ACCURACY MEASURES
AND THE EVALUATION OF FORECASTS

Working Paper No. 463

Essam Mahmoud
University of Michigan-Flint

Accuracy Measures and The Evaluation of Forecasts

Many forecasters and decision makers such as executive managers, planners, production managers, sales managers, and inventory managers have different needs in terms of the following:

- The timing of an event (e.g., when the next recession will start);
- The magnitude of a variable (e.g., sales volume next month);
- The timing and quantities of some variables (e.g., when and how many raw materials to order); and
- The monitoring of some quantity (e.g., market share).

Managers need the above predictions and are faced with the problem of having to select forecasting techniques among the many that are available. Forecasting techniques range from naive models, moving average, exponential smoothing (single, double, etc.), adaptive techniques and econometric models to sophisticated techniques (Box-Jenkins, Parzen's Method, etc.). In addition, forecasts can be made judgmentally. The obvious question is what is the best way of predicting the future.

This paper deals with the different accuracy measures and forecasting evaluation. First, a review of research studies in the area of forecasting measures and their application in evaluating forecasts is provided. Second, the reliability of the data sources available for forecasting is discussed. The different accuracy measures and their use are summarized. The paper then considers the selection of the proper parameters and the adjustment of the forecasts through the monitoring of forecast accuracy on a continuous basis. Finally, there is a general discussion before conclusions are drawn and future directions for research suggested.

A REVIEW OF RESEARCH STUDIES IN THE AREA OF FORECASTING ACCURACY AND EVALUATION

There have been many research studies which summarize the accuracy and the performance of quantitative and qualitative forecasting techniques. For example, numerous studies have indicated that quantitative techniques perform better than qualitative techniques while others have found the opposite result or that their performance is about the same. Other research has evaluated the performance of a particular model relative to other models. Many studies have indicated that simple forecasting techniques do as well as sophisticated techniques and in some cases they do better. Other researchers have showed the importance of using combining forecasting techniques and the impact on improving the accuracy whether using a simple combining approach or a weighted approach. Detailed information on many of the findings is summarized in an article by Mahmoud (1984). Table 1 summarizes briefly some of the most important findings.

Insert Table 1 about here

On the whole, past research suggests that quantitative methods out-perform qualitative methods. This is of obvious significance to practitioners wishing to improve their forecasting accuracy. Forecasters or practitioners must, however, be aware of the particular circumstances under which empirical research has demonstrated the superiority of quantitative methods. Only where the circumstances are similar in practice can more accurate forecasts using quantitative techniques be expected. For instance, when the forecaster is dealing with a limited number of past observations, the applicability of qualitative methods might be more appropriate. Armstrong (1985) suggested that it is advantageous to experiment with more than one qualitative method as some are more accurate than others.

Another interesting conclusion from the practitioners' point of view that many past studies have revealed is that simple forecasting methods perform equally as accurately as do sophisticated methods. This has been illustrated by a variety of studies. Examples are Makridakis and Hibon (1979) and Makridakis et. al. (1982). The implication of these findings is to encourage

practitioners to view forecasting methodologies as a set of methods within their ability to understand and use them. This may be especially so in the case of managers who wish to predict and cope with future uncertainties but do not have the training or expertise to deal with the very complex forecasting techniques. For theorists, the implications are to concentrate their efforts on the development and refining of simpler forecasting models, and on the simplification of more complex techniques.

THE RELIABILITY OF THE DATA SOURCES

A major consideration in the selection of a forecasting method for a particular application is the type of pattern in the data. Normally, there are four different data patterns: horizontal, seasonal, cyclical and trend. However, one may find that one or more of these patterns could exist in a particular time-series. Identifying the type of data would enable the forecaster to concentrate on a group of methods which is more suitable to a particular data pattern. However, before a data pattern is identified, it is important that the forecaster recognizes the dependence of any forecasting method upon a reliable database. For example, Mahmoud (1982) and Rice and Mahmoud (1985) provided information on a variety of data bases available which would be useful for organizations or international businesses. The lists identified the type of data available and its applicability in forecasting. Also, they discussed the importance of measuring the accuracy of the databases and how one would identify their reliability for a particular source. Proper operation and maintenance of an accurate and timely data system gives the forecaster an instrument with which to control and minimize the shortcomings of various forecasting methods. It is, therefore, essential to evaluate the databases available to verify the reliability of the data before analyzing the data pattern.

Finally, from a practical standpoint, if valuable results are to be obtained from applying forecasting models, managers and forecasters must remember that a forecast is only as accurate as the data set upon which is based.

MEASURES OF FORECASTING ACCURACY

Accuracy plays an important role in evaluating forecasting methods. Accuracy can refer to "goodness of fit" which in turn measures how well the forecasting model is able to reproduce the data that were used to develop the forecasting model. Most importantly, however, it should refer to the future (post-sample), that is, for data that have not been used to develop the forecasting model. Perceived accuracy varies from one application to another or from one decision maker to another as described by Wheelwright and Makridakis (1985). For some decision situations, plus or minus 10% may be sufficient; in others, a variation of as little as 5% could spell disaster. Thus, being familiar with the different accuracy measures and their pros and cons would enable those decision makers seeking high levels of accuracy to acheive more accurate forecasts.

