THE DYNAMICS OF ADAPTIVE FORECASTING MODELS

by

RICHARD NOLAN DAY

B.S., KANSAS STATE UNIVERSITY, 1966

A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1969

Approved by:

L. E. Grosh

Major Professor

TABLE OF CONTENTS

INTRODUCTION

One of the biggest problems in industry today is to obtain a forecasting model that conforms to the demand data. Instead of forecasting actual data from some operation, we have designed a controlled experiment in which we generate a known demand series according to a particular model. The proglem then is to test how well each of the forecasting models can estimate this known demand pattern. The more accurate the forecasting model, the better job it will do in predicting this generated demand. In particular, it would be of benefit if we could evaluate a forecasting model and its parameters, and readily ascertain if the real process is adequately represented. By investigating the response of the forecasting model and its parameters to the demand model, we will make recommendations that should simplify the forecasting process, and provide the forecaster with more accurate results, obtained with less time and effort.

In forecasting the demand (generated) series we will use several different combinations of two different models; model #1 consists of a linear model with superimposed sinusoidal and model #2 consists of a quadratic model with superimposed sinusoidal.

The following combination will be used in making this study.

		(Generation) Demand	Forecasting
Model Combination	1	Model #1	Model #1
	2	Model #2	Model #1
	3	Model #1	Model #2

There are several methods that can be used to forecast the demand series; the first group includes the static forecasting methods, subjective estimates, graphical curve fitting, and regression analysis. The second group includes the dynamic methods, simulation, moving averages, discounted regression, adaptive smoothing and exponential smoothing.

Simulation involves formulating a model that represents a real process and by studying the reaction of this real process, we can simulate it as to future time.

The moving average is the more elementary form of the dynamic forecasting methods and can be represented by

$$\hat{X}(t+1) = \frac{X(t) + X(t-1) + \ldots + X(t-n+1)}{n} ,$$

where $\hat{X}(t+1)$ = estimate of the process at time t+1 based on the moving average of the last n observations.

In this method each of the last n observations has a weight of 1/n, and the observations t-n and older have a weight of zero.

Discounted regression and adaptive smoothing are methods which use the discounted sum of the squared residuals in forecasting the demand (generated) series.

Exponential smoothing is the more complex form of dynamic forecasting methods and can be represented by

$$S_x(t) = \alpha X(t) + \beta S_x(t-1),$$

where $S_x(t)$ = the smoothed value of X at time t,

$S_x(t-1)$ = the smoothed value of X at time t-1,

α = smoothing constant, (weight given the current observation),

β = discount factor, (weight given all past observations),

$= X - \alpha$

$X(t)$ = observed value of X at time t.

All of the dynamic methods have one thing in common; they all discount data by giving older data less weight. This is a much more realistic approach; when a company finds that its sales have taken a rather large increase in the past few periods, it would make sense that the more recent data should have more influence in making the next forecast than data that is three years old.

There are several advantages in using exponential smoothing over moving averages. Essentially exponential smoothing fits a least squares curve to the weighted time series; that is, it minimizes the discounted sum of squared residuals, or fitting errors. The data is weighted by giving data taken k periods in the past, a weight of $\alpha\beta^k$. It also provides for updating the forecasting model's coefficients at each new sampling interval to correct for past forecasting errors. Because of its ability to react to changes in the observed process and its ability to update its coefficients at each new sampling interval we will use exponential smoothing in forecasting all of the time series used in this thesis (Chart 1, 5).

In doing this research we will evaluate methods of determining when a model is no longer representative of the observed process, and study how the following factors affect the ability of models to respond to different demand data:

1. The effect of the discount factor on the response time,
2. The effect of using different combinations of both similar and dissimilar demand generating and forecasting models,
3. The distribution of errors encountered in making the forecasts.

The reader should be cautioned that the matrix notations and techniques used in both the demand generator and the forecasting process are very similar in nature; the terminology is the same, but the values of different parameters may differ slightly.

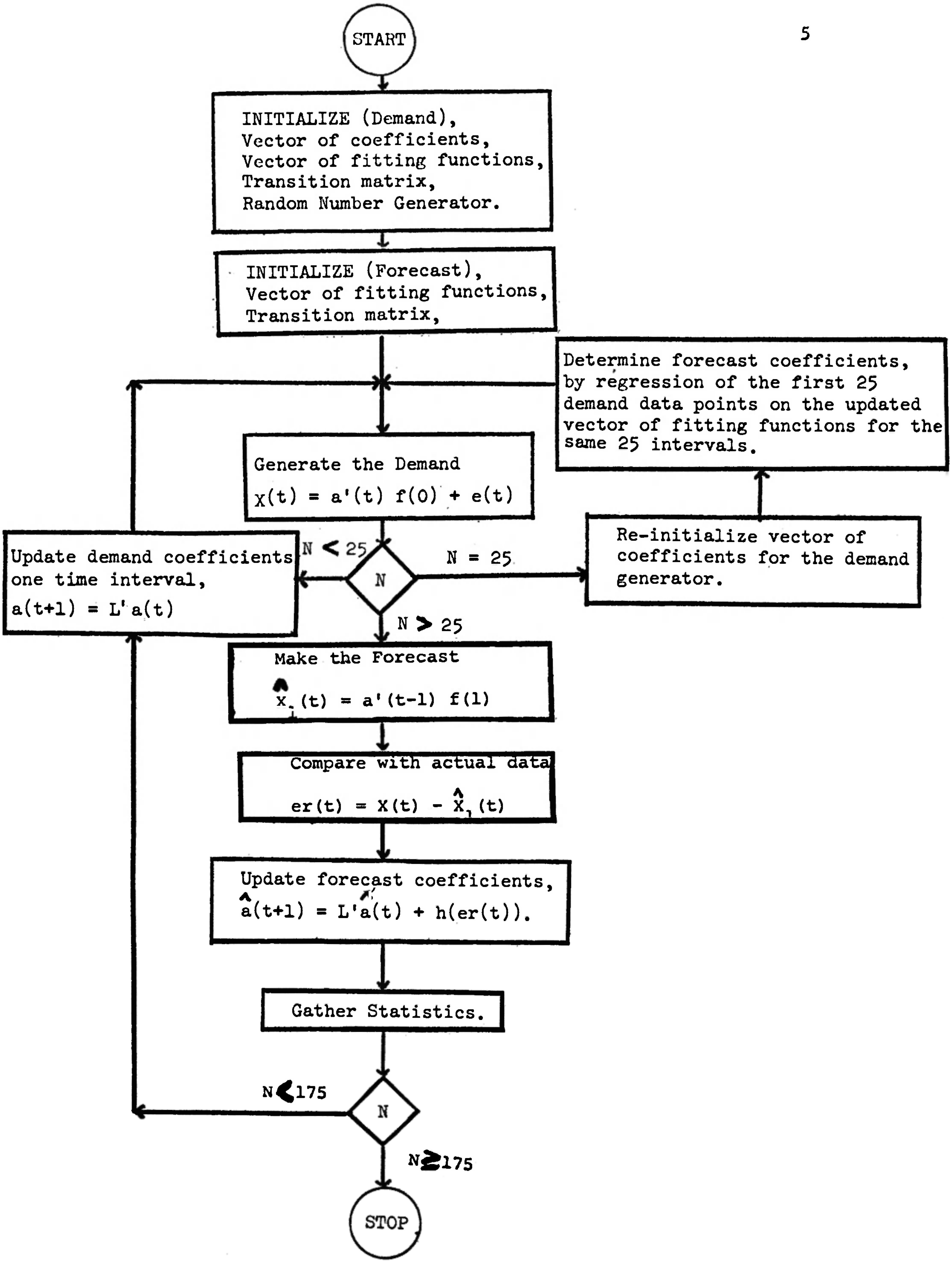

START
INITIALIZE (Demand),
Vector of coefficients,
Vector of fitting functions,
Transition matrix,
Random Number Generator.
INITIALIZE (Forecast),
Vector of fitting functions,
Transition matrix,
Determine forecast coefficients, by regression of the first 25 demand data points on the updated vector of fitting functions for the same 25 intervals.
Generate the Demand
$X(t) = a'(t) f(0) + e(t)$
Update demand coefficients one time interval,
$a(t+1) = L'a(t)$
N < 25
N
N = 25
Re-initialize vector of coefficients for the demand generator.
N > 25
Make the Forecast
$\hat{x}_1(t) = a'(t-1) f(1)$
Compare with actual data
$er(t) = X(t) - \hat{X}_1(t)$
Update forecast coefficients,
$\hat{a}(t+1) = L'\hat{a}(t) + h(er(t))$.
Gather Statistics.
N<175
N
N≥175
STOP

CHART I

1. REPRESENTING THE TIME SERIES

1.1 Time Series Model

A useable time series is a set of observations taken at equal time intervals. To simplify the analysis, the time series can be represented as the sum of two components:

1. The process that generates the time series,
2. Random noise that is superimposed.

Random noise is an unpredictable (or not to be predicted variation, which can be described by a probability distribution having a zero mean.

Mathematically the time series can be represented as:

$$X(t) = \xi(t) + \varepsilon(t),$$

where $X(t)$ = the actual time series at time t,

$\xi(t)$ = the systematic component of the observation at time t,

$\varepsilon(t)$ = random component with

$$E(\varepsilon\ (t)) = 0,$$

$$E(\varepsilon\ (t_i)\ \varepsilon(t_j)) = 0 \text{ for } i \neq j,$$

$$= \sigma^2 \text{ for } i = j.$$

The above equations imply that the random noise components are serially independent, and have a variance of σ^2.

We will attempt to describe the systematic component and defer the random component until later. Once we are given a past history of a time series we may employ any of the following methods to attack this forecasting problem:

1. Subjective estimates,
2. Graphical curve fitting,
3. Statistical regression analysis,
4. Simulation,
5. Moving averages,
6. Discounted regression,
7. Adaptive smoothing,
8. Exponential smoothing.

Static methods of time series representation involve the first three methods mentioned above. In subjective estimation a person will rely on his own experience, hunch, intuition, etc., to estimate the value $X(t)$ at a time t in the future. Curve fitting is very similar to subjective estimation; once a person sees a plot of past data, he will attempt to draw a representative curve through it. By continuing it on into future time, he will attempt to forecast using the predictions based on this "curve". Neither of these above mentioned methods has any analytical foundation.

Statistical regression analysis is a common analytical tool used in static time series analysis. This method attempts to formulate a model that will describe the systematic component of the time series, $\xi(t)$. The model may contain any independent variables that are known over the total time span; it is up to the investigator, however, to choose the proper terms and their form in the model. The result of statistical regression techniques is the proper coefficients of the terms in the model. These coefficients are selected so as to minimize the sum of squared residuals, Rs, for the time series values given (Appendix A).

$$Rs = \sum_{t=0}^{N} [X(t) - \hat{X}(t)]^2$$

where N = number of data values given,

$\hat{X}(t)$ = forecasted value of X at time t,

$X(t)$ = actual time series value of X at time t.

Although this method minimizes the sum of the squared fitting errors, it makes no provision to re-evaluate the model coefficients at each point in time to reflect the forecasting error made in the previous time period. We therefore have to repeat the regression process before each forecast is made; this is very time-consuming and expensive, especially when we are dealing with large amounts of data.

1.2 Dynamic Methods of Time Series Representation

Simulation techniques attempt to formulate a model that will react like the actual process over a given time period. By studying the reaction of the actual process we can simulate this process as to future time period.

The remaining four methods that we will discuss have one unique feature that separates them from all of the previously discussed methods. In attempting to forecast a time series we would expect the most recent values to be more representative of the actual process; we should therefore give the more recent observations more weight than the older ones. In other words, we should discount past data as new data becomes available.

In moving averages, the simplest form of dynamic series representation, we will keep a moving average based on the N most recent time periods.

$$M(t) = \frac{X(t) + X(t-1) + \dots + X(t-N+1)}{N},$$

$$\hat{X}(t+1) = M(t),$$

$$= M(t-1) + \frac{(X(t) - X(t-N))}{N},$$

where $M(t)$ = moving average of X over the N most recent observations.

When the model that we are using is a constant model,

$$\hat{X}(t+1) = a(t) + \varepsilon(t)$$

$\hat{X}(t+1)$ = estimated value of true process at time t+1

$\varepsilon(t)$ = random noise

$a(t)$ = constant process

M(t) minimizes the sum of squares of the differences between the N most recent observations and the estimate $\hat{X}(t+1)$ of the process model (Appendix A). If the model is a constant process, the average is an unbiased estimate of the coefficients a(t), but if the process is changing, it will take N observations for the moving average to yield estimates relevant to this new level. Note that when N is large we are giving more weight to older data than when N is small. The three disadvantages of moving averages are:

1. It gives equal weight to the observations that it does consider,
2. It does not consider all of the past data,
3. It is very costly to change the response rate, 1/N

It is desirable that we consider all past data; however, we are faced with the problem of storing this data, especially when there are several hundred different "items". Problems can also arise when we want to change the rate of response, based on N, the number of observations in our moving average. To alter this response rate would entail a complete revision of our record system.

The desire to discount the older data brings us to the topic of smoothing the data. Smoothing is a process like curve fitting; but there is a distinction that is perhaps more than just a different point of view. In a curve fitting problem, we have a set of data to which some curve is to be fitted; the computations are done once, and the curve should fit equally well to the entire set of data.

A smoothing problem starts the same way, with good data and a model to represent the process being forecast. The model is fitted to the data; that is, the coefficients in the model are estimated from the data available to date. The differences between curve fitting and smoothing are two:

1. The model should fit current data very well, but it is not important that the data obtained a long time ago fit as well,
2. The computations are repeated with each new observation.

The differences stem from the fact that the values of the coefficients in the true process may be changing slowly and at random.

Let us consider the following example where we have assumed that we have lost all of our past data except for the moving average M(t-1). We would expect our best estimate for $\hat{X}(t+1)$ to be,

$$\hat{X}(t+1) = \hat{M}(t) = M(t-1) + \frac{(X(t) - M(t-1))}{N}$$

$$= \frac{1}{N} X(t) + (1 - \frac{1}{N}) M(t-1)$$

where $\hat{M}(t)$ is an estimate of M(t).

This process is known as 'smoothing', where

$\alpha = 1/N$,

$\beta = 1-1/N$,

$S_x(t) = M(t)$.

We now have

$$S_x(t) = \alpha X(t) + \beta S_x(t-1),$$

$X(t)$ = observed value of process at time t,

where $S_x(t)$ = smoothed value of X at time t,

$S_x(t-1)$ = smoothed value of X at time t-1,

α = smoothing constant, $\alpha<1$,

β = discount factor, $\beta<1$,

$\alpha + \beta = 1$.

α and β are not the same as 1/N and 1-1/N, but are very similar.

When we apply these smoothing techniques to a time series we get:

$$S_x(t) = \alpha X(t) + \beta S_x(t-1)$$

$$= \alpha X(t) + \beta[\alpha X(t-1) + \beta S_x(t-2)]$$

$$= \sum_{i=0}^{t} \alpha\beta^{i} X(t-i)$$

As we go back in time the exponent i increases arithmetically; at time i, the weight given X(t-i) is β^{i}. We therefore call this exponential smoothing. This form of smoothing offers several advantages over moving averages.

1. Every data point, regardless of age, is used in determining $S_x(t)$, rather than the last N points as used in moving averages.
2. The response rate may be changed by varying α or β; the smaller the β, the less weight we are giving to past data. Because of the problems involved in changing the complete record system, it is impractical to change the response rate when using moving averages.

3. All of the past data may be stored as one value, $S_x(t)$. In moving averages we must keep a record of all of the data points so that each point may be added and subtracted in determining the moving average.
4. Exponential smoothing minimizes the weighted sum of squared residuals.

We have, up until the present, combined all of the last three dynamic methods of forecasting under the title of smoothing. There are some very important differences that should be pointed out.

Discounted multiple regression is very similar to multiple regression analysis; the main difference is that the older the data, the less weight it receives in making the next forecast. After each forecast is made and the actual value observed, all of the dependent variables are again regressed on the independent variables to determine the least squares coefficients. This is a continuous process and is very time consuming.

Adaptive smoothing is similar to discounted regression analysis except that the origin of time is the time of the most recent observation. It is dissimilar in that it is limited to functions that have transition matrices; because of the transition matrix, there are recursive relationships that can be used instead of the continuous least squares regression analysis.

General exponential smoothing is the last and most complex of the dynamic forecasting methods; it is almost identical to adaptive smoothing in its recursive relationships except for the addition of a method of correction the coefficients for the error made in the previous forecast.

Because it contains the good qualities of the two previous methods plus the ability to correct for the previous forecast error, general exponential smoothing is a most useful tool.

1.3 Choice of Fitting Functions

As mentioned in section 1.1, the time series can be represented by a systematic component $\xi(t)$ and a random component $\varepsilon(t)$. $\xi(t)$ is composed of a set of coefficients $a_i(t)$ and their corresponding fitting functions $f_i(t)$, which, in matrix notation, are defined as:

$$X(t) = a'(t)\ f(t),$$

where $X(t)$ = forecasted value of X at time t,

$a'(t)$ = transposed vector of coefficients,

$f(t)$ = vector of fitting functions.

The time series that we will be using will determine the form of the fitting function to be used; we must keep in mind that fitting functions must be known exactly for future time.

The simplest model that Brown (1, 57) talks about is the polynomial model; a polynomial of degree zero can be represented by the following constant model.

$$\begin{aligned} X(t) &= a'(t)\ f(t), \quad f(t) = 1, \\ &= a'(t)\ 1, \\ &= a'(t). \end{aligned}$$

Going one step further we can consider a polynomial of degree 1, in other words, a linear model where:

$$f_1(t) = 1, \; f_2(t) = t,$$

$$X(t) = \sum_{i=1}^{2} a_i'(t)\, f_i(t)$$

$$= a_1'(t)\, 1 + a_2'(t)\, t$$

$$= a'(t)\, f(t)$$

$$a(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \end{bmatrix}, \qquad f(t) = \begin{bmatrix} 1 \\ t \end{bmatrix}.$$

Brown (1,119) recommends that the number of observations gathered over a past period should be at least five times the number of terms (degrees of freedom); naturally, too few observations would be unreliable in making an accurate forecast. In some cases the proper degree polynomial to be used can be determined from visual inspection of the plot; in other cases, statistical regression analysis can be used.

Brown's second classification of time series models is sinusoidals. Whenever we attempt to describe a periodic process, it is more convenient to use sinusoidal models; a periodic process is one that repeats itself over a given time period p.

The general form of this model is

$$\xi(t) = a \cos \frac{(2\pi)}{p} (t-t_0) + C,$$

where a, C, and t_0 are to be estimated. Rewriting this equation in terms of sines and cosines, Brown (1, 68),

$$\xi(t) = A\sin\left(\frac{2\pi t}{p}\right) + B\cos\left(\frac{2\pi t}{p}\right) + C,$$

where $A = a\sin\left(\frac{2\pi t_0}{p}\right)$,

$B = a\cos\left(\frac{2\pi t_0}{p}\right)$,

$C = c$, have to be estimated.

In final form, the forecasts can be expressed as:

$$X_{t+\tau} = \hat{A}_t\sin\left(\frac{2\pi t}{p}\right) + \hat{B}_t\cos\left(\frac{2\pi t}{p}\right) + \hat{C}_t.$$

The third classification of models is exponential - in other words, the rate of change of the process is a fixed percentage.

$$X(t) = ka^t,$$

$$X(t-1) = ka^{t-1},$$

$$\frac{X(t)}{X(t-1)} = a \text{ (constant)}.$$

When it is plotted on a semi-logarithmic paper, it will define a straight line with a slope of log a,

$$\log \xi(t) = \log k + t\log a.$$

The general form of the model is,

$$\xi(t) = k_1\binom{t}{0} a^t + k_2\binom{t}{1} a^{t-1}b + \ldots + k_N\binom{t}{N-1} a^{t-N+1}b^{N-1}$$

where $\binom{t}{k} = \frac{t!}{(t-k)!k!}$ is the binomial coefficient of the k+1 st term of the expansion of $(a+b)^t$.

Generally when we are attempting to forecast a given process we will have more than one of these functions present. The process will probably contain a constant and a linear term to show the starting point and the growth trend; it may also contain a periodic, or sinusoidal, term to represent seasonal variation,

$$\hat{X}(t) = a_1 + a_2t + a_3\sin(\frac{2\pi t}{p}) + a_4\cos(\frac{2\pi t}{p})$$

$\xi(t)$ will equal $\hat{X}(t)$ if the coefficients $a_i(t)$ are accurately estimated and our model is correct.

The last model that Brown mentions is the autoregression model; this model covers a wide range of linear forecast models of the form,

$$\xi(t) = a_1\xi(t-1) + a_2\xi(t-2) + \ldots + a_n\xi(t-n),$$

where $\xi(t-i)$ can be any combination of independent fitting functions of the three previous models. Many of the economic forecasts are based on 'leading' series where the observed series $X(t)$ is related to several other time series represented by $\xi(t-i)$. The use of regression models can, however, be very dangerous. If past sales have shown a significant correlation with the Dow-Jones Index, the better the past correlation, the greater will be our confidence in these forecasts; hence, the greater the fiasco that results when, sooner or later, sales will turn in the opposite direction from the index. Another problem is the fact that there appears to be no statistical test that will tell whether the model is a good one. However, we shall discuss some tests in Chapter 4, which we feel gives us a reasonable method of evaluating a good, or bad, model. A second difficulty is that even if two series are related, one doesn't know the independent series far enough in advance to be useful in a forecast; an exception to this is the case of population related phenomena - the number of people age Y in year T is known quite accurately in year T-Y.

Note: the fitting functions $f_i(t)$ are subject to the restriction that the value of the function is known exactly both at the time of the current observation and at a time in the future for which the forecast is required.

