1968

A comparison of spectral analysis and adaptive exponential smoothin forecasting models

Steven M. Brown
Lehigh University

A COMPARISON OF SPECTRAL ANALYSIS AND ADAPTIVE EXPONENTIAL SMOOTHING FORECASTING MODELS

by

Steven M. Brown

A Thesis

Presented to the Graduate Committee

of Lehigh University

in Candidacy for the Degree of

Master of Science

in

Industrial Engineering

Lehigh University

1968

CERTIFICATE OF APPROVAL

This thesis is accepted and approved in partial fulfillment of the requirements for the degree of Master of Science.

May 14, 1968
Date

Sutton Monro
Professor in Charge

Chairman of the Department
of Industrial Engineering

ACKNOWLEDGMENTS

The author wishes to express his gratitude for the advise and guidance which he received from his faculty advisor, Professor S. Monro.

The author also wishes to express thanks to Mr. K. E. Larson of the Western Electric Engineering Research Center for advice and aid given in the formulation of the thesis topic, and encouragement during thesis work.

Finally, the author wishes to thank his wife, Marilyn, for encouragement and patience during the preparation of this paper.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont.)

ABSTRACT

Two basically different approaches to forecasting, spectral analysis and adaptive exponential smoothing, are compared by measuring squared error of forecasts on controlled computer-generated time series. Three adaptive models, each using three variations of exponential smoothing are compared with three variations of spectral analysis. The resulting squared forecast errors are used in a series of F-tests to detect those parameter combinations for which each forecasting technique is not satisfactory. All of the techniques are then compared for significant forecasting differences via another series of F-tests. The spectral analysis techniques are found to be significantly better over most of the parameter choices.

CHAPTER I

INTRODUCTION

The accuracy of forecasts is one of the principal factors influencing the effectiveness of management control. For most companies, decisions in the areas of production planning, purchasing, employment rates and employee utilization, inventory levels, sales effort, etc. are all dependent upon forecasted requirements. Large inaccuracies in these forecasts will often result in considerable additional costs from additional inventory carrying costs, unsellable goods, lost sales, loss of goodwill from backorders, and reduced facility utilization.

In order to improve forecasting speed and accuracy, a number of techniques are available which permit machine computation of projected values of time series. A large number of items can be forecast in a short time by computer, and a variety of considerations may be introduced into the forecasting technique to take advantage of any additional available information. The results of these forecasts can then be used directly in computerized policy optimization systems if so desired.

There is a considerable range of complexity available in forecasting techniques. One basic method which has found considerable acceptance and is widely used is exponential smoothing. The ease and speed of computation together with the very small amount of stored information required to project historical data into the future makes exponential forecasting a highly useful tool. Brown[3,4] treats the basic properties of exponential smoothing models and develops several variations which remedy certain shortcomings of the basic model. A whole group of models has been discussed by Fleisher.[12]

The presence of various properties in the time series data has been exploited in the design of some models. Cohen[5] considers the affect of autocorrelation in the input data. Hinich and Farley[23] consider the problem of model design for data with nonstationary mean. One particular type of model which has attracted attention is the adaptive exponential smoothing model. An adaptive model adjusts the nature of the forecast according to the accuracy of prior forecasts. Bossons[2] discusses the properties of adaptive forecasts under certain conditions of error and when nonstationary data is being handled. Hester[21] looks at the adaptation to data which has step jumps in the mean. Nerlove and Wage[32] and Theil and Wage[40] consider some of the factors determining an optimal adaptive forecasting technique.

Chapter 2 will consider some of the general properties of exponential smoothing, and will consider the adaptive exponential smoothing model of Trigg and Leach[41] and two of the models of Salmon.[39]

Another forecasting technique which has had considerable theoretical treatment is spectral analysis. This procedure has evolved out of research in the areas of power spectra analysis and communication theory. A number of works have treated with the theory of spectral analysis for time series analysis, including the books of Blackman and Tukey[1] and Grenander and Rosenblatt,[17] and articles by Parzen.[33,34,35]

A number of additional works dealing with the theory of spectral analysis appear in the bibliography.

Application of spectral analysis has been made in a wide variety of areas. Granger and Hatanaka[16] discuss the use of spectral analysis in handling economic time series data. Nerlove[31] applies it to analysis of seasonally adjusted economic data. Elmaghraby[8] uses it in forecasting for inventory control. Examples of the application of spectral analysis in analyzing data are found in Fishman and Kiviat's[11] work with simulation models and Larson's[28] application to sequences of numbers generated by a random number generator. Analysis of oceanographic and meteorlogical data has been another area of application.

In order for spectral analysis theory to apply, certain assumptions are required to be met by the data. Most "live" data will not completely fulfill these requirements. In addition, the technique is far more difficult to apply and requires considerable data and computations. The theory of spectral analysis, together with the associated limitations and drawbacks of the technique, will be discussed in Chapter 3.

The intent of this study is to compare several adaptive exponential smoothing techniques with spectral analysis via a controlled set of simulation experiments. In order to do this, a time series generator, described in Chapter 4 was designed. The selection of values for various parameters in this model permitted evaluation of the forecasting capabilities of the chosen procedures. The techniques of Fiddleman[10] and Hester[21] in the construction of forecast comparison tests were considered with respect to nature of the time series generator and evaluation of the results.

The parameter choices, experimental data, and statistical methods used to analyze it are discussed in Chapter 5.

The results of the analysis and conclusions drawn are treated in Chapter 6.

Some additional areas of investigation were suggested by certain problems encountered during the experimentation. Other possible extensions of the work also seemed of interest. These are discussed in Chapter 7.

CHAPTER 2

EXPONENTIAL SMOOTHING MODELS

2.1 Moving Averages and Elementary Exponential Smoothing Models

Exponential smoothing has found great acceptance as a forecasting technique. The properties of this technique can be shown by first considering moving averages. Given a discrete time series $X(t_1)$, $X(t_2), \ldots, X(t_i), \ldots, X(t_n)$ one procedure available for prediction of the next period is the average of the latest K periods, i.e.

$$f_{n+1} = \frac{1}{K} \sum_{i=0}^{K-1} X_{n-i}$$

where f_{n+1} is the forecast for period n+1,

K is an integer, $K > 0$,

X_{n-i} is the value for time n-i, i.e. $X_{n-i} = X(t_{n-i})$.

The response of such a model can be quite satisfactory for a stable process with little or no seasonal or trend effect. The magnitude of K determines the rate of response to changes in the time series data. A large value for K causes reaction to a change in the time series to be slow, but yields a more stable forecast when there is present a large random error. A small value for K causes more rapid reaction to changes, but at the expense of stability. When a trend in the mean is present in the time series, the moving average will constantly lag behind the trend. The presence of seasonal components in the time series can have different results dependent upon K. For small values of K, the model responds to the seasonal component, but lags behind the variation caused by the component. For K of magnitude about equal to the period of a seasonal component, the effects of the seasonal variation will be averaged out, and no response to the component will result.

Exponential smoothing can be viewed as a form of weighted moving average. As presented by Brown[3,4] the basic model for exponential smoothing is:

(2.1) $f_{n+1} = \alpha X_n + (1-\alpha) f_n$

where f_i is the forecast for period i,

α is the smoothing constant, $0 < \alpha \leq 1$,

X_n is the historical value for period n.

By application of relation (2.1) recursively the following relationship is found:

(2.2) $$f_{n+1} = \alpha X_n + \alpha(1-\alpha)X_{n-1} + \alpha(1-\alpha)^2 X_{n-2} + \ldots + \alpha(1-\alpha)^k X_{n-K} + \ldots$$

It can be seen in equation (2.2) that the forecast f_{n+1} uses all historic data, but reduces the effect of data that is K+1 periods past by the exponential weight $\alpha(1-\alpha)^k$. For a constant (infinite) series $X_1 = X_2 = \ldots = X_2 = \ldots = X$, the forecast would be X, since:

$$f_{n+1} = \alpha X + \alpha(1-\alpha)X + \ldots + \alpha(1-\alpha)^k X + \ldots$$
$$= \alpha X \left[1 + (1-\alpha) + (1-\alpha)^2 + \ldots \right]$$
$$= \alpha X \left[\frac{1}{1-(1-\alpha)} \right] = \frac{\alpha X}{\alpha} = X.$$

The effect of applying the technique to a finite constant time series of length n is as follows:

$$\begin{aligned} f_{n+1} &= \alpha X + \alpha(1-\alpha)X + \ldots + \alpha(1-\alpha)^{n-1}X \\ &= \alpha X\left[1 + (1-\alpha) + \ldots + (1-\alpha)^{n-1}\right] \\ &= \alpha X\left[\frac{1-(1-\alpha)^n}{1-(1-\alpha)}\right] = X\left[1-(1-\alpha)^n\right]. \end{aligned}$$

Since $\alpha > 0$, $(1-\alpha) < 1$ and for moderate n, the size of $(1-\alpha)^n$ will be small. Hence, the error of this forecast will normally be small when using a finite amount of history data, e.g.:

		n = 10	n = 25
α	0.1	.349	$7.6\cdot10^{-2}$
	0.5	$9.7\cdot10^{-4}$	$3\cdot10^{-8}$

Sample of errors for α given and n.