While accuracy is a significant factor in evaluating forecasts, it is difficult to define it. The difficulty is associated with the absence of a single universally accepted measure of accuracy (Gardner and Dannenbring, 1980; Mahmoud, 1984; Makridakis and Wheelwright, 1979; Makridakis and Winkler; 1985. This is due to the fact that specific accuracy measures are appropriate for different types of forecasting applications. For example, accuracy measures are defined by Granger (1969) as loss functions and can be in the form of linear, quadratic, or non-symmetric functions. Suppose a forecasting model could be best fitted using a quadratic model and an accuracy model such as Mean Absolute Error (MAE) was used which is more suitable for measuring linear or non-symmetric functions. In this case an accuracy measure is not appropriate for the type of data used. This problem can be avoided with a clear understanding of the different accuracy measures. Unfortunately, there is no single accuracy measure that can be implemented in every forecasting situation. Also, it has been shown in many studies that the best model fitted (ex-ante) in terms of accuracy does not necessarily provide the best forecast in the forecasted phase (ex-post) (see Mahmoud, 1982; Makridakis and Wheelwright, 1979). For example, Table 2 shows the performance of thirteen forecasting models

tested with a representative series of weekly sales data
covering a 104 week time horizon, in which 12 periods were used
for the ex-post phase (Mahmoud, 1984). The thirteen forecasting
models ranged from the simplistic naive forecasting method to
the complex Box-Jenkins approach. The thirteen forecasting
models are shown in Table 2 and are listed in the order of the
Mean Square Error (MSE) of forecasting accuracy. The rank order
of the MSE is from low to high. Note that the rank order of
the Mean Square Error (ex-ante) and the U-Statistic are usually
closely related. There is little association between the rank
order of the accuracy measures at the ex-ante and ex-post phases.
The cost of forecasting error does not appear to be related to
any of the other accuracy measures used at the forecasted phase.

Insert Table 2 about here

In this section some of the most widely applied measures
will be discussed to show their advantages and disadvantages.

It should be noted that one common goal is to minimize
the error in the forecast. Thus, the error is defined as:

Error = Actual - Forecast

or

$$e_t = A_t - F_t$$

where e_t represents the error at period t

A_t represents the actual value at period t

F_t represents the forecasted value at period t.

For a time series of a variable such as the sales of product
A, Figure 1 represents the actual value of the monthly sales of the
product from January 1980 to December 1985, that is, 72 periods.
By identifying the data pattern and choosing the appropriate model,
the forecaster can measure the performance of the model by
calculating the total errors from January 1980 to December 1985
(fitted phase). The difference between the two values (actual-
forecast) is a measure of the error in forecasting this variable
for each period. In this fashion, t_1 = January 1980 and t_n =
December 1985. Remember that December 1985 represents the current

period. In Figure 1, it should also be noted that the forecaster can consider the fitted phase as starting form January 1980 which is represented by period t_1, through t_n and the forecasted phase is from period t_{n+1} to period t_{n+m}.

The forecaster would like to forecast sales for the next six months of 1986. These periods are defined as t+1, t+2, to t+m where m = 6. In other words, F_{t+1} represents the sales forecast for January 1986, F_{t+2}, February 1986 and F_{t+m} represents the forecast for June 1986. A clear distinction is needed between the errors of fitting the model to the data from January 1980 to December 1985 (the fitted phase) and the errors of forecasting from January 1986 to June 1986 (the forecasted phase). Total errors from January 1980 to December 1985 (fitted phase) :

$$= \sum_{t=1}^{n} (A_t - F_t)$$

or

$$= \sum_{t=1}^{n} e_t$$

where t = 1, 2, ..., n, from January 1980 to December 1985. The right hand side is known as "the sum of the error term." The total errors of the forecast for January 1986 to June 1986 can be calculated as follows, after the actual sales values for those months are known :

$$\text{Sum of the error} = \sum_{t=n+1}^{m} e_t$$

where e_1 = the error of January 1986; e_m represents the error of June 1986 in this example.

Summary of Measures

A summary of accuracy measures is presented below based on sources such as Makridakis et. al. (1982), Armstrong (1985), Steece

(1982), and Mahmoud (1982). Some of these measures are more
widely used than others. However, it is important to know
what type of measures are available. Following the same definition
mentioned earlier, we will summarize and discuss some of these
measures.

1. **Error = Actual - Forecast**
   or
   $$e_t = A_t - F_t$$
   This represents an individual error for a given time t.

2. **Mean Error (ME)**

$$ME = \frac{\sum_{t=1}^{n} (A_t - F_t)}{n}$$

3. **Mean Absolute Error (MAE)**

$$MAE = \frac{\sum_{t=1}^{n} \left| e_t \right|}{n} = \frac{\sum_{t=1}^{n} \left| (A_t - F_t) \right|}{n}$$

This measure is also known as Mean Absolute Deviation (MAD).
The measure gives an equal weight to the individual error of
each period, while not offsetting the positive and negative
values of the individual error. MAE is an appropriate measure
whenever the loss function is linear and symmetric. In the
case of a linear cost function, the MAE of 10 units is twice
as costly as an error of 5 units.

4. **Percentage Error (PE$_t$)**

$$PE_t = \frac{A_t - F_t}{A_t} (100)$$

The error is determined based on a weighted value which is the

- 7 -

actual value of each period.

5. Mean Percentage Error (MPE)

$$MPE = \frac{\sum\limits_{t=1}^{n} PE_t}{n} = \frac{\sum\limits_{t=1}^{n} \dfrac{A_t - F_t}{A_t}}{n}$$

If the percentage errors are simply added together, positive values will offset negative values and the average percentage error will be small, even though the individual error may be substantial. MPE assumes a linear cost function.

An alternative approach to MPE is the Mean Absolute Percentage Error (MAPE) which combines the individual percentage errors without offsetting the negative and the positive values.

6. Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{\sum\limits_{t=1}^{n} \left| PE_t \right|}{n}$$

This measure is similar to the Mean Absolute Error (MAE) or MAD. However, MAPE treats each error equally without taking account of the sign. It is useful in comparing different forecasting models. MAPE assumes that the cost of errors is more closely related to the percentage error than to the unit error.

