what alternative force structures can do. Task 86 has demonstrated that crime predictions can be used in conjunction with operational experiments to assist in the evaluation of the effectiveness of certain new and old tactics and equipment.

Tactical commanders can more easily deploy their forces in an effective manner if they know when, where, and how much of their forces will be needed. Prediction models can help supply this information. In this application, the capability to predict specific crimes might be of the greatest benefit.

If, as anticipated, activities of the police have an influence on the amount of crime, then a tool that would allow police planners to estimate the effects of a number of alternative possible actions could be of considerable value. Development of such a tool requires a better description of social forces and processes than is currently available to analysts. In addition, since these relationships are probably also dependent upon social conditions (such as educational levels, unemployment, housing conditions, etc.), such a postulated tool would potentially be useful beyond the police system - by legislators, social workers, city planners, and others. (In this regard, it may be noted that the

limitation of the current methodology is that it is "open loop". That is, there is no way to incorporate into the prediction model actions taken by the Police Department or other social agencies. A causal model would presumably contain such feedback loops.)

B. CURRENT APPLICATIONS

Corresponding to the broad spectrum of potential uses for crime prediction methodology, several rather different approaches have been proposed and are under study. The next few paragraphs will describe some of these applications briefly. It should be noted, however, that there are some gaps - no quantitative research was uncovered focusing on some of the potential uses.*

Predictions for use in decisions concerning the tactical deployment of police resources are concerned more with calls for police service than with actual crime data. To be useful, such predictions must be concerned with the distribution of needs for the police by jurisdiction, by hour of the day, by day of the week, and by season. LEMRAS (Law Enforcement Manpower Resource Allocation System), which has been in use in several cities and in the Van Nuys Division of Los Angeles for about 1 year, produces such predictions by use of exponential smoothing techniques.

Prediction of individual crimes requires a different approach. Conceivably, a comprehensive study of the causes of crime could provide sufficient understanding that such predictions could be made. As a first step, Philadelphia, in conjunction with the Franklin Institute, has been studying the correlations between about three dozen variables and the occurrence of crime in order to identify conditions under which crime is likely.

*In particular, there appears to be no quantitative research relating crime to controllable social factors. Changes in housing conditions, unemployment rates, welfare rules or costs, police deployment policies, and the like are apparently not being investigated in terms of their quantitative effects on crime. As a result, there is not enough information to construct closed-loop control models or long term prediction models.

Another approach to the prediction of individual crimes is under study in Los Angeles. Called PATRIC, it is an attempt to isolate the crimes committed by the same criminal or gang by recognition of patterns in modi operandi (methods of operation). This approach has met with two major implementation problems, both of which can probably be resolved by sufficient time and experience. One problem is the difficulty in describing crimes in such a way that the computer can identify patterns. This problem has two facets, determination of suitable descriptors and quality controls. The other major problem is that of collecting and processing data fast enough to be useful.

The determination of effectiveness of new tactical alternatives is exemplified by Task 86. Since this application is discussed throughout this report, it will not be elaborated upon here.

The scientifically most appealing technical approach to crime prediction is that of causal modeling. Models of this type are considerably beyond the current state of the art: A great deal of research into the forces interacting within our society must be conducted before such a model will be feasible. A good causal model would give insight into the probable effectiveness of various possible gross social actions and changes, and is the only hope for reliable long term predictions. But it is quite possible that extrapolative models would continue to give more precise short term predictions.

SECTION V

CONCLUSIONS

1) Crime statistics can be predicted with sufficient accuracy for some of the possible applications by extrapolation of historical data.

2) The potential applications of crime prediction methodology are sufficiently diverse that no single technical approach is appropriate to all.

3) Current qualitative and quantitative understanding of the causes of crime is grossly insufficient to permit the construction of usable causal crime prediction models.

4) Extrapolative crime models rely upon the assumption that trends will continue as in the immediate past. For example, changes* in public policies (especially, but not exclusively, by police agencies), in economic or social conditions, or in public moral or philosophical attitudes can invalidate this assumption.

*That is, extraordinary changes in these factors beyond those that have occurred during the period of time covered by the data base.

APPENDIX A

PREDICTION MODELING

There are three levels on which modeling decisions must be made: the kind of model, the choice of independent variables and functional relationships, and the selection among many available parameter estimation techniques. Each of these topics is discussed below.

A. TYPES OF MODELS

The first consideration of any modeling effort must be the user. What kinds of information will be most useful? Are some kinds of information required? How will it be used? How accurate must it be? What are the payoffs for accuracy and the costs of errors? What are the users' data resources, constraints, and computational capabilities? Will the user need more, better, or different information later than he needs now? Questions such as these are prerequisite to an intelligent choice of analytical emphases.

A classification of model types follows:

1. Extrapolative (Time-Series) Models

Historical data is simply extrapolated into the immediate future. Sophistication can range from simple "eyeball" extrapolation of a plot of the historical data to complex manipulation dealing with cycles, trends, and seasonal variations.

2. Associative Models

If two objects behave similarly, it is not unreasonable to anticipate that this similarity extends beyond the data used for the comparison. That is, objects that are associated in some way with the phenomena of interest can be used to predict those phenomena. For example, if the divisions of a city are assumed to be alike in some sense, crime data from some divisions could be used to predict crime rates in other divisions.

3. Quasi-Causal Models

A limited approach to consideration of the factors that influence crime is to seek those whose time-series are highly correlated with the crime time-series of interest. Predictions are then based on the assumption that the observed correlations will continue and the expectation that the highly correlated variables are at least proxies for the real causes. Since the supposed cause-effect relationships are not known, the parameters associated with the genuine causal factors used are, at best, point estimates of the partial derivatives of those relationships. The danger of identifying false causes is particularly high in a model of this kind.

4. Causal Models

If some or most of the factors that influence the variable of interest and the ways in which these influences occur are either known or theorized, causal models can be constructed to express this knowledge or theory. Causal models that include factors under the control of one or more of the users are clearly of the greatest potential value.

5. Pattern Recognition Models

Models in this category are aimed at isolating the crimes committed by single criminals or gangs, and using this information to identify likely crime targets and likely suspects.