2. DEVELOPING A DEMAND GENERATOR

In order to use exponential smoothing to forecast a time series we must first be able to furnish past data on the time series. If past data is not available we must generate this data. In our particular case, we must develop some method of generating random normal deviates with a known mean and variance.

2.1 Random Generator

Lehmer (Naylor, 7, 42) has developed the congruential method for generating random numbers on the interval 10, 1; this method generates random numbers that are:

1. Uniformly distributed,
2. Statistically independent,
3. Reproducible,
4. Nonrepeating for a maximum length,
5. Capable of being generated at high rates of speed,
6. Require a minimum of computer memory capacity.

Congruential methods are based on a fundamental congruence relationship which may be expressed as the following recursive formula: (Naylor, 7, 48)

$$n_{i+1} = an_i + c(\mathrm{modM}),$$

$$n_i = a^i n_0 + c\frac{(a^i-1)}{(a-1)} (\mathrm{modM}),$$

where n_0 = initial starting value,

a = constant multiplier,

c = additive constant.

Three basic congruential methods have been developed for generating pseudo-random numbers using different versions of the above equation. Because of its rapid speed we have chosen the multiplicative congruential method, in which c is set equal to zero, and

$$n_{i+1} = an_i (\text{mod}M),$$

where $M = p^e$

p = the base of the number system,

= 2, for the binary system,

e = number of binary bits in a word,

= 31,

therefore $M = 2^{31}$,

and 1. The longest sequence (period) is 2^{e-2}, or 2^{29},

2. a is relatively prime to M,

3. $a = \pm 3(\text{mod}8)$,

$= 8t \pm 3,\ t \geq 0.$

2.2 Derivation of a Normal Generator

Since we have already developed a method for generating random numbers on the interval (0,1), we would now like to modify these random numbers to be normally distributed with a mean of zero and a variance of σ^2. To do this, we let U1 and U2 be independent random variables, from the above random generator on the interval (0,1), (Naylor, 7, 90), where

$$X1 = (-2\log_e U1)^{\frac{1}{2}} \cos(2\pi U2),$$
$$X2 = (-2\log_e U1)^{\frac{1}{2}} \sin(2\pi U2),$$

then (X1, X2) will be a pair of independent random normal deviates from the same normal distribution with mean = 0, and variance = 1.

The above method is an exact method (Naylor, 6, 95) which yields normal deviates that are reliable in the tails of the distribution. Whereas the accuracy of other methods is not easy to analyze, or to change, the accuracy obtainable here depends essentially on the precision of the necessary function subroutines. Further, since most computing centers have library programs to compute values of sinusoidal functions, logarithms, and square roots, this approach requires little additional program writing.

Another method of generating the normal distribution is called the 'central limit' approach; the Central Limit theorem states that the random variable Y_N, defined as the arithmetic mean of a sequence of N independent random variables X_i, has a distribution whose shape tends to a limiting shape that is independent of the distribution of X_i, as long as X has a finite variance as N becomes infinite. With N approaching infinity,

$$NY_N = X_1 + X_2 + \ldots + X_N.$$

has a mean $(N\mu)$, and a variance $(N\sigma^2)$ that becomes infinite with N; the shape of the distribution becomes difficult to follow because it flattens out and moves off to infinity. The drawbacks to this theorem's approximation to the normal distribution are:

1. Poor distribution in the tails,
2. The number of observations must approach infinity if the function is to be continuous.

3. FORECASTING THE TIME SERIES

As mentioned in Chapter 1, the forecasted value $X(t)$ of any time series can be represented by

$$X(t) = a(t)' f(t) \text{ where}$$

1. the time series can be represented by

   $$X(t) = L(t) + P(t) + E(t) + \varepsilon(t)$$

   $L(t)$ = linear trend

   $P(t)$ = periodic component

   $E(t)$ = exponential growth trend

   $\varepsilon(t)$ = random noise

   L, P, E are deterministic values of time

2. The coefficients $a(t)$ can be determined from regression analysis.
3. The fitting functions are fixed for a given model.

3.1 Discounted Multiple Regression

Since the previous models we have talked about have been rather elementary, we must develop the smoothing techniques for the more complicated models, as discussed in Chapter 1.

Over some period of time, the coefficients $a_i(t)$ can be treated as constant. Two sequences of observations of the same process at widely separated intervals of time may require different values of one or more of the coefficients. The underlying process $\xi(t)$ can be going through a slow random walk with respect to one or more of the coefficients. We therefore want to develop a simple iterative procedure for revising the estimate of the coefficient values with each new observation, and the revision should in some way discount information received to date. The resulting forecast of X for τ periods into the future will be:

$$\hat{X}(t+\tau) = a_1(t)f_1(t+\tau) + a_2(t)f_2(t+\tau) + \ldots + a_n(t)f_n(t+\tau),$$

$$= a_i(t)f_i(t+\tau),$$

where $\hat{X}(t+\tau)$ = forecasted value of X at time $(t+\tau)$

$a_i(t)$ = least squares estimate of $a_i(t)$ based on weighted data, evaluated from time 0 to t,

$f_i(t+\tau)$ = the ith fitting function evaluated for a time period of $(t+\tau)$.

If the current model, as of time t, were used to describe past data, the residual would be defined as $X(t-j) - \hat{X}(t-j) = e(t-j)$. Note: $X(t-j)$ stands for the model where the coefficients are evaluated with all the data through time t, but the model is evaluated at a time j periods earlier.

The usual least squares multiple regression seeks to determine the coefficients $a_i(t)$ so that the sum of squares of these residuals will be a minimum over the finite number of observations t. We shall therefore determine values for the coefficients so as to minimize the discounted sum of squares residuals,

$$\sum_{j=1}^{t} \beta^j [X(t-j) - \sum_{i=1}^{n} a_i(t)\ f_i(t-j)]^2,$$

where t = present time.

It is easier to use matrix notation in estimating these least squares techniques. By definition,

$X = [X(1), X(2), X(3), \ldots X(t)]$ = $(1 \times t)$ row vector of the t observations up to the present,

$$f(t) = \begin{bmatrix} f_1(t) \\ f_2(t) \\ f_3(t) \\ \vdots \\ f_n(t) \end{bmatrix}$$ = $(n \times 1)$ column vector of n linearly independent fitting functions at time t,

$$a(t) = \begin{bmatrix} a_1(t) \\ a_2(t) \\ a_3(t) \\ \vdots \\ a_n(t) \end{bmatrix}$$ = $(n \times 1)$ column vector of coefficients at time t,

$$F = \begin{bmatrix} f_1(1) & f_1(2) & \ldots & f_1(t) \\ f_2(1) & f_2(2) & \ldots & f_2(t) \\ \cdots & \cdots & \cdots & \cdots \\ f_n(1) & f_n(2) & \ldots & f_n(t) \end{bmatrix} = \text{(n x t) matrix of fitting functions,}$$

$$W(t) = \begin{bmatrix} w(1) & & & 0 \\ & w(2) & & \\ & & w(3) & \\ & & & \\ 0 & & & w(t) \end{bmatrix}$$

where w(t) is the t-th diagonal element of a txt matrix W(t), and

$$w^2(t-j) = \beta^j, \quad j = 0, 1, 2, \ldots t - 1,$$

If we define a row vector of residuals, e(t),

$$e(t) = X(t) - a'(t)\ F(t),$$

then the discounted sum of squared residuals is,

$$Rs = [e(t)W(t)][e(t)W(t)]'$$

$$= w^2(t)e^2(t)$$

for a particular set of coefficients a(t). To minimize this sum, we need to select a(t) such that

$$\frac{d\ Rs}{d\ a(t)} = 0.$$

This expression gives us n simultaneous linear equations which can be solved for the n coefficients $a_i(t)$, where,

$$a_i(t) = X(t)W'(t)\mathcal{F}(t)W(t)F^{-1}(t),$$

$$= g(t)F^{-1}(t),$$

$F(t) = \mathcal{F}(t)W(t)W'(t)\mathcal{F}(t) =$ (n x n) symmetrical matrix of weighted fitting functions,

If $t \geq n$, and if the $F_i(t)$ are linearly independent, $F^{-1}(t)$ exists.

Note: the F(t) matrix does not depend on the data; it does however depend on time, and must be updated at each sampling interval,

$$F_{ik}(t+1) = f_i(t+1)f_k(t+1) + \sum_{j=0}^{t-1} \beta^j f_i(t-j)f_k(t-j),$$

$$F_{ik}(t) = \sum_{j=0}^{t-1} \beta^j f_i(t-j)f_k(t-j),$$

where t = present time.

The n component data vector g(t), mentioned above, is

$$g(t) = \begin{bmatrix} g_1(t) \\ g_2(t) \\ \vdots \\ g_n(t) \end{bmatrix} = XW'FW,$$

and $$g_i(t) = X(t)f_i(t) + \beta\ g(t-1),$$

$$= \sum_{j=0}^{t-1} \beta^j X(t-j)\ f_i(t-j).$$

Initially the value of the i-th component of the data vector would be the product of the first observation X(1) and the value of the i-th fitting function at time $f_i(1)$. Referring to the expression for the i-th data vector indicates that it can be updated at each sampling interval by,

$$g(t) = X(t)\ f(t) + \beta g(t-1),$$

where t = present time

t-1 = previous time period.

Once $F^{-1}(t)$ can be calculated, a'(t) may be estimated by

$$\hat{a}'(t) = g(t)F^{-1}(t),$$

and future observations may be forecast by

$$\hat{X}(t+\tau) = \hat{a}'(t)\ f(t+\tau).$$

3.2 Functions with Fixed Transition Matrices

By placing a restriction on the class of fitting functions we can simplify the procedures of the previous section. The vector of values of these functions at time t+1 must be a linear combination of the values of the same functions at the previous time t - that is, there is a set of coefficients L_{ij} that do not depend on time, such that

$$\begin{array}{lcccccc} f_1(t+1) & = L_{11}f_1(t) & + L_{12}f_2(t) & + & \ldots & + L_{1n}f_n(t) & \\ f_2(t\pm1) & = L_{21}f_1(t) & + L_{22}f_2(t) & + & \ldots & \pm L_{2n}f_n(t) & \\ \cdot & \cdot & \cdot & & \cdots & \cdot & \\ \cdot & \cdot & \cdot & & \cdots & \cdot & \\ f_n(t+1) & = L_{n1}f_1(t) & + L_{n2}f_2(t) & + & \ldots & + L_{nn}f_n(t) & ; \end{array}$$

the above matrix is called a transition matrix and is denoted as L. Its relationship with f(t) is

$$f(t+1) = L\ f(t).$$

The transition matrix L, in general, is not symmetrical, but it must have an inverse. The only set of functions for which such a transition matrix exists are:

1. Polynomials,
2. Exponentials,
3. Sinusoidals.

The transition matrices for these functions can be written down by inspection from the class of functions used in the model. Once this matrix is found, the vector of fitting functions at any other time can be found by

$$f(t) = L^t f(0).$$

For a polynomial of degree n, the transition matrix is an (n + 1) x(n + 1) matrix with ones on the diagonal, ones in the first element to the left of the diagonal, and zeros everywhere else,

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \qquad f(0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where L has n+1 rows and n+1 columns and the fitting functions have the form

$$\begin{aligned} f_n(t) &= \binom{t}{n}, \\ &= t!/(t-n)!n! \\ &= \frac{t(t-1)(t-2)(t-3)\ldots(t-n+1)}{n!}. \end{aligned}$$

If the fitting functions are trigonometric functions, both the sine and the cosine for each harmonic must be included, so that we can allow the origin of time and the mean to be arbitrary. If $f_1(t) = \sin(wt)$, and $f_2(t) = \cos(wt)$, then

$$L = \begin{bmatrix} \cos w \sin w \\ -\sin w \cos w \end{bmatrix}, \qquad f(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

The procedure can cope with growing amplitudes or shifting phase angles by including the functions $f_3(t) = t\sin(wt)$, and $f_4(t) = t\cos(wt)$,

$$L = \begin{bmatrix} \cos w & \sin w & 0 & 0 \\ -\sin w & \cos w & 0 & 0 \\ \cos w & \sin w & \cos w & \sin w \\ -\sin w & \cos w & -\sin w & \cos w \end{bmatrix}, \quad f(0) = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

The transition matrix may always be formed by stacking the simple matrix shown along the diagonal of L. For the construction of exponential transition matrices, see (Brown, 1, 166).

3.3 Adaptive Smoothing

Observing the restrictions placed on f(t) in the previous section, and moving the origin of time to the most recent observation, we write

$$\hat{X}(t+\tau) = f'(\tau)\ \hat{a}(t).$$

Since we count time relative to the present, this is reflected in using $f(\tau)$ instead of $f(t+\tau)$. The forecast still shows the actual time. Now the error criterion that is to be minimized is

$$\sum_{j=0}^{t} \beta^j [f'(-j)a(t) - X(t-j)]^2.$$

The n simultaneous equations that result from setting the partial derivatives of the error criterion equal to zero can be written in matrix form as

$$F(t)\ \hat{a}(t) = g(t).$$

If there have been at least n observations, the symmetrical matrix F(t) is positive and has an inverse; therefore the coefficients in the forecast equation can be estimated, as before, by

$$\hat{a}(t) = F^{-1}(t)\ g(t),$$

where $$F_{ik}(t) = \sum_{j=0}^{t} \beta^j f_i(-j) f_k(-j),$$

$$g_i(t) = \sum_{j=0}^{t} \beta^j f_i(-j)\ X(t-j).$$

With each new observation that data vector g(t) and the function matrix F(t) must be modified. Because of the change of time origin in evaluating the fitting functions, the recursive computations will be different than with discounted regression,

$$g(t) = X(t)\ f(0) + \sum_{j=1}^{t} \beta^j f(-j)\ X(t-j).$$

If successive values of the fitting functions vectors are generated by a transition matrix L, then $f(-j) = L^{-1}\ f(-j+1)$ and

$$g(t) = X(t)\ f(0) + \sum_{j=1}^{t} \beta^j L^{-1} f(-j+1)\ X(t-j),$$

$$F(t) = \sum_{j=0}^{t} \beta^j f(-j)\ f'(-j),$$

$$= F(t-1) + \beta^j f(-t)\ f'(-t).$$

Since for polynomials and sinusoidal functions, β^{τ} tends to zero faster than f(t) can grow, the matrix of fitting functions reaches a steady state value, and its inverse can be computed once and for all. The only problem appears when one of the fitting functions is a decreasing exponential of the form,

$$f(t) = e^{-at};$$

F(t) will reach a steady state if, and only if, $\beta < e^{-2a}$ (Brown, 1, 170); that is, past data must be discounted rapidly. If the fastest growing function in the model is t^n, then the number of periods taken for this convergence is

$$T = \frac{7 + (5.1)n}{(1-\beta)^{.95}} .$$

The criterion for this convergence is that the largest quantity added to any element of F is less than 10^{-6} times the previous value of that element.

Once the steady state has been reached, we do not need the time identification of the matrix of fitting functions, but can simply call it F, and its inverse F^{-1}. The coefficients can be estimated by $\hat{a}'(t) = g'(t)F^{-1}$ and the forecast of future observations will be given by

$$\begin{aligned} \hat{X}(t+T) &= \hat{a}'(t)\ f(\tau), \\ &= g'(t)\ F^{-1}\ f(\tau), \quad (F^{-1}f(\tau) = c(\tau), \\ &= g'(t)\ c(\tau), \end{aligned}$$

where g'(t) is the transpose of the current data vector and c(τ) is a column vector of coefficients that depend only on the values of fitting functions for a forecast period of τ units in the future, but not on absolute time.

When the steady state conditions are to be used,

$$\hat{a}'(t) = g'(t)\ F^{-1},$$

which suggests that the data vector can be given initial conditions for some estimate of the model. Thereafter, the data vector can be updated with each new observation and the forecast $\hat{X}(t+\tau)$ can immediately be obtained.

3.4 General Exponential Smoothing

We have shown under certain conditions, that the matrix of weighted fitting functions F(t) reaches a steady state, F = F(∞); these conditions are:

1. Successive values of the fitting functions can be generated by a fixed transition matrix, L,
2. The forecast is based on these functions with the origin of time at the most recent observation.

These results will be used to minimize the computational effort required in revising the successive estimates of the coefficients a(t) used in the forecasting equation,

$$\hat{X}(t+\tau) = a'(t)\ f(\tau).$$

Since there is noise added to the actual observations, the forecasted values will not be exactly the same as the model; there will be an error, $e_1(t)$, which is the difference between the actual demand value at time t, $X(t)$, and the forecasted value made in the previous time period t-1, $\hat{X}_1(t-1)$. Note: the subscript on the forecasted value is used to indicate the number of periods in the past that the forecast was made. The difference will therefore be

$$e_1(t) = X(t) - \hat{X}_1(t-1).$$

The main advantage of General Exponential Smoothing is the development of an n component h vector of program constants from which the estimates of the forecast coefficients can be revised with each new observation to reflect:

1. The change in the origin of time to the end of the next sampling interval.
2. The error in forecasting the last observation.

The coefficients will be revised by,

$$a(t) = L' \, a(t-1) + h(e_1(t)),$$

where $a(t)$ = coefficients to be used in making next forecast,

$a(t-1)$ = coefficients used in making last forecast,

L' = transpose of transition matrix,

$e_1(t)$ = error made in making last forecast,

h = vector of constants for revising the vector of coefficients.

To derive the above equation, consider the following:

$$F(t) = \sum_{j=0}^{t} \beta^j f(-j)\ f'(-j),$$

$$g(t) = \sum_{j=0}^{t} \beta^j X(t-j) f(-j),$$

$$= X(t)\ f(0) + \beta L^{-1} g(t-i),$$

$$f(t) = L\ f(t-1).$$

We want to minimize the discounted squared residual sum

$$\sum_{j=0}^{t} \beta^j [X(t-j) - a'(t)\ f(-j)]^2.$$

This sum is obtained when

$$F(t)\ a(t) = g(t),$$

and since F has an inverse F^{-1} when its steady state is reached,

$$a(t) = F^{-1} g(t).$$

Substituting the minimum steady-state solution into the recursive equation of the data vector g(t), we get:

$$Fa(t) = X(t)\ f(0) + \beta L^{-1} Fa(t-1),$$

and premultiplying by the inverse of F, F^{-1},

$$a(t) = X(t)F^{-1}f(0) + \beta F^{-1}L^{-1}Fa(t-1)$$

$$= hX(t) + H\ a(t-1)$$

$$h = F^{-1}\ f(0),$$

$$H = \beta F^{-1}L^{-1}F,$$

We shall attack $L^{-1}F$ first by post-multiplying the definition of the F matrix by $L'^{-1}L'$,

$$(L^{-1}F)\ L'^{-1}L' = \sum_{j=0}^{\infty} \beta^{j}[L^{-1}f(-j)][L^{-1}f(-j)]'L',$$

$$= \frac{1}{\beta}[F - f(0)f'(0)]L',$$

therefore $H = F^{-1}L^{-1}F,$

$$= [I - F^{-1}f(0)f'(0)]L',$$

but $h = F^{-1}f(0)$ by definition,

so $H = L' - h[Lf(0)]'$

$$= L' - hf'(1),$$

and $a(t) = h\ X(t) + Ha(t-1)$

$$= h\ X(t) + L'a(t-1) - hf'(1)a(t-1)$$

$$= L'a(t-1) + h[X(t) - X_1(t-1)]$$

4. DISCUSSION OF RESULTS

In making a forecasting model study, we first had to determine the following values for both the demand generator and the forecasting model:

1. The choice of the model,
2. The choice of the initial conditions.

NOTE: We would like to point out that the parameters used in both the demand generator and the forecasting model are identical in terminology; the values, however, may be different.

4.1 Choice of the Model

As stated in the introduction, the purpose of this thesis is to study the reaction and response of the forecasting model when combined with the generating model as follows:

		Demand (Generating)	Forecast
Model Combination	1	Model #1	Model #1
	2	Model #2	Model #1
	3	Model #1	Model #2

The two basic models that were used for these three studies were:

Model #1. A linear trend with superimposed sinusoidal,

$$X(t) = a_0 + a_1(t) + a_2\sin(\frac{2\pi t}{p}) + a_3\cos(\frac{2\pi t}{p}),$$

Model #2. A quadratic with superimposed sinusoidal,

$$X(t) = a_0 + a_1(t) + a_2(t^2) + a_3\sin(\frac{2\pi t}{p}) + a_4\cos(\frac{2\pi t}{p}).$$

We feel that these two models are fairly representative of a time series with a growth component plus a seasonal variation.

Note: Since Model 1 is a 1st degree polynomial, it is less complex than the 2nd degree polynomial, Model 2.

4.2 Choosing the Initial Conditions for the Demand Generator.

Once we have determined which model we will use to generate the demand series, the following parameters must be initialized:

1. Vector of fitting functions and the transition matrix,
2. Vector of coefficients,
3. The initial value for the random number generator.

In referring to 3.2, the vector of fitting functions and the transition matrix can be determined by inspection of the model. The vector of coefficients will be determined to reflect the desired growth and periodic properties; the initial value for the random number generator must be selected so as to minimize first-order serial correlation, while at the same time, must observe the restrictions placed on 'modulo congruency',(see 2.1.)

4.3 Generating the demand series.

When the above parameters have been read into the main program of the computer, the generation of the demand series can begin; the following steps occur in the main program.

1. The vector of coefficients a(0) is updated 1 time period ahead to arrive at a(1); the updating is accomplished by multiplying a(0) by the transpose of the transition matrix, L',

$$a(t) = L' \, a(t-1)$$

2. The vector of fitting functions $f(\tau)$, for a $\tau = 0$, is multiplied by the transpose of the coefficient vector, $a'(t)$, to arrive at a pure demand, apf(t); since we are not making a forecast, but simply arriving at a product, apf(t), the vector of fitting functions is evaluated for a $\tau = 0$,

$$apf(t) = a'(t) \, f(\tau), \quad \tau = 0.$$

3. Normally distributed random noise, with a mean = 0 and a variance $\sigma^2 = 1$, is generated and added to the pure demand to arrive at the simulated actual demand.

$$X(t) = apf(t) + norm(ebar, sigmae, IX)$$

where apf(t) = pure demand

X(t) = simulated actual demand

norm = subroutine for generating normally distributed random noise with a mean = 0, and variance = 1.

ebar = mean of random noise,

= 0,

sigmae = variance of random noise,

= 1,

IX = initial value for the random number generator.