7. Adjusted Mean Absolute Percentage Error ($\overline{MAPE}$)

$$\overline{MAPE} = \left[\frac{\sum\limits_{t=1}^{n} \dfrac{\left| A_t - F_t \right|}{1/2(A_t + F_t)}}{n} \right] (100)$$

The MAPE is similar to the MAPE. It does not weigh the error
based on the actual value only but on both the actual and
forecasted values for the same period.

8. Mean Squared Error (MSE)

$$MSE = \frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}$$

MSE is one of the most commonly used measures of accuracy.
Forecasters usually choose the models which minimize MSE.
However, there are two shortcomings of using MSE as a
measurement of accuracy as discussed by Makridakis et. al.
(1983). First, a comparison of the MSE developed during the
fitted phase may give little indication of the accuracy of
the model at the forecasting phase. Secondly, MSE as a
measure of forecasting accuracy is limited by the fact that
different methods use different procedures in the fitting
phase. For example, smoothing methods are highly dependent
upon initial forecasting estimates, whereas regression methods
minimize the MSE by giving equal weight to all observations.
Furthermore, Box-Jenkins minimizes the MSE of a non-linear
optimization procedure. Thus comparisons are difficult and
highly dependent on absolute units which makes comparisons
among series practically impossible.

9. Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n}}$$

It is similar to the MSE measure, but the associated cost
function is quadratic. The disadvantage of using the RSME, as
with MSE, is that it is an absolute measure of the errors.

10. Standard Deviation of Error (SDE)

$$ SDE = \sqrt{\frac{\sum_{t=1}^{n} (A_t - F_t)^2}{n - 1}} $$

It is similar to the RSME. The only difference is that the total sum squared of the errors is divided by n - 1 instead of n.

11. Coefficient of Variation (CV)

It is similar to the statistical inference coefficient of variation. It relates either SDE or RMSE to the average of the actual data. The smaller the value the better the performance of the model.

$$ CV = \frac{SDE}{\left(\sum_{t=1}^{n} A_t \right) / n} $$

or

$$ CV = \frac{RMSE}{\left(\sum_{t=1}^{n} A_t \right) / n} $$

12. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum\limits_{t=1}^{n} (A_t - F_t)^2}{\sum\limits_{t=1}^{n} (A_t - \bar{A})^2}$$

where $\bar{A}$ is the average of A_t. R^2 is commonly used in regression analysis. It can also be used as a measure of accuracy for time series models. R^2 ranges from 0 to 1. The closer the value of R^2 to 1 the better the forecast of the model. However, one should be familiar with the interpretation and the use of R^2. Armstrong (1985) and Nelson (1974) discussed the use of R^2.

13. Theil's U-Statistic (U)

$$U = \sqrt{\frac{\sum\limits_{t=1}^{n-1} \left(\frac{F_{t+1} - A_{t+1}}{A_t} \right)^2}{\sum\limits_{t=1}^{n-1} \left(\frac{A_{t+1} - A_t}{A_t} \right)^2}}$$

Theil (1971) explains in detail the use of the U-Statistic as a relative accuracy measure. U as a measure of accuracy allows a relative comparison of formal forecasting methods with the naive approch and also squares the error involved so that large errors are given much more weight than small errors.

When the accuracy of a naive method and the formal forecasting model are compared, the interpretation of the U-Statistic is as follows :

U = 1 the naive method is as good as the forecasting model being evaluated.

U < 1 the forecasting model being used is better than the naive method.

U > 1 the naive method produces better results than the forecasting model.

14. The Durbin-Watson Statistic (D-W)

$$D\text{-}W = \frac{\sum\limits_{t=2}^{n} (e_t - e_{t-1})^2}{\sum\limits_{t=1}^{n} e_t^2}$$

Makridakis et.al. (1983) detailed the computation and its use. As a rule of thumb, a good fitted forecasted model would reveal a value of the D-W statistic around 2.

15. Loss-cost Function for Measuring Accuracy

Accuracy measures are known as cost functions. Granger (1969) indicated that the actual function of the error could be estimated by standard accounting procedures but that in many cases it will not be symmetric. He discussed different cases and the consequences of using the generalized cost function. Many cases can be considered such as the quadratic, linear and non-symmetric cost functions. Granger focused on the error term and the shape of the function of the error. We assume that for every forecast error, there is an associated loss function. Granger assumes that for each forecast error (e) at time t, e_t, there is an associated loss $\ell(e_t)$. He discussed many different functions associated with forecasting error. Mahmoud (1982, 1985) also defined the accuracy of forecast errors in terms of a loss-cost function which is measured in dollars.

The previously discussed measures of accuracy suffer from several shortcomings in addition to giving no indication of the dollar value of the forecasting error. It is useful to analyze the trade-off between the cost of using a method

and the accuracy of that method. Mahmoud (1982) and Mahmoud et. al. (1985) developed several accuracy measures that managers or forecasters can use in measuring the opportunity cost of an inaccurate prediction in terms of dollars for various practical situations. Thus, managers or forecasters will be able to evaluate the total cost of using a forecasting model including its accuracy in terms of dollars. Applying a comparative analysis among different methods also enables them to determine the best alternative. The cost consists of the amount of money invested in a special program or a partial amount if the forecasting method is a part of a comprehensive or integrated package, the storage cost of the program and data, running costs and cost of human resources (for more details see Mahmoud, 1982). Furthermore, determining the total cost of different forecasting methods in terms of dollars enables managers or forecasters to choose the method which provides them with the accuracy they would like to achieve within their financial constraints and according to the level of technology and forecasting abilities that they may have. The loss-cost functions enable managers to choose a model based on the trade-off between the amount of money that could be saved and the extra money required to implement a more sophisticated method. For example, it is possible to compare the cost of combining two simple forecasting techniques versus the use of a more sophisticated technique. Studies have shown that the former in combination is more accurate than the latter (for more details on combining forecasting methods, see Makridakis and Winkler, 1983 and Mahmoud, 1984).