B. VARIABLES AND FUNCTIONAL RELATIONSHIPS

With the exception of models designed to test theories, the selection of appropriate factors to be included as independent variables and the functional forms to be used is a difficult question. Resource limitations generally dictate that the variables be restricted to those data types for which records are available. Some insight into the choice of functional forms may be gained by consideration of plots of the variable to be predicted against each of the candidate factors. Consideration of the fraction of the variance which is

eliminated by a fit obtained when using a functional expression under consideration can also be enlightening; however, the variables and functions that give the best fit do not necessarily provide the best predictions.

In this task, time was selected as proxy for all other variables in the expectation that it would serve as an adequate proxy. Functional forms were chosen on the basis of how well they performed when 6 years of data were used to predict the next 2 years.

C. PARAMETER ESTIMATION TECHNIQUES

The final level on which modeling decisions must be made is the technical one of choosing techniques for determining the prediction parameters. Two techniques were discussed in Section II: regression and exponential smoothing. There are other techniques, such as moving averages and optimal filtering, that might also be considered. When a manageable number of variables are to be predicted, and especially when unmodeled changes are known to have occurred, consideration should be given to "eyeball" fitting and extrapolation, coupled with the use of experienced judgement.

APPENDIX B

ARREST-OFFENSE VECTORS

The anticipated effects of using helicopters as patrol vehicles include a reduction in the number of offenses (in at least some categories of crime) and an increase in the proportion of offenders caught. It has been observed in the past that (in most crime categories) arrests are positively correlated with offenses*. Hence, the data types dealt with here, arrests and offenses, are not independent, and a reduction in the number of offenses can cause a reduction in arrests that might mask, if the data types are treated separately, an increase in the fraction of offenders caught. This appendix presents the development of a technique for determining the statistical significance of test results when both data types are considered simultaneously.

The first step is to consider the relationship between arrests and offenses for a particular crime type in a particular division. This is done by combining the arrest and offense time-series in the arrest-offense plane (see Fig. B-1, the details of which will be discussed presently). Each time point then provides two coordinates (one from each time-series), which may be used to prepare an arrest-offense scattergram. The resulting "cloud" of historical points can be represented by a probability distribution, two contours of which are shown (labeled "1σ" and "2σ") in Fig. B-1. The contour lines shown are ellipses, a consequence of assuming that the appropriate probability distribution is bivariate Gaussian. Mathematically, the parameters of the ellipses are given by

*Since a larger number of offenses usually means that more criminals are working in the area, this is not a surprising observation.

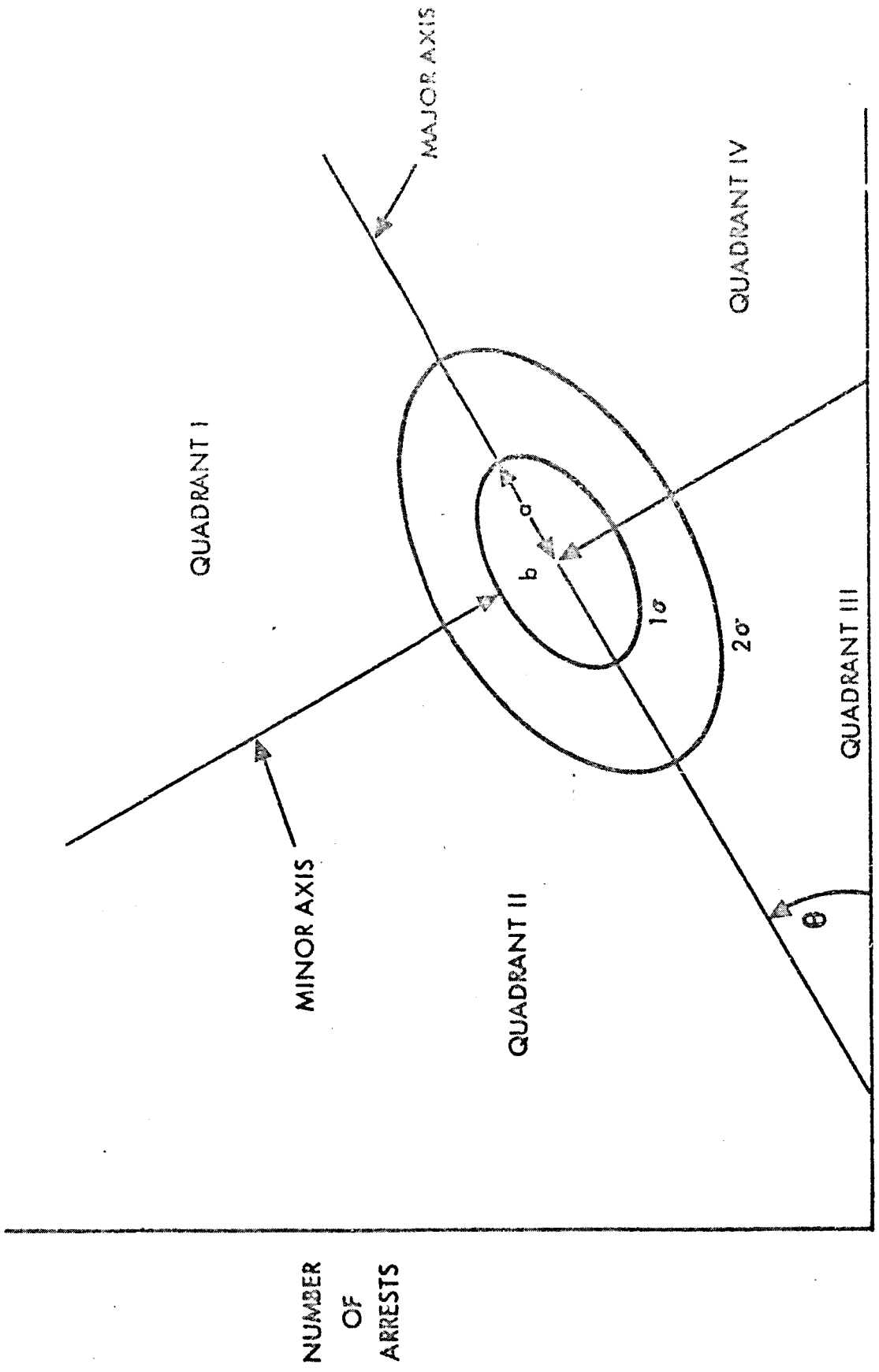

Fig. B-1. The arrest-offense plane

$$a = \left[\frac{\sigma_a^2 + \sigma_o^2 + \sqrt{\left(\sigma_a^2 - \sigma_o^2\right)^2 + 4\rho^2 \sigma_a^2 \sigma_o^2}}{2}\right]^{1/2}$$