The result of using exponential smoothing is to weigh the most recent historical value by α, and every other historical value by lesser weight. One measure of the performance of the technique is its response to specific perturbances in the time series data. One such perturbance is a jump in mean to a new value which persists. The exponential smoothing model reacts by a change of α times the difference in the two levels in the first forecast following the jump. Subsequent forecasts approach the new value along a geometric curve.

The response to an impulse (a one period perturbance) is the same in the first period following as for the jump in mean. The original value is approached during subsequent periods along a geometric curve.

The exponential smoothing model does not react in as satisfactory a fashion to a time series with a trend in the mean. When a trend is introduced into a constant time series, the subsequent forecasted values will fall for several periods farther and farther behind the actual values until reaching a point where the forecasts change at a rate equal to the trend, but lag a constant amount behind. In order to correct the lag behind trend components, several variations of the basic exponential smoothing model have been developed. One of these is the trend adjusted model which adds to the forecasted value an adjustment term representing the estimate of the trend. The model is as follows:

$$f_{n+1} = y_n + \frac{1-\alpha}{\alpha} \cdot b_n$$

and $$y_n = \alpha X_n + (1-\alpha)y_{n-1}$$

$$b_n = \alpha(y_n - y_{n-1}) + (1-\alpha)b_{n-1}$$

where y_i the smoothed average for the i^{th} period

b_i is the smoothed trend estimate for the i^{th} period.

This model reacts to a trend by detecting the change in the smoothed average and using the resulting estimate of the trend to adjust the forecasted value.

Another model which is also trend adjusting is the second order exponential smoothing model. This model has the following recursive description:

$$f_{n+1} = \alpha X_n + 2(1-\alpha)f_n - (1-\alpha)f_{n-1}$$

Hence, the model uses the most recent present value, together with the forecasts for the latest two periods and uses a smoothed value of the error and direction of change in these forecasts to calculate the next forecasted value. The response to changes in time series data is fairly rapid (dependent upon the value of α), but the forecasts will overshoot any continuing perturbance before stabilizing at the correct value.

2.2 Advantages and Disadvantages of the Models

Several problems exist with the moving average approach. As mentioned previously, the moving average model lacks proper response to trended data. Depending upon the choice of K, (the number of periods used in the average) the model may be either overly sensitive to single period perturbances, or overly slow in reacting to persisting changes in the data. Another drawback of the moving average approach is that all of the past K periods must be stored so that as each new point becomes available, the K+1 point in the past can be deleted from the average.

Most of the problems associated with moving averages are to some degree remedied by exponential smoothing techniques. First order exponential smoothing most closely resembles the moving average approach in effect, but has the advantage that only one value need be stored for use in the next forecast. This permits economy of storage.

The trend adjusted model has the further advantage of correctly responding to the presence of a trend, but at the expense of storing three values instead of one, and requires more calculating to achieve the forecast. The trend adjusted model has the property of overshooting a change in mean before finally settling down to the proper value. In turn, however, it is also more responsive than the first order model to change.

The second order model requires the storing of the two previous forecasted values. The calculation is not very much more extensive than the first order model. The trend tracking is satisfactory, but the model tends to overshoot continuing perturbances before finally stabilizing. The speed of response is good.

Exponential smoothing models hence satisfactorily replace moving average models since the response of a K period average may be emulated by an exponential smoothing model with proper α . None of these models, however, will adjust correctly to seasonal fluctuations.

2.3 Adaptive Models

A technique for improving the response rate of an exponential smoothing model is to adjust the smoothing constant as the properties of the time series are detected to change. These types of models are generally termed adaptive models. A number of techniques are available for determining the nature and duration of the adjustment in α . Three such models were used for the comparison with spectral analysis. These are:

1. Trigg-Leach model (Reference 41)
2. Ratio model (Reference 39)
3. Panic model (Reference 39)

2.3.1 Trigg-Leach Model Description

The Trigg-Leach model uses a tracking signal to monitor the performance of the model. The appropriate smoothing value α is determined from the tracking signal. For each forecast a new α is calculated.

The tracking signal is calculated as follows:

$$ts_n = \frac{se_n}{sae_n}$$

and $$se_n = \gamma(X_n - f_n) + (1-\gamma)se_{n-1}$$

$$sae_n = \gamma\left|(X_n - f_n)\right| + (1-\gamma)sae_{n-1}$$

where ts_n is the tracking signal for period n,

se_i is the smoothed error for period i,

sae_i is the smoothed absolute error for period i,

γ is the smoothing constant for the error.

For the experiments, γ was assigned the value 0.1.

The smoothing value for period n, α_n, is taken as $|ts_n|$. It is clear that so long as the system is tracking properly, the xmoothed absolute error will remain near zero, and the tracking signal will fluctuate about zero. Should the error increase in a direction, the ratio of the smoothed absolute error and smoothed error will move either toward +1 or -1, and the resulting α_n will cause the most recent history values to be weighed more heavily in the forecast. As the forecasts again decrease in error, the corresponding smoothing values will again drop toward zero. The smoothing done to the error measurements prevents over-reaction to random fluctuations in the time series.

2.3.2 Ratio Model Description

The ratio model uses two measures of the absolute error of the forecast, a "fast" measure which responds to changes quickly, and a "slow" measure which is much slower in reacting to such changes. Each measure is calculated via the following first order smoothing formula:

$$a_{n,j} = \alpha_j \left|(X_n - f_n)\right| + (1 - \alpha_j)a_{n-1,j}$$

where $a_{i,j}$ is the j^{th} measure for period i,

j is fast or slow,

α_j is the smoothing constant for the j^{th} measure.

The values used for α_{fast} and α_{slow} were 0.25 and 0.05 respectively. These were the values used by Salmon and were derived from simulation trials.

The value of the smoothing value α_n for the exponential smoothing model is determined from $a_{n,fast}$ and $a_{n,slow}$ as follows:

For $a_{n,fast} > a_{n,slow}$

$$\alpha_n = \alpha_{n-1} + \left(1 - \frac{a_{n,slow}}{a_{n,fast}}\right)(\alpha_{max} - \alpha_{n-1})$$

and for $a_{n,fast} < a_{n,slow}$

$$\alpha_n = \alpha_{n-1} + \left(1 - \frac{a_{n,fast}}{a_{n,slow}}\right)(\alpha_{n-1} - \alpha_{min})$$

where α_i is the smoothing constant for the i^{th} period,

α_{max} is the upper limit which α_i may attain,

and α_{min} is the lower limit which α_i may attain.

For the experiment α_{max} and α_{min} were assigned the values 0.5 and 0.025 respectively. These values again were selected so as to agree with values used by Salmon.

2.3.3 Panic Model Description

The panic model uses consecutive period measures of the absolute smoothed error to detect when the system is no longer in control. The chosen measure is an estimated error of more than three standard deviations from forecasted value. When such an error is detected, the smoothing value α is adjusted to a larger value. The α is subsequently permitted to return to its original (low) level as the system stays within the controlled limits. The method as described by Salmon has the following relations:

$$a_n = \alpha_{n-1} \left| (X_n - f_n) \right| + (1 - \alpha_{n-1}) a_{n-1}$$

where a_i is the absolute smoothed error for period i,

α_i is the smoothing value for period i.

$r_n = (2.75 \ \alpha_{n-1} + 1) a_{n-1} / a_n$

where r_n is the measure of error.

A value of $r_n = 1$ represents an error of three standard deviations (3σ), $r_n < 1$ is an error greater than 3σ, $r_n > 1$ is an error less than 3σ.

For $r_n < 1$,

$$\alpha_n = \alpha_{n-1} + (1 - r_n)(\alpha_{max} - \alpha_{n-1})$$

For $r_n > 1$,

$$\alpha_n = \alpha_{n-1} + (1 - \frac{1}{r_n})(\alpha_{n-1} - \alpha_{min})$$

where α_{max} and α_{min} are as described in section 2.3.2.

To show ~~that~~ r = 1 corresponds to an error of 3σ, the following development suffices:

The ratio of the smoothed absolute error to the standard deviation is about 0.8 for a number of distributions. Hence, $a_{n-1} = 0.8$ is a reasonable estimate to use for a stable process. Assume that an error of 3σ occurs in a forecast. Hence

$$a_n = \alpha_{n-1} \cdot 3\sigma + (1-\alpha_{n-1}) \cdot 0.8\sigma = 0.8\sigma + 2.2\sigma \cdot \alpha_{n-1}.$$

$$r = \frac{(2.75\,\alpha_{n-1}+1) \cdot 0.8\sigma}{2.2\sigma \cdot \alpha_{n-1} + 0.8\sigma} = \frac{2.2\sigma\,\alpha_{n-1} + 0.8\sigma}{2.2\sigma\,\alpha_{n-1} + 0.8\sigma} = 1.$$

The above formulation adjusts the smoothing value to correct for increase in error. When operating at less than 3σ error, the smoothing value is stabilized at α_{min}. It then takes an error in excess of 3σ to trigger a modification of the smoothing value.