A particular example of a loss-cost function can be implemented only under a specific assumption. This assumption is that management requires that the total amount available in a period must be equal to the forecast for the period. Thus, the forecast is defined as F_t and A_t represents the actual value at the same period. The cost of forecasting error at period t can be defined as :

$$C_t = h \ Z_t(F_t - A_t) + s \ (1 - Z_t)(A_t - F_t)$$

where h represents the stock holding cost per item per period,

s represents the shortage cost per item and $Z_t = 1$ when $F_t \geq A_t$, otherwise Z_t is zero. Thus, the total loss-cost function over the 'n' forecasting periods is determined as follows :

$$TC = \sum_{t=1}^{n} [h \, Z_t(F_t - A_t) + s(1 - Z_t)(A_t - F_t)]$$

Some other practical cases for different types of inventory systems can be found in Mahmoud (1982) and Mahmoud et. al. (1985).

Relative Accuracy Measures

There is a need to compare the accuracy of two or more models in order to select the most accurate model or models. While Theil's U-Statistic is considered a relative accuracy measure, it is limited in that it only compares the formal forecast method being evaluated relative to the naive method. Gardenfors (1980) also introduced a relative accuracy measure known as the 'I' value for comparing the naive model with the formal forecasted model. The 'I' value is determined as follows :

$$I = \log NSS - \log ESS = 2 \, (\log RMS_N - \log RMS_S)$$

where NSS is the sum squared of the difference between the actual values of the variable and the values obtained from the naive model, and ESS is the sum squared of the difference between the actual values of the variables and the values obtained from the forecasting model. RMS_N is the root mean square of using the naive model and RMS_S is the root mean square of using the forecasting model (for more details see Gardenfors, 1980). Mahmoud showed that the 'I' value is identical to the MSE. Table 3 shows the relative performance of thirteen different forecasting methods and their accuracy according to the use

---
Insert Table 3 about here
---

of both the 'I' value and the Mean Squared Error. It is clear that the rank order is identical. The greater the value of 'I'

the more accurate the model will be. One should notice the value
of the Time-Series Multiple Regression model which was -0.03. As
Gardenfors indicated, one should not consider a given model as long as
its value by applying 'I' is negative. One can conclude, as long as
the MSE of a particular model is greater than the MSE of the naive
model the forecaster should not consider the model.Furthermore, one
would be able to determine a relative accuracy measure by using MSE
rather than using the 'I' value which would require more computational
efforts from the forecaster. Mahmoud's study concluded the following
I. The accuracy measures of both the forecasting method and the naive
method can be defined by the following ratios :

$$\frac{\text{Accuracy Measure of Forecasting Method}}{\text{Accuracy Measure of Naive Method}} \quad \left[\text{ i.e. } \frac{MSE_S}{MSE_N} \text{ or } \frac{MPE_S}{MPE_N} \right.$$

$$\left. \text{or } \frac{MAPE_S}{MAPE_N} \text{ or } \frac{SDE_S}{SDE_N} \right]$$

where all the numerators (MSE_S, MPE_S, $MAPE_S$, SDE_S) represent the
accuracy measures of the forecasting method (a smoothing technique,
for example) and all the accuracy measures in the denominators
(MSE_N, MPE_N, $MAPE_N$, SDE_N) represent the accuracy measures of the
naive model.

The following comments can be made :

A.   The forecasting model (i.e. the smoothing model) performs
better than the naive model if any of the MSE_S, MPE_S, or
SDE_S is less than the associated measure for the naive model
(MSE_N, MPE_N, $MAPE_N$, or SDE_N) or the ratio is less than 1.

B.   The naive model performs better than the forecasting model
if the MSE_S, MPE_S, $MAPE_S$, or SDE_S of the forecasting model
is greater than the associated accuracy measure for the naive
model or the ratio is greater than 1.

C.   The naive model performs the same as the forecasting model if
the MSE_S, MPE_S, $MAPE_S$, or SDE_S is as the same as the
associated accuracy measure the naive model or the ratio is
equal to 1. It is clear that these rules are similar to the
rule indicated by the U-Statistic.

It should be clear that in some situations where the data are very stable and do not fluctuate, using a naive method would be appropriate.

II. In the case of evaluating any two forecasting methods, the same rules can be followed. For example, if a manager would like to determine the relative performance of two methods, such as Single Exponential Smoothing and Box-Jenkins method, he or she would define the ratio as follows :

$$\frac{MSE_S}{MSE_{B-J}} \quad \text{or} \quad \frac{MSE_1}{MSE_2}$$

If the MSE_S or MSE_1 (Mean Squared Error of using Single Exponential Smoothing) is less than MSE_{B-J} or MSE_2 (Mean Squared Error of using Box-Jenkins method), this indicates that the exponential smoothing model is more accurate than the Box-Jenkins model and the ratio is less than 1. One could conclude the same as A, B, C.

The above list does not represent an exhaustive list of accuracy measures. Some more accuracy measures are discussed by Armstrong (1985) and Makridakis and et. al. (1983).

Structural Change and Bias

Forecasters should find systematic methods that improve the forecast performance. Thus, it is desirable to have a forecasting system that corrects potential biases prior to the integration of a forecast in the organization. It is useful to rely on an approach that helps forecasters detect bias and measure it. Thus, once bias is measured, the opportunity to improve future forecasts arises through correction of the bias. This could either be achieved through applying a test of structural stability as explained by Tiao et. al. (1975) or by determining bias decomposition as discussed by Theil (1971).