$$b = \left[\frac{\sigma_a^2 + \sigma_o^2 - \sqrt{\left(\sigma_a^2 - \sigma_o^2\right)^2 + 4\rho^2 \sigma_a^2 \sigma_o^2}}{2}\right]^{1/2}$$

$$\theta = \frac{1}{2} \tan^{-1}\left[\frac{2\,\sigma_a \sigma_o}{\sigma_a^2 - \sigma_o^2}\right]$$

where

a = semimajor axis of the 1σ ellipse

b = semiminor axis of the 1σ ellipse

θ = orientation angle

and

ρ = correlation coefficient obtained from a simple regression of past arrests on past offenses

σ_a = standard deviation of arrest forecasts

σ_o = standard deviation of offense forecasts

Assuming that the bivariate Gaussian distribution is appropriate, the 1σ ellipse can be expected to contain 39% of the data points, the 2σ ellipse to contain 86%, and a 3σ ellipse to contain 99%. Further, data points are equally likely to fall into each of the quadrants.

In light of the previous discussion, it may be noted that points falling in Quadrants I or II suggest improved effectiveness in apprehending offenders, while points in Quadrants II or III suggest successful repression of crime. Only those points falling in Quadrant II suggest improved effectiveness in both areas, and only points in Quadrant IV suggest decreased effectiveness in both areas.

The next step is to determine what results are needed to conclude (at some statistical confidence level) that a change in the system (such as using

helicopters for patrol) has caused a bias toward some quadrant or pair of quadrants to occur*.

The problem may be stated mathematically as follows:

Given N statistically independent trials of some event R. What is the number n such that the probability is at least p (the confidence level) that the number of times the event occurs, k, will be less than n, under the hypothesis that the probability of occurrence of the event is P?

or

Find the smallest integer n such that $Pr(k < n \mid P, N) \geq p$

Since each trial is independent (by assumption) and has a probability P of resulting in R, the probability obeys the Bernoulli distribution law:

$$Pr(k = x \mid P, N) = \binom{N}{x} P^x (1 - P)^{N-x} \qquad \text{(B-1)}$$

Consequently,

$$Pr(k < n \mid P, N) = \sum_{x=0}^{n-1} \binom{N}{x} P^x (1 - P)^{N-x} \qquad \text{(B-2)}$$

*The following "crosswind" analogy has been suggested: Consider an ideal archer shooting along a long thin line. Ignore the few times his arrow actually lands on the line. This archer is ideal in that his shots are unbiased and statistically independent. If there is no wind, approximately half his shots will go to the left of the line and half to the right (since he is unbiased). It is not likely, however, that exactly half will go to each side. If there is a right-to-left crosswind, considerably more than half can be expected to go to the left of the line. If he shoots N arrows, how many must go on one side of the line before the existence of a crosswind has been (statistically) demonstrated?

Then, the desired value of n is that which satisfies

$$\sum_{x=0}^{n-1} \binom{N}{x} P^x (1-P)^{N-x} < p \leq \sum_{x=0}^{n} \binom{N}{x} P^x (1-P)^{N-x} \qquad \text{(B-3)}$$

To simplify notation, let

$$f_x = Pr(k = x \mid P, N) \qquad \text{(B-4)}$$

$$G_r = Pr(k \geq r \mid P, N) = \sum_{x=r}^{N} f_x \qquad \text{(B-5)}$$

$$q = 1 - p \qquad \text{(B-6)}$$

With this notation, Eq. (B-3) may be rewritten, after multiplying by minus one (which reverses the inequalities) and adding one, as Eq. (B-7).

$$G_n \leq q < G_{n-1} \qquad \text{(B-7)}$$

The value of n that satisfies these conditions cannot be found analytically, but must be determined by computation of the G_r or by table search. When N is large, the Gaussian approximation to the Bernoulli distribution may be used, with Eq. (B-7) becoming

$$\Phi\left(\frac{n - NP}{\sqrt{NP(1-P)}}\right) \geq p > \Phi\left(\frac{n - NP - 1}{\sqrt{NP(1-P)}}\right) \qquad \text{(B-8)}$$

where

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-s^2/2)\,ds$$

To recap, when n has been found from Eq. (B-7) or Eq. (B-8), its meaning is

n = the minimum number of occurrences of event R in N trials to reject, at a confidence level of p, the hypothesis that the probability that event R will occur is no greater than P.

This statistical test rests upon the assumption that differences between predicted and actual crime data are statistically independent from one time period to the next. Unfortunately, inspection of the crime-time plots for the non-test divisions showed that this assumption could not be relied upon. Thus, though application of the test to LAPD data during the helicopter test period gave results suggesting that the use of helicopters did indeed give improvement in both areas of effectiveness, this statement could not be given a statistical foundation.

APPENDIX C

ALGORITHMS FOR TIME-SERIES ANALYSIS

The extrapolation of time-series has been used for some time by statisticians, economists, scientists and others interested in forecasting the future. This appendix presents some discussion of the technical problems involved and two algorithms, multiple regression and modified multiple exponential smoothing, for producing predictions from streams of historical data. Since regression is an older technique, with a longer history of development, a more complete set of theoretical results is presented, but the analytical development is abbreviated. Exponential smoothing is a considerably newer technique: Theoretical results are less extensive, and more attention is devoted to the analytical development. It should be noted that both techniques are members of the general class of weighted regression techniques.