One problem was found to be present in the above formulation. When an error of 3σ occurs, the value a_{n-1} for use in the next period has been increased as a result of smoothing the absolute error. The use of the 2.75 value in calculating r_n requires an error larger than 3σ for triggering a response. If the first error was 3σ, the second error E' must be 3.875σ as shown below ($\alpha_i = 0.1$):

$$a_n = 0.1 \cdot 3\sigma + (1-0.1) \cdot 0.8\sigma = 1.02\sigma$$

$$a_{n+1} = 0.1 \cdot E' + 0.9 \cdot 1.02\sigma = 0.1E' + .918\sigma$$

$$r_{n+1} = \frac{(2.75 \cdot 0.1+1)1.02\sigma}{0.1E' + .918\sigma} = \frac{1.275 \cdot 1.02\sigma}{0.1E' + .918\sigma}$$

For a response, $r_{n+1} \leq 1$. Hence the minimum triggering value is E' such that $r_{n+1} = 1$. Hence:

$$r_{n+1} = \frac{1.275 \cdot 1.02\sigma}{0.1E' + .918\sigma} = 1, \text{ or}$$

$$E' = \frac{1.275 \cdot 1.02\,\sigma - .918\,\sigma}{.1} = 3.825\sigma.$$

Hence, the test for errors is desensitized if the 2.75 test value is used. The procedure was modified to adjust this test value when the smoothing value was adjusted. The adjustment used was determined as follows:

An error E' equal to or in excess of $3\,\sigma$ causes

$$a_n = 0.1E' + 0.9 \cdot 0.8\,\sigma = .72\,\sigma + 0.1E'$$

Let 2.75 be the test value K. We now wish to change K so that an error of $3\,\sigma$ causes r_{n+1} to equal 1. But

$$a_{n+1} = 0.1 \cdot 3\,\sigma + 0.9\,(.72\,\sigma + 0.1E') =$$

$$.3\,\sigma + .648\,\sigma + .09E' = .948\,\sigma + .09E'$$

Thus $r_{n+1} = \dfrac{(K \cdot 0.1 + 1)(.72\,\sigma + 0.1E')}{(.948\,\sigma + .09E')} = 1$, or

$$K = \frac{2.28\,\sigma - 0.1E'}{.72\,\sigma + 0.1E'} = -1 + \frac{3\,\sigma}{.72\,\sigma + 0.1E'} = -1 + \frac{3}{.72 + \frac{0.1E'}{\sigma}}$$

But it was assumed that $a_{n-1} = 0.8\,\sigma$, or $\sigma = \dfrac{a_{n-1}}{0.8}$

Hence $K = -1 + \dfrac{3}{.72 + \frac{0.1E' \cdot 0.8}{a_{n-1}}} = -1 + \dfrac{3}{.72 + \frac{.08E'}{a_{n-1}}}$

Since K is initially at 2.75, thus $K = K_{init} - 3.75 + \dfrac{3}{.72 + \frac{.08E'}{a_{n-1}}}$

In order to restore K after adjustment, the following was used ($r_{n+1} > 1$)

$$K_{new} = K_{prev} + \left(1 - \frac{1}{r_{n+1}}\right)(2.75 - K_{prev})$$

This revised formulation was used during the tests.

2.4 Experimental Models and Initialization Values

The three adaptive techniques discussed in Section 2.3 were each tested in the experiment. The resulting smoothing values were used in first order, first order with trend adjustment and in second order exponential smoothing models.

The models are somewhat sensitive to starting values for limited history forecasting. In order to start each model with similar values, it was decided to have the forecast for period 2 equal to the period 1 value, and the smoothing value equal to 0.1 for this initial forecast. The forecast initial value was the first period value. The proper values to assign to the other parameters for each model were determined as follows:

2.4.1 Initialization by Method

Trigg-Leach method:

$$se_1 = 0.1(X_1-f_1) + 0.9se_o = 0.1(X_1-X_1) + 0.9se_o = 0.9se_o,$$

$$sae = 0.1\left|(X_1-f_1)\right| + 0.9sae_o = 0.1\left|(X_1-X_1)\right| + 0.9sae_o$$

$$= 0.9sae_o,$$

$$ts_1 = \frac{0.9se_o}{0.9sae_o} = \frac{se_o}{sae_o}$$

$$a_1 = \left|\frac{se_o}{sae_o}\right|$$

For $a_1 = 0.1$, se_o was assigned the initial value 0.1 and sae_o was given the value 1.

Ratio method:

$$a_{1,fast} = 0.25 \left|(X_1 - f_1)\right| + (0.75)a_{o,fast} =$$

$$0.25 \left|(X_1 - X_1)\right| + (0.75)a_{o,fast} = 0.75a_{o,fast}$$

Similarly $a_{1,slow} = 0.95\ a_{o,slow}$

Since when $a_{1,fast} = a_{1,slow}$, α_o is not changed, the values chosen were $a_{o,fast} = 0.95$ and $a_{o,slow} = 0.75$, $\alpha_o = 0.1$.

Panic method:

$$a_1 = \alpha_o \left|(X_1 - f_1)\right| + (1-\alpha_o)a_o$$

$$= \alpha_o \left|(X_1 - X_1)\right| + (1-\alpha_o)a_o = (1-\alpha_o)a_o.$$

$$r_1 = \frac{(2.75\,\alpha_o + 1)a_o}{(1-\alpha_o)a_o} = \frac{2.75\,\alpha_o + 1}{1-\alpha_o} \geq 1.$$

$$0.1 = \alpha_o - (1 - \frac{1-\alpha_o}{2.75\alpha_o+1})(\alpha_o - .025)$$

$$.275\,\alpha_o + 0.1 = 2.75\,\alpha_o^2 + \alpha_o -$$

$$(2.75\,\alpha_o + 1 - 1 - \alpha_o)(\alpha_o - .025) =$$

$$2.75\,\alpha_o^2 + \alpha_o - 1.75\,\alpha_o^2 + .06875\,\alpha_o$$

$$\alpha_o^2 + .79375\,\alpha_o - 0.1 = 0$$

$$\alpha_o = .11$$

a_o was arbitrarily assigned the value 1.

2.4.2 Initialization by Model

First Order model:

$$f_2 = \alpha_1 X_1 + (1-\alpha_1)f_1 = \alpha_1 X_1 + (1-\alpha_1)X_1 = X_1$$

No additional values to be initialized.

First Order with Trend Adjustment model:

$$f_2 = \bar{X}_1 + \frac{1-\alpha_1}{\alpha_1} \cdot b_1$$

where $\bar{X}_1 = \alpha_1 X_1 + (1 - \alpha_1)\bar{X}_o$

and $b_1 = \alpha_1(\bar{X}_1 - \bar{X}_o) + (1 - \alpha_1)b_o$

$= \alpha_1(\alpha_1 X_1 + \bar{X}_o - \alpha_1\bar{X}_o - \bar{X}_o) + (1 - \alpha_1)b_o$

$= \alpha_1(\alpha_1 X_1 - \alpha_1\bar{X}_o) + (1 - \alpha_1)b_o .$

Hence $f_2 = \alpha_1 X_1 + (1 - \alpha_1)\bar{X}_o + \dfrac{(1 - \alpha_1)}{\alpha_1}(\alpha_1^2 X_1 - \alpha_1^2\bar{X}_o + b_o - b_o\alpha_1)$

$= \alpha_1 X_1 + \bar{X}_o - \alpha_1\bar{X}_o + \alpha_1 X_1 - \alpha_1\bar{X}_o + \dfrac{b_o}{\alpha_1} - b_o - \alpha_1^2 X_1$

$+ \alpha_1^2\bar{X}_o - b_o + b_1 \ \alpha_1 = 2\alpha_1 X_1 - 2\alpha_1\bar{X}_o + \bar{X}_o$

$+ \dfrac{b_o}{\alpha_1} - b_o - \alpha_1^2 X_1 + \alpha_1^2\bar{X}_o - b_o + b_o \ \alpha_1$

For b_o assigned the value 0, $\bar{X}_o$ assigned value X_1,

$f_2 = 2\alpha_1 X_1 - 2\alpha_1 X_1 + X_1 - \alpha_1^2 X_1 + \alpha_1^2 X_1 = X_1 .$

Second Order model:

$f_2 = \alpha_1 X_1 + 2(1 - \alpha_1)f_1 - (1 - \alpha_1)f_o$

For $f_2 = X_1$ and $f_1 = X_1$, thus

$f_1 = \alpha_1 X_1 + 2(1 - \alpha_1)X_1 - (1 - \alpha_1)f_o$

Hence $f_o = \dfrac{\alpha_1 X_1 + 2X_1 - 2\alpha_1 X_1 - X_1}{1 - \alpha_1} = \dfrac{X_1 - \alpha_1 X_1}{1 - \alpha_1} = X_1 .$

CHAPTER 3

SPECTRAL ANALYSIS

The forecasting procedures previously described disregard the presence of any autocorrelation in the time series data in determining the forecast. The technique of spectral analysis attempts to utilize the information present in the estimate of the autocovariance function in determining a forecast. This is done by both determining any periodic components indicated in the time series data, and by determining "optimal" weights to assign to the historical data.

In order for the theory of spectral analysis to be applicable, certain assumptions about the nature of the time series are necessary. Chief among these assumptions is that of weak stationarity which requires that the following hold: $E(X_n \cdot X_{n+}) = f(\tau)$, i.e. the expected value of the product of two points separated by an interval τ be a function of τ only. This assumption is very stringent in its requirements upon the time series data. One of the goals of the experiment was to determine the accuracy of spectral analysis in case of violation of this assumption.