It is important to realize that almost all forecasts can be expected to contain some error. However, one would select the model which minimizes the errors.

Theil (1971) showed the usefulness of using three different

bias attributes known as the Mean Different Error, U^M, the Regression Pattern Error, U^R, and the Random Error expressed in terms of the population parameters, U^D. Expected Squared Error can be decomposed into three components as follows:

$$MSE = E \ (A_t - F_t)^2$$

$$= (\mu_A - \mu_F)^2 + (\sigma_F - \rho \ \sigma_A)^2 + (1 - \rho^2)\sigma_A^2$$

where μ_A and μ_F are the population means of actual value, A_t, and the forecasted value F_t, σ_A and σ_F are the population standard deviations of and ρ is the population correlation between A_t and F_t.

Dividing the previous equation by MSE provides the following :

$$1 = \frac{(\mu_A - \mu_F)^2}{MSE} + \frac{(\sigma_F - \rho \ \sigma_A)^2}{MSE} + \frac{(1 - \rho^2)\sigma_A^2}{MSE}$$

| Mean Differnce Error | The Regression Pattern Error | The Random Error |

Thus, the three components derived by Theil (1971) can be defined as follows :

$$\text{Mean Different Error} = \quad U^M = \frac{(\overline{A} - \overline{F})^2}{MSE}$$

$$\text{Regression Pattern Error} = U^R = \frac{(S_F - rS_A)^2}{MSE}$$

$$\text{Random Error} = \quad U^D = \frac{(1 - r^2)S_A^2}{MSE}$$

The bias attributed to differences in sample average levels of actual values A_t and forecasted values F_t is measured by U^M. In a

regression of the form :

$$A_t = F_t + \varepsilon_t$$

U^R measures the deviation of the sample regression slope from 1 and U^D measures the sample variance of the Regression Error term ε_t. To achieve perfect forecasts, $U^M = U^R = 0$ and $U^D = 1$. Thus, one could conclude in practice that if both values U^M and U^R are close to zero and U^D is close to one, the forecaster has achieved an unbiased forecast. For more detailed information related to the use and the application of different cases readers are referred to Theil (1971). and Moriarty (1985).

Forecasting Alternatives

The issue of considering alternative models, measures or approaches is an important one. The work of Makridakis et. al. (1982), known as the M-Competition, indicated the importance of three factors (time horizon, the type of data, and the accuracy measure) that affect the forecasting accuracy of various methods.

Accuracy depends on the application considered. Managers should consider alternative accuracy measures when these are to be used for a variety of forecasting applications. For example, Makridakis (1979) indicated that in case of forecasting inventories, large errors are undesirable, thus the use of the MSE accuracy measure would be appropriate. In budget forecasting the MAPE accuracy measure is commonly used. In situations requiring a single forecast (e.g. in bidding for a large contract in the futures market), Average Rankings of a particular accuracy measure must be used. Where only two methods are considered and the size of the error is not important, the Percentage Better method should be employed.

Some forecasting techniques require a minimum number of data points. For example, applying Census II and Decomposition forecasting methods using some computer packages requires a minimum number of 78 data points. Thus, if the database does

not include the minimum number of data points that a particular
model requires it would affect the availability of this
alternative. Different models also are appropriate for
different time horizons. For example, it is advisable to
consider deseasonalizing single exponential smmothing and
Holt's method in the case of forecasting for one-period ahead.
In the case of forecasting for four to six periods ahead,
that is longer forecasting horizons, sophisticated methods
such as Parzen's and Lewandowski's methods are recommended
by Makridakis et. al (1982).

Fildes (1982) indicated the importance of considering
alternative variables to be included in econometric or regression
models. The inclusion of different variables will affect the
explanatory power and predictive accuracy of such models. By
considering a checklist of variables and their likely impact
on the decision being contemplated, it becomes possible to
identify those variables that most require attention. Thus,
preparing alternatives would be very useful in addressing this
issue. In addition, the availability of such alternatives
depends on the information available and the number of data
points.

It is important to consider the alternative of combining
two or more forecasting methods instead of relying on only one
method. Several studies of combining forecasting methods have
shown the value of combining forecasts in improving accuracy
(Makridakis et. al., 1982; Makridakis and Winkler, 1983;
Mahmoud, 1982 and Zarnowitz, 1984). Users can consider
one of the three forms of combining forecasting techniques
(see Makridakis and Winkler, 1983). The first form takes a
simple average of two or more forecasts. The second approach
is known as historical weightings, in which each forecast
generated from each model should be weighted by the ratio of
one minus its Mean Squared Error to the total Mean Square for
all the forecasts. The third method uses subjective weighting,
in which managers apply weights based upon their personal
judgments to the forecasts as to which methods more closely
reflect the changing reality.

Managers should be encouraged to consider even a simple

average of two or more forecasting models.

In conclusion, being aware of the conditions under which some techniques perform better than others would enable managers to prepare for different alternatives. By monitoring which alternative works best, managers would be able to achieve their goals effectively.

MONITORING THE PERFORMANCE OF FORECASTING METHODS

The manager should monitor the environment closely and should constantly attempt to adjust the parameters of the forecasting model to incorporate any environmental changes. This would provide the manager with a predicted figure which would be closer to the actual value. Makridakis (1986) and Gardner (1969) indicated that monitoring the changes is an extremely important task to ensure that the sytem remains in control. Thus, quantitative forecasts can be modified to account for non-random changes. The constant monitoring of forecasting methods can be achieved easily through a good forecasting system. As computers and many good forecasting packages are now widely available, it is the responsibility of managers or forecasters to select the best program(s) or the system that provides them with many of these guidelines.