In general, time-series data exhibit four components: secular trend, cyclical variation, seasonal variation, and irregular fluctuations. The trend component is probably the most familiar; its existence is usually easily recognized when the series is graphed. This component commonly increases with time since many factors (for which time is used as a substitute, or proxy) grow as a result of population increases, technological advances, etc. Since, in the main, these underlying factors change smoothly with time, the trend components of time-series usually change smoothly (but not necessarily linearly) with time. The second, or cyclical component, has been chiefly observed and studied in economic time-series: As a result of feedback loops in the economic system, business conditions tend to vary between the extremes of boom and recession over intervals of several years. Cyclic fluctuations have also been observed in meteorological and sunspot data and undoubtedly exist in many other types of time-series. Seasonal effects are also evident in many time-series. Because of their regularity, it is common for published data to be seasonally adjusted to provide more readily understandable statistical information. The irregular component of time-series is the most difficult to interpret. These fluctuations result from factors that do not change smoothly with time. When the fluctuations

are small enough or when the processes giving rise to them are poorly understood, they are usually assumed to occur randomly (though perhaps with some autocorrelation). From a sampling of the time-series, an estimate of the expected variation can usually be inferred and the fluctuations may be approximated by a probability distribution. Occasionally, fluctuations occur which are unlikely to be associated with this distribution, possibly caused by unusual, nonrecurring events such as riots, natural disasters, etc. The difficulty in distinguishing between these two types of fluctuations makes the irregular component of time-series the most difficult to interpret and to deal with.

A common approach to time-series analysis has been to model first the trend by fitting a polynomial or other function of time to the data and then to search for periodic fluctuations that result from seasonal and cyclical influences. The seasonal effects are often quantified by indices obtained by averaging monthly or quarterly values of the series. Cyclical effects are often determined subjectively through observations of the residuals after the removal of trend and seasonal effects. A sophisticated technique for analysis of cyclical effects is spectral analysis, originally developed in research on telecommunication systems. This approach, which has been applied to economic time-series (see Ref. C-1), consists basically of the determination, by means of the spectral density function of the series, of the period and phase of cycles that account for a statistically significant portion of the series variance. This technique has the advantage of verifying the existence of suspected cycles and even uncovering cyclic behavior that would otherwise go unrecognized.

A. MULTIPLE REGRESSION

1. Preliminaries

The problem to be solved is the estimation of the coefficients (and related statistics) of a model relating one dependent variable, denoted y, to one or more (K, say) independent variables, denoted by x_k. The regression is <u>linear</u> if the only power any x_k takes in the model is unity. It is <u>multiple</u> if K is 2 or more. The dependent and independent variables all vary with some index variable (such as time); hence, a subscript, t, is added to the notation: y_t and x_{tk}. The model is then represented by

$$y_t = \sum_{k=1}^{K} \beta_k x_{tk} + e_t \qquad t = 1, 2, \ldots, T \tag{C-1}$$

where e_t is the residual, representing the "irregular fluctuations".

In the time-series application where the model is a polynomial in time plus seasonal constants, $K = n + L - 1$, where

n = the degree of the polynomial
L = the number of seasons in a year.

Further, the independent variables are

$$x_{it} = \delta_{it} \quad \text{for } i = 1 \text{ to } L$$

where

$$\delta_{it} \begin{cases} 1 & \text{if } t \text{ corresponds to season } i \\ 0 & \text{otherwise} \end{cases}$$

and

$$x_{L+k} = t^k \quad \text{for } k = 1 \text{ to } K$$

For convenience in notation, the various time-series are defined as vectors and matrices:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{pmatrix}, \quad X = \begin{pmatrix} x_{11}\, x_{12} \cdots x_{1K} \\ x_{21}\, x_{12} \cdots \\ \cdot \cdot \cdot \cdot \cdot \cdot \cdot \\ x_{T1}\, x_{T2} \cdots x_{TK} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_K \end{pmatrix}, \text{ etc.}$$

Then, the model, Eq. C-1, can be succinctly written as Eq. C-2.

$$Y = X\beta + E \tag{C-2}$$

2. Determination of Parameters

It can be shown (see, for example, Ref. C-2) that the choice of parameters that minimizes the sum of the squares of the residuals (hence the term "least squares"), thereby minimizing the variance, is given by Eq. (C-3).

$$B = (X'X)^{-1}X'Y \tag{C-3}$$

where the prine (') denotes a matrix transpose, and B is the vector of estimates of the coefficients, β_k. Then, the vector of smoothed values of Y, denoted $\hat{Y}$, is

$$\hat{Y} = XB \tag{C-4}$$

and the vector of estimated residuals, E, is estimated by

$$\hat{E} = Y - \hat{Y} = Y - XB \tag{C-5}$$

3. Coefficient of Determination and Multiple Correlation Coefficient

The coefficient of determination, R^2, is defined as the ratio of the amount of variation "explained" by the regression to the variation of the original series of the dependent variable:

$$R^2 = 1 - \frac{Y'Y - BX'Y}{\sum_{t=1}^{T} (y_t - \bar{y})^2} \tag{C-6}$$

where

$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t$, the mean value of the original series.

The multiple correlation coefficient, R, commonly used as a measure of the "goodness" of fit, can take on values between -1 and +1, with the extreme values representing perfect negative and perfect positive correlation, respectively, between the dependent and independent variables. Multiple correlation coefficient values near zero represent regressions with little or no correlation.

4. Statistical Inference

The estimated standard deviation of the fit, denoted s, may be found from Eq. (C-7).

$$s^2 = \hat{E}'\hat{E}/(T - K - 2) \qquad \text{(C-7)}$$

The estimated covariance matrix of the coefficient vector, denoted S_{bb}, is given by Eq. (C-8).

$$S_{bb} = s^2 (X'X)^{-1} \qquad \text{(C-8)}$$

5. Forecasting

Suppose a forecast is desired at some time $t = \tau$, and estimated values of the independent variables, $\hat{X}_\tau$, are known. Then the forecast for the dependent variable is simply

$$\hat{y}_\tau = \hat{X}'_\tau B, \qquad \text{(C-9)}$$

and the estimated variance of the forecast, s_f^2, is*

$$s_f^2 = s_{\hat{y}_T}^2 + s^2 + s_{\hat{X}_T}^2 \qquad \text{(C-10)}$$

where

$$s_{\hat{y}_T}^2 = \hat{X}_T' S_{bb} \hat{X}_T \qquad \text{(C-11)}$$

6. Confidence Intervals

Upper and lower 100γ percent confidence limits on the model parameters are given in Eq. (C-12).

$$\left.\begin{matrix} U \\ L \end{matrix}\right\} = B \pm z_{\nu p} S_b \qquad \text{(C-12)}$$

where S_b is a K x 1 vector of the square roots of the diagonal elements of S_{bb},

$z_{\nu p}$ is from a Student's t distribution of cumulative probability,

$p = (1 + \gamma/2)$, and

$\nu = T - K - 2$ degrees of freedom

*If the independent variables are known exactly, then $s_{\hat{X}_T}^2$ is, of course, zero.