3.1 Continuous History

The theory of spectral analysis derives from consideration of a continuous function f(t). The autocovariance function, given infinite history, may be calculated. The resulting autocovariance function $c(\tau)$ is a function of the lag τ alone, as a result of the assumption of stationarity. This function gives sufficient information to derive an "optimal" weighting function for weighting the historical function f(t) for forecasting purposes.

The spectral function derived from the autocovariance is a transformation from the time domain into the frequency domain. The following relation defines this transformation:

$$p(f) = \int_{-\infty}^{\infty} c(\tau) \cos (2\pi f\tau)d\tau$$

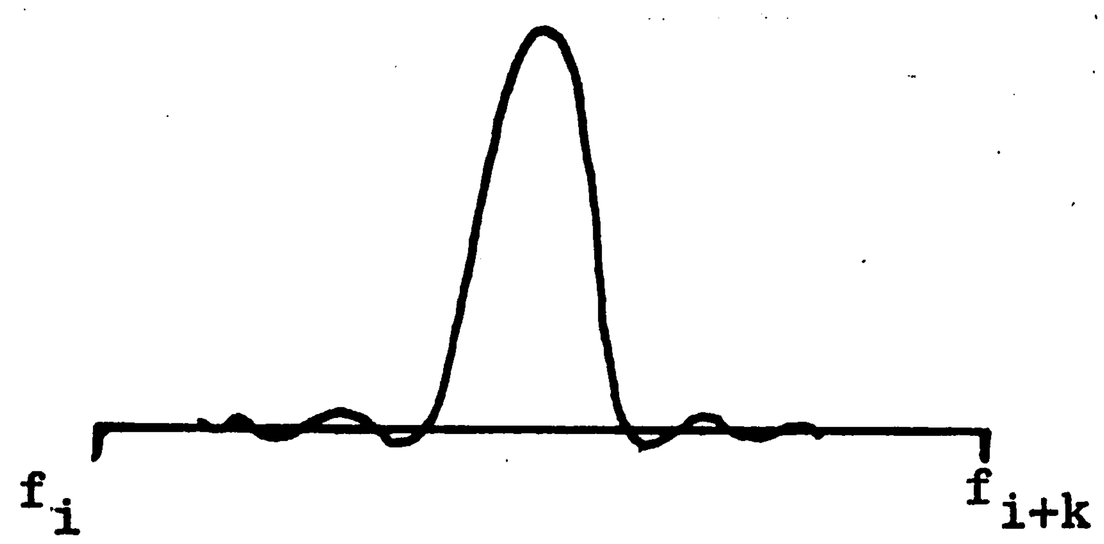
where p(f) is the spectral transform of the autocovariance function $c(\tau)$.

An inverse transform also exists, i.e.:

$$c(\tau) = \int_{-\infty}^{\infty} p(f) \cos (2\pi f\tau)df.$$

The spectral function which results from the first transform permits analysis of the contributions of components at various frequencies to the variance of the source function f(t). The presence of periodic components may be detected from such an analysis.

In order to detect frequency components present, but slightly diffused or obscured, a smoothing function is usually applied to the spectral function before analysis (akin to a power filter). The resulting smoothed function has the property that high frequency components are reduced in effect, permitting better determination of less prominent features. The smoothing functions, usually called spectral windows, apply weighting functions to intervals in the range of the function. These weighting functions have a peak in the center of each interval and rapidly decrease on either side of the center. Unfortunately, there will be present side loops which permit a small amount of leakage from other frequencies. The following figure is an example of how such a smoothing function would appear:



Spectral window centered on (f_i, f_{i+k})

A number of spectral windows have been discussed in the literature. In some cases the choice may be made so as to minimize the leakage from other frequencies, especially when large contributions at certain frequencies are known to be present. The smoothed function resulting from the application of a spectral window can be analyzed for contributions at various frequencies. A reverse transformation into the time domain yields a smoothed autocovariance function which can be used in determining forecasts.

3.2 Discrete Case

Given discrete data, estimates can be made of the autocovariance function. For a time series $X_1, X_2, \ldots, X_n$, an estimate of the autocovariance function is obtained from the lag products C_r defined:

$$C_r = \frac{1}{n-r} \sum_{q=1}^{n-r} X_q X_{q+r} \text{ for } r = 0, 1, \ldots, m$$

where C_r is the estimate of the autocovariance function for lag r (C_0 is an estimate of the variance),

and m is the maximum lag for which the estimate is made $(0 < m < n)$.

For forecasting purposes, the estimated autocovariance function could be used in calculating the weights to assign to previous data points. It is advantageous, however, to smooth the autocovariance function so as to be able to determine less prominent contributing features by reduction of the effects of the largest components. Smoothing of the raw autocovariance function directly also introduces problems in statistically placing confidence intervals on the accuracy of the resulting function. By performing the spectral transformation prior to smoothing, the problems concerning confidence intervals are eliminated since well established statistical procedures may be used. This fact has been one of the more important motivations toward using spectral analysis.

For the discrete case, a discrete spectral transform is required. the transformation is defined by:

$$S_r = \tfrac{1}{2}\left(C_o + 2\sum_{j=1}^{m-1} C_j \cos\frac{\pi j r}{m} + C_m \cos r\pi\right) \text{ for } r = 0, 1, \ldots, m$$

where S_r is the r^{th} discrete spectral value. The inverse relation is:

$$C_r = \frac{1}{m}\left(S_o + 2\sum_{j=1}^{m-1} S_j \cos\frac{\pi j r}{m} + S_m \cos r\pi\right) \text{ for } r = 0, 1, \ldots, m$$

A number of discrete smoothing functions may be applied to the spectral function. The one used in the experiments was the Hamming function defined by:

$$U_o = .5\, s_o + .5\, s_1$$

$$U_r = .25\, s_{r-1} + .5\, s_r + .25\, s_{r+1} \qquad 1 \leq r \leq m - 1$$

$$U_m = .5\, s_{m-1} + .5\, s_m$$

where U_i is the i^{th} term of the smoothed spectral function.

From the smoothed spectral function it is possible to check for the presence of various periodic components in the time series data. This feasure of spectral analysis has been used in the analysis of meteorological, oceanographic and economic time series data.

In order to formulate a forecast, the smoothed spectral function can be converted back into a smoothed autocovariance function. The resulting function may then be used to calculate the weights which to assign to the previous history values in forecasting.

The following derivation shows the way in which the m values of the autocovariance function can be used to determine the weights for the previous m-1 periods:

Given: Time series $X_1, X_2, \ldots, X_n$ with mean zero, and autocovariance function $C_o, C_1, C_2, \ldots, C_m$, and assuming weak stationarity (i.e. $E(X_i X_{i+\tau}) = C_\tau$).

To Find: $a_o, a_1, \ldots, a_{m-1}$ such that $f_{n+1} = a_o X_n + a_1 X_{n-1} + \ldots + a_{m-1} X_{n-m+1}$, the forecast for period n+1, has minimum expected value for squared error, i.e. $E\left[(f_{n+1} - X_{n+1})^2\right]$ is a minimum.

$$S.E. = E\left[(f_{n+1} - X_{n+1})^2\right] = E\left[(a_o X_n + a_1 X_{n-1} + \ldots + a_{m-1} X_{n-m+1} - X_{n+1})^2\right]$$

$$= E\left[\left(\sum_{i=o}^{m-1} \sum_{j=o}^{m-1} a_i a_j X_{n-i} X_{n-j} - 2 \sum_{i=o}^{m-1} a_i X_{n-i} X_{n+1} + X_{n+1}^2\right)\right]$$

$$= \sum_{i=o}^{m-1} \sum_{j=o}^{m-1} a_i a_j \, E(X_{n-i} X_{n-j}) - 2 \sum_{i=o}^{m-1} a_i \, E(X_{n-i} X_{n+1}) + E(X_{n+1}^2)$$

$$= \sum_{i=o}^{m-1} \sum_{j=o}^{m-1} a_i a_j C_{|i-j|} - 2 \sum_{i=o}^{m-1} a_i C_{i+1} + C_o$$

(as a result of the assumption of stationarity).

Taking the partial derivatives with respect to $a_o, a_1, \ldots, a_{m-1}$, and setting the resulting expressions equal to zero yields:

$$\frac{\delta(S.E.)}{\delta a_o} = 2 \sum_{i=0}^{m-1} a_i C_i - 2 C_1 = 0$$

$$\frac{\delta(S.E.)}{\delta a_1} = 2 \sum_{i=0}^{m-1} a_i C_{|i-1|} - 2 C_2 = 0$$

$$\vdots$$

$$\frac{\delta(S.E.)}{\delta a_j} = 2 \sum_{i=0}^{m-1} a_i C_{|i-j|} - 2 C_{j+1} = 0$$

$$\vdots$$

$$\frac{\delta(S.E.)}{\delta a_{m-1}} = 2 \sum_{i=0}^{m-1} a_i C_{|i-m+1|} - 2 C_m = 0$$

Transposing the C_i's, dividing by two and writing the resulting system of equations in matrix form yields:

$$\begin{bmatrix} C_o & C_1 & C_2 & C_3 & \cdot & \cdot & \cdot & C_{m-1} \\ C_1 & C_o & C_1 & C_2 & \cdot & \cdot & \cdot & C_{m-2} \\ C_2 & C_1 & C_o & C_1 & \cdot & \cdot & \cdot & C_{m-3} \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ C_{m-1} & C_{m-2} & C_{m-3} & C_{m-4} & \cdot & \cdot & \cdot & C_o \end{bmatrix} \begin{bmatrix} a_o \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{m-1} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \cdot \\ \cdot \\ \cdot \\ C_m \end{bmatrix}$$

Solving for the a_i's yields:

$$\begin{bmatrix} a_o \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{m-1} \end{bmatrix} = \begin{bmatrix} C_o & C_1 & C_2 & C_3 & \cdot & \cdot & \cdot & C_{m-1} \\ C_1 & C_o & C_1 & C_2 & \cdot & \cdot & \cdot & C_{m-2} \\ C_2 & C_1 & C_o & C_1 & \cdot & \cdot & \cdot & C_{m-3} \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ \cdot & \cdot & \cdot & \cdot & & & & \cdot \\ C_{m-1} & C_{m-2} & C_{m-3} & C_{m-4} & & & & C_o \end{bmatrix}^{-1} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ \cdot \\ \cdot \\ \cdot \\ C_m \end{bmatrix}$$

The resulting a_i's are the weights to apply to the historical values X_{n-i} for $i = 0, 1, 2, \ldots, m-1$.