4.4 Choosing the Initial Conditions for the Forecasting Model.

Since we have already developed a simulation procedure for generating the demand time series, the next step involves using General Exponential Smoothing to arrive at a forecast value. As we mentioned in the introduction to this chapter, the following parameters must be initialized after we have chosen our forecasting model:

1. Vector of fitting functions and the transition matrix,
2. Vector of coefficients,
3. Smoothing constant α and discount factor β.

We may choose any realistic model to forecast a time series; the accuracy of the forecast will naturally depend upon our choice of the model and the initial value of its parameters. Since we have limited our choice of models to two, the initial value of the vector of fitting functions and the transition matrix can readily be found (appendix C, Tables 1.1, 2.1). The vector of coefficients can be determined by regressing the independent variable, the vector of fitting functions, updated for 25 time intervals, against the dependent variable, the generated demand data for the first 25 observations. By using simple regression analysis we can fit a least squares curve to the generated data, and arrive at the least squares coefficients, $\hat{a}(t)$.

In choosing the discount factor, β, we must consider its effect on the smoothing process. By varying β we can vary the weight of the older data; we would expect

1. β to be very large if the forecasting model is the same as the demand model (indicating a high degree of confidence in past data).
2. β to be small if the demand model is changing faster than the

forecasting model (indicating less confidence in older data).

3. β to be very large if the forecasting model is growing faster than the demand model.

Thus for a large β the system response rate is slow; for a small β the response rate is high.

4.5 Forecasting the demand series.

Once the values of the input parameters have been initialized, we can evaluate the other variables necessary to forecast the demand series using the following computer subroutines which are listed in more detail in Appendix D.

The F matrix, the matrix of fitting functions, is calculated in subroutine, RAY,

$$F(t) = \sum_{j=0}^{t} B^{j} f(-j) f'(-j)$$

$$= F(t-1) + \beta^{t} f(-t)\ f'(-t)$$

The F matrix will converge to a steady state when the largest quantity added to any element of F is less than 10^{-6} times the previous value of that element.

In subroutine MATINV the F matrix is inverted, and the h vector is calculated in subroutine HVEC, where

$$h = F^{-1} f(0),$$

F^{-1} = inverse of F matrix,

f(0) = vector of fitting functions for a $\tau = 0$.

In subroutine FORCST the forecasting process begins, after the parameters are fed into the routine.

1. The vector of coefficients a(t) obtained from our regression analysis is multiplied by the vector of fitting functions, $f(\tau)$, $(\tau=1)$, to obtain $X_1(t)$, the forecast made in time t for a forecasting interval of 1,

   Thus

   $$\hat{X}_1(t) = a'(t)f(1).$$

2. This forecast is then compared to the actual demand value X(t+1) at the next interval.
3. The difference between the actual demand X(t+1) and the forecast $\hat{X}_1(t)$ is the error in forecasting the demand.
4. In updating the coefficients a(t) for the next forecasting interval t+1, the vector of coefficients is multiplied by the transpose of the transition matrix, as in the demand generator. The error, $(X(t+1) - \hat{X}_1(t))$, made in making the last forecast is now multiplied by the h vector to correct the model coefficient vector a(t+1) for the last error,

   $$a(t+1) = L'a(t) + h(X(t+1) - \hat{X}_1(t)).$$

This procedure for making the forecasts is repeated N times before it is terminated.

In most cases N was set at 150 time units; this was done because

1. It allowed the forecasting process to be come stable,
2. It was long enough to avoid any large affect due to errors in estimating the initial coefficients, and, at the same time, was short enough to keep the variables small enough to show distinct differences when plotted on the same time scale.

In only one study was N different from 150; when studying the normality of the forecasting errors, we set N equal to 1000 to get as large a sample as was reasonable to run our Chi-squared test.

4.6 Error Criterion

We have developed the following statistics for evaluating the different combinations of models:

1. The forecast error, er(t),
2. The cumulative error, suma,
3. The variance of the forecasts,
4. The mean error,
5. The mean absolute error,
6. The variance of the error,
7. The Smoothed Mean Absolute Deviation, SMAD,
8. The tracking signal, TS.

The first two error criterion use the difference between the actual demand data and the forecasted value of the data.

The cumulative error, which is simply the sum of the forecast errors to date, indicates if the forecasting model is adapting to the demand data. Theoretically, if the forecasting model is doing a good job of forecasting the real process, the cumulative sum of the errors should fluctuate around zero; if the cumulative error is growing in either the positive or negative direction, the forecasting model is not doing its job, and should be revised.

The variance of the forecasts is not acually the variance of the forecasts, but the variance of the fitting errors, er(t). The computations that are involved in finding the true variance of the forecasts are quite lengthy; we have therefore used this simpler approximation. The variance of the forecast errors is computed by:

$$\sigma^2_{er} = \sum_{t=1}^{N} \frac{(er(t))^2}{N}$$

where $er(t) = X(t) - \hat{X}_1(t-1)$,

$X(t)$ = actual demand data at time t,

$\hat{X}_1(t-1)$ = forecast of the t-th observation made at time t-1,

N = total number of points forecasted.

In a similar manner, the standard deviation of the forecast errors is

$$\sigma_{er} = \sqrt{\sigma^2 er} \quad .$$

The mean error is computed by

$$\bar{e} = \sum_{t=1}^{N} \frac{er(t)}{N} ,$$

and the variance of the mean error is a true error and is computed by

$$\sigma^2_{\bar{e}} = \sum_{t=1}^{N} \frac{(e(t) - \bar{e})^2}{N} ;$$

the standard deviation of the mean error is simply the square root of the variance

$$\sigma_{\bar{e}} = \sqrt{\sigma^2_{\bar{e}}} .$$

If there were no noise in our data, and if our forecasting model exactly represented the demand model, $\sigma_{\bar{e}}$ would equal σ_{er}.

The computation of the standard deviation of the forecasts, as mentioned before, is a somewhat complicated process, involving sum of squares and square roots. Therefore the mean absolute deviation is a more meaningful statistic to consider in systems where there are a large number of forecasts to be made. Brown (1,282) derives the proof which shows that the mean absolute deviation (MAD) is equal to .79788 times the true standard deviation. By definition,

$$MAD = \sum_{t=1}^{N} \frac{|er(t)|}{N} .$$

Another method of evaluating the goodness of a model is by the tracking signal. Brown (1,296) defines the tracking signal as the ratio of the sum of the forecast errors to the Mean Absolute Deviation, or

$$TS = \frac{\text{Sum of the Forecast Errors}}{\text{MAD}},$$

where MAD = Mean Absolute Deviation.

The tracking signal is therefore a measure of whether the sum of the forecast errors is reasonably close to zero. From the above definition the tracking signal will vary from $-\infty$ to $+\infty$. The main rationale of using the tracking signal is if the forecast is good, then the average error will be zero; if the average error is zero, then the sum of the errors to date should be zero.

Trigg (8) alters the tracking signal of Brown to

$$TS = \frac{\text{Sum of the Forecast Errors}}{\text{SMAD}},$$

where SMAD = the smoothed mean absolute deviation,

= (β) previous SMAD + (1-β) latest absolute error.

Initially the SMAD is set equal to the MAD, and the following observed absolute errors are smoothed. The tracking signal was altered because of the following reasons:

1. Once the tracking signal has gone out of control, it will not necessarily return within limits even though the forecasting system itself comes back into control. Consequently, intervention is necessary to set the sum of the errors back to zero if we want to avoid future false claims; these interventions can become very tedious and expensive when several hundred items are involved.

2. Once the system starts to give exceptionally accurate forecasts, the tracking signal may go out of control. If perfect forecasts begin to occur, the MAD will tend to zero while the sum of the errors will remain unaltered; this leads to the tracking signal tending to infinity.

To prevent the tracking signal from tending to infinity when the SMAD goes to zero, Trigg (8) recommends still another change in the tracking signal, where smoothed errors are used,

$$TS = \frac{\text{Smoothed Error}}{\text{SMAD}} ,$$

where Smoothed error = (β) previous smoothed error + ($1-\beta$) latest error.

Brown (2, 161) states that it is a characteristic of the smoothed error that for very small values of the smoothing constant β, the distribution of the tracking signal is nearly normal (.25% of the random samples lie outside the range from minus to plus three standard deviations). When β decreases below .90, however, the distribution becomes noticeably bimodal with an obvious tendency for the ratio to be nearly either +1 or -1 (Brown, 2, 197), it cannot go outside of these limits.

4.7 Normality of Errors

We would like to investigate the normality of errors involved in the forecasting process. To do this we ran a Chi-squared test which involved the generation of one thousand observations with both the generation model and the forecasting model the same as model 1, section 4.1. The initial conditions for both models, except for the coefficient vectors, were the same. The initial values for the vector of coefficients for the forecasting model were determined by regressing the first 25 generated demand data on the forecasting model's vector of fitting functions, updated for each of the 25 observations, where

$f(t) = L^t f(0)$

$f(t)$ = vector of fitting functions evaluated for a forecast interval of t,

$f(0)$ = vector of fitting functions for a $T = 0$,

L = transition matrix.

After each forecast,

1. The vector of four coefficients for the demand model were subtracted from the vector of four coefficients for the forecasting model, to arrive at the four coefficient errors, $A_1 - A_4$.
2. The forecast for each time interval $X_1(t-1)$ was compared with the generated demand value for that interval $X(t)$, and the difference $X(t)-X_1(t-1)$ was tabulated as the forecast error, $Er(t)$,
3. The tracking signal, although not an error, was also tabulated as T.S.

After the one-thousandth observation, all of these variables were tabulated into 18 intervals by using the IBM scientific subroutine TAB1 and a Chi-squared test was run (Tables 10.1 to 10.6). The theoretical frequency in each interval was not less than 5; in some runs, the frequencies in the tails were combined to yield this minimum frequency.

The Chi-squared tests resulted in six Chi-squared values for each model combination. Since we will be satisfied if no more than one in twenty observations are outside our acceptance limits, we have set a .95 significance level on the Chi-squared values; at this level, the tracking signals were significant for both runs when Beta was equal to .70, and with a Beta of .80, one of the tracking signals was significant (Appendix C, Tables 10.1 to 10.6).

These results imply that as Beta decreases from .90 to .70, the distribution of the tracking signal becomes less normal: contrary to what Brown (2,197) states, we find that the tracking signal is not bimodal, but has a very poor distribution in its tails (Appendix C, Tables 9.1 to 9.6).

4.8 Evaluation of the Model Combinations

We have divided this section into three parts, the number of model combinations,

1. Linear Demand Model, Linear Forecast Model,
2. Quadratic Demand Model, Linear Forecast Model,
3. Linear Demand Model, Quadratic Forecast Model.

Note: In the first combination, because it is the simplest of the three combinations, we use only two different initial values for the random number generator. The second combination, since it is the most common, was run with four different initial values. The third and last combination had but one initial value since it is the least common of the three; it is seldom that we know enough about the demand model to use this combination. All of these initial values for the random number generator satisfy the conditions for the modulo congruency relationships of section 2.2, (Appendix C, Table 1.2).

4.8.1 Linear Demand Model, Linear Forecast Model

This combination of demand and forecasting models is as close to identical as we can get. The initial conditions for both models are the same, except for the vector of coefficients. Because the only difference between the models is the vector of coefficients, we have tabulated the coefficient errors, besides the forecast errors and the tracking signal, to test their normality. In running this Chi-squared test we assume that there is no serial correlation that enters into the tabulation (Appendix B).

We would expect, since our models are identical, the larger discount factor, β of .90, to give much better error statistics than the smaller betas of .80 and .70. As β decreases from .90 to .70, all of the error statistics (Appendix C, Table 6.1) indicate that decreasing β

1. Increases the cumulative error,
2. Increases the variances of the forecasts,
3. Decreases the mean error,
4. Increases the mean absolute error,
5. Decreases the variance of the error,
6. Increases the Smoothed Mean Absolute Deviation (SMAD).

These results indicate that when both the demand model and the forecasting model are the same, a large β will give a more realistic weighting factor for the older data. The mean error and its variance indicate that with a large β, the forecast errors become negative and fluctuate less than with the smaller Betas of .80 and .70; this is confirmed by the cumulative error, which becomes increasingly more negative.

The tracking signal for all three Betas fluctuates between -1 and +1; the larger the β, however, the smaller the deviations from the zero line (Appendix C, Table 3.1 to 3.3).

From these above results, it is clear that when both our forecasting model and our generating model are the same, a large β will minimize our error criterion and produce a tracking signal that fluctuates very close to zero.

4.8.2 Quadratic Demand Model, Linear Forecast Model

This combination of models introduces us to a forecasting process that is most common in industry today. We will attempt to forecast a demand model which we do not completely understand; we will forecast a quadratic model with a linear model. The choice of the initial conditions is the same, except for the coefficient of the quadratic term (Appendix C, Table 1.1). From studying the output statistics we see that decreasing β

1. Decreases the cumulative error,
2. Increases the variance of the forecasts,
3. Decreases the mean error,
4. Increases the mean absolute error,
5. Decreases the variance of the error,
6. Increases the Smoothed Mean Absolute Deviation (SMAD),

As in the previous section, the β of .90 gives better error statistics;

except for the mean error. The mean error and its variance indicated that the larger the β , the larger the cumulative error becomes. This is verified by examining the cumulative error statistics.

As the discount factor Beta decreases from .90 to .70, the tracking signal reflects the increase in the forecasting errors and responds accordingly. For a β of .90, the tracking signal remains predominantly positive, approaching 1.0; for a Beta of .70, the tracking signal fluctuates on both sides of zero. This fluctuation of the tracking signal gives us one method of evaluating the adequacy of the model; the more accurate the forecasting model, the smaller the error, and the closer the tracking signal will be to zero.

We have chosen the initial value of the quadratic term so small that it will take more than 150 time intervals before its real effect on the forecasting process becomes evident. This is somewhat implied by considering the cumulative error and the mean error. With a β of .90, the cumulative error becomes very large as we near the 150th observation; the quadratic term is starting to react and the β of .90 is too large to correct for it. At the same time the mean error is the largest for a β of .90, while the mean absolute error is the smallest for the same value of Beta.

4.8.3 Linear Demand Model, Quadratic Forecast Model

With this combination of models we are attempting to forecast a linear process using a quadratic forecast model. Once we have determined the initial coefficients of the forecasting model through regression analysis, we can make the 150 forecasts and gather the statistics. From studying the output statistics, we see that decreasing Beta

1. Increases the cumulative error,
2. Increases the variance of the forecasts,
3. Increases the mean error,
4. Increases the mean absolute error,
5. Decreases the variance,
6. Increases the Smoothed Mean Absolute Deviation (SMAD).

As Beta decreases below .85 the statistics exceed their output formats and the forecasting model "blows up" (Appendix C).

When β is .90, the tracking signal fluctuates initially, but soon reaches a steady state value of +1. As β decreases the tracking signal periodically varies between +1 and -1; this suggests that there is strong serial correlation present.

It would appear from the above results that using a more complex forecasting model than generating model can lead to particularly disastrous results, unless the discount factor β is between .90 and 1.00. For Beta less than .90 the forecasting model introduces errors due to the extra quadratic term and over-corrects for the quadratic coefficient at each new sampling interval.

5. CONCLUSION

From our discussion in the previous chapter we may conclude that there are definite relationships between the choice of the different model combinations and the value of the test statistics.

When the forecasting model is the same as the demand model, we find that the best error statistics are obtained. At the same time our choice of Beta should be large (.90) to reflect our confidence in the demand model.

When using a quadratic demand model and linear forecasting model, we find that we obtain the best error statistics when Beta decreases as the magnitude of the quadratic coefficient increases; thus, the greater the difference between our demand model and our forecasting model, the less weight we want to put on our older data.

In the final combination of models (linear demand, quadratic forecast), we observe some interesting results. As Beta decreases, the weight given to the older data decreases, but the forecasting process over-corrects for the errors, causing the forecasting model to go out of control. This implies that by reading in erroneous variables as forecasting parameters we must use a large Beta (.95-.90) or the forecasting model 'blows up'.

The tracking signal indicates, in each case, the adequacy of the forecasting model to respond to random changes in the demand data. For the optimal model combinations to minimize the error statistics, the tracking signal fluctuates very close to zero; the larger the fluctuation from zero, the poorer the choice of the forecasting model. As the tracking signal becomes predominantly positive (or negative), the forecasting model leads (lags) the demand model at an increasing rate proportionate to the magnitude of the tracking signal.

Generally the exponential smoothing process reacted very favorably as a forecasting process. In one instance when the initial values had been erroneous, the smoothing process took only a matter of a few observations to correct itself when Beta was small; for large Betas the process is slower to correct itself. In another instance the smoothing process was able to quickly adapt to an error made in initializing the phase angle of the sinusoidal term; only through thorough scrutinizing of the time series plot were we able to detect this error.

BIBLIOGRAPHY

1. Brown, Rogert G. Smoothing Forecasting and Prediction of Discrete Time Series. New Jersey: Prentice Hall, Inc., 1963.
2. Brown, Rogert G. Decision Rules for Inventory Management. New York: Holt, Rinehart, and Winston, Inc., 1967.
3. Draper, Norman R., and Smith, Harry Applied Regression Analysis. New York: John Wiley and Sons, Inc., 1966.
4. Graybill, Franklin A. An Introduction to Linear Statistical Models. New York: McGraw-Hill Book Co., Inc., 1961.
5. Irvine, William F. Developing of a Demand Forecasting Model for Floral Data. (Unpublished Master's Thesis, Kansas State University, 1968.)
6. Miller, Raymond C. The Application of Exponential Smoothing to Forecasting a Time Series. (Unpublished Master's Thesis, Kansas State University, 1967.)
7. Naylor, Thomas H., Balintfy, Joseph L., Burdick, Donald S., and Chu, Kong Computer Simulation Techniques. New York: John Wiley and Sons, Inc., 1966.
8. Trigg, D. W. "Monitoring a Forecast System", OPERATIONAL RESEARCH QUARTERLY, Volume 15 (1964), 271-274.

ACKNOWLEDGEMENTS

The author would like to express his deep appreciation to Dr. L. Eugene Grosh for his counsel and guidance in doing this research. The author is also grateful to Mr. William F. Irvine for many of the computer routines used in this research.

APPENDIX A DERIVATION OF THE MATRIX REGRESSION EQUATION

Let us assume we have a set of n discrete observations Y(t) which we want to describe by some model. We also know a set of independent variables $X_i(t)$ with which we can estimate Y(t).

$$Y(t) = \hat{Y}(t) + e(t)$$

where $Y(g)$ = actual value of time series at time t.

$\hat{Y}(t)$ = estimated value of Y(t) at time t.

$e(t)$ = error in estimation.

β_0 = constant

β_i = coefficients of $X_i(t)$

$$Y(t) = \beta_0 + \beta_1 X_i(t) + \beta_2(t)X_2(t) + \cdots\cdots + \beta_n X_n(t) + e(t)$$

We can represent the above equation in matrix form by,

$$Y = X\beta + \varepsilon \quad \text{where}$$

$$Y(t) = \begin{bmatrix} Y_1(t) \\ Y_2(t) \\ \cdot \\ \cdot \\ Y_n(t) \end{bmatrix} = (n \times 1) \text{ vector of observations}$$

$$\dot{X}(t) = \begin{bmatrix} X_1(1) & \ldots & X_{k-1}(1) \\ X_1(2) & & X_{k-1}(2) \\ & & \\ X_1(n) & \ldots & X_{k-1}(n) \end{bmatrix} = (n \times k) \text{ matrix of independent variables}$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_{k-1} \end{bmatrix} = (k \times 1) \text{ vector of coefficients}$$

$$\varepsilon = \begin{bmatrix} \varepsilon(1) \\ \varepsilon(2) \\ \cdot \\ \cdot \\ \varepsilon(n) \end{bmatrix} = (n \times 1) \text{ vector of errors}$$

We want to determine the vector of coefficients to minimize the sum of squared errors;

$$\varepsilon'\varepsilon = \varepsilon_1^2 + \varepsilon_2^2 +++ \varepsilon_n^2,$$

where ε' is the transpose of ε, and

$$\varepsilon = Y - X\beta$$

$$S = \varepsilon'\varepsilon$$

$$= (Y' - X'\beta')(Y - X\beta)$$

$$= Y'Y - 2\beta'X'Y + \beta'X'X\beta$$

$$= Y'Y - 2(Y'X)\beta + \beta'X'X\beta$$

Taking the partial derivative with respect to β and setting it equal to zero (Graybill, 11), we get

$$\frac{\partial S}{\partial \beta} = - 2X'Y + 2X'X\beta$$

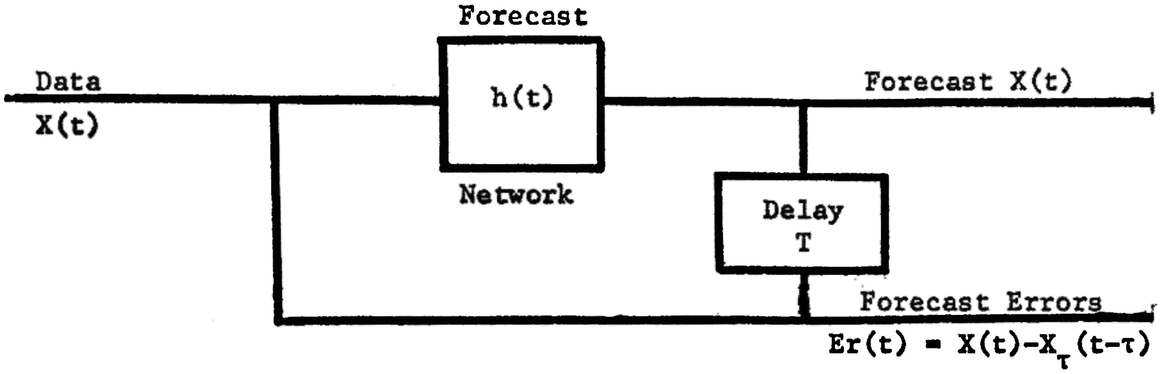
$$X'X\beta = X'Y$$

$$\beta = (X'X)^{-1}X'Y$$

= least square residual estimate of B.