The statistical significance of individual components of B is demonstrated at the 100γ percent confidence level if

$$\frac{b_i}{S_{b_i}} > z_{\nu p} \qquad \text{(C-13)}$$

where b_i and S_{b_i} are the i-th elements of the B and S_b vectors, respectively.

Similarly, 100γ percent confidence levels for a forecast are given by Eq. (C-14).

$$\left.\begin{matrix} U \\ L \end{matrix}\right\} = Y_T \pm z_{\nu p}\, s_f \qquad \text{(C-14)}$$

B. EXPONENTIAL SMOOTHING

Exponential smoothing has emerged as a competitor* to simple regression as a technique for estimating the parameters of a model fit to a nonstationary discrete time-series. Theoretically speaking, exponential smoothing** and simple regression are members of the more general class of weighted regression techniques: In simple regression, data points are weighted equally; in exponential smoothing, they are weighted by an exponential decreasing with the "ages" of the data points. As a result of giving more weight to the more recent data, exponential smoothing often performs better in extrapolation, particularly when the process is modeled imperfectly or is nonstationary.

*There are, of course, other techniques, such as moving averages and Fourier analysis, which are also competitive.

**This statement holds for exponential smoothing as developed by Brown in Ref. C-3. (It is proved in his Appendix A.) In some recent generalizations, such as presented in Ref. C-4, the weight given to each data point may be different for computation of estimates of different parameters in the model being fit.

The exponential smoothing technique requires initial estimates of the model parameters. After a sufficiently long time, depending on the smoothing constant (that is, the rate at which the weighting factors decay), the initial estimates do not influence the estimates produced. In the meantime, a poor set of initial estimates can lead to poor predictions. Here, Brown's development (Ref. C-3, Chapter 9) of multiple exponential smoothing is paralleled to derive a modified multiple exponential smoothing procedure which does not require initial estimates of the parameters. The result is analogous to computation of a cumulative average, such as is used to report season-to-date batting averages for baseball players.

1. Preliminaries

The problem to be solved is the estimation of the coefficients of an nth-degree polynomial in time* to model a given discrete time-series.

Let

y_t for $t = 1, 2, \ldots, T$ represent the uniformly spaced (input data) values of the discrete time-series to be modeled,

$\hat{y}_t$ for any (integral) t represent the predicted or smoothed values of the time-series,

c_k for $k = 1, 2, \ldots, L$ and any t represent additive seasonal effects for each of L seasons, and

x_t for $t = 1, 2, \ldots, T$ represent the seasonally adjusted time-series,

so that

$$x_t = y_t + c_{k(t)} \quad \text{and} \quad \hat{y}_t = \hat{x}_t - \hat{c}_{k(t)} \tag{C-15}$$

*The index variable, denoted here by t, could refer to some other quantity instead of time, but it is convenient to use ther term "time" rather than the more precise, but awkward term "the index variable."

where k(t) is the season corresponding to time t and $\hat{x}_t$ is the estimate of x_t. The assumed nth-degree polynomial model can then be expressed as

$$x_{t+\tau} = \sum_{r=0}^{n} \frac{a_r}{r!} (t+\tau)^r + e_{t+\tau} \qquad \text{(C-16)}$$

or as

$$x_{t+\tau} = \sum_{r=0}^{n} \frac{b_r}{r!} \tau^r + e_{t+\tau} \qquad \text{(C-17)}$$

where

a_r, b_r are alternative expressions for the coefficients of the polynomial depending upon whether the polynomial is expanded about time 0 or time t, respectively, and

e_t is the residual (that is, the difference between the fit and the data), and is often assumed to be an independent, Gaussian random variable.

Several of the variables defined above will be estimated at different points in time. The following conventions will be used to indicate such estimates: When the variable does not already have a subscripted t, a subscript t will be added. (For example, b_{rt} will represent an estimate of $_r$ made from data available at time t.) When the variable does already have a subscripted t (as does y_t), then a circumflex (or "hat") will be added (as in $\hat{y}_t$).

Often, the seasonal effects must also be estimated prior to their removal. Since

$$\hat{x}_t = \sum_{r=0}^{n} b_{r,\,t-1}/r!$$

this can be done as follows:

$$\hat{c}_{k(t)} = \hat{c}_{k(t-L)} + c_{kt} \left[\sum_{r=0}^{n} \frac{b_{r,\,t-1}}{r!} - \left(y_t + \hat{c}_{k(t-L)} \right) \right] \qquad \text{(C-18)}$$

where single exponential smoothing is applied to each seasonal constant separately* with a possibly different smoothing constant for each. Time-varying values of the smoothing constants, α_k, as shown later**, can be found from

$$\alpha_{kt} = \frac{\alpha_k}{1 - (1 - \alpha_k)^t} \qquad \text{(C-19)}$$

where α_k is the ultimate value of the kth smoothing constant.

With these preliminaries out of the way, it is possible to proceed to the next subsection with the seasonally adjusted time-series x_t for which the polynomial coefficients b_{rt} will be estimated.

If it is desired to study the variations in the fitting constants, it may be more convenient to deal with the a_{rt}, since their values do not nominally vary with t. (That is, the $a_{rt} = 2_r$ for all t if the data is from a noise-free polynomial of degree less than or equal to n.) The a_{rt} and b_{rt} are related by the following equations:

$$a_{rt} = \sum_{i=0}^{n-r} (-t)^i \, b_{r+i,t}/i! \; ; \; r = 0, 1, \ldots, n \qquad \text{(C-20)}$$

*Cf. Pegels' models in Ref. C-4.