It is crucial to detect errors as quickly as possible. The forecasting model can then be refitted to the data or changed to a more appropriate model to prevent any serious production or inventory problems. In inventory control, for example, forecast monitoring is essential because of the need to take action when there is a significant change in demand. If the forecast model suggests an increase in demand, new orders should be placed on a priority basis. If demand is expected to fall, any unneeded orders should be cancelled to prevent excess inventory investment.

Monitoring devices (tracking signals) are used to keep watch for signs of bias in the forecast errors. Gardner (1983) also discussed three warning signs that can be used to show when a forecasting system goes out of control.

1. The first indicator is the simple cumulative sum (cusum)

of the forecast errors, which can be computed and tested
in several different ways. This indicator compares the
cumulative sum of the errors at the end of each period
to the smoothed Mean Absolute Deviation (MAD). Cusum is
determined as follows :

$$e_t = A_t - F_t$$
$$SUM_t = e_t + SUM_{t-1}$$
$$MAD_t = \alpha \mid e_t \mid + (1 - \alpha) \, MAD_{t-1}$$
$$C_t = \mid SUM_t \, / \, MAD_t \mid$$

where e_t represents the error for period t, A_t is the
actual value and F_t is the forecast value. SUM_t
represents the sum of the error at period t, and MAD_t
is the smoothed Mean Absolute Deviation at period t.
The smoothing parameter α should have a value between
zero and one. C_t is the tracking signal (cusum) for
period t. Cusum should fluctuate around zero when the
system is in control. Biased errors occur when the
cusum departs from zero or the system is out of control
if the C_t exceed the valueof the smoothing MAD_t.

2. The second indicator is the smoothed-error tracking
   signal. The tracking signal T_t is measured by using
   the following set of equations :

$$E_t = \alpha \mid e_t \mid + (1 - \alpha) \, E_{t-1}$$
$$MAD_t = \alpha \mid e_t \mid + (1-\alpha) \, MAD_{t-1}$$
$$T_t = \mid E_t \, / \, MAD_t \mid$$

where E_t represents the updated value of the error value
e_t.

3. The third indicator is the first-order autocorrelation
   in forecast errors. It is more complex than the previous
   two indicators. The existence of any significant positive

autocorrelation indicates lack of control. For more
details and signals, see Gardner (1985).

Applying any of the three methods is recommended. They are
easy to use. The only problem, however, that a user faces is
determining the starting values. For example, MAD_{t-1} and
SUM_{t-1} represent the values at the starting point for

the previous period t-1. SUM_{t-1} can start with a value of zero.
MAD_{t-1} can be set equal to its expected value.

Monitoring forecasts constantly by using one of the tracking
signals discussed previously alerts the manager to any problem
in his forecasting system. This may lead to managerial
intervention in the system by actions such as refitting the
model, or searching for another model or a combination of
models. Also, monitoring the error and adjusting the forecast
constantly enables the manager to take some other courses
of action related to the decisions that are affected by the
forecasts such as inventory and production adjustment.

DISCUSSION AND FUTURE DIRECTIONS

This paper has outlined some of the important issues related
to forecasting evaluation. However, organizations or individuals
implement the process of evaluation differently. This is due
to the knowledge of forecasting available, training facilities,
time, data, the computer system, the degree of integration, and
software used.

The level of knowledge that a particular forecaster has
will reflect on his forecasting performance given a forecasting
situation. Having more knowledge enables the forecaster to apply
more techniques and to apply more tests to check the accuracy
and the basic assumptions under which a particular method can
be implemented. However, one should realize that it does not
take a great deal of time or effort to gain this knowledge.
Once a particular person or team starts, he or she or it will
be able to gain more knowledge, especially with the help
of the many forecasting software packages available for a

variety of computers.

Time is an important factor, especially for those who handle many forecasts in a given time period and for those who forecast a limited number of variables during the same time. It is obvious that the latter would be able to apply more accuacy measures and set different alternatives and check thoroughly for model specification parameter performance, etc. For example, forecasters seeking to achieve better forecasting performance are encouraged to spend more time applying different computer runs for their models. This would enable them to specify the parameters that improve the accuracy of the forecasting model. Figures 2, 3, and 4 illustrate the actual sales values of a personal care product, the fitted forecast

---

Insert Figures 2,3,4 about here

---

model (ex-ante) and the forecast (ex-post) using Winter's Exponentional Smoothing model. Figures 3 and 4 show the results achieved by using the parameter values (α = 0.06 , β = 0.17 and δ = 0.10) and (α = 0.06 , β = 0.06 and δ = 0.90) respectively. Note that the former parameter values were determined by the software package using the search guide for the optional values of α and β given that δ is set at 0.10. One should realize that spending more effort and time would enable the forecaster to achieve better results by using the value judgmentally as applied in Figure 2 . In this way the forecaster can identify any seasonal factors and the degree of variability in the data. Studying the data structure carefully will indicate whether the values of the parameters should be increased or decreased.

Among many other studies, Mahmoud (1982) showed that using Single Exponential Smoothing and changing the parameter from 0.10 to 0.11 improved the accuracy of the model. Thus, to summarize, it is important that the forecasters or managers spend sufficient time in applying different attempts (runs) with different values of parameters (i.e. α, β, δ, etc.) to choose the estimated parameters which provide the better forecast.

Applying different combining forecasting methods enables the forecaster to achieve more accurate results than relying on an individual method.

Integrating forecasting applications as part of the planning process of the organization would help the forecasters to monitor closely any changes and their impact on forecasting. Wright (1986) showed the importance of considering forecasting as part of Decision Support Systems and not as a self-contained activity.