APPENDIX B SERIAL CORRELATION OF THE FORECAST ERRORS

In examining serial correlation of the forecast errors, let us look at the correlation of successive errors produced by some smoothing system. The block diagram (Brown, 1, 309) of the process is shown below:



The output is the forecast X(t) made as of time t. The forecast is delayed T periods and then compared with the current observation. The forecast error is

$$er(t) = X(t) - \hat{X}_\tau(t-\tau),$$

and the forecast can be expressed in terms of the original data by the convolution

$$\hat{X}(t) = \sum_{n=0}^{\infty} h(n)X(t-n).$$

All the information that is relevant to forecasting future errors is contained in the autocorrelation function,

$$R_{ee}(k) = \overline{e(t)e(t+k)},$$

where the bar over the lagged product indicates the average of all such products over the time t. From Brown (1,393) by direct substitution, we get

$$R_{ee}(k) = R_{xx}(k) - \sum_{n=0}^{\infty} h(n)[R_{xx}(k+\tau+n) + R_{xx}(k-\tau-n)]$$

$$+ \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} h(m)h(n)\ R_{xx}(k+n-m)$$

where $R_{xx}(k)$ is the autocovariance for the original data, sampled k periods apart in the sequence.

If we assume that the input data are merely random noise, so that $R_{xx}(0) = \sigma_x^2$, and $R_{xx}(k) = 0$ for $k \neq 0$, then there is no serial correlation between any pairs of noise samples.

When we substitute this assumption into the preceding equations we find the autocovariance of the errors in forecasting for a lead time of τ is

$$R_{ee}(0) = \sigma_x^2[1 + \sum_{n}^{\infty} h^2(n)],$$

$$R_{ee}(k) = \sigma_x^2 \sum_{n=0}^{\infty} h(n)\ h(k+n), \quad 0 < k < \tau,$$

$$R_{ee}(h) = \sigma_x^2 \sum_{n=0}^{\infty} [h(n)h(k+n) - h(k-T)], \quad k \geq \tau.$$

In particular, if we are using single exponential smoothing filter for which the impulse responses is $h(t) = \beta^t$, then

$$R_{ee}(0) = \frac{2}{2-\alpha} \sigma_x^2$$

$$R_{ee}(k) = \frac{\alpha\beta^k}{2-\alpha} \sigma_x^2, \quad 0 < k < \tau,$$

$$= \frac{\alpha\beta^k}{2-\alpha} \sigma_x^2[1 - (2-\alpha)\beta^{-\tau}], \quad k \geq \tau.$$

Since we are using a $\tau = 1$, when

$\beta = .90$	$R_{ee}(1) = - .0053$
$= .80$	$= - .1111$
$= .70$	$= - .1765.$

These results would indicate that there is more serial correlation present than desirable; thus the forecast errors are serially correlated, contrary to our assumptions.

APPENDIX C. TABLES AND CHARTS

A Model 1 - LINEAR TREND WITH SUPERIMPOSED SINUSOIDAL

$$X(t) = a_o + a_1(t) + a_2 \sin \frac{(2\pi t)}{p} + a_3 \cos \frac{(2\pi t)}{p}$$

a) INITIAL VECTOR OF COEFFICIENTS

$$\hat{a}(t) = [1.50000 \quad .50000 \quad 1.50000 \quad 1.50000]$$

b) INITIAL VECTOR OF FITTING FUNCTIONS = f(t)

$$f(t) = [1.00000 \quad 0.00000 \quad 0.00000 \quad 1.00000]$$

B Model 2 - QUADRATIC TREND WITH SUPERIMPOSED SINUSOIDAL

$$X(t) = a_o + a_1(t) + a_2(t^2) + a_3 \sin \frac{(2\pi t)}{p} + a_4 \cos \frac{(2\pi t)}{p}$$

a) INITIAL VECTOR OF COEFFICIENTS

$$\hat{a}(t) = [10.00000 \quad 0.10000 \quad 0.00100 \quad 1.50000 \quad 1.50000]$$

b) INITIAL VECTOR OF FUNCTIONS

$$f(t) = [1.00000 \quad 0.00000 \quad 0.00000 \quad 0.00000 \quad 1.00000]$$

TABLE 1.1

INITIAL CONDITIONS FOR DEMAND MODELS

1 LINEAR DEMAND MODELS, LINEAR FORECAST MODEL

IX(1) = 55003
IX(2) = 54995

2 QUADRATIC DEMAND MODEL, LINEAR FORECAST MODEL

IX(1) = 55003
IX(2) = 54995
IX(3) = 55011
IX(4) = 55019

3 LINEAR DEMAND MODEL, QUADRATIC FORECAST MODEL

IX(1) = 55003

TABLE 1.2

TABLE OF INITIAL VALUES FOR NORMAL RANDOM GENERATOR

A Model 1

$$
TM = \begin{bmatrix} 1.00000 & 0.00000 & 0.00000 & 0.00000 \\ 1.00000 & 1.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.86600 & 0.50000 \\ 0.00000 & 0.00000 & -0.50000 & 0.86600 \end{bmatrix}
$$

B Model 2

$$
TM = \begin{bmatrix} 1.00000 & 0.00000 & 0.00000 & 0.00000 & 0.00000 \\ 1.00000 & 1.00000 & 0.00000 & 0.00000 & 0.00000 \\ 0.00000 & 1.00000 & 1.00000 & 0.00000 & 0.00000 \\ 0.00000 & 0.00000 & 0.00000 & 0.86600 & 0.50000 \\ 0.00000 & 0.00000 & 0.00000 & -0.50000 & 0.86600 \end{bmatrix}
$$