**See also Ref. C-5.

2. The Fundamental Theorem of Modified Exponential Smoothing

To proceed with the development of exponential smoothing, the next step is to define multiple smoothing of orders 1 through p. Let

$$S_t = \alpha_t x_t + \beta_t S_{t-1}$$

$$S_t^{(2)} = \alpha_t S_t + \beta_t S_{t-1}^{(2)} \qquad \text{(C-21)}$$

. . .

$$S_t^{(p)} = \alpha_t S_t^{(p-1)} + \beta_t S_{t-1}^{(p)}$$

where α_t is the smoothing constant to be used at time t, and is so defined that the ratio of the weights assigned to successive data points is $(1 - \alpha)$, and the asymptotic value of α_t is α, the smoothing constant chosen from considerations not discussed here.

$\beta_t = 1 - \alpha_t$ for convenience, since it occurs often. Similarly, $\beta = 1 - \alpha$.

Starting with $S_1 = x_1$, these difference equations can easily be solved for the $S_t^{(p)}$ in terms of the x_t to provide the following results:

$$S_t = \frac{x_t + \beta x_{t-1} + \beta^2 x_{t-2} + \ldots + \beta^{t-1} x_1}{1 + \beta + \beta^2 + \ldots + \beta^{t-1}} = \frac{\alpha}{1 - \beta t} \sum_{\tau_1=0}^{t-1} \beta^{\tau_1} x_{t-\tau_1}$$

$$S_t^{(2)} = \frac{\alpha}{1 - \beta^t} \sum_{\tau_1=0}^{t-1} \beta^{\tau_1} \left[\frac{\alpha}{1 - \beta^{t-\tau_1}} \sum_{\tau_2=0}^{t-1-\tau_1} \beta^{\tau_2} x_{t-\tau_1-\tau_2} \right] \qquad \text{(C-22)}$$

. . .

$$S_t^{(p)} = \frac{\alpha}{1 - \beta^t} \sum_{\tau_1=0}^{t-1} \beta^{\tau_1} \left[\frac{\alpha}{1 - \beta^{t-\tau_1}} \sum_{\tau_2=0}^{t-1-\tau_1} \left(\cdots \sum_{\tau_p=0}^{t-1-\sum_{i=1}^{p-1} \tau_i} \beta^{\tau_p} x_{t-\sum_{i=1}^{p} \tau_i} \right) \right]$$

From these equations, it is evident that the time-varying smoothing constant, α_t, must be given by

$$\alpha_t = \frac{\alpha}{1 - \beta^t} \quad \text{or, equivalently,} \quad \alpha_t = \frac{\alpha_{t-1}}{\beta + \alpha_{t-1}} \qquad \text{(C-23)}$$

If it is now assumed that the observations are taken from an nth-degree polynomial in time, so that

$$x_{t-\rho} = b_{0t} - b_{1t}\rho + \frac{1}{2!} b_{2t} \rho^2 - \ldots + \frac{1}{n!} b_{nt} (-\rho)^n = \sum_{j=1}^{n+1} \frac{(-\rho)^{j-1}}{(j-1)!} b_{j-1,t} \qquad \text{(C-24)}$$

then the fundamental theorem of modified exponential smoothing states that the smoothed series $S_t, \ldots, S_t^{(p)}$ can be expressed as linear combinations of the coefficients $b_{0t}, \ldots, b_{nt}$ and that consequently (if $p = n + 1$) the coefficients can be expressed as linear combinations of the smoothed series. More succinctly, the fundamental theorem states that there exists a matrix $\underset{\sim}{M}_t$ such that

$$\underset{\sim}{S}_t = \underset{\sim}{M}_t \underset{\sim}{b}_t \qquad \text{(C-25)}$$

with (if $p = n + 1$) an inverse $\underset{\sim}{M}_t^{-1}$, so that

$$\underset{\sim}{b}_t = \underset{\sim}{M}_t^{-1} \underset{\sim}{S}_t \qquad \text{(C-26)}$$

where

$$\underset{\sim}{b}_t \triangleq \begin{pmatrix} b_{0t} \\ b_{it} \\ \cdots \\ b_{nt} \end{pmatrix} ; \quad \underset{\sim}{S}_t \triangleq \begin{pmatrix} S_t \\ S_t^{(2)} \\ \cdots \\ S_t^{(p)} \end{pmatrix} ;$$

and $\underset{\sim}{M}_t$ is the upper left hand corner of

$$\begin{bmatrix} m_{11t} & m_{12t} & m_{13t} & \cdots \\ m_{21t} & m_{22t} & m_{23t} & \cdots \\ m_{31t} & m_{32t} & m_{33t} & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

The theorem may be demonstrated by substituting Eq. (C-24) into Eq. (C-22), which gives, for r = 1, ..., p,

$$S_t^{(r)} = \frac{\alpha}{1-\beta^t} \sum_{\tau_1=0}^{t-1} \beta^{\tau_1} \left[\frac{\alpha}{1-\beta^{t-\tau_1}} \sum_{\tau_2=0}^{t-1-\tau_1} \beta^{\tau_2} \left(\cdots \right.\right.$$

$$\left.\left. \left(\frac{\alpha}{1-\beta^{t-1-\sum_{i=1}^{r-1}\tau_i}} \right)^{t-1-\sum_{i=1}^{r-1}\tau_i} \sum_{\tau_r=0} \beta^{\tau_r} \left\{ \sum_{j=1}^{n+1} \frac{b_{j-1,t}}{(j-1)!} \left(-\sum_{i=1}^{r} \tau_i \right)^{j-1} \right\} \right) \right] \qquad \text{(C-27)}$$

Comparison of Eqs. (C-25) and (C-27) shows that

$$m_{rjt} = \frac{(-)^{j-1}\alpha^r}{(1-\beta^t)(j-1)!} \sum_{\tau_1=0}^{t-1} \frac{\beta^{\tau_1}}{1-\beta^{t-\tau_1}} \sum_{\tau_2=0}^{t-1-\tau_1} \frac{\beta^{\tau_2}}{1-\beta^{t-\tau_1-\tau_2}}$$

(C-28)

$$\cdots \sum_{\tau_{r-1}=0}^{t-1-\sum_{i=1}^{r-2}\tau_i} \frac{\beta^{\tau_{r-1}}}{1-\beta^{t-\sum_{i=1}^{r-1}\tau_i}} \sum_{\tau_r=0}^{t-1-\sum_{i=1}^{r-1}\tau_i} \beta^{\tau_r} \left(\sum_{i=1}^{r} \tau_i \right)^{j-1}$$