Finally, the availability of a good forecasting system would have a great impact on the forecaster or the decision maker. Surely a good forecasting package or program would be very useful so the forecaster would be able to address many of the issues discussed in this paper. For example, Wheelwright and Makridakis (1985) describe in detail two complete comprehensive forecasting systems, SIBYL-RUNNER and FUTURCAST. In the latter the system consists of a wide range of forecasting techniques. The system allows the forecaster to use at least five different accuracy measures and different techniques for model structure such as autocorrelation. It also permits combining more than one forecasting technique. The system allows the forecaster to monitor the environmental change and include the change in the model. For more details about forecasting software and systems see Beaumont et al. (1985) and Mahmoud et al. (1986).

In order to achieve a better accuracy performance the forecaster applies more than one accuracy measure to ensure a thorough evaluation. This can be achieved by either including more accuracy measures in the forecasting package that is currently available or considering having such facilities when buying a package or a program. Using an integrated package such as a spreadsheet would enable the forecaster to apply more accuracy measures. Furthermore, a package which includes comprehensive forecasting methods would enable the forecaster to test the performance of a variety of models. A package or system that provides the forecaster with different tools for diagnostic checking, model specification and monitoring the forecast error as was

mentioned earlier would be more useful.

CONCLUSIONS

Managers or forecasters face a great deal of difficulty in selecting and evaluating their forecasts. A systematic procedure should be integrated with the planning activities and as a part of the Decision Support System. However, a great understanding of the state of the art of forecasting techniques, awareness of the most important findings, the availability of different accuracy measures and their use, the availability of reliable databases, the contribution of different alternatives, monitoring environmental changes, and the availability of good computer forecasting systems are important criteria to be considered when evaluating the forecasts. The value and the outcome of the evaluation process depends on the organization's databases and the forecasters' experience, their knowledge of certain forecasting models, and their ability to understand the past and current changes.

ACKNOWLEDGEMENTS

<div align="center">Table 1</div>

Area of Application	Main Results	Literature Sources
Quantitative Methods versus Judgmental Methods	These studies indicated that quantitative methods provided better forecasts than judgmental methods.	Armstrong (1985) and Fildes and Fitzgerald (1981).
Box-Jenkins vs. Exponential Smoothing	In a comprehensive study of 111 time series for determining the accuracy of different time series forecasting techniques, the results indicated that Box-Jenkins models were less accurate than moving average and smoothing methods.	Makridakis and Hibon (1979).
Combining Forecasts Using Simple or Weighted Average	It was found that combining methods either using simple or weighted average methods provided a good forecast which resulted in a total overall better performance on the average than the individual methods included.	Makridakis et. al. (1982), Winkler and Makridakis (1983).
Combining Corresponding Sets of Individual Forecasts	The study investigated the accuracy of combining corresponding predictions from different sources and the corresponding sets of individual forecasts. The results showed that there are gains to the forecast users from combining predictions from different sources.	Zarnowitz (1984)
Forecasting Systems for Reduced Bias in Forecasts	The study illustrated different design features of forecasting systems that can be used to improve the performance of any forecasting method.	Moriarty (1985)
Assessment of the State of the Art of Forecasting	The study provided an overall assessment of the state of the art of forecasting and also suggested guidelines for forecasters.	Makridakis (1986)

Table 2

Forecasting Model and Associated Accuracy Measures

Rank Order of Forecasting Models According to MSE	Fitted Phase Ex-Ante				Forecasted Phase (12 Periods Ahead) Ex-Post			
	MSE	MPE	MAPE	U-Stat	MSE	MPE	MAPE	Cost of Forecast. Error
1. Harrison's Harmonic	4,760	- 8.64	23.8	.23	849	9.04	18.2	$ 54,112
2. Classical Decomposition	5,585	-10.01	26.2	.28	1,214	13.25	21.3	$ 75,280
3. Census II	5,677	- 9.01	26.8	.29	1,299	-19.17	24.2	$120,433
4. Linear Exp. Smooth.	7,276	- 7.62	30.4	.37	949	13.90	19.5	$ 68,618
5. General Adaptive	7,290	-11.58	30.7	.34	889	12.53	18.0	$ 64,636
6. Quadratic Exp. Smooth.	7,357	- 8.39	30.1	.39	1,194	12.32	19.8	$ 70,003
7. Single Exp. Smooth.	7,758	-14.01	31.8	.34	1,395	28.81	24.2	$ 68,395
8. Winters' Trend and Seasonal	8,045	- 5.85	30.0	.36	1,298	11.38	23.4	$120,705
9. Adaptive Response	8,336	-15.68	33.2	.40	945	14.81	20.5	$ 69,171
10. Box-Jenkins	9,877	-11.87	35.0	.46	1,416	17.07	27.2	$ 76,694
11. Holt's-Modified	10,445	- 8.03	35.2	.57	1,095	12.66	19.8	$ 69,785
12. Time Series Multiple Regression	11,930	- 9.73	34.0	.66	1,410	19.22	23.2	$155,501
13. Naive Method	13,324	-12.61	40.8	1.00	1,420	- 2.31	26.2	$168,426

Notes : MSE   = Mean Squared Error
         MPE   = Mean Percentage Error
         MAPE   = Mean Absolute Percentage Error
         U-Stat = Theil's U-Statistic

Source : Mahmound (1984)

Table 3

Rank Order of MOdels According to "I: Value and MSE Series A

| Rank Order of Models According to "I" Value	"I" Value using Naive Model	"I" Value using Harrison's Harmonic Model	Mean Squared Error MSE	Rank Order Of Models According in MSE
1. Harrison's Harmonic	0.85		4,952.6	1
2. General Adaptive	0.53	-0.32	6,816.3	2
3. Classical Decomposition	0.51	-0.34	6,957.4	3
4. Quadratic Ex. Smooth.	0.48	-0.37	7,209.9	4
5. Single Exp. Smooth.	0.43	-0.42	7,515.3	5
6. Adaptive Response Rate	0.33	-0.52	8,333.6	6
7. Box-Jenkins	0.31	-0.54	8,530.6	7
8. Linear Exp. Smooth.	0.15	-0.70	10,004.8	8
9. Linear and Seasonal Exp. Smooth.	0.06	-0.79	10,904.9	9
10. Holt's Exp. Smooth.	0.03	-0.82	11,199.3	10
11. Naive Model		-0.85	11,595.8	11
12. Time-Series Multiple Regression	-0.03	-0.88	11,905.0	12