3.3 Problems in the Use of Spectral Analysis

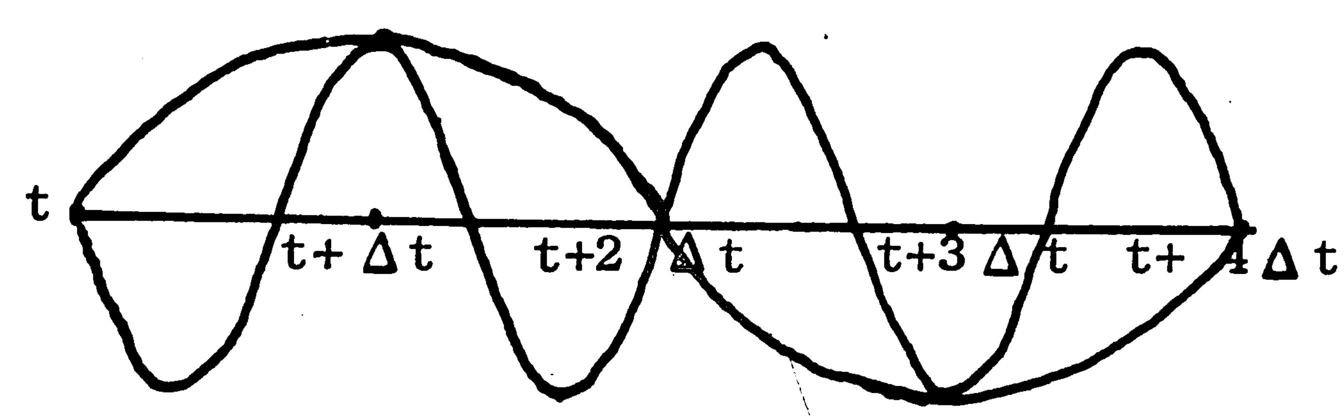
Several problems are encountered in the practical application of spectral analysis to time series. One major problem is in determining whether the assumption of stationarity is met. Ordinarily, the time series will not be stationary. The applicability of the technique appears then to be dependent upon the extent of the deviation of the time series from stationarity, although few practical results in the area are available. Granger[16] does discuss the effects of certain types of deviation upon the results. Associated with the problem of non-stationarity is the desirability of trend removal and adjustment of the mean to zero. The performing of these adjustments on the source data may yield a time series sufficiently near to stationary for the techniques of section 3.2 to be applied. One of the goals of the experiments was to determine how severely the presence of highly non-stationary data affects the resulting forecast when using spectral analysis.

Intrinsically bound up in the theory of spectral analysis is the requirement for large amounts of time series data if the resulting analysis is to be meaningful. In order for the results to be stable without too high a variance, it is suggested that m, the number of lags, not exceed 10% of the total historical data used. In addition, it is desirable that the number of history points used in forming the forecast be sufficient to cover the full cycle of cyclic components.

Hence, if a component of period 12 is present in the time series, at least 13 lag products should be calculated for estimates of the autocovariance function, so that 12 historical values will be used in the forecast. In order that no more than 10% of the total history be used in the final forecast, it is required that at least 130 time series values be available. Another goal of the experiments was to evaluate the effects of limited amounts of data and use of more than 10% of the data on the resulting forecasts.

Large amplitude periodic components can obscure other features of the time series during analysis. One procedure that helps reduce this problem is smoothing, which was discussed in section 3.2. A second technique, which can be applied to the source data, also helps in reducing this problem. The technique is prewhitening, which consists of crudely estimating the spectral function, determining strongly contributing components, and filtering these components from the source data. The resulting filtered data can be processed in the usual way, and otherwise secondary components will be more readily detected. For use in forecasting, however, the use of prewhitening has the drawback of requiring either additional adjustment of the final smoothed autocovariance function, or addition of adjustment terms, in forming the forecast.

A problem, heretofore unmentioned is that of aliasing. When performing the spectral transform, the autocovariance function is mapped into a set of discrete frequency intervals. It is not possible, however, to determine frequencies greater than $\frac{1}{2}$ cycle per interval of measurement. Frequencies greater than this value (which is called the Nyquist frequency) are mapped into lower frequencies. When an attempt is made to analyze the frequency properties of the time series the high frequency components are indistinguishably confounded with the lower frequencies. In the following figure an example is given of the impossibility of distinguishing the two sinusoidal components with a sampling interval of Δt:



Confounding of two components given a sampling interval Δt.

In order to prevent the loss of significant components in analysis of time series data, it is hence necessary that the sampling interval be of sufficiently short length to permit detection of the component. If this is not done, any high frequency components will result in enlarged values being determined for corresponding lower frequency elements.

A final significant problem in the use of spectral analysis is the considerable number of calculations required to form a forecast. The calculation of lag products, performing of two finite Fourier transforms, smoothing, matrix formation, inversion and multiplication, and final forecast calculation require a vast number of calculations and considerable time. It is hence necessary to assess whether any additional accuracy possible in the final forecast is worth the additional cost associated with the extensive computational requirements of spectral analysis.

CHAPTER 4

TIME SERIES GENERATOR

In order to compare the forecasting techniques under examination, a time series generator was designed which permitted selection and control of several parameters while still permitting the basic variables of the time series to be random variables. The basic model for the time series values used an additive model consisting of a term for mean with trend, a term for periodic components, and a term for error. Hence, the time series value TS has the following form:

$$TS_n = M_n + P_n + E_n$$

where TS is the time series value,

M is the trended value for the mean,

P is the total periodic component,

E is the error,

and all values are calculated for period n.

Each of the above values will be discussed in detail subsequently.

In order to select the form of the time series generator several works were considered. Gross and Ray[18] mention briefly a set of parameters which control their time series generator. Fiddleman[10] used a rather simple model in the generation of his time series test values. Hester[21] used an additive model similar to Fiddleman's, with the additional property of discrete jumps in mean. The construction of the generator for this experiment was aimed at permitting control of all parameters considered of interest for examination, while permitting emulation of most features present in the models discussed by the above authors. An additional aim was to permit generation of time series which were highly non-stationary. The model which was finally designed seemed to be a reasonable compromise of any conflicts in the several desired goals.

The subroutines RANDU and GAUSS (IBM-for Models 1130 and 360) were used to generate the uniform and normal random deviates respectively required in the course of the experiments. A modification of GAUSS was used in generating autocorrelated values for evaluation of autocorrelation on time series analysis.

4.1 Mean, Autocorrelation and Trend

The mean generated by the model consisted of several components. The initial value for the mean was fed to the program, together with an autocorrelation number and standard deviation. Subsequent values for the mean were derived from the current value by selection (using GAUSS) from a normal distribution with the current value as mean and with the given standard deviation. The autocorrelation number was used to select a degree of autocorrelation between the current value of the mean and the new value. The autocorrelation was achieved by reuse of values used by GAUSS is calculating the normal deviate. The value for untrended mean was thus calculated:

MO = N(M1, SD, A)

where SD is the standard deviation,

M1 is the current value for the mean,

A is the autocorrelation number,

N(a,b,c) is a normal distribution with parameters a, b, c

The calculation for the normal deviate was made by the following technique:

$$S = \sum_{i=1}^{12} R_i$$

$$MO = (S - 6) \cdot SD + M1$$

where S = sum of twelve random numbers R_i (R_i uniform on the interval (0,1).)

The reuse of one or more R_i causes a degree of autocorrelation between successive values. The degree of this autocorrelation was controlled by A, which specified the number of R_i's to be reused in the calculation of successive means.

The trend components were linear and quadratic. The coefficient for the linear element was selected from a uniform distribution specified by parameters. The quadratic coefficient was obtained from a normal distribution with mean zero and parameter specified standard deviation. The durations of these trend values were also treated as random variables, selected from a uniform distribution whose limits were specified by input parameter. When the length of time specified for trend components to be in effect was exceeded, the current value for mean and the trend coefficients were changed. The choice for new value of the mean was selected from a normal distribution with current mean as the distribution mean, and parameter specified standard deviation. Hence, the mean plus trend was calculated by the following:

$$D = U(D1,\ D2)$$

where D is the duration of the trend components,

U is the uniform distribution on interval (D1, D2).

$$M = MO + t \cdot L + t^2 Q$$

where M is the value for the mean,

MO is the new mean without trend,

t is the number of periods for which the current trend values have been in effect ($0 \leq t \leq D$),

L is the linear trend coefficient,

Q is the quadratic trend coefficient.