3. Computational Sequence

The algorithm for use of modified exponential smoothing is essentially the same as that used by Brown (Ref. C-2). Explicitly, the following sequence of steps may be used:

1) Step 1: Initialization

a) A priori guesses of the values of the coefficients, b_r, are not required, but starting values are needed. One way to obtain the first set would be to temporarily ignore the seasonal constants, c_k, and fit the first L (or the first n + 1) data points to the polynomial. Thus, if $L > (N + 1)$, set $b_{t=L} = (\underset{\sim}{F}'\underset{\sim}{F})^{-1}\underset{\sim}{F}\underset{\sim}{Y}$, where

$$\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_L \end{pmatrix} \quad \text{and} \quad \underset{\sim}{F} = \begin{pmatrix} 1 & -(L-1) & (L-1)^2 & \cdots & (-)^n (L-1)^n \\ 1 & -(L-2) & (L-2)^2 & \cdots & (-)^n (L-2)^n \\ \cdot & \cdots & \cdots & \cdots & \cdots \\ 1 & -1 & 1 & \cdots & (-)^n \; 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

If $L \leq (n + 1)$, then simply fit the first n + 1 points to the polynomial by $\underset{\sim}{b}_{t=n+1} = \underset{\sim}{F}^{-1}\underset{\sim}{y}$, where

$$\underset{\sim}{y} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_{n+1} \end{pmatrix} \quad \text{and} \quad \underset{\sim}{F} = \begin{pmatrix} 1 & -n & n^2 & \cdots & (-)^n n^n \\ 1 & -(n-1) & (n-1)^2 & \cdots & (-)^n (n-1)^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & -1 & 1 & \cdots & (-)^n \; 1 \\ 1 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

b) With starting values of the coefficients in hand, the starting values of the smoothed series (starting at t = L or n + 1) may be obtained from the fundamental theorem, Eq. (C-25):

$$\underset{\sim}{S}_{t = L \text{ or } n+1} = \underset{\sim}{M}_{t = L \text{ or } n+1} \; \underset{\sim}{b}_{t = L \text{ or } n+1}$$

c) Next, it is necessary to obtain starting values for the seasonal constants. If $L > (n + 1)$, the residuals during the initial period can be used, so that

$$c_{k(t-q)} = \begin{cases} b_{0t} - y_t & \text{for } q = 0 \\ \sum_{r=0}^{n} b_{rt}(-q)^r / r! - y_q & \text{for } q = 1, \ldots, L-1 \end{cases}$$

If, on the other hand, $L \leq (n + 1)$, the polynomial fits the data points exactly, so the next L points may be used to find initial values of the seasonal constants. Perhaps a better technique is to use 2L or 3L data points in finding the initial set of b_{rt}.

d) At completion of the initialization, the smoothed series, $\underset{\sim}{S}_t$, and the seasonal constants, c_k, are available as of some time t.

2) Step 2: Increment the time index.

a) Increment t by 1. Find the seasonally adjusted value of the time-series from $x_t = y_t + c_{k(t-L)}$. Compute the new value of the smoothing constant, α_t, from Eq. (C-23).

b) Compute the smoothed series from Eq. (C-21).

c) Compute new values of the coefficients from Eq. (C-26), and update the estimate of the appropriate seasonal constant by Eq. (C-18).

d) If desired, use Eq. (C-15) to compute the smoothed value of the time-series, $\hat{y}_t$, and compile any statistics of interest.

3) Step 3: Exit.

Has the available data been exhausted? If not, return to Step 2. If so, the $b_{t_{max}}$ and c_k are now available for prediction by Eqs. (C-15) and (C-17).

4. Elements of the M_t and M_t^{-1} Matrices

It can be shown that

$$S_n(t, x) \equiv \sum_{k=1}^{t=1} k^n x^k = S_n(\infty, x) - x^t t^n D_n(t, x)/(1 - x)^{n+1}$$

where $S_n(\infty, x) = \sum_{j=1}^{n} A_{nj} x^j$, and

$$D_n(t, x) = (1 - x)^n + \sum_{r=1}^{n} \sum_{j=1}^{n} C_{rj}(n) x^j / t^r$$

The A_{nj} are Eulerian numbers and the $C_{rj}(n)$ are another set of numbers, given by

$$C_{rj}(n) = \binom{n}{r} \sum_{i=0}^{j-1} (-)^i \binom{n+1}{i} (j - i)^r, \text{ and}$$

$$A_{nj} = C_{nj}(n)$$

Recursion* and symmetry relations to simplify calculation of these numbers can be demonstrated:

$$A_{nj} = \begin{cases} 1 & \text{for } j = 1 \\ jA_{n-1,j-1} + (n-j+1)A_{n-1,j} & \text{for } 1 < j \le n/2 \\ A_{n,n-j+1} & \text{for } n/2 < j \le n \end{cases}$$

$$C_{rj}(n) \begin{cases} \binom{n}{r} & \text{for } j = 1,\ 1 \le r \le n \\ (-)^{j-1} n \binom{n-1}{j-i} & \text{for } i < j \le n/2,\ r = 1 \\ \Big[(n+1-j)C_{r-1,j-1}(n-1) + jC_{r-1,j}(n-1) & \\ \qquad - C_{r,j-1}(n-1) + C_{rj}(n-1)\Big] & \text{for } 1 < j \le n/2,\ i < r < n \\ jC_{n-1,j-1}(n-1) + (n+1-j)C_{n-1,j}(n-1) & \text{for } 1 < j \le n/2,\ r = n \\ (-)^{n-r} C_{r,n+1-j}(n) & \text{for } n/2 < j \le n,\ 1 \le r \le n \end{cases}$$

*The author is indebted to Dr. Harry Lass for discovery of the recursion relation for the $C_{rj}(n)$.