TABLE 2.1

TRANSITION MATRICES

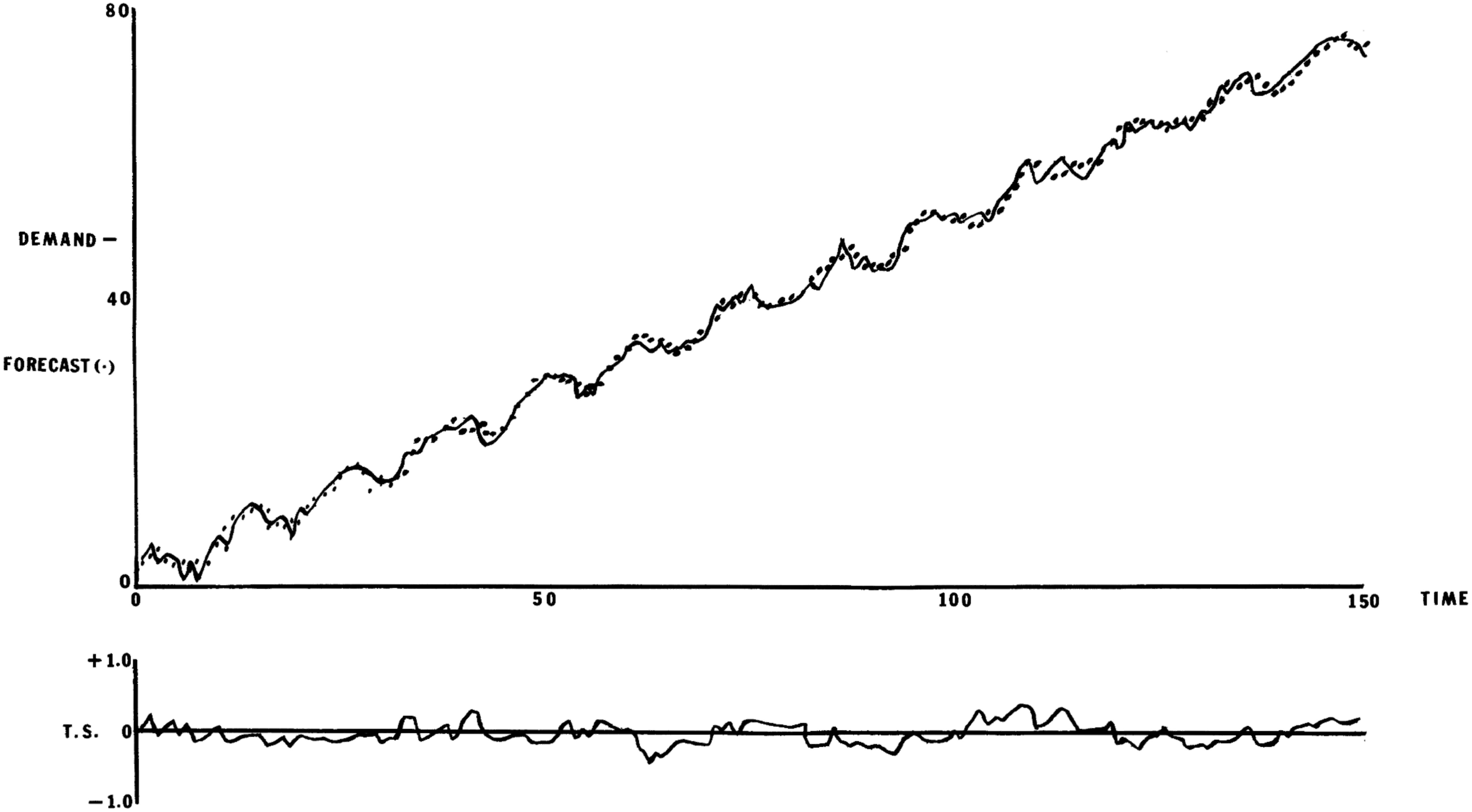


CHART 3.1

beta : .90

linear/linear

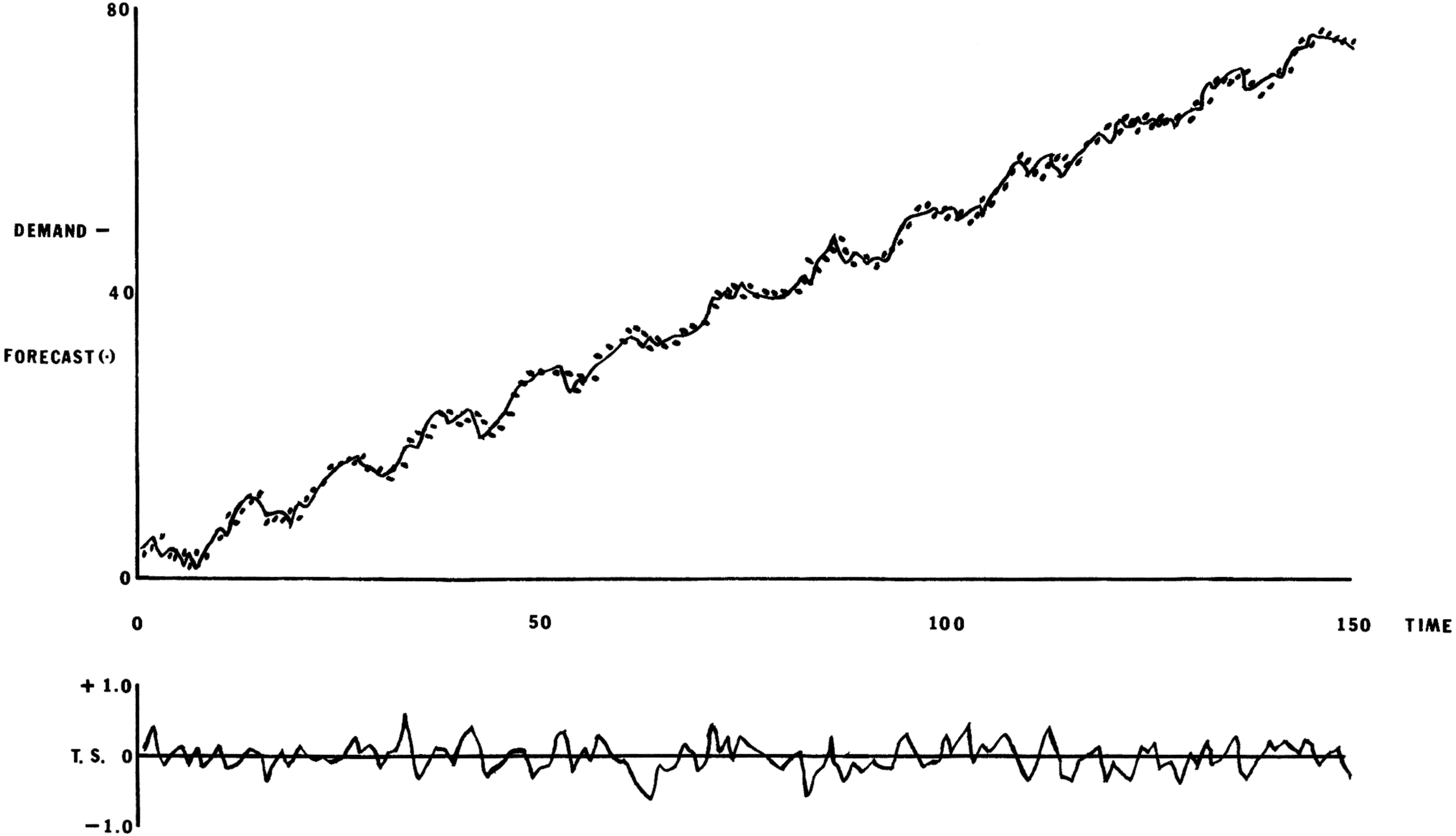


CHART 3.2

beta : .80

linear/linear

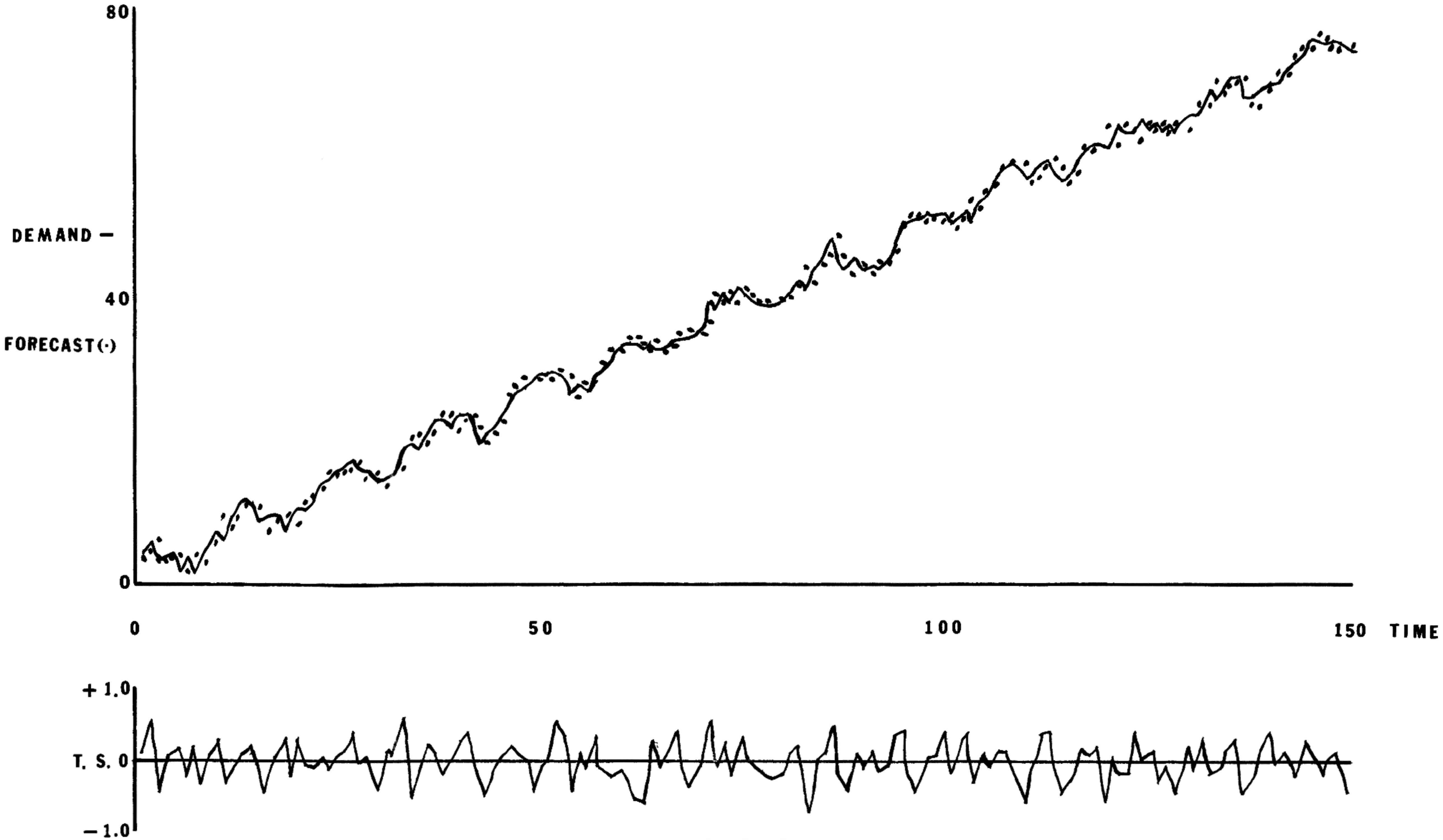


CHART 3.3

beta : .70

linear/linear

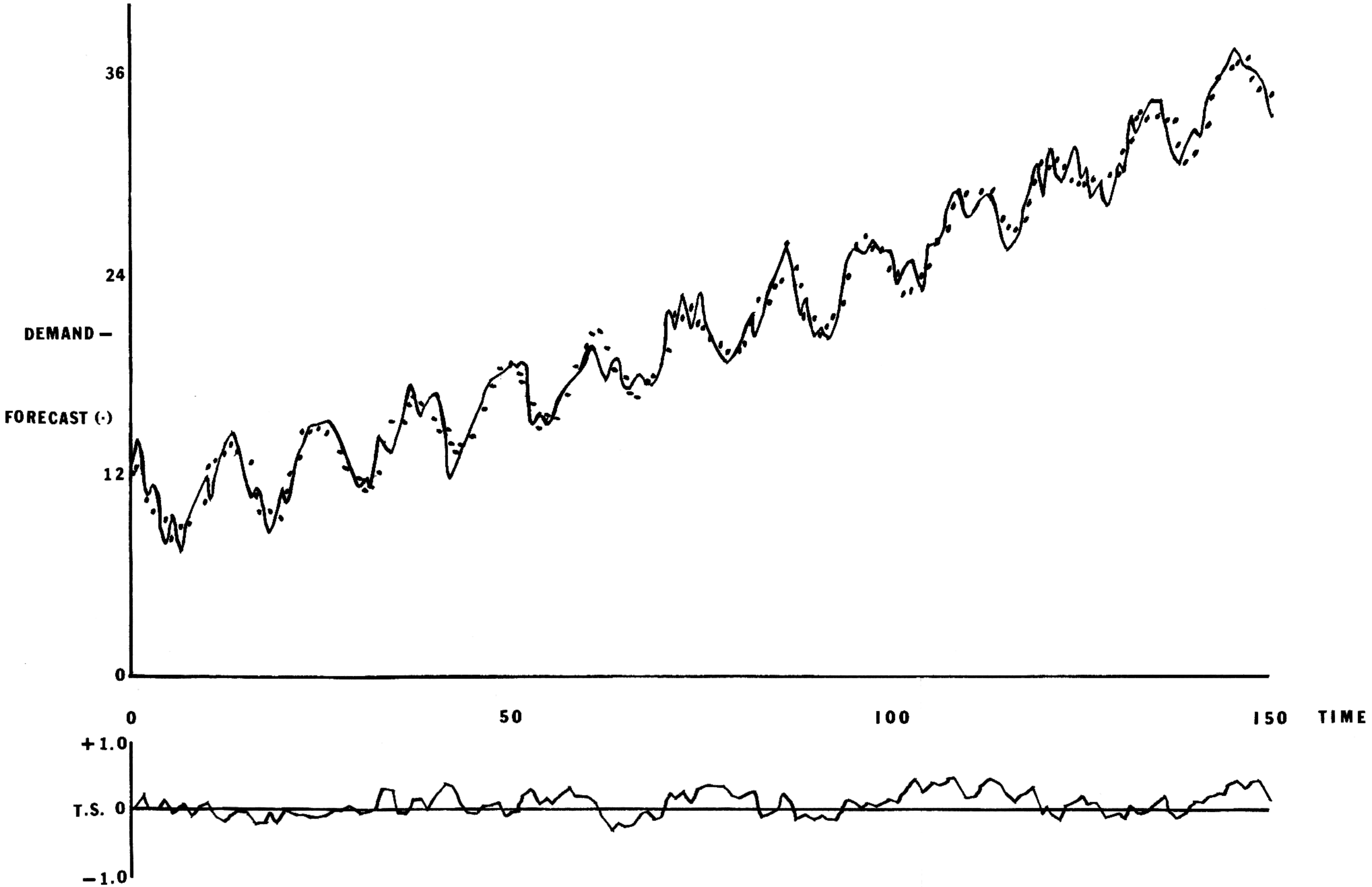


CHART 4.1

beta : .90

quad/line

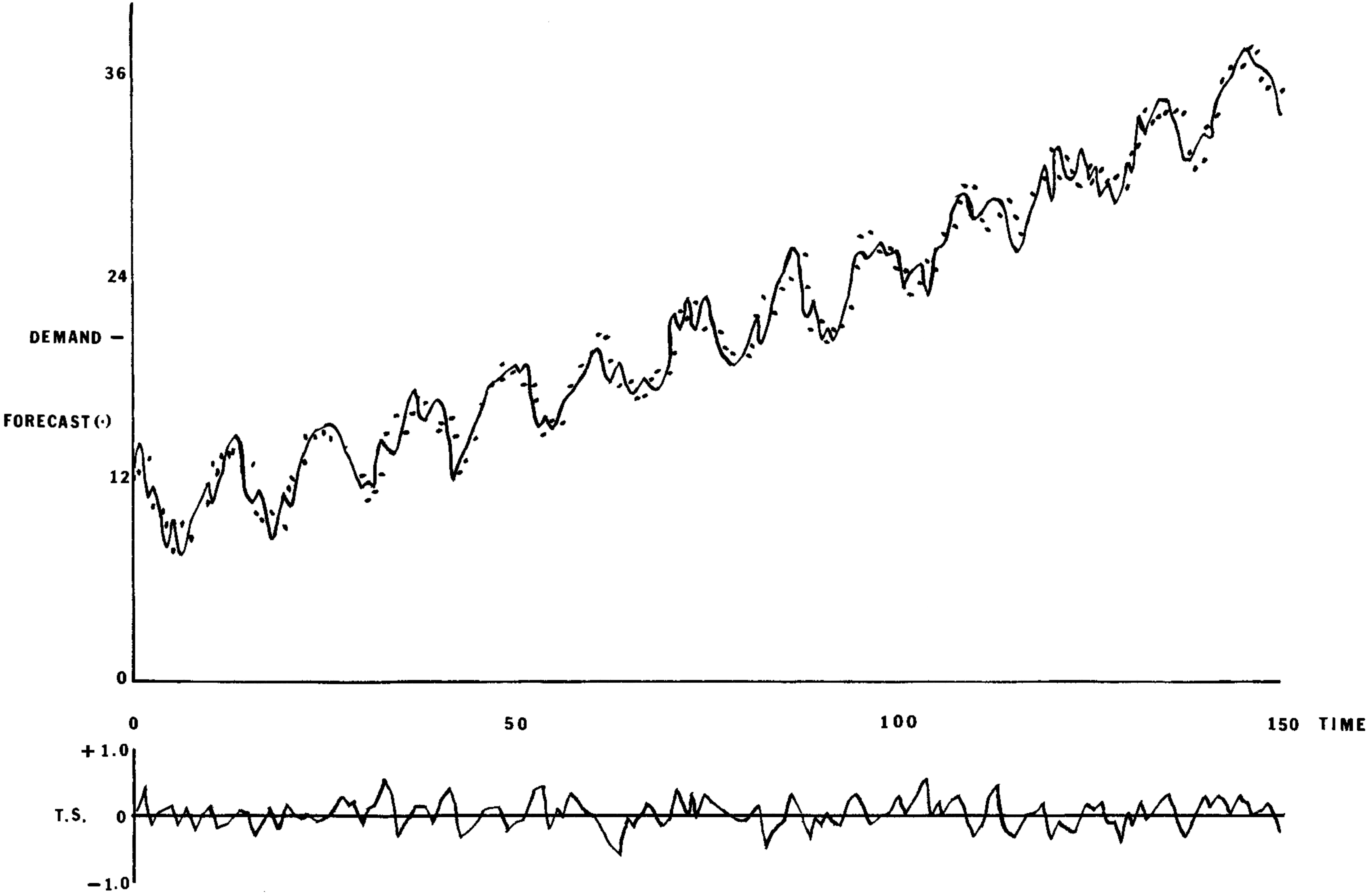


CHART 4.2

beta: .80

quad/ line

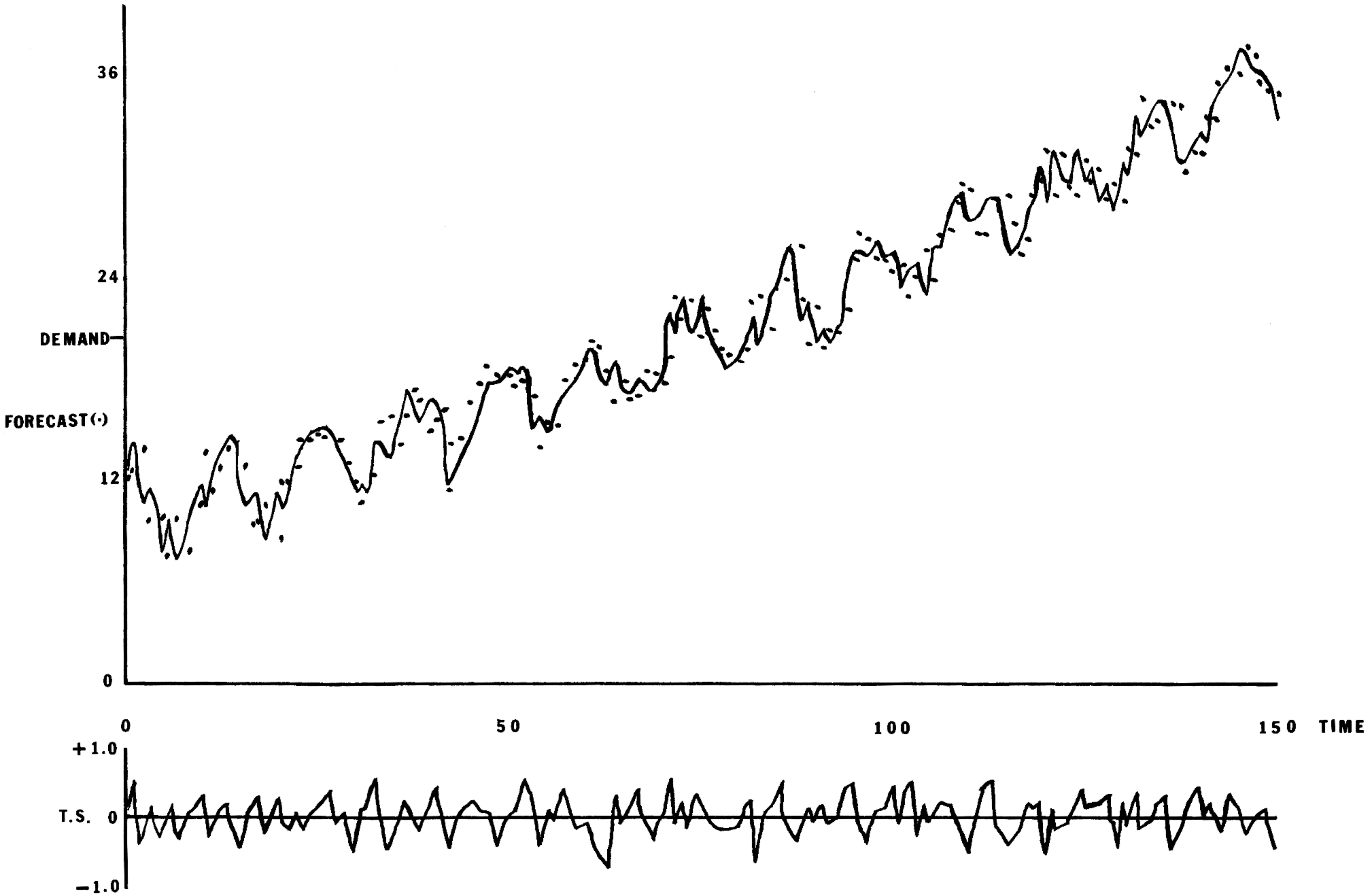


CHART 4.3

beta: .70

quad/line

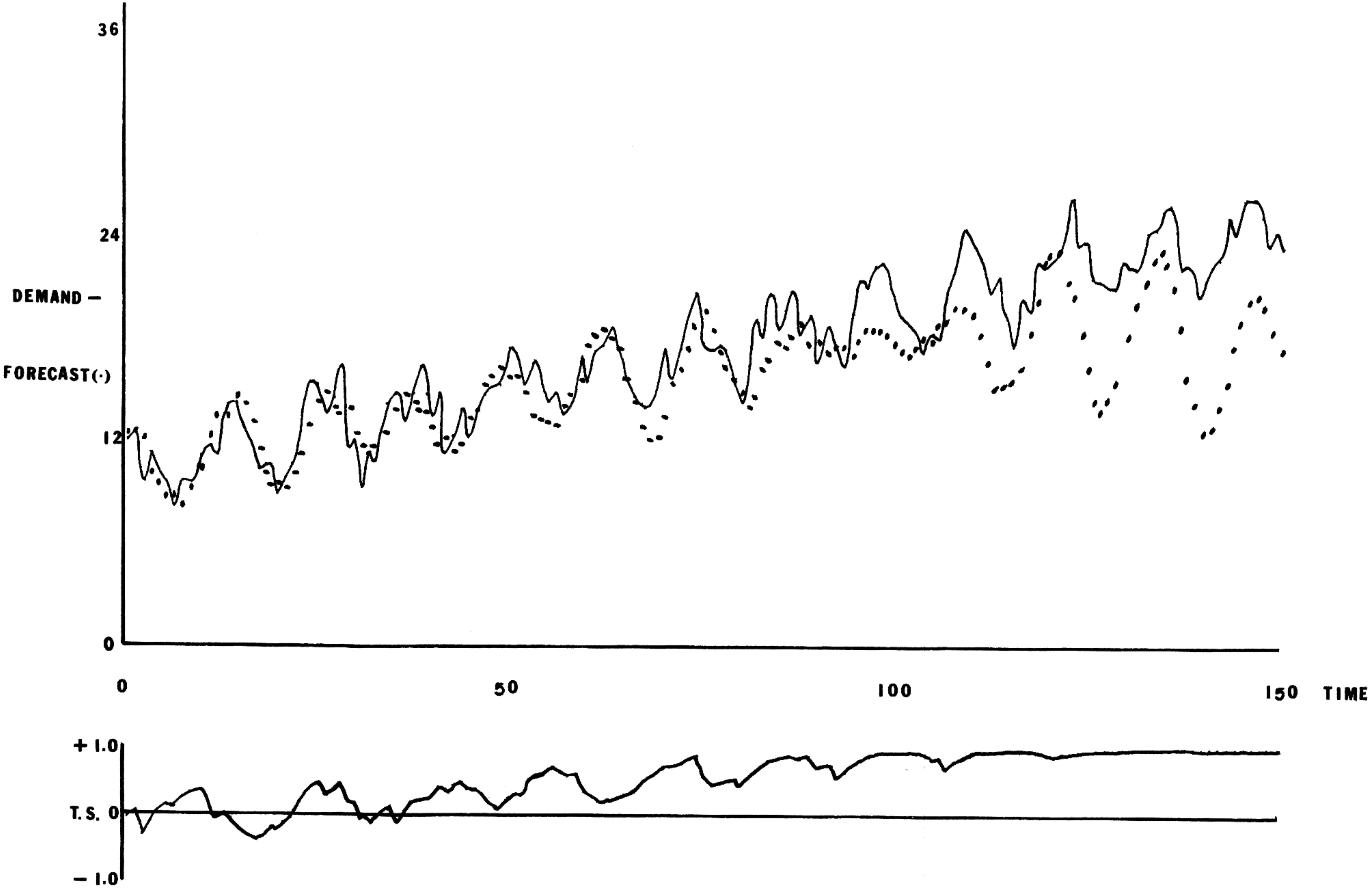


CHART 5.1

beta : .90

linear/quad

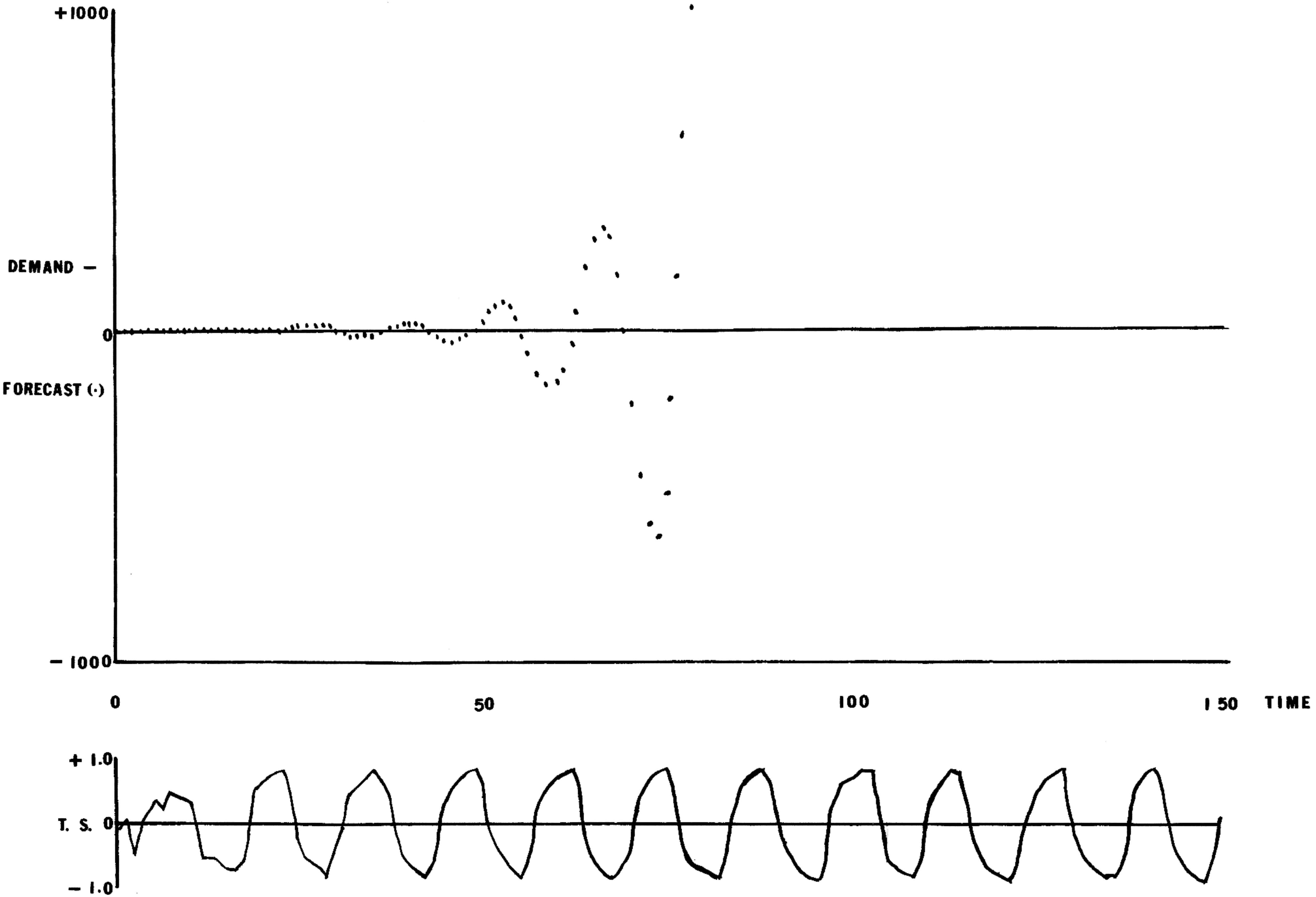


CHART 5.2

beta : .80

linear/quad

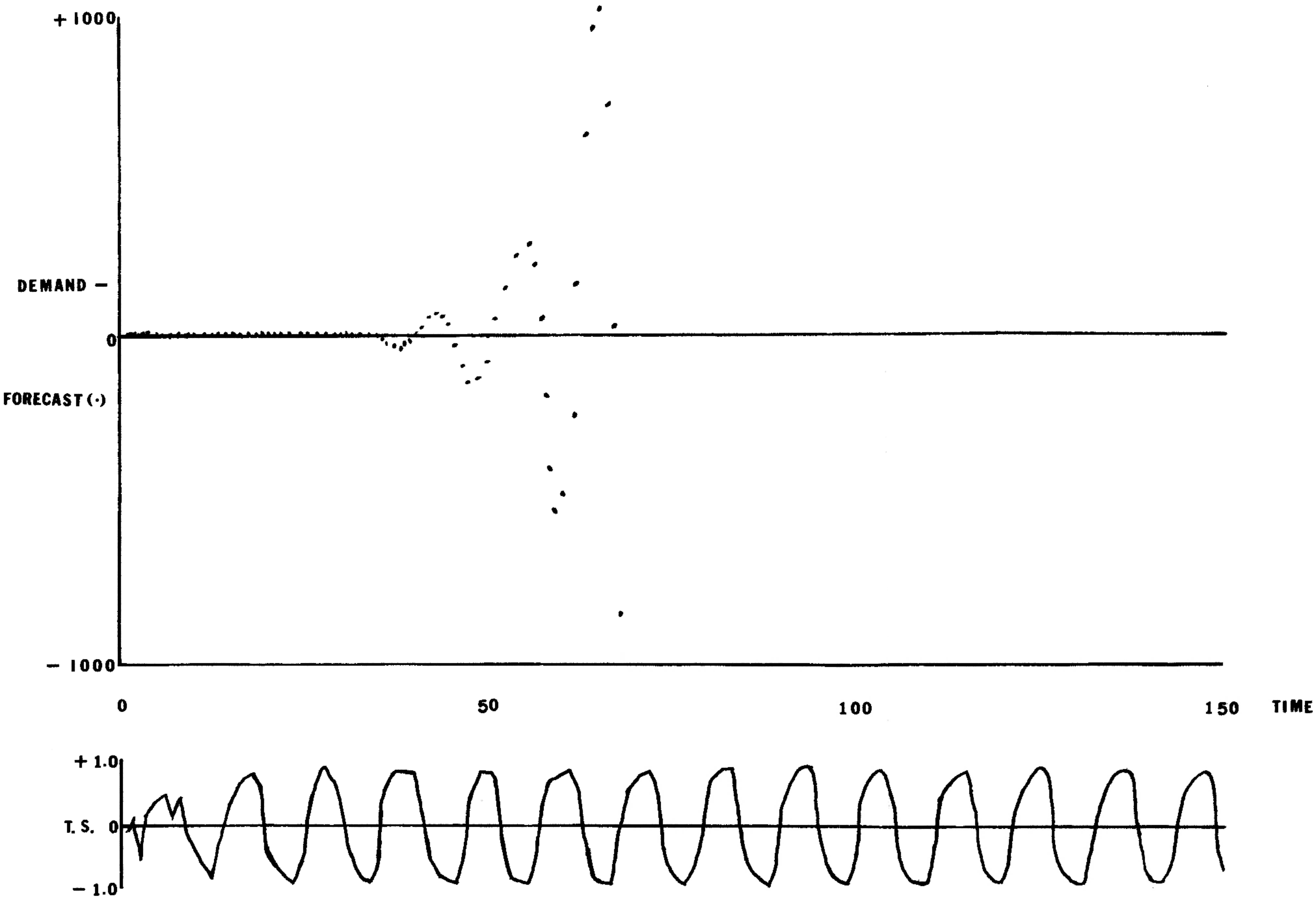


CHART 5.3

beta: .70

linear/quad

		BETA	
	.90	.80	.70
1. Sum of Square of Forecast Error	165.301	222.059	313.014
2. Variance of Forecasts	1.132	1.5210	2.144
3. Cumulative Error*	$\frac{.17}{-6.16}^{(-)}$	$\frac{2.30}{-3.07}^{(-)}$	$\frac{2.07}{-2.72}$
4. SMAD**	$\frac{.51}{1.10}$	$\frac{.49}{1.64}$	$\frac{.52}{2.16}$
5. Mean Error	-0.0174	-0.0076	- 0.0078
6. Variance of the Error	1.1017	1.4803	2.0870
7. Mean Absolute Error	.8161	.9816	1.1898

*Numerator indicates largest positive value; denominator indicates largest negative value; parenthesis indicates the predominate sign.

**Numerator indicates the smallest value, denominator indicates largest value.

TABLE 6.1

LINEAR DEMAND MODEL, LINEAR FORECAST MODEL

		BETA		
		.90	.80	.70
1.	Sum of Square of Forecast Error	166.505	222.201	313.075
2.	Variance of Forecasts	1.1404	1.5219	2.1444
3.	Cumulative Error*	$\frac{11.13}{-2.80}^{(+)}$	$\frac{4.67}{-2.34}^{(+)}$	$\frac{3.07}{-2.32}^{(+)}$
4.	SMAD**	$\frac{.52}{1.11}$	$\frac{.52}{1.64}$	$\frac{.50}{2.16}$
5.	Mean Error	0.06632	0.01475	0.00148
6.	Variance of Error	1.1056	1.4811	2.0871
7.	Mean Absolute Error	.8228	.9828	1.1903

*Numerator indicates largest positive value; denominator indicates largest negative value; parenthesis indicates the predominate sign.

**Numerator indicates the smallest value, denominator indicates largest value.

TABLE 7.1

QUADRATIC DEMAND MODEL, LINEAR FORECAST MODEL

	BETA		
	.95	.90	.85
1. Sum of Square of Forecast Error	1009.650	1480.192	74265940.
2. Variance of the Forecasts	6.963	10.208	512178.
3. Cumulative Error*	$\frac{293.58^{(+)}}{-.292}$	$\frac{274.54^{(+)}}{-3.53}$	$\frac{6566.}{6145}$
4. SMAD**	$\frac{.60}{3.93}$	$\frac{.63}{5.57}$	$\frac{.59}{2144.}$
5. Mean Error	1.957	1.830	43.777
6. Variance of the Error	2.900	6.518	493189.7
7. Mean Absolute Error	2.1471	2.3166	289.750

*Numerator indicates largest positive value; denominator indicates largest negative value; parenthesis indicates the predominate sign.