When the trend components had reached the end of their period of effect ($t > D$), the following changes were made:

$$M_n = N(M_o,\ SM)$$

where M_n is the adjusted value of the current mean,

M_o is the former value of the current mean,

and SM is the jump-point standard deviation of the mean.

$$L = U(P1,\ P2)$$

where P1, P2 are the end points of the uniform interval U.

$$Q = N(0,\ SQ)$$

where SQ is the quadratic component standard deviation.

A new value for D was also selected from its uniform distribution.

The rationale for the choices made in the construction of the mean is the following: for some processes it is reasonable to suppose that on occasion, the factors controlling the process will suddenly alter, causing a change in the mean to some new level. For processes such as sales or usage, it is also reasonable to suppose that previous trend values will no longer be appropriate when a change in level occurs. Hence, the changes in mean and trend have been chosen to occur simultaneously. The choices of distributions from which to draw new values were primarily arbitrary, but to some extent selected for reasonableness and similarity to distributions used in the other data generators previously mentioned.

4.2 Periodic Perturbances

In order to permit some flexibility in the periodic components of the model, three separate periodic elements were permitted. Each had individual parameters available for specification of its properties. The periodic terms were each sinusoidal with specifiable period and amplitude. In order to permit greater flexibility of these terms two additional sets of parameters were used. One set controlled the period location of each component at time zero. The other set of parameters specified a standard deviation for each periodic component. These standard deviations were used in a normal distribution in order to vary the exact periodicity of the components. The reason for the inclusion of such an element was to permit some testing of the effectiveness of smoothing the spectral function in detecting diffused periodic components. The periodic components were calculated as follows:

$$P = A1 \cdot \sin(2\pi P11) + A2 \cdot \sin(2\pi \cdot P22) + A3 \cdot \sin(2\pi \cdot P33)$$

where P is the periodic component,

Ai is the amplitude of the i^{th} component, i = 1, 2, 3 and $P_{ii} = (M + C_i + V_i)/P_i$, $i = 1, 2, 3$

where M is the period number

$C_i = N(0, S_i)$, a perturbance normally distributed with mean zero and standard deviation S_i,

V_i is the period at time $M = 0$,

and P_i is the period of component i.

4.3 Error Term

The error term was chosen to correspond to the common assumptions about the nature of the error, namely that the errors are normally distributed with mean zero and have no autocorrelation. Hence, the error term had the following form:

$$E = N(0, SDE)$$

where E is the error term with standard deviation SDE.

4.4 Summary

The complete model, as described, permitted specification and control of the properties of the time series generated. In order to test the forecasting ability of the various models, the parameters were adjusted so as to increase the contribution to the time series of certain components. The forecasting models were then tested in forecasting the resulting series. The flexibility of the generator in emulating many different models permitted a number of characteristics of time series to be evaluated as to effect on forecasts.

CHAPTER 5

PARAMETERS, DATA, AND ANALYSIS PROCEDURE

In conducting the experiment, certain of the available parameters were chosen as those for which evaluation of the various forecasting techniques was of interest. These parameters were then assigned levels and experiments were performed. The resulting data were in the form of squared forecast error for the last twenty-four periods of the generated time series. These data were then analyzed by χ^2 tests for significance of each techique over combinations of treatments and for significance of the techniques when compared. From these analyses, the conclusion of Chapter 6 were drawn.

5.1 Choices of Parameters and Levels

Of the numerous parameters available in the time series model for adjustment, six parameters (or groups of parameters) were chosen for evaluation, each at two levels. These were:

1. Number of points generated in the time series. The low level was 75; the high was 150. In each case, the last 24 were forecasted.
2. Autocorrelation of the mean. For each new value of the mean, twelve uniform random numbers were used to calculate the new value from the previous value. (See section 4.1). The autocorrelation was determined by the number of values used in common from one period to the next. The low level

was chosen as 2; the high level was 10.

3. The error standard deviation. The errors were drawn from a normal distribution with mean 0. The low value for the standard deviation was .01, the high value .05.

4. The standard deviation of the mean. After a base value for the mean was selected, an amended base value was selected from a normal distribution with mean equal to the original base value, and assigned standard deviation. The low level of this standard deviation was 1, and the high level 5.

5. Trend components (See section 4.1). Two portions of the trend were varied together. The first was the length of the trend, which was uniformly distributed over an assigned range. The low level was set at 2 to 6 periods, and the high at 9 to 15 periods. In conjunction with these values was the choice of standard deviation of the quadratic component coefficient. Associated with the low level of the trend length was a standard deviation for the quadratic component of 1, and with the high level, a value of .05. The smaller value for the high level was necessary so that the larger number of periods for which a particular trend was in effect did not result in the quadratic component dominating the trend.

6. The periodic components. Three periodic components

were permitted in each time series generated. For the experiments, these components were assigned periods of twelve, six and four. The maximum amplitudes of these three components were assigned two levels, a low level of 5, 0, and 0 and a high level of 5,2, and 1.

All other variables were held at fixed levels. There were:

Starting random number = 127.

Beginning mean = 100.

Coefficient of linear component of the trend, uniform on 1-5.

Periodic component periods = 12, 6, 4.

Standard deviations of periodic components = 0 for each.

Points in periods of periodic components at t = 0 were 1, 0, 2.

A fuller description of the meaning of each of these parameters is found in Chapter 4.

5.2 Nature of Experimental Data

It was assumed that the forecast errors for a forecasting technique were normally distributed. Under this assumption, each squared forecast error would be χ^2 distributed with 1 degree of freedom. Twenty-four consecutive periods, and their associated squared forecast errors were used as the evaluation for each technique and combination of treatments. The sum of these twenty-four values for a technique in each experiment should then be χ^2

with 24 degrees of freedom.

5.3 Statistical Procedure Used to Analyze the Data

Two types of information were desired from the data. Information concerning the effectiveness of a technique in forecasting a time series with a feature or set of features present was desired. In addition, an overall comparison of the forecasting techniques over those values for which a technique was found to be effective was of extreme interest. In order to obtain these types of information, the analysis was carried out in two parts. The first portion of this analysis looked at each technique separately over the 64 (6 factors at 2 levels each) treatments on which the technique was tested. For each treatment (representing a particular fixed level for each of the parameter sets), the data was divided into two equal sets of 32 points. One of these sets represented a high level measure of the treatment, and the other a low level. An F-test was performed for these values by dividing the larger by the smaller. The degrees of freedom for the numerator and denominator were both equal to 32 x 24 = 768. A ratio in excess of 1.33 represented a significant difference at the 99% level. This denoted a failure of the technique in successfully handling the treatment. In this fashion, each of the techniques over all of its treatments could be so classified.

In order to then compare the performance of the techniques, another series of F-tests was performed. The experimental data was adjusted on the basis of the results of the first F-tests by removal of those points for each technique for which it did not satisfactorily perform. The remaining points for each technique were then summed. The resulting values were then χ^2 with C_i degrees of freedom where $C_i = 24\ n_i$, and n_i = number of points used in the sum for technique (i), i = 1-12. These values were then evaluated in pairs with F-tests, dividing the values by their respective degrees of freedom and then checking the ratios against the corresponding F-value for the number of degrees of freedom of numerator and denominator. The values here were tested at the 99% level, with those values in excess of 1.19 being considered significant. Significance indicated a real difference of the two techniques in forecasting ability over those factors for which each was satisfactory.

CHAPTER 6

EXPERIMENTAL RESULTS AND CONCLUSIONS

The data which resulted from the experiments consisted of values for each of the 12 forecasting techniques at each of the 64 treatment points (representing each of the 6 factor groups at 2 levels). Each of these data points was assumed χ^2 with 24 degrees of freedom. (See 5.2) The results of the two stages of analysis (See 5.3) follow in the next two sections.

6.1 Results of Technique Evaluation over Each Treatment

Each of the techniques tested was found to fail on at least two treatments. The results, by technique, follow:

1. Ratio with First Order Smoothing. Failed on two treatments:
   a. Trend alone.
   b. Interaction of Number of Points, Trend, and Mean Standard Deviation.
2. Ratio with Trend Adapted First Order Smoothing. Failed on three treatments:
   a. Number of Points alone.
   b. Trend alone.
   c. Mean Standard Deviation alone.
3. Ratio with Second Order Smoothing. Failed on two treatments:
   a. Mean Standard Deviation alone.
   b. Autocorrelation Level alone.

4. Panic with First Order Smoothing. Failed on eight treatments:

   a. Number of Points alone.

   b. Interaction of Trend and Number of Points.

   c. Trend alone.

   d. Mean Standard Deriation alone.

   e. Interaction of Autocorrelation Level and Trend.

   f. Interaction of Autocorrelation Level, Trend and Number of Points.

   g. Interaction of Autocorrelation Level and Number of Points.

   h. Autocorrelation Level alone.