In particular, the first few sums are:

$$S_0(t, x) = (x - x^t)/(1 - x)$$

$$S_1(t, x) = [x - t\,x^t + (t - 1)\,x^{t+1}]/(1 - x)^2$$

$$S_2(t, x) = [x + x^2 - t^2 x^t + (2t^2 - 2t - 1)\,x^{t+1} - (t - 1)^2\,x^{t+2}]/(1 - x)^3$$

$$S_3(t, x) = [(1 - x^t)(x + 4x^2 + x^3) - t^3 x^t (1 - x)^3 - x^{t+1}\{(3t^2 + 3t + 1) + x(-6t^2 + 4) + x^2(3t^2 - 3t + 1)\}]\ /(1 - x)^4$$

These sums can be used to obtain simpler expressions for some of the elements of the $\underset{\sim}{M}_t$ matrix given in Eq. (C-28). In particular, with $j = 1$,

$$m_{rlt} = 1 \quad \text{for all } r, t \ .$$

For $r = 1$, $j > 1$,

$$m_{1jt} = \frac{(-)^{j-1}}{(j-1)!}\left(\frac{\alpha}{1-\beta^t}\right) S_{j-1}(t, \beta)$$

$$= \frac{(-)^j}{(j-1)!}\left(\frac{1}{1-\beta^t}\right)\left[\beta^t t^{j-1} - \frac{1}{\alpha^{j-1}} \sum_{k=1}^{j-1} \beta^k \left(A_{j-1,k} - \beta^t \sum_{r=1}^{j-1} C_{rk}(j-1)\, t^{j-1-r}\right)\right]$$

Hence,

$$m_{12t} = -\frac{\beta}{\alpha} + \frac{t\beta^t}{1-\beta^t}$$

$$m_{13t} = \frac{\beta(1+\beta)}{2\beta^2} - \frac{t\beta^t}{2(1-\beta^t)}\left(t + \frac{\beta}{\alpha}\right)$$

Similarly, for $r = 2$, $j > 1$,

$$m_{2jt} = \frac{(-)^{j-1}}{(j-1)!}\left(\frac{\alpha^2}{1-\beta^t}\right)\left[\sum_{q=1}^{j=1}\binom{j-1}{q}\sum_{\tau_1=0}^{t-1}\frac{\beta^{\tau_1}\tau_1^{j-1-q}}{1-\beta^{t-\tau_1}}S_q(t-\tau_1,\beta) + \frac{S_{j-1}(t,\beta)}{\alpha}\right]$$

In addition to the $\underset{\sim}{M}_t$ matrix used in Eq. (C-25), its inverse, $\underset{\sim}{M}_t^{-1}$, is also required, as may be seen in Eq. (C-26).

For exponential smoothing where the data base is assumed to be (practically speaking) infinite, the M matrix is constant*, and is the upper left hand corner of

*See Brown (Ref. C-3), p. 135.

$$\underset{\sim}{M}_{\infty} = \begin{pmatrix} 1 & -\frac{\beta}{\alpha} & \frac{\beta(1+\beta)}{2\alpha^2} & \cdots \\ 1 & -\frac{2\beta}{\alpha} & \frac{2\beta(1+2\beta)}{2\alpha^2} & \cdots \\ 1 & -\frac{3\beta}{\alpha} & \frac{3\beta(1+3\beta)}{2\alpha^2} & \cdots \\ \cdot & \cdots & \cdots & \cdots \end{pmatrix}$$

Inverses of this matrix for n = 2 and 3 are presented here, even though computation is trivial, for the convenience of the reader.

When n = 2, $\underset{\sim}{M}_{\infty}^{-1} = \begin{pmatrix} 2 & -1 \\ \alpha/\beta & -\alpha/\beta \end{pmatrix}$

When n = 3, $\underset{\sim}{M}_{\infty}^{-1} = \begin{pmatrix} 3 & -3 & 1 \\ \frac{\alpha(6-5\alpha)}{2\beta^2} & -\frac{\alpha(10-8\alpha)}{2\beta^2} & \frac{\alpha(4-3\alpha)}{2\beta^2} \\ \alpha^2/\beta^2 & -2\alpha^2/\beta^2 & \alpha^2/\beta^2 \end{pmatrix}$

The inverses, for n = 2 and 3, of the more general $\underset{\sim}{M}_t$ matrix are given below. The subscripted t has been omitted for improved readability.

If n = 2,

$$\underset{\sim}{M} = \begin{pmatrix} 1 & m_{12} \\ 1 & m_{22} \end{pmatrix} \text{ then } \underset{\sim}{M}^{-1} = \frac{1}{m_{22} - m_{12}} \begin{pmatrix} m_{22} & -m_{12} \\ -1 & 1 \end{pmatrix}$$

while if $n = 3$,

$$\underset{\sim}{M} = \begin{pmatrix} 1 & m_{12} & m_{13} \\ 1 & m_{22} & m_{23} \\ 1 & m_{32} & m_{33} \end{pmatrix}$$

then

$$\underset{\sim}{M}^{-1} = \frac{1}{D}\begin{bmatrix} (m_{22}m_{33} - m_{32}m_{23}) & (m_{32}m_{13} - m_{12}m_{33}) & (m_{12}m_{23} - m_{22}m_{13}) \\ (m_{23} - m_{33}) & (m_{33} - m_{13}) & (m_{13} - m_{23}) \\ (m_{32} - m_{22}) & (m_{12} - m_{32}) & (m_{22} - m_{12}) \end{bmatrix}$$

where

$$D = \det \underset{\sim}{M} = m_{12}(m_{23} - m_{33}) + m_{22}(m_{33} - m_{13}) + m_{32}(m_{13} - m_{23})$$

With modern high-speed computers, computation of the elements of these matrices is a simple matter. It should be noted that they depend upon α and t, but not upon the time-series data, and need not be recalculated for each time-series.

C. REFERENCES

C-1 Granger, C. W. J., Spectral Analysis of Economic Time Series. Princeton University Press, 1964.

C-2 Goldberger, A. S., Econometric Theory. John Wiley and Sons, 1964.

C-3 Brown, R. G., Smoothing, Forecasting and Prediction of Discrete Time Series. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1963.

C-4 Pegels, C. C., "Exponential Forecasting: Some New Variations," Management Science, Vol. 15, No. 5, January 1969.

C-5 Chamberlain, R. G., "Elimination of Starting Values in Exponential Smoothing," JPL SPS 37-59, Vol. III, pp. 264-5.