**Numerator indicates the smallest value, denominator indicates largest value.

TABLE 8.1

LINEAR DEMAND MODEL, QUADRATIC FORECAST MODEL

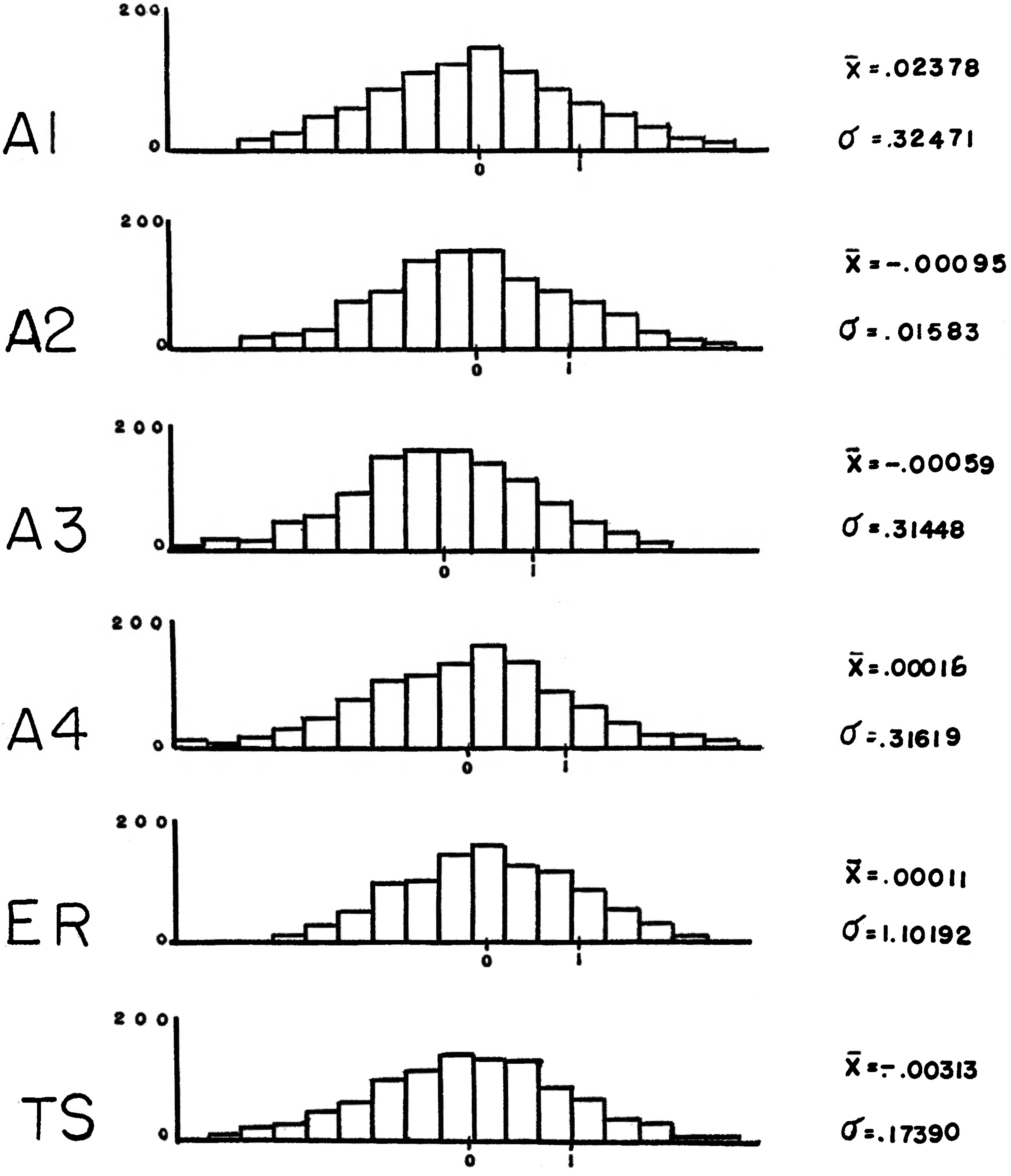

A1
200
0
0
1
$\bar{X}$=.02378
σ=.32471
A2
200
0
0
1
$\bar{X}$=-.00095
σ=.01583
A3
200
0
0
1
$\bar{X}$=-.00059
σ=.31448
A4
200
0
0
1
$\bar{X}$=.00016
σ=.31619
ER
200
0
0
1
$\bar{X}$=.00011
σ=1.10192
TS
200
0
0
1
$\bar{X}$=-.00313
σ=.17390

CHART 9.1

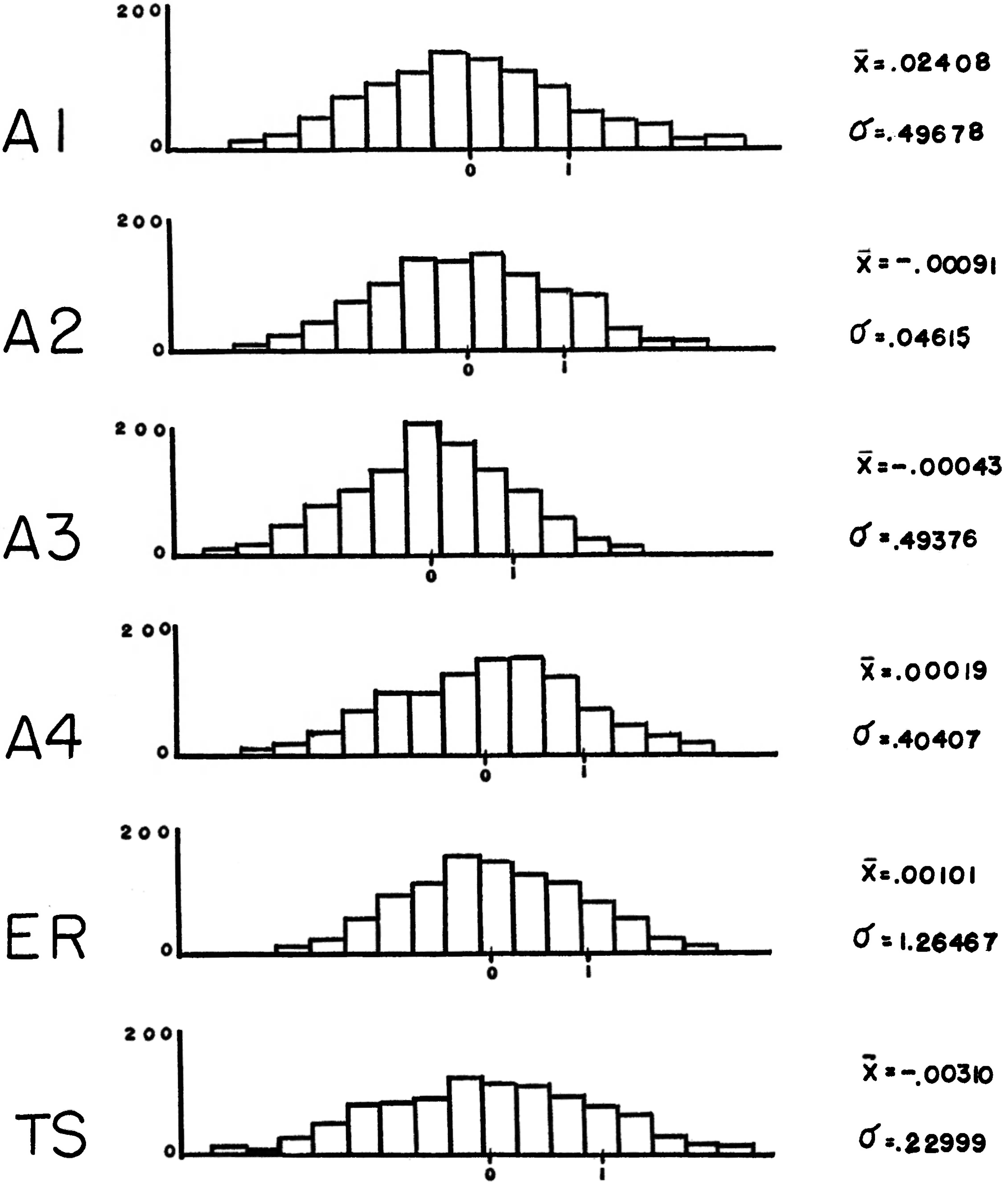

200
0
A1
0
1
$\bar{X}$=.02408
σ=.49678
200
0
A2
0
1
$\bar{X}$=-.00091
σ=.04615
200
0
A3
0
1
$\bar{X}$=-.00043
σ=.49376
200
0
A4
0
1
$\bar{X}$=.00019
σ=.40407
200
0
ER
0
1
$\bar{X}$=.00101
σ=1.26467
200
0
TS
0
1
$\bar{X}$=-.00310
σ=.22999

CHART 9.2

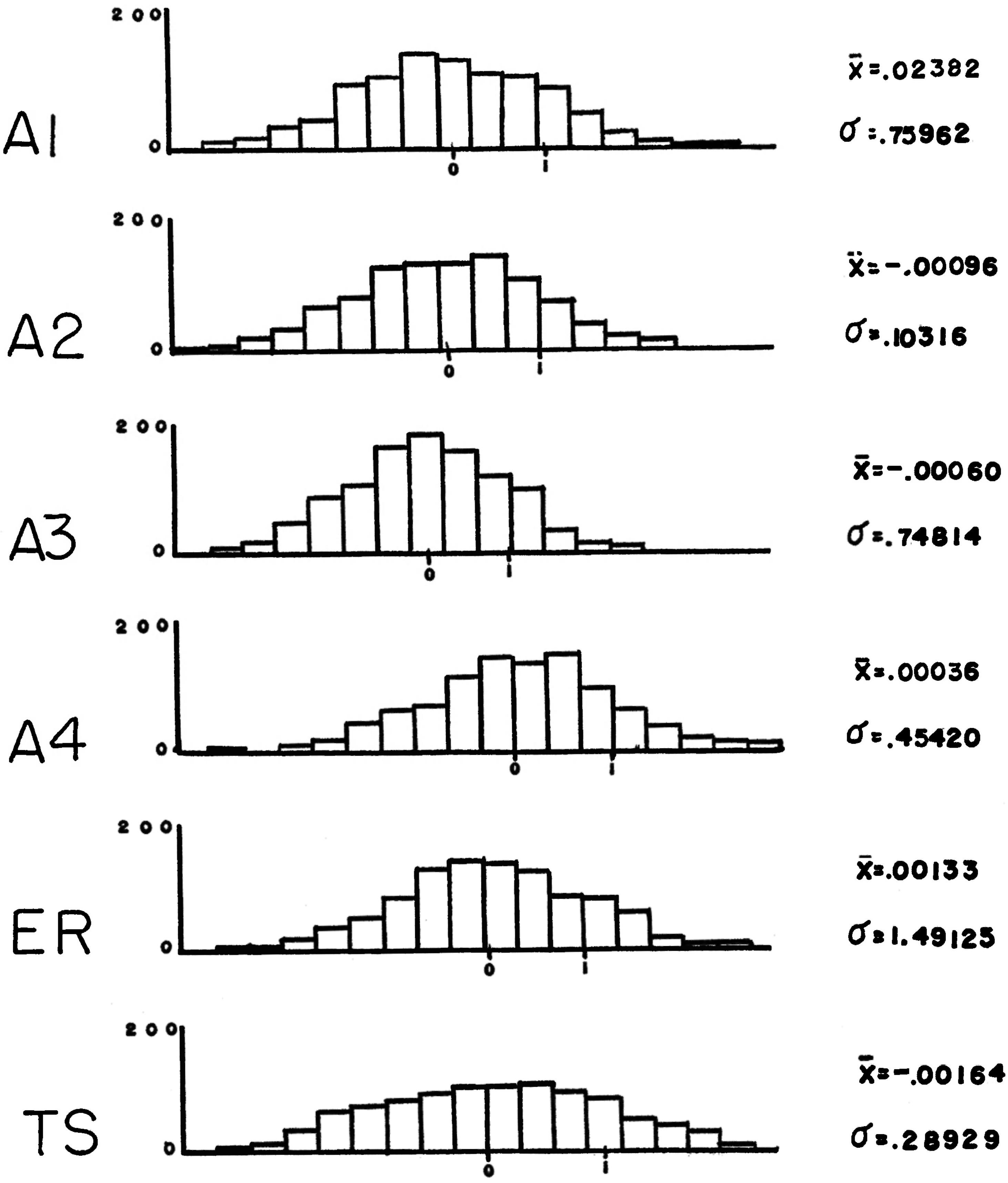

200
0
A1
0
1
x̄=.02382
σ=.75962
200
0
A2
0
1
x̄=-.00096
σ=.10316
200
0
A3
0
1
x̄=-.00060
σ=.74814
200
0
A4
0
1
x̄=.00036
σ=.45420
200
0
ER
0
1
x̄=.00133
σ=1.49125
200
0
TS
0
1
x̄=-.00164
σ=.28929

CHART 9.3

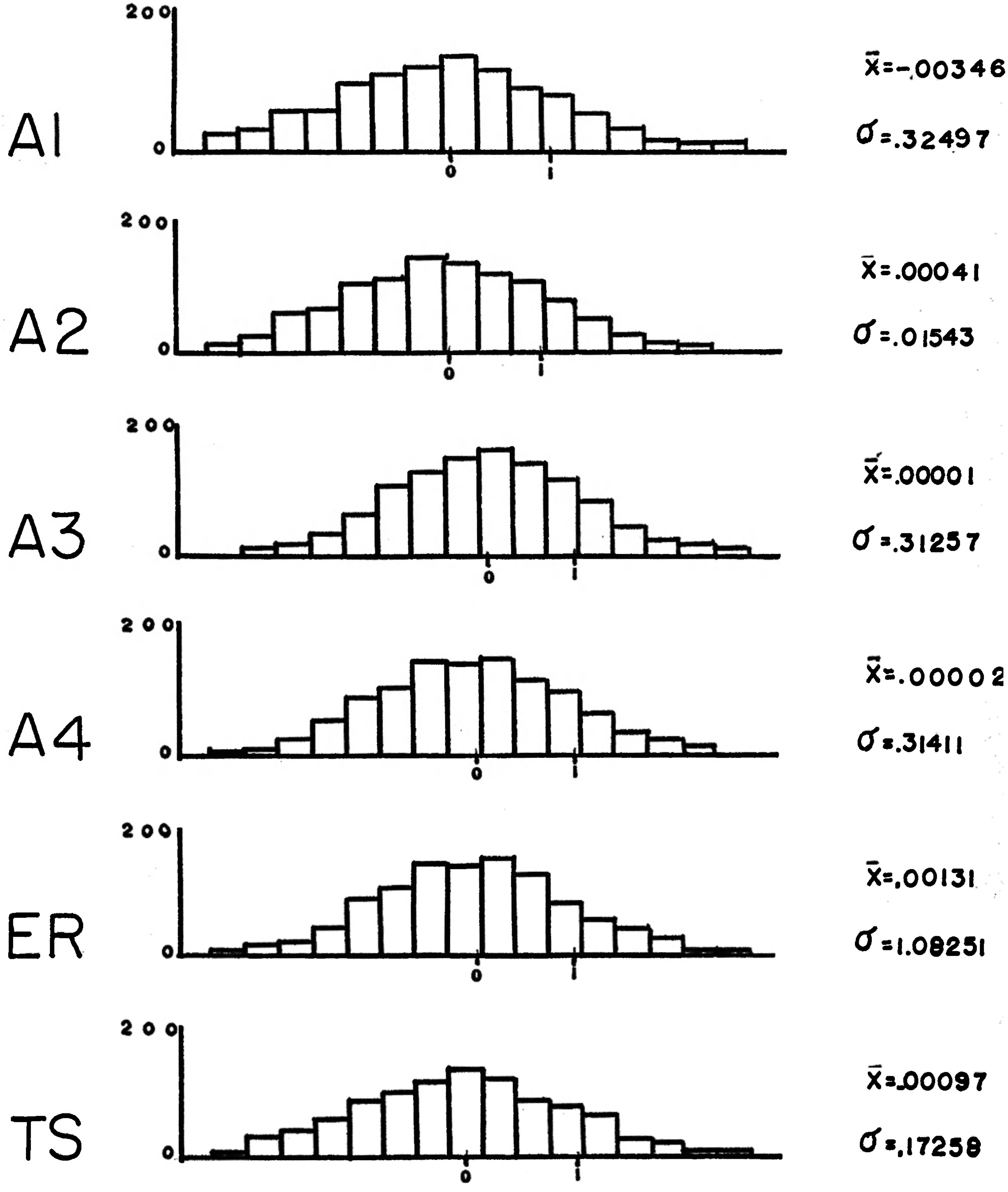

200
0
A1
0
1
x̄=-.00346
σ=.32497
200
0
A2
0
1
x̄=.00041
σ=.01543
200
0
A3
0
1
x̄=.00001
σ=.31257
200
0
A4
0
1
x̄=.00002
σ=.31411
200
0
ER
0
1
x̄=.00131
σ=1.08251
200
0
TS
0
1
x̄=.00097
σ=.17258

CHART 9.4

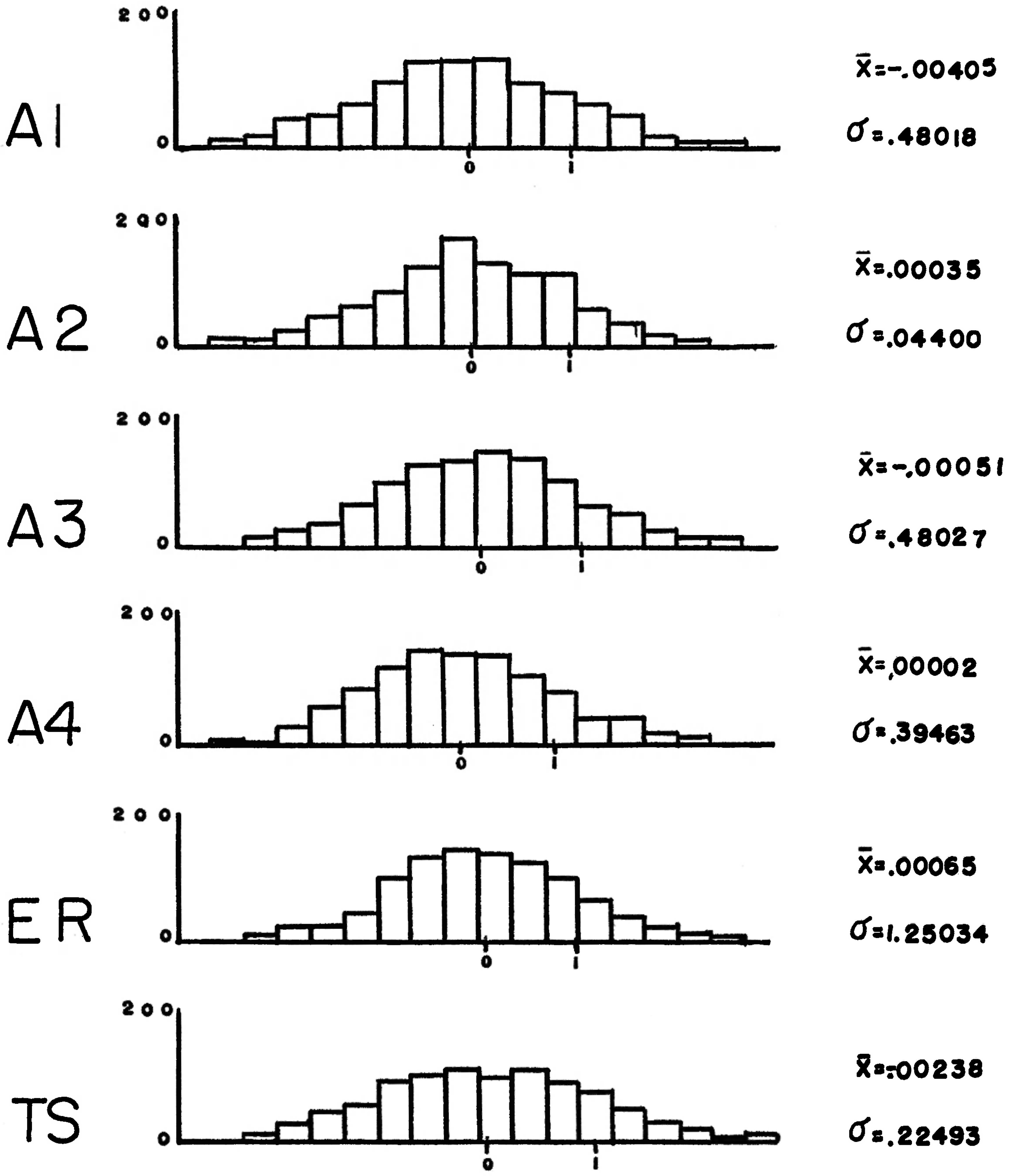

A1
200
0
0
1
$\bar{X}$=-.00405
σ=.48018
A2
200
0
0
1
$\bar{X}$=.00035
σ=.04400
A3
200
0
0
1
$\bar{X}$=-.00051
σ=.48027
A4
200
0
0
1
$\bar{X}$=,00002
σ=.39463
ER
200
0
0
1
$\bar{X}$=.00065
σ=1.25034
TS
200
0
0
1
$\bar{X}$=.00238
σ=.22493