Source : Mahmoud, Essam (1984), Gardenfors' 'I' Value : A Comment on the Measurement of the Relative Accuracy of Sales Forecasting Models. Technological Forecasting and Social Change, Vol 25, No. 4, p. 357.
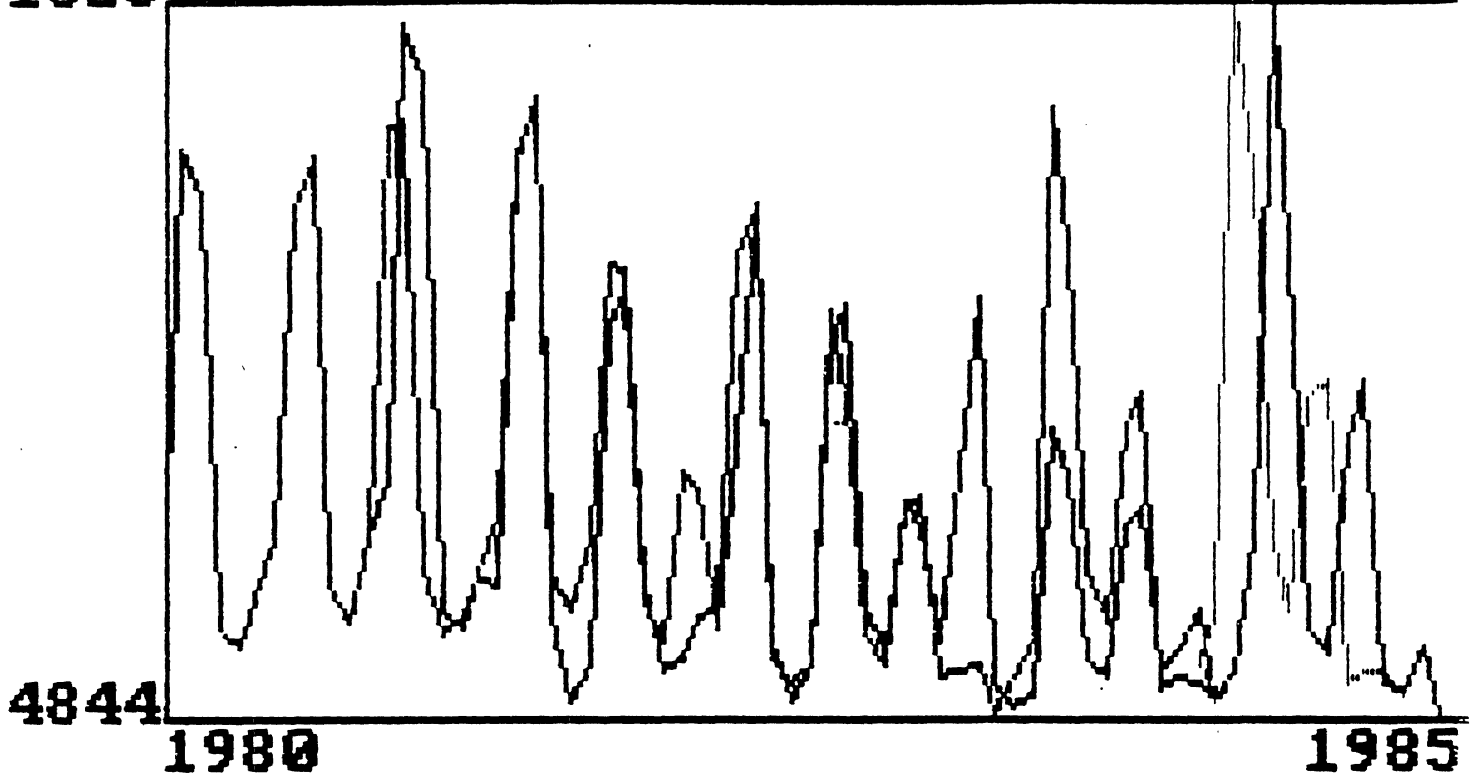
Figure 1
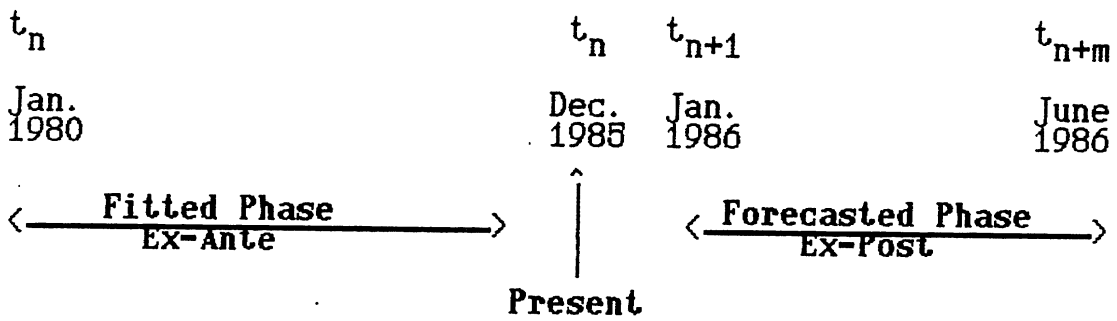