5. Panic with Trend Adapted First Order Smoothing. Failed on eleven treatments:

   a. Number of Points alone.

   b. Interaction of Periodic Component and Number of Points.

   c. Interaction of Trend and Periodic Component.

   d. Interaction of Trend and Number of Points.

   e. Interaction of Mean Standard Deviation, Trend, and Number of Points.

   f. Mean Standard Deviation alone.

   g. Interaction of Autocorrelation Level and Mean Standard Deviation.

   h. Interaction of Autocorrelation Level, Mean Standard Deviation, Trend, and Number of Points.

i. Interaction of Autocorrelation Level and Trend.

j. Interaction of Autocorrelation Level, Trend, and Number of Points.

k. Autocorrelation Level alone.

6. Panic with Second Order Smoothing. Failed on eleven treatments:

a. Number of Points alone.

b. Interaction of Trend and Number of Points.

c. Trend alone.

d. Interaction of Mean Standard Deviation, Trend, and Number of Points.

e. Interaction of Mean Standard Deviation and Periodic Component.

f. Interaction of Autocorrelation Level, Mean Standard Deviation and Periodic Component.

g. Interaction of Autocorrelation Level, Mean Standard Deviation, Trend, and Periodic Component.

h. Interaction of Autocorrelation Level, Mean Standard Deviation, Trend, and Number of Points.

i. Interaction of Autocorrelation Level, Mean Standard Deviation, and Trend.

j. Interaction of Autocorrelation Level, Trend, and Number of Points.

k. Interaction of Autocorrelation Number and Number of Points.

7. Trigg-Leach with First Order Smoothing. Failed on four treatments:

   a. Number of Points alone.

   b. Trend alone.

   c. Mean Standard Deviation alone.

   d. Autocorrelation Level alone.

8. Trigg-Leach with Trend Adapted First Order Smoothing. Failed on <u>all</u> treatments.

9. Trigg-Leach with Second Order Smoothing. Failed on six treatments:

   a. Number of Points alone.

   b. Trend alone.

   c. Mean Standard Deviation alone.

   d. Interaction of Autocorrelation Level and Mean Standard Deviation.

   e. Interaction of Autocorrelation Level, Mean Standard Deviation, and Number of Points.

   f. Autocorrelation Level alone.

10. Spectral Analysis - 7% of History. Failed on seven treatments:

   a. Interaction of Trend and Number of Points.

   b. Trend alone.

   c. Interaction of Mean Standard Deviation and Number of Points.

   d. Mean Standard Deviation alone.

e. Interaction of Autocorrelation Level and Mean Standard Deviation.

f. Interaction of Autocorrelation Level and Trend.

g. Autocorrelation Level alone.

11. Spectral Analysis - 10% of History. Failed on six treatments:

a. Trend alone

b. Interaction of Mean Standard Deviation and Number of Points.

c. Mean Standard Deviation alone.

d. Interaction of Autocorrelation Level and Mean Standard Deviation.

e. Interaction of Autocorrelation Level and Trend.

f. Autocorrelation Level alone.

12. Spectral Analysis - 20% of History. Failed on six treatments.

a. Trend alone.

b. Interaction of Mean Standard Deviation and Number of Points.

c. Mean Standard Deviation alone.

d. Interaction of Autocorrelation Level and Mean Standard Deviation.

e. Interaction of Autocorrelation Level and Trend.

f. Autocorrelation Level alone.

These results indicate which techniques should be avoided for time series dominated by certain types of errors.

6.2 Results of Technique Comparisons with Points of Failure Removed

The results presented in 6.1 were used to delete those points for each technique for which the technique had failed. The remaining points were then compared as described in 5.3 with the following results:

Technique (Numbered as in 6.1)	Sum of Points	Degrees of Freedom
1	179,048	1488
2	84,820	1464
3	94,897	1488
4	3,742,480	1344
5	445,396	1272
6	712,126	1272
7	76,301	1440
8	--	--
9	77,348	1392
10	64,579	1368
11	59,374	1392
12	63,281	1392

The F-tests conducted on the above values yielded the following results.

Technique 11 better than 1,2,3,4,5,6,7,9.

Technique 12 better than 1,2,3,4,5,6,9.

Technique 7 better than 1,3,4,5,6,

Technique 9 better than 1,4,5,6.

Technique 2 better than 1,4,5,6.

Technique 3 better than 1,4,5,6.

Technique 1 better than 4,5,6.

Technique 5 better than 4,6.

Technique 6 better than 4.

Technique 8 was not included in the analysis as a result of having been found ineffectual for the factors under consideration.

6.3 Observations and Conclusions

Based upon the results discussed in 6.1 and 6.2, certain conclusions can be drawn. Spectral analysis, using 10% of the available history data in forming the forecast, performed with higher accuracy overall than any of the adaptive forecasting techniques tested. The three spectral analysis forecasting models all performed better than the panic and ratio adaptive techniques. In addition, the three spectral analysis techniques all failed to perform satisfactorily when certain components dominated the time series. These components were trend or standard deviation of the mean or autocorrelation of successive means, and some of the interactions of these components. The problems with trend and mean deviation could have been anticipated, since spectral analysis is not designed to deal with these types of variation. The lack of proper response to consecutive period autocorrelation is surprising in that the technique relies upon autocorrelation estimates to operate. The success of spectral analysis on the number of points factor was also of interest. It is normally suggested in the analysis of stationary time by spectral analysis that as many as 200 points of history are desirable for successful operation. Here it was found that as few as 51 points of history apparently could be used. (The first forecast for 75 points used only that amount). Using the ranges from 51-74 and 126-149 did not seem to result in significantly different forecast accuracies.

The complete failure of the Trigg-Leach with trend adapted first order smoothing model, taken with the good performance of the other two Trigg-Leach models is very interesting. It appears that some degree of interference must arise between the adaptive technique and hybrid smoothing model. The problems which may arise in combining two techniques performing different function into a single unit is indicated.

Another interesting result of the experiments was the failure of all but two of the adaptive models to handle satisfactorily the change in number of history points. There appears to develope, over the course of operation of these models, a build up of inaccuracy. This may be a result of the continuous modification of the smoothing value, which causes the exponential smoothing model no longer to be representable even remotely by a geometrically weighed series. Since the smoothing value is permitted to vary toward 1. at times, the historical data can be very rapidly discounted and thus lost. This property of adaptive models should be investigated if the model is to be used over a long period of time.

The similarity of response of variations of each technique (excluding the Trigg-Leach model with trend adapted first order smoothing) was evidenced by the grouping which occurred in the final comparisons, and the general similarities in points of failure. The spectral techniques, in general, dominated. The two satisfactory Trigg-Leach models outperformed nearly all of the other adaptive models. The panic models, despite the changes to improve sensitivity, still gave unsatisfactory performances, with fewer successful treatment handlings, and larger overall errors. The ratio model was much less successful when used with first order smoothing than with the other two smoothing models. For the panic and ratio models the best performance was found in using the first order trend adaptive model. But, as noted, the Trigg-Leach model completely failed when using the model.

In general, it would appear that spectral analysis, for accuracy, gives the best forecasting accuracy for series not dominated by short variable trends, wide variations in the values of the mean, and autocorrelation between the means of successible periods. This accuracy is obtained, however, at the expense of much greater computational and storage requirements and much more time than the adaptive exponential smoothing techniques. The best adaptive model tested was the Trigg-Leach with first order smoothing. It outperformed all but the Trigg-Leach with second order smoothing model and the ratio model with trend adapted first order smoothing, which were not found to be significantly poorer in the F-tests performed. The Trigg-Leach model is the most simple of the adaptive models used and requires the least computation. Care is indicated as necessary, however, in selecting a model for use. The complete failure of the one Trigg-Leach model indicates that lack of proper evaluation may result in a totally unsatisfactory forecasting choice.

With regards to the six factor groups evaluated, the following selection of models is indicated:

1. Autocorrelation Between Successive Means dominating: Ratio model with trend adapted first order model.
2. Error dominating: Values selected gave no indication as to best method.
3. Variation of Mean dominating: Ratio model with first order smoothing.
4. Trend dominating: Ratio model with second order smoothing.
5. Periodic Components dominating: Spectral analysis.
6. Number of Available History Points (assumed at least 50): Spectral analysis.

All interaction series save those dominated by trend, autocorrelation and mean variation may be forecasted by spectral analysis. Those series with those interactions dominating may be forecasted by the Trigg-Leach with first order smoothing model.

Chapter 7

SUGGESTIONS FOR FURTHER INVESTIGATION

Several areas of investigation, associated with the procedures involved in the paper, appear to be interesting and worthy of further investigation. One of these is an extension of the ranges of the parameters used in the experiments so far performed, with the purpose of further clarifying and characterizing the properties of the forecasting models. Incorporation of some of the parameters held constant during the experiments of this work would also expand the characterization of the forecasting techniques as to application to certain types of time series.

Another mode of analysis could well be applied to the experimental data resulting from this type of experiment. In the present work, the measure of error was chosen to be squared error. This choice was made because the formulation of spectral analysis for choosing optimal weights was based upon the assumption that the error measure would be squared error. Had error itself been the criterion, analysis of variance could have been used for data analysis, and some ranking technique such as Duncan's multiple range test used to finally evaluate the techniques. Had absolute error been chosen as the error measure, the data would have been half-normal, and other statistical techniques could have been used.

A problem encountered in applying spectral analysis existed with respect to trend removal. It would have been most desirable to detect and remove the short range trend contributions. Several approaches were investigated, but none was found to be effective. Among the approaches looked at (although not in depth) were first order and second order differences of the time series data, two period differences, and first order and second order smoothing values. None of these series of values would properly detect the short range trend components of the test series. Should some effective tool for trend detection be found, the ability to forecast accurately would be well advanced.