CHART 9.5

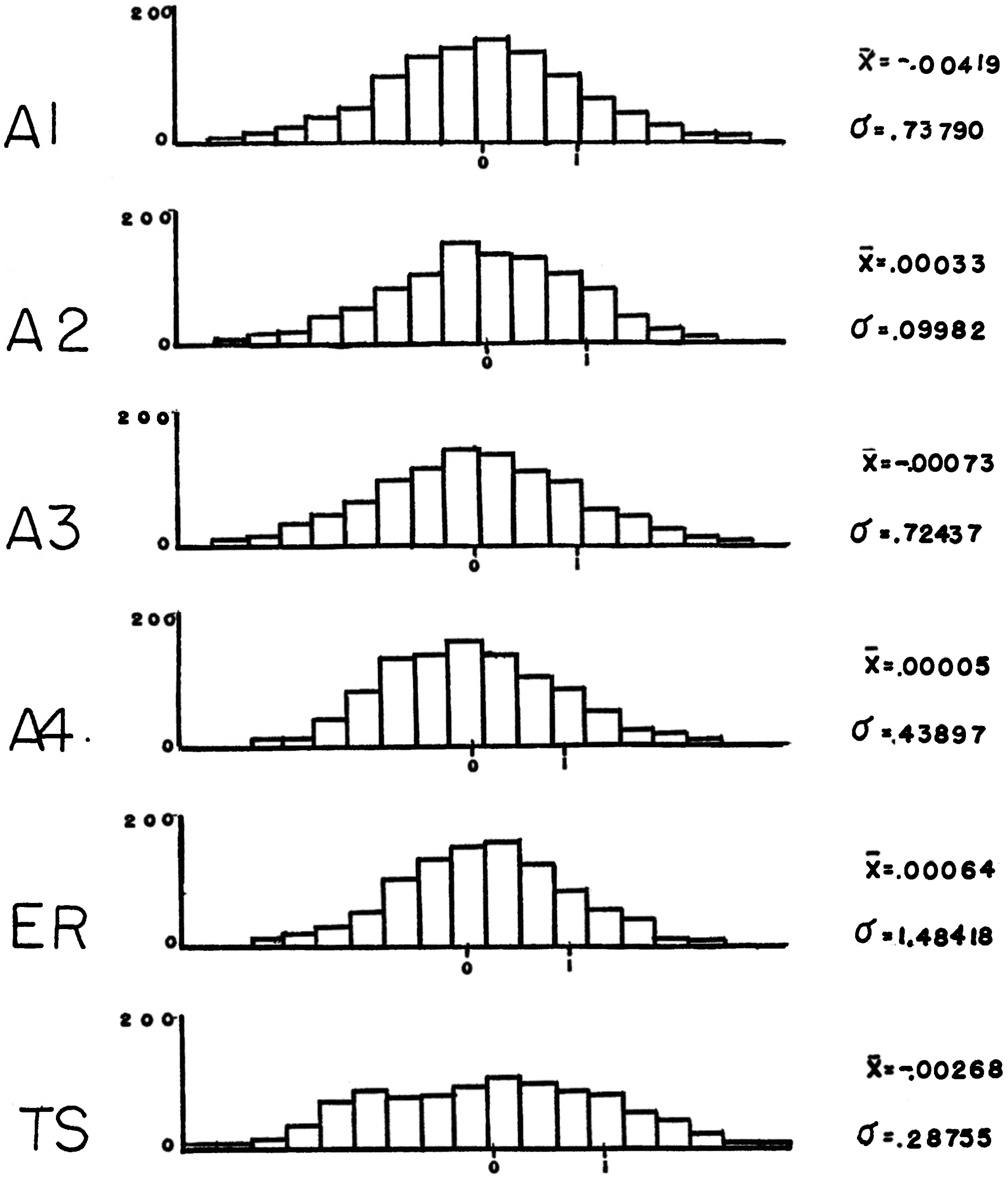

200
0
A1
0
1
X̄ = -.00419
σ = .73790
200
0
A2
0
1
X̄ = .00033
σ = .09982
200
0
A3
0
1
X̄ = -.00073
σ = .72437
200
0
A4
0
1
X̄ = .00005
σ = .43897
200
0
ER
0
1
X̄ = .00064
σ = 1.48418
200
0
TS
0
1
X̄ = -.00268
σ = .28755

CHART 9.6

Coef. Errors		CLASS INTERVALS																		df	χ^2	Sig. at .95	Sig. at .99
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18				
A_1	F_1	0	0	12	22	45	57	86	112	127	150	114	90	68	52	34	19	12	0	16	25.765	NO	NO
	F_2	0	0	18	20	37	40	87	112	128	133	122	102	74	50	29	16	7	5				
A_2	F_1	0	0	18	21	26	71	86	132	147	148	104	87	69	53	25	13	6	0	15	23.926	NO	NO
	F_2	0	0	14	20	37	63	93	121	140	141	125	99	67	41	22	10	7	0				
A_3	F_1	0	16	10	40	49	81	139	149	144	129	101	66	43	23	10	0	0	0	15	21.352	NO	NO
	F_2	0	11	16	33	59	92	123	144	148	130	101	79	40	20	9	5	0	0				
A_4	F_1	0	9	15	26	45	70	103	110	122	152	128	79	61	38	18	19	5	0	16	14.987	NO	NO
	F_2	0	10	13	26	45	71	98	121	135	132	116	91	63	39	22	11	7	0				
A_5	F_1	0	0	12	17	32	51	96	97	139	149	119	104	80	52	29	12	11	0	15	7.967	NO	NO
	F_2	0	0	12	17	32	55	84	112	133	136	133	107	77	50	28	14	9	0				
A_6	F_1	0	8	24	27	52	64	95	108	135	125	120	81	68	38	31	12	12	0	16	22.685	NO	NO
	F_2	0	13	15	28	48	70	96	117	118	127	112	91	65	42	25	13	11	0				

F_1 = observed frequency, F_2 = theoretical frequency

TABLE 10.1
BETA = .90
IX = 55003

Coef. Errors		CLASS INTERVALS 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	df	χ^2	Sig. at .95	Sig. at .99
A_1	F_1	0	11	11	21	44	74	93	113	140	133	116	90	56	46	33	8	11	0	16	12.766	NO	NO
	F_2	0	8	15	24	44	68	94	118	135	127	128	93	67	43	24	13	9	0				
A_2	F_1	0	8	9	24	42	73	98	134	131	140	112	88	80	32	15	8	6	0	16	4.321	NO	NO
	F_2	0	8	11	24	43	70	100	126	139	137	119	91	61	37	19	9	6	0				
A_3	F_1	0	6	16	42	72	92	121	193	163	124	90	53	19	9	0	0	0	0	13	15.363	NO	NO
	F_2	0	7	15	34	66	108	146	168	160	128	85	48	22	13	0	0	0	0				
A_4	F_1	0	8	10	18	32	67	91	90	119	139	142	116	67	45	27	17	6	6	17	15.793	NO	NO
	F_2	0	7	9	20	35	58	86	111	129	135	124	103	76	50	30	15	7	5				
A_5	F_1	0	0	8	16	29	52	88	108	148	144	120	108	78	58	25	12	6	0	15	6.030	NO	NO
	F_2	0	0	10	15	29	54	83	114	137	145	133	108	77	88	26	13	8	0				
A_6	F_1	0	18	10	29	54	83	83	89	118	110	109	90	75	65	29	20	18	0	17	30.559	YES	NO
	F_2	0	16	17	29	46	66	87	106	117	119	109	93	72	51	33	19	11	9				

F_1 = observed frequency, F_2 = theoretical frequency

TABLE 10.2
BETA = .80
IX = 55003

		CLASS INTERVALS																				Sig.	Sig.
Coef. Errors		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	df	χ^2	at .95	at .99
A_1	F_1	0	13	15	33	45	99	105	140	133	112	112	91	52	26	12	7	5	0	16	12.545	NO	NO
	F_2	0	12	17	31	54	83	111	133	139	131	107	78	50	29	15	6	4	0				
A_2	F_1	6	7	19	33	66	82	122	130	129	138	108	77	40	24	13	6	0	0	17	8.312	NO	NO
	F_2	5	8	18	34	59	88	117	138	141	129	102	73	44	25	11	5	3	0				
A_3	F_1	0	8	17	47	71	102	158	175	152	118	96	33	16	7	0	0	0	0	13	11.456	NO	NO
	F_2	0	9	17	39	72	117	153	170	156	121	77	41	18	10	0	0	0	0				
A_4	F_1	0	7	2	12	20	46	66	70	115	141	132	147	96	64	38	20	14	10	17	16.763	NO	NO
	F_2	0	3	5	11	22	40	65	93	118	134	135	121	96	69	43	24	12	9				
A_5	F_1	0	5	12	26	43	56	82	127	138	135	121	83	81	59	18	5	9	0	16	21.820	NO	NO
	F_2	0	7	10	21	39	64	92	118	135	137	122	98	69	43	25	12	8	0				
A_6	F_1	0	8	21	41	70	77	83	95	100	100	100	92	81	51	41	30	10	0	17	44.449	YES	YES
	F_2	0	19	19	32	48	68	89	106	115	116	106	90	70	50	32	20	10	10				

F_1 = observed frequency, F_2 = theoretical frequency

TABLE 10.3
BETA = .70
IX = 55003

Coef. Errors		CLASS INTERVALS																			χ^2	Sig. at .95	Sig. at .99
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	df			
A_1	F_1	0	23	26	53	53	91	104	118	135	119	86	83	50	29	12	7	11	0	16	11.589	NO	NO
	F_2	0	24	24	52	63	88	109	124	127	116	97	74	49	31	17	9	6	0	16			
A_2	F_1	9	9	17	58	59	95	103	139	130	113	103	75	49	22	11	8	0	0	16	15.938	NO	NO
	F_2	8	11	23	40	65	93	118	132	135	120	97	69	43	24	13	9	0	0	16			
A_3	F_1	0	0	10	13	27	54	99	116	137	156	131	110	75	38	20	9	5	0	15	4.168	NO	NO
	F_2	0	0	8	14	29	54	87	121	146	152	137	105	73	42	20	9	5	0	15			
A_4	F_1	0	7	3	19	48	81	98	138	134	142	108	94	60	30	24	14	0	0	15	15.376	NO	NO
	F_2	0	6	11	22	41	69	101	128	143	141	120	92	60	35	18	12	0	0	15			
A_5	F_1	0	14	16	21	41	75	98	138	131	142	118	73	52	38	28	8	7	0	16	15.873	NO	NO
	F_2	0	10	14	27	47	75	100	124	137	131	115	88	60	36	20	10	6	0	16			
A_6	F_1	7	13	35	38	57	86	96	113	129	114	86	77	60	29	26	15	13	6	18	18.549	NO	NO
	F_2	13	14	25	39	59	80	100	113	118	113	99	79	58	39	25	13	7	6	18			

F_1 = observed frequency, F_2 = theoretical frequency

TABLE 10.4
BETA = .90
IX = 54995

Coef. Errors		CLASS INTERVALS																					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	df	χ^2	Sig. at .95	Sig. at .99
A_1	F_1	0	11	14	39	47	64	95	128	130	133	97	85	65	49	21	12	10	0	16	10.896	NO	NO
	F_2	0	13	16	28	48	72	98	118	130	127	111	89	64	41	23	13	9	0				
A_2	F_1	0	10	8	25	47	62	82	120	166	128	115	105	56	38	21	17	0	0	16	26.686	YES	NO
	F_2	0	7	12	21	42	68	97	124	138	137	120	93	63	39	21	10	6	0				
A_3	F_1	0	0	14	23	33	60	93	119	129	140	136	101	60	50	22	10	10	0	15	8.665	NO	NO
	F_2	0	0	15	20	37	61	91	118	135	138	125	100	71	44	25	12	8	0				
A_4	F_1	0	8	6	27	55	85	118	143	134	136	109	78	36	40	14	11	0	0	15	16.030	NO	NO
	F_2	0	9	13	28	51	80	113	134	146	135	111	80	50	28	13	9	0	0				
A_5	F_1	0	0	15	24	25	48	102	132	140	140	124	100	68	40	22	11	9	0	15	10.324	NO	NO
	F_2	0	0	13	19	35	60	92	119	140	141	128	101	69	43	22	11	7	0				
A_6	F_1	0	8	14	30	51	60	99	105	113	103	113	93	80	58	33	21	9	10	17	10.331	NO	NO
	F_2	0	13	15	26	43	63	86	106	118	121	113	96	74	53	34	20	10	9				

F_1 = observed frequency, F_2 = theoretical frequency

TABLE 10.5
BETA = .80
IX = 54995

		CLASS INTERVALS																					
Coef. Errors		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	df	χ^2	Sig. at .95	Sig. at .99
A_1	F_1	0	0	11	20	34	57	76	131	134	151	131	103	65	43	23	13	8	0	15	5.878	NO	NO
	F_2	0	0	12	17	33	58	89	118	139	143	130	103	72	44	24	11	7	0				
A_2	F_1	0	6	12	15	41	53	83	104	151	133	130	107	79	39	25	10	12	0	16	10.448	NO	NO
	F_2	0	5	9	18	36	59	87	116	135	139	127	102	73	47	25	13	9	0				
A_3	F_1	0	9	14	33	51	66	95	116	142	136	109	99	53	38	21	10	8	0	16	7.675	NO	NO
	F_2	0	11	14	27	48	73	100	123	134	130	114	89	61	38	21	11	7	0				
A_4	F_1	0	0	10	14	42	78	132	133	159	138	101	88	54	26	15	10	0	0	14	16.476	NO	NO
	F_2	0	0	14	16	41	73	108	138	154	147	120	85	52	28	12	7	0	0				
A_5	F_1	0	0	15	22	32	55	104	130	153	155	127	87	53	39	16	12	0	0	14	9.927	NO	NO
	F_2	0	0	12	19	49	53	100	130	149	147	126	94	60	34	16	11	0	0				
A_6	F_1	6	8	16	34	72	89	87	85	95	105	96	82	83	55	46	25	9	7	18	38.850	YES	YES
	F_2	11	11	20	32	49	68	88	103	112	113	105	89	70	51	34	21	12	11				

F_1 = observed frequency, F_2 = theoretical frequency

TABLE 10.6
BETA = .70
IX = 54995

APPENDIX D. COMPUTER PROGRAMS

1. MAIN PROGRAM - Generates 150 demand data points. D.1
2. REG - Routine runs a regression analysis on the updated vector of fitting functions for 25 time intervals. D.2
3. EXPOS - Routine sets up following routines to do the actual forecasting. D.3
4. RAY - Routine computes the F - matrix for a given Beta. D.4
5. MATINV - Routine inverts F matrix. D.5
6. HVEC - Routine computes the h vector. D.6
7. FORCST - Routine forecasts 150 time intervals and computes the statistics. D.7
8. UPDATE - Routine updates vector of fitting functions of the forecasting model 25 time intervals. D.8
9. RAND & NORM - Routines generate normally distributed random variables. D.9

```
          RND,RUN=CHECK,TIME=12,LINES=60,PAGES=50
   N=NUMBER OF GENERATED POINTS
   A = VECTOR OF CONSTANTS
   PF = VECTOR OF FITTING FUNCTIONS FOR GENERATING MODEL
   CHK = CHANGE VECTOR
   LL = TRANSITION MATRIX FOR GENERATING MODEL
   FP = VECTOR OF FITTING FUNCTIONS FOR FORECASTING MODEL
   LI = TRANSITION MATRIX FOR FORECASTING MODEL
   MN = MODEL NUMBER
   NTOU = PERIOD OF PERIODIC COMPONENTS
   NC = NUMBER OF COMPONENTS IN GENERATION MODEL
   NV = NUMBER OF COMPONENTS IN FORECASTING MODEL
   IX = INITIAL VALUE OF RANDOM NUMBDR GENERATOR
   SF = SCALE FACTOR
   X(I) = ACTUAL GENERATED VALUES
   V(I) = COEFFICIENTS - A VECTOR OF CONSTANTS- DETERMINED FROM REG
  DIMENSION A(20), APF(1), AL(10), L1(100), LL(10,10)
  DIMENSION X(150), PLT(600), RESID(150), NN(10), IY(150)
  DIMENSION CHK(10), PF(20,1), P1(20), FCST(150), PLS(750)
  DIMENSION XX(800),Y(200),XPX(100),YPY(1),XPY(40),B(10),JWV1(10)
  DIMENSION JWV2(10), BPXPY(1), COF(100)
  DIMENSION LI(I0,10), L2(100), FP(20,1)
  DIMENSION TLS(600)
  DIMENSION V(20)
  DIMENSION A2(10,1)
  DIMENSION A0(20,1)
  DIMENSION A1(20)
  COMMON IC
  REAL NORM, LL, L1, LI
  INTEGER T, F, TAU
1 FORMAT (I4,I5,I2,F5.2,I2,I2,F5.3)
2 FORMAT (20F4.2)
3 FORMAT (2F2.0)
4 FORMAT (10F10.5)
5 FORMAT ('0', ' VALUES OF X(T) '//)
6 FORMAT ('0', 10F10.5)
7 FORMAT (' MEAN OF THE ERROR TREMS=',F10.5/)
8 FORMAT (' STANDARD DEVIATION OF THE ERROR =',F10.5/)
9 FORMAT (' ERROR OF THE GENERATED X VALUES'/)
10 FORMAT ('0A VECTOR ',10F10.5)
11 FORMAT ('0F VECTOR ',10F10.5)
12 FORMAT (10F7.4)
14 FORMAT ('0FINAL VALUE OF GENERATOR=',I12/)
15 FORMAT ('1','MODEL # =',I2/' ','NUMBER OF PARAMETERS =',I2/' ','IN
  1ITIAL VALUE OF THE GENERATOR =',I5/' ',' # OF POINTS GENERATED =',
  2I4/' ',' SCALE FACTOR =',F5.2/' ',' TAU =',I2)
16 FORMAT ('0',15I5,1X,2I2)
17 FORMAT (15I5,1X,2I2)
21 FORMAT (2I3,F5.3)
24 FORMAT (' TRANSITION MATRIX FOR GENERATION MODEL')
25 FORMAT (' INITIAL VALUES OF GENERATION MODEL'/)
26 FORMAT (IH ,10FI2.7)
32 FORMAT (F8.4,I2,I2)
37 FORMAT (' MEAN ABS DEVIATE = ',F8.4/' # TERMS IN FORECASTING MODEL
  1=',I5)
45 FORMAT (' FP VECTOR',10F10.5)
46 FORMAT (' TRANSITION MATRIX FOR FORECASTING MODEL')
56 FORMAT (' MEAN OF GEN DIST=',F4.2/' VAR OF GEN DIST=',F6.4)
   M=20
13 READ (1,1) N, IX, MN,SF, NC, TAU, BBETA
```

```
   IF (N) 55, 55, 18
18 READ (1,32) D, NV, NM
   READ (1,I2) (A(I),I=1,NC)
   READ (1,2) (PF(I,1), I=1,NC)
   READ (1,2) (CHK(I),1=1,NV)
   READ (1,3) EBAR,SIGMAE
   DO 33 I=1,NC
33 READ (1,4) (LL(I,J),J=1,NC)
   READ (1,21) NTOU, NL, ABETA
   DO 34 I=1,NV
34 READ (1,4)(LI(I,J),J=1,NV)
   READ (1,2) (FP(I,1),1=1,NV)
   WRITE (3,15) MN, NC, IX, N, SF, TAU
   WRITE (3,56) EBAR, SIGMAE
   WRITE (3,25)
   WRITE (3,10) (A(I),I=1,NC)
   WRITE (3,11) (PF(I,1),I=1,NC)
   WRITE (3,24)
   DO 27 I=1,NC
27 WRITE(3,26) (LL(I,J),J=1,NC)
   WRITE (3,46)
   DO 28 I=1,NV
28 WRITE (3,26) (LI(I,J),J=1,NV)
   WRITE (3,45) (FP(I,1),I=1,NV)
   WRITE (3,37) D, NV
   IC=0
   ERRSUM = 0.0
   ERRSQ = 0.0
   IM = IX
   N1=NM+1
   DO 98 I=1,NC
98 A1(I) = A(I)
   DO 30 I=1,NV
30 AO(I,1) = FP(I,1)
   CALL ARRAY (2,NC,NC,10,10,L1,LL)
   ID = 1
   IE = NM
43 DO 100 I=ID,IE
   CALL GTPRD (A,L1,AL,NC,1,NC)
   DO 20 J=1,NC
   A(J) = AL(J)
20 CONTINUE
   CALL ARRAY (2,NC,1,M,1,P1,PF)
   CALL GMPRD (AL,P1,APF,1,NC,1)
    GENERATE ACTUAL PROCESS VALUES
   X(I) = APF(1) + NORM(EBAR,SIGMAE,IX)
   ERRSUM = ERRSUM + (X(I) - APF(1))
   ERRSQ = ERRSQ + (X(I) - APF(1))**2
00 CONTINUE
   IF(I-N) 41,42,42
41 CONTINUE
   DO 60 I=ID,IE
60 Y(I) = X(I)
    DETERMINE VALUES OF COEF FOR FORECASTING MODEL FROM REGRESSION
   CALL REG(NM,FP,X,Y,B,NV,LI,XX,SMAD)
   DO 31 I=1,NV
31 V(I) = B(I)
   ID = 1
   IE = N
   DO 99 I=1,NC
```

```
99 A(I) = A1(I)
   GO TO 43
42 ERRBAR = ERRSUM/N
   ERRVAR = ERRSQ/N - ERRBAR**2
   ERRSIG = SQRT(ERRVAR)
   WRITE (3,9)
   WRITE (3,7) ERRBAR
   WRITE (3,8) ERRSIG
   WRITE (3,14) IX
    CALL ROUTINES TO FORECAST PROCESS
   CALL EXPOS(N,NL,NV,MN,NTOU,V,FP,CHK,LI,X,ABETA,FCST,PLS,TAU,TLS,BB
  1ETA,XX,D,S,SMAD)
   GO TO 13
55 STOP
   END
```

```
      SUBROUTINE REG(NM,AO,X,Y,B,NC,LL,XX,SMAD)
       ROUTINE DOES A REGRESSION OF XX  ON Y
      DIMENSION XX(200),Y(100),XPX(100),YPY(1),XPY(40),B(10),JWV1(10)
      DIMENSION JWV2(10),BPXPY(1),COF(100),AO(1,1),LL(1,1),X(1)
      REAL LL
   33 FORMAT(1H0/24X,20HANALYSIS OF VARIANCE/)
   34 FORMAT(1H0,5X,19HSOURCE OF VARIATION,7X,9HDEGREE OF,7X,6HSUM OF,9X
     1,4HMEAN,10X,1HF/33X,7HFREEDOM,8X,7HSQUARES,7X,6HSQUARE,7X,5HVALUE,
     24X,'CORR COEF')
   35 FORMAT (20H0   DUE TO BO         ,12X,I6,F17.5)
   36 FORMAT (32H    DUE TO RESIDUAL              ,I6,F17.5,FI4.5/)
   37 FORMAT (20H0   DUE TO REG|BO     ,12X,I6,F17.5,F14.5,F13.5,F14.5)
   39 FORMAT ('    TOTAL UNCORRECTED SUM SQ    ',I6,F17.5/)
   44 FORMAT (100H   B(0)       B(1)       B(2)       B(3)       B(4)       B
     1(5)       B(6)       B(7)       B(8)       B(9)    )
   45 FORMAT (100A1)
   46 FORMAT ('0VALUES OF COEF DETERMINED FROM REGRESSION ANAL'/100A1)
   47 FORMAT (1H ,10F10.5/)
      NO = NC*10
      CALL UPDATE(AO,XX,LL,NM,X,NC)
      CALL GTPRD(XX,XX,XPX,NM,NC,NC)
      CALL GTPRD(Y,Y,YPY,NM,1,1)
      CALL GTPRD(XX,Y,XPY,NM,NC,1)
      CALL MINV2(XPX,NC,DET,JWV1,JWV2)
      CALL GTPRD(XPX,XPY,B,NC,NC,1)
      CALL GTPRD(B,XPY,BPXPY,NC,1,1)
      SUMY = 0.0
      YSUM = 0.0
      DO 101 K=1,NM
      SUMY = SUMY + (Y(K))**2
  101 YSUM = YSUM + Y(K)
      CFAC = (YSUM**2)/NM
      CSSDR = BPXPY(1) - CFAC
      RGMS = CSSDR/(NC-2)
      CSSRL = YPY(1) - BPXPY(1)
      RLMS = CSSRL/(NM-NC)
      SMAD = .8*SQRT(RLMS)
      JRG = NC-1
      JRB = 1
      JRL = NM-NC
      FVAL = RGMS/RLMS
      RSQ = CSSDR/(YPY(1)-CFAC)
      WRITE (3,33)
      WRITE (3,34)
      WRITE (3,35) JRB,CFAC
      WRITE (3,37) JRG,CSSDR,RGMS,FVAL,RSQ
      WRITE (3,36) JRL,CSSRL,RLMS
      WRITE (3,39) NM, SUMY
      REWIND 4
      WRITE(4,44)
      REWIND 4
      READ(4,45) (COF(I),I=1,NO)
      REWIND 4
      WRITE(3,46) (COF(I),I=1,NO)
      WRITE (3,47) (B(I),I=1,NC)
      RETURN
      END
```

ROUTINE REG

D.2

```
      SUBROUTINE EXPOS(ND,NL,N,MN,NTOU,COI,AO,CHK,TM,X,ABETA,FCST,PLS,TA
     1U,TLS,BBETA,XX,D,S,SMAD)
       ROUTINE CALLS MINOR ROUTINES TO DO THE ACTUAL FORECASTING
      DIMENSION TLS(600), A3(10,1)
      DIMENSION FCST(150),PLS(750)
      DIMENSION A1(10,1), FK(10,10), F(10,10), H(10,1)
      DIMENSION COI(20), AO(20,1), X(150)
      DIMENSION  CI(10),A(10,1),TM(10,10)
      DIMENSION  ER(200), PCTER(200)
      DIMENSION CHK(10)
      DIMENSION FL(10,10), FF(10,10), A11(10,1), HH(10,1)
      DIMENSION XX(1)
      INTEGER TAU
      DOUBLE PRECISION F,FK,FL,FF
    5 FORMAT ('OTRANSITION MATRIX FOR FORECASTING MODEL')
    6 FORMAT (1H ,10F12.7)
    7 FORMAT (' ABETA=',F6.3/' OBETA=',F6.3)
    8 FORMAT (' BBETA=',F6.3/' PBETA=',F6.3/)
      NNN = 1
      TOU = NTOU
      PI = 3.1415927
SAVE THE INITIAL VALUES OF THE CONSTANTS
          SAVE THE INITIAL VALUES OF THE FITTING FUNCTIONS
11 DO 16  I=1,N
16 CI(I) = COI(I)
      CALL RAY (ND,NTOU,N,A,X,CI,MN,CHK,TM,OBETA,FK,F,AO,A1,ABETA,COI)
      CALLMATINV(F,N)
      CALL HVEC(F,AO,N,H,OBETA,ABETA)
      CALLVMARIX(FK,F,N)
      CALL RAY(ND,NTOU,N,A,X,CI,MN,CHK,TM,PBETA,FL,FF,AO,A11,BBETA,COI)
      CALL MATINV(FF,N)
      CALL HVEC(FF,AO,N,HH,PBETA,BBETA)
***FORM THE V MATRIX
      CALL VMARIX (FL,FF,N)
      CALL FORCST(N,ND,CI,A1,X,MN,TM,F,H,FCST,PLS,TLS,HH,ABETA,BBETA,D,S
     1,SMAD)
      CALL PLOT(1,TLS,ND,4,ND,0)
      NNN=NNN+1
12 CONTINUE
      RETURN
      END
```

ROUTINE EXPOS

D.3

```
    SUBROUTINE RAY (ND,NTOU,N,A,X,CI,MN,CHK,TM,OBETA,FK,F,AO,A1,ABETA,
   1COI)
     THIS ROUTINE COMPUTES THE F MATRIX FOR A GIVEN BETA
    DIMENSIONAA(10,1)
    DIMENSIONCHK(I0)
    DIMENSION FK(10,10),C11(10,10)
    DIMENSION A(10,1),B(1,10),C1(10,10),F(10,10),COI(10)
    DIMENSION X(200),CI(10),A1(10,1),H(10,1),AO(10,1),TM(10,10)
    DIMENSION AF(10,1)
    DOUBLE PRECISION F,FK,C1,C11
  1 FORMAT (/' F-MATRIX',10X,'BETA=',F5.3,10X,'EFFECTIVE BETA=',F5.3)
  2 FORMAT (1H ,10D13.5)
    TN=N
    AP=(1./TN)
STARTING HERE THE F MATRIX IS RECALCULATED WITH A NEW BETA
 67 OBETA=(ABETA)**AP
          REINITIALIZE THE VECTOR OF FITTING FUNCTIONS AND CONSTANTS
    DO42I=1,N
    A(I,1) = AO(I,1)
 42 AA(I,1)=AO(I,1)
COMPUTE THE F MATRIX FOR ONE MODEL AND ONE BETA
    DO8I=1,N
    DO8J=1,N
    FK(I,J)=0.0
  8 F(I,J)=0.0
    BETA=1.0
    CNT=1.
    K=1
 99 DO30I=1,N
 30 B(1,I)=A(I,1)
    DO40I=1,N
    M=I
    DO40J=M,N
    C1(I,J)=A(I,1)*B(1,J)*(BETA)
 40 CONTINUE
    DO43I=1,N
    M=I
    DO43J=M,N
 43 C11(I,J)=C1(1,J)*BETA
    IF(K-1)66,16,66
 ***CHECK THE CONVERGENCE OF THE F MATRIX AT EACH 50TH ITERATION
 ***THIS ROUTINE CHECKS UNTIL IT FINDS AN ELEMENT WHICH IS NOT STABIL-
 ***IZED
 66 IF(CNT-50.)16,44,16
 44 CNT = 0.0
    DO 80  I=1,N
    M=I
    DO80J=M,N
    DENM=F(I,J)
    ANUM=C1(I,J)
    RATIO=ANUM/DENM
    IF(RATIO-.000001)80,80,16
 80 CONTINUE
 ***NORMAL EXIT FROM THIS LOOP OCCURS WHEN ALL OF THE ELEMENTS OF
 ***THE F MATRIX MEET THE CONVERGENCE CRITERIA WHEN THIS OCCURS THE
 *** F MATRIX IS COMPLETED BY SYMMETRY AND THE PROGRAM CHECKS
 ***TO SEE IF THE RANGE OF BETA HAS BEEN COVERED
    DO87I=2,N
    L=I-1
    DO87J=1,L
```

```
      FK(I,J)=FK(J,I)
 87   F(I,J)=F(J,I)
      WRITE (3,1) ABETA,OBETA
      DO 88  I=1,N
 88   WRITE(3,2)  (F(I,J), J=1,N)
      GO TO 45
****THE PROGRAM BRANCHES TO HERE WHEN THE CONVERGENCE
****CRITERIA IS NOT MET AND NORMAL ITERATION CONTINUES
 16   DO50I=1,N
      M=I
      DO50J=M,N
      FK(I,J)=FK(I,J)+C11(I,J)
 50   F(I,J)=F(I,J)+C1(I,J)
****NOW UPDATE THE VECTOR OF FITTING FUNCTIONS USING THE L MATRIX
      DO31I=1,N
 31   AF(I,1)=0.0
      DO90I=1,N
      DO90J=1,N
 90   AF(I,1)=TM(I,J)*AA(J,1)+AF(I,1)
      DO100I=1,N
      AA(I,1)=AF(I,1)
100   A(I,1)=AF(I,1)
****THIS IS A ROUTINE FOR CONVERTING THE VECTOR OF FITTING FUNCTIONS
****FOR T TO THE VECTOR EVALUATED FOR -T
      DO63I=1,N
      CK=CHK(I)
      IF(CK-1.)63,55,63
 55   A(I,1)=-A(I,1)
 63   CONTINUE
SAVE F(1) FOR FORECAST
      IF(K-1)62,22,62
 22   DO 61 I=1,N
 61   A1(I,1)=AA(I,1)
 62   CONTINUE
      BETA = ABETA*BETA
      CNT=CNT+1.
      K=K+1
      GO TO 99
 45   CONTINUE
      RETURN
      END
```

```
      SUBROUTINEMATINV(C,N)
C       THIS ROUTINE INVERTS THE F MATRIX
      DIMENSION C(10,10),V(10),F(10,10)
      DOUBLE PRECISION F,FK,C1,C11,C
      NM1=N-1
      DO 6660 L=1,N
      ALL=C(1,1)
      DO 4440 J=1,NM1
 4440 V(J)=C(1,J+1)/ALL
      V(N)=1./ALL
      DO 9990 I=1,NM1
      IP1=I+1
      CIP11=C(IP1,1)
      DO 5550 J=1,NM1
 5550 C(I,J)=C(IP1,J+1) - CIP11*V(J)
 9990 C(I,N)=(-1.0)*CIP11*V(N)
      DO 6660 J=1,N
 6660 C(N,J)=V(J)
      RETURN
      END
```

ROUTINE MATINV

D.5

```
      SUBROUTINE HVEC(F,AO,N,H,OBETA,ABETA)
       ROUTINE COMPUTES THE H VECTOR = F MATRIX INVERSE TIMES FITTING
      DIMENSION F(10,10),AO(10,1),H(10,1),HF(10,1)
      DOUBLE PRECISION F,FK,C1,C11
  666 FORMAT (' F(0)     ',10F11.6)
  727 FORMAT ('OH VECTOR',10F11.6//)
  789 FORMAT (' ',4F14.6)
  444 FORMAT (1H ,10D13.5)
  888 FORMAT(' F-MATRIX INVERSE  ','BETA=',F5.3,10X,'EFFECTIVE BETA=',F5
     1.3)
      DO32I=1,N
   32 H(I,1)=0.0
      DO543I=1,N
      DO543J=1,N
  543 H(I,1)=H(I,1)+ F(I,J)*AO(J,1)
  626 CONTINUE
  ****WRITE THE EFFECTIVE BETA
  ****WRITE F INVERSE
      WRITE (3,888) ABETA, OBETA
      DO333I=1,N
  333 WRITE(3,444)(F(I,J),J=1,N)
  WRITE OUT THE H VECTOR
      WRITE (3,727) (H(I,1),I=1,N)
      WRITE (3,666) (AO(I,1),I=1,N)
      RETURN
      END
```

ROUTINE HVEC
D.6

```
      SUBROUTINE FORCST(N,ND,CI,A1,X,MN,TM,F,H,FCST,PLS,TLS,HH,ABETA,BBE
     1TA,D,S,SMAD)
       ROUTINE MAKES THE ACTUAL FORECAST
      DIMENSION TLS(600), CII(10), HH(10,1)
      DIMENSION FCST(150) ,ER(200),F(10,10),COI(10),A1(10,1),A(10,1)
      DIMENSIONX(200),H(10,1),CI(10),TM(10,10),CIF(10),CIL(10)
      DIMENSION OUT(70), ANG(10)
      DIMENSION TTM(10,10), PCTER(200), PLS(750)
      DIMENSION CON(70)
      DIMENSION TS(1000)
      REAL MAD
      DOUBLE PRECISION F,FF
    1 FORMAT (' STD DEV FORECASTS IS ',F12.6/)
    2 FORMAT (' MEAN ERROR IS ',F12.6/)
    3 FORMAT (' MEAN ABSOLUTE ERROR IS ',F12.6/)
    4 FORMAT (' VARIANCE OF THE ERROR IS ',F15.6/)
    5 FORMAT (' STD DEV ABOUT MEAN OF THE ERROR IS ',F12.6/)
    6 FORMAT (' STD DEV ABOUT ZERO OF THE ERROR IS ',F12.6/)
    7 FORMAT (' MEAN PERCENT ERROR IS ',F12.6/)
    8 FORMAT ('ONEW BETA =',F5.3/)
   44 FORMAT (70A1)
  101 FORMAT (1H1,'MODEL NO. =',I2,20X,'BETA =',F5.3/)
  104 FORMAT (1H ,I4,4F9.2,10F7.2)
  106 FORMAT(24H SUM OF SQUARE OF ERROR=,F15.3/)
  107 FORMAT(1X,F20.7,F20.4)
  108 FORMAT(1X,15H NEW FORECAST    )
  198 FORMAT (//' **** THE ERROR IS ****'//)
  200 FORMAT (' *** THE TAU ERROR IS ***'//)
  222 FORMAT (' THE VARIANCE OF THE FORECASTS =',F15.6/)
  973 FORMAT (' PERIOD DEMAND   FORCST    ERROR  CUM.ERR  TRACK    SMAD
     1 ',56A1)
  974 FORMAT (70HA(1)   A(2)   A(3)   A(4)   A(5)   A(6)   A(7)   A(8)
     1 A(9)   A(10)   )
            PRINT OUT THE MODEL NO. AND BETA FOR THIS F MATRIX
      WRITE(3,101)MN,ABETA
MAKE THE TRANSPOSE OF THE TRANSITION MATRIX
      ALPHA = 1.0-ABETA
      SER = 0.0
      SUM=0.0
      SUMA=0.0
      SUMX=0.0
      SUMAB = 0.0
      SEDSQ = 0.0
      KK=0
      DO 55 I=1,N
   55 CII(I) = CI(I)
      NN=0
      DO429I=1,N
      DO429J=1,N
  429 TTM(I,J)=TM(J,I)
MAKE THE FORECAST
      REWIND 4
      WRITE (4,974)
      REWIND 4
      READ (4,44) (CON(I),I=1,70)
      REWIND 4
      NO=N*7
      WRITE (3,973) (CON(J),J=1,NO)
      DO191K=1,ND
      Z=0.0
```

```
    DO711I=1,N
711 Z = CII(I)*A1(I,1) + Z
    FCST(K)=Z
CALCULATE THE ERROR
    ER(K) = X(K) - Z
    SMAD = ABETA*SMAD + ALPHA*ABS(ER(K))
    SER = ABETA*SER + ALPHA*ER(K)
    TS(K) = SER/SMAD
    TLS(K) = 1.0*K
    TLS(K+ND) = TS(K)
    TLS(K+2*ND) = X(K)
    TLS(K+3*ND) = FCST(K)
CALCULATE THE CUMULATIVE ERROR
    SUM = SUM + ER(K)*ER(K)
    SUMA = SUMA + ER(K)
    SUMAB = SUMAB + ABS(ER(K))
****CALCULATE THE SUM OF THE OBSERVATIONS
    SUMX=SUMX+X(K)
UPDATE THE CONSTANTS
    DO449I=1,N
449 CIL(I)=0.0
    DO944I=1,N
    DO944J=1,N
944 CIL(I) = CIL(I) + TTM(I,J)*CII(J)
    DO926I=1,N
    CII(I) = CIL(I) + H(I,1)*ER(K)
926 CONTINUE
    WRITE (3,104) K,X(K),Z,ER(K),SUMA,TS(K),SMAD,(CII(I),I=1,N)
191 CONTINUE
    CALL RUN(ER,ND)
    ERM = SUMA/ND
    ABERM = SUMAB/ND
    DO 20  I=1,ND
 20 SEDSQ = SEDSQ + (ER(I) - ERM)**2
    VARE = SEDSQ/ND
    SDEVE = SQRT(VARE)
    SDEV = SQRT(SUM/ND)
    WRITE(3,198)
    WRITE(3,106) SUM
****CALCULATE THE VARIANCE OF THE FORECASTS
    VAR = SUM/(ND-N)
    SDFST = SQRT(VAR)
    WRITE(3,222)VAR
    WRITE(3,1)  SDFST
    WRITE(3,2)  ERM
    WRITE(3,3)  ABERM
    WRITE(3,4)  VARE
    WRITE(3,5)  SDEVE
    WRITE(3,6)  SDEV
    RETURN
    END
```

```
   SUBROUTINE UPDATE(AO,XX,TM,NM,X,N)
    ROUTINE UPDATES FITTING FUNCTION VECTOR NM TIMES
   DIMENSION AO(20,I), A2(10,1), TM(10,10)
   DIMENSION X(1), XX(1), A3(10,1)
 1 FORMAT ('OGEN X VALUES--UPDATED FITTING FUNCTIONS',I6,2X,'PERIODS
  1AHEAD')
 2 FORMAT (1H ,F10.5,2X,10F10.5)
   DO 10 I=1,N
   A2(I,1) = 0.0
10 A3(I,1) = AO(I,1)
   WRITE (3,1) NM
   DO 50 K=1,NM
   DO 20 I=1,N
   DO 20 J=1,N
20 A2(I,1) = TM(I,J)*A3(J,1) + A2(I,1)
   DO 21 I=1,N
   I1=I-1
   A3(I,1) = A2(I,1)
   XX(K+I1*NM) = A3(I,1)
21 A2(I,1) = 0.0
   WRITE (3,2) X(K),(A3(I,1),I=1,N)
50 CONTINUE
   RETURN
   END
```

ROUTINE UPDATE
D.8

```
      FUNCTION RAND (IX)
      IX = IX * 65579
      IF (IX) 5,6,6
    5 IX = IX + 2147483647 + 1
    6 RAND = IX * .4656613E-9
      RETURN
      END
```

```
      FUNCTION NORM (EBAR, SIGMAE, IX)
         PROGRAM GENERATES NORMAL GENERATOR WITH MEAN=EBAR AND VARIANCE
         EQUAL TO SIGSQE.
      REAL NORM
      COMMON IC
      IF(IC.EQ.1) GO TO 18
      U1=RAND(IX)
      U2=RAND(IX)
      T1=SQRT(-2*ALOG(U1)) * COS(6.28*U2)
      NORM = T1 * SIGMAE + EBAR
      IC = 1
      RETURN
   18 T2=SQRT(-2*ALOG(U1)) * SIN(6.28*U2)
      NORM = T2 * SIGMAE + EBAR
      IC = 0
      RETURN
      END
```

ROUTINES RAND & NORM
D.9

THE DYNAMICS OF ADAPTIVE FORECASTING MODELS

by

RICHARD NOLAN DAY

B.S., KANSAS STATE UNIVERSITY, 1966

AN ABSTRACT OF A MASTER'S THESIS

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Industrial Engineering

KANSAS STATE UNIVERSITY

Manhattan, Kansas

1969

ABSTRACT

The ability to obtain a forecasting model that conforms to the true demand data is of importance to all those involved in the forecasting process. It would be of particular benefit to us if we can evaluate our forecasting model and its parameters, and readily detect when the real demand process has changed. Through this research we have investigated the response of the forecasting model and its parameters to the demand model, and made recommendations that should simplify the forecasting process and provide the forecaster with more accurate results, obtained with less time and effort.

In this thesis three different forecasting studies were made where the forecasting model was

1. The same as the demand model,
2. Simpler than the demand model,
3. More complex than the demand model.

In order to control the demand data, a computer simulation program was developed that will generate a demand time series having a normal random error with a mean of zero and a variance of 1.

There are basically two different groups of methods that can be used in making the forecasts. The first group consists of the static forecasting methods - subjective estimates, graphical curve fitting, and regression analysis. This group is termed 'static' in that it gives equal weight to all of the past data.

The second group, the dynamic group, includes moving averages and exponential smoothing. Moving averages is the more elementary

form of this group in that it considers only the n most recent observations; each of the last n observations has a weight of 1/n, and the older observations have a weight of zero. Exponential smoothing is the more complex form of this group; in this form all of the past is considered. Essentially exponential smoothing fits a least squares curve to the time series much like statistical regression analysis; it does this by regressing the vector of fitting functions for the forecasting model so as to minimize the sum of the weighted squared residuals. The main difference between regression analysis and exponential smoothing is that exponential smoothing uses weighted data, and the fitting function coefficients are recomputed at each new sampling interval to reflect the error made in the previous observation.

There are several advantages in using this method of smoothing over the other methods of forecasting:

1. The response of the forecasting model can be controlled,
2. All of the past data is considered, and retained as one value,
3. The model coefficients are recomputed at each new forecasting interval to reflect the error made in the previous interval.

The only restrictions placed on exponential smoothing is that the vector of fitting functions must consist of polynomials, sinusoidals, exponentials, or mathematical combinations of the above three types.

In evaluating the results of this research we have made recommendations that will affect the response of the model to the different demand models, methods of evaluating a model to determine when it is no longer able to accurately forecast the demand data, and a study of the normality of the errors encountered in the forecasting process.