The panic forecasting models used in this work still did not perform well despite changes made to improve their sensitivity. More work on these models would perhaps yield an improved adaptive model of this type.

The whole area of adaptive models would appear to merit further study. Some of the failures of most of these models in the experiments performed were unexpected. Both study of these models for explanations of these failures and investigations of alterations of these models or formulations of new models with the goal of removing these faults could prove fruitful.

The use of spectral analysis for forecasting non-stationary time series gives rise to a number of possible investigations. The number of history points used in this work was found to cause insignificant differences. A further investigation of the range of points is warrented. Another such extension could involve the amount of history data to be used in formulating the final forecast.

The spectral analysis procedure used in this work utilized a standard smoothing technique (spectral window). A number of spectral windows have been suggested, most giving excellent results under certain time series conditions. An investigation of these windows over the classes of time series possible with the time series model used might help further clarify the relations between time series nature and "optimal" smoothing function.

An alternate approach to forecasting by the use of spectral analysis is possible. When the estimated autocorrelation function has been transformed into the spectral function, certain information concerning period components of the time series is available. The use of this information for formulation of a model of the time series is possible. The resulting model could then be used as a forecasting tool directly. Many problems are associated with the implementation of this approach, but successful construction of such models could yield a powerful new forecasting technique.

BIBLIOGRAPHY

1. Blackman, R. B. and J. W. Tukey, The Measurement of Power Spectra, Dover Publications, Inc., New York, 1958.

2. Bossons, John, "The Effect of Parameter Misspecification and Non-Stationarity on the Applicability of Adaptive Forecasts," Management Science, Vol. 12, No. 9, May, 1966, pp. 659-669.

3. Brown, Robert G., Statistical Forecasting for Inventory Control, McGraw-Hill, New York, 1959.

4. Brown, Robert G., Smoothing, Forecasting and Prediction of Discrete Time Series, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1962.

5. Cohen, Gerald D., "A Note on Exponential Smoothing and Auto-correlated Inputs," Operations Research, Vol. 11, No. 3, May-June, 1963, pp. 361-367.

6. Cooley, James W. and John W. Tukey, "An Algorithm for Machine Calculation of Complex Fourier Series," Mathematics of Computation, Vol. 19, 1965, pp. 297—301.

7. Dubman, M. R., The Spectral Characterization and Comparison of Nonstationary Processes, Rocketdyne, Canoga Park, Calif., 1965, U. S. Department of Commerce, AD 619 178.

8. Elmaghraby, Salah E., Research Leader, Western Electric Co., private communication.

9. Federer, Walter T., Experimental Design, The Macmillion Co., New York, 1955.

10. Fiddleman, Richard H., Research Engineer, Western Electric Co., private communication.

11. Fishman, George S. and Philip J. Kiviat, Spectral Analysis of Time Series Generated by Simulation Models, The Rand Corporation, Memorandum RM-4393-PR, Santa Monica, Calif., 1965.

12. Fleisher, R. E., Research Engineer, Western Electric Co., private communication.

BIBLIOGRAPHY (cont.)

13. Goodman, N. R., "Some Comments on Spectral Analysis of Time Series," Technometrics, Vol. 3, No. 2, May, 1961, pp. 221-228.

14. Goodman, N. R. and M. R. Dubman, Research on Spectral Analysis Summary of Results, Rocketdyne, Canoga Park, Calif., 1965, U.S. Department of Commerce, AD 619 265.

15. Goodman, N. R. and M. R. Dubman, Research on Spectral Analysis Final Report, Rocketdyne, Canoga Park, Calif., 1967, U. S. Department of Commerce, AD 649 631.

16. Granger, C. W. F. and M. Hatanaka, Spectral Analysis of Economic Time Series, Princeton University Press, Princeton, N. J., 1964.

17. Grenander, U. and M. Rosenblatt, Statistical Analysis of Stationary Time Series, John Wiley & Sons, Inc., New York, 1957.

18. Gross, Donald and Jack L. Ray, "A General Purpose Forecast Simulator," Management Science, Vol. 11, No. 6, April, 1965, pp. B119-B135.

19. Hannan, E. J., Time Series Analysis, John Wiley & Sons, Inc., New York, 1960.

20. Harris, Bernard (ed.), Spectral Analysis of Time Series, John Wiley & Sons, Inc., New York, 1967.

21. Hester, Stephen D., Prediction of Non-Stationary Time Series with the Adaptive Step Method, Master's Thesis in Industrial Engineering, Lehigh University, Bethlehem, Pa., 1967.

22. Hicks, Charles R., Fundamental Concepts in the Design of Experiments, Holt, Rinehart and Winston, New York, 1964.

23. Hinich, Melvin and John U. Farley, "Theory and Application of an Estimation Model for Time Series with Nonstationary Means," Management Science, Vol. 12, No. 9, May, 1966, pp. 648-658.

24. Jenkins, G. M., "General Consideration in the Analysis of Spectra," Technometrics, Vol. 3, No. 2, May, 1961, pp. 133-166.

25. Jenkins, G. M., "Comments on the Discussions of Messrs. Tukey and Goodman," Technometrics, Vol. 3, No. 2, May, 1961, pp. 229-232.

26. Johnson, Norman L. and Fred C. Leone, Statistics and Experimental Design Vol. II, John Wiley & Sons, Inc., New York, 1964.

BIBLIOGRAPHY (cont.)

27. Jones, Richard H., "A Reappraisal of the Percodogram in Spectral Analysis," Technometrics, Vol. 7, No. 4, Nov., 1965, pp. 531-542.

28. Larson, K. E., Spectral Analysis of Number Sequences Formed by Random Number Generators, Master's Thesis in Industrial Engineering, Lehigh University, Bethlehem, Pa., 1964.

29. Levin, M. J., A Method for Power Spectrum Parameter Estimation, M.I.T. Lincoln Laboratory, Lexington, Mass., 1965, U. S. Department of Commerce, AD 612 796.

30. Mohsen, A., Senior Engineer, Western Electric Co., private communication.

31. Nerlove, Marc, "Spectral Analysis of Seasonal Adjustment Procedures," Econometrica, Vol. 32, No. 3, July, 1964, pp. 241-286.

32. Nerlove, M. and S. Wage, "On the Optimality of Adaptive Forecasting," Management Science, Vol. 10, No. 2, January, 1964, pp. 207-224.

33. Parzen, Emanuel, The Role of Spectral Analysis in Time Series Analysis, Stanford University, Stanford, Calif., 1965, U. S. Department of Commerce, AD 617 937.

34. Parzen, Emanuel, "Mathematical Considerations in the Estimation of Spectra," Technometrics, Vol. 3, No. 2, May, 1961, pp. 167-190.

35. Parzen, Emanuel, "Comments on the Discussions of Messrs. Tukey and Goodman," Technometrics, Vol. 3, No. 2, May, 1961, pp. 232-234.

36. Priestley, M. B., "Basic Considerations in the Estimation of Spectra," Technometrics, Vol. 4, No. 4, Nov., 1962, pp. 551-564.

37. Priestley, M. B. and C. H. Gibson, "Estimation of Power-Spectra by a Wave-Analyser," Technometrics, Vol. 7, No. 4, Nov., 1965, pp. 553-560.

38. Rosenblatt, Murray (ed.), Time Series Analysis, John Wiley & Sons, Inc., New York, 1963.

39. Salmon, Richard L., Adaptive Exponential Smoothing for Forecasting Non-Stationary Time Series, Master's Thesis in Industrial Engineering, Lehigh University, Bethlehem, Pa., 1965.

40. Theil, H. and S. Wage, "Some Observations on Adaptive Forecasting," Management Science, Vol. 10, No. 2, Jan., 1964, pp. 198-206.

BIBLIOGRAPHY (cont.)

41. Trigg, D. W. and A. G. Leach, "Exponential Smoothing with an Adaptive Response Rate," Operational Research Quarterly, Vol. 18, No. 1, March, 1967, pp. 53-59.

42. Tukey, John W., "Discussion, Emphasizing the Connection Between Analysis of Variance and Spectrum Analysis," Technometrics, Vol. 3, No. 2, May, 1961, pp. 191-219.

43. Wonnacott, Thomas H., "Spectral Analysis Combining a Bartlett Window with an Associated Inner Window," Technometrics, Vol. 3, No. 2, May, 1961, pp. 235-243.

VITA

PERSONAL HISTORY

Name:	Steven Michael Brown
Date of Birth:	May 26, 1939
Place of Birth:	St. Louis, Missouri
Parents:	Albert J. and Yetta P. Brown
Wife:	Marilyn D. Brown
Children:	Hal David

EDUCATIONAL BACKGROUND

Washington University of St. Louis Bachelor of Arts in Mathematics	1957 - 1962
Bachelor of Science in Engineering Science - Applied Mathematics	1957 - 1962
Lehigh University Candidate for Master of Science in Industrial Engineering	1966 - 1968

PROFESSIONAL EXPERIENCE

Computer System Design and Development as Development Engineer	1962 - 1966

SCHOLASTIC HONORS

Phi Eta Sigma - Freshman Honorary	1957
Phi Beta Kappa	1962
Alpha Pi Mu - Industrial Engineering Honorary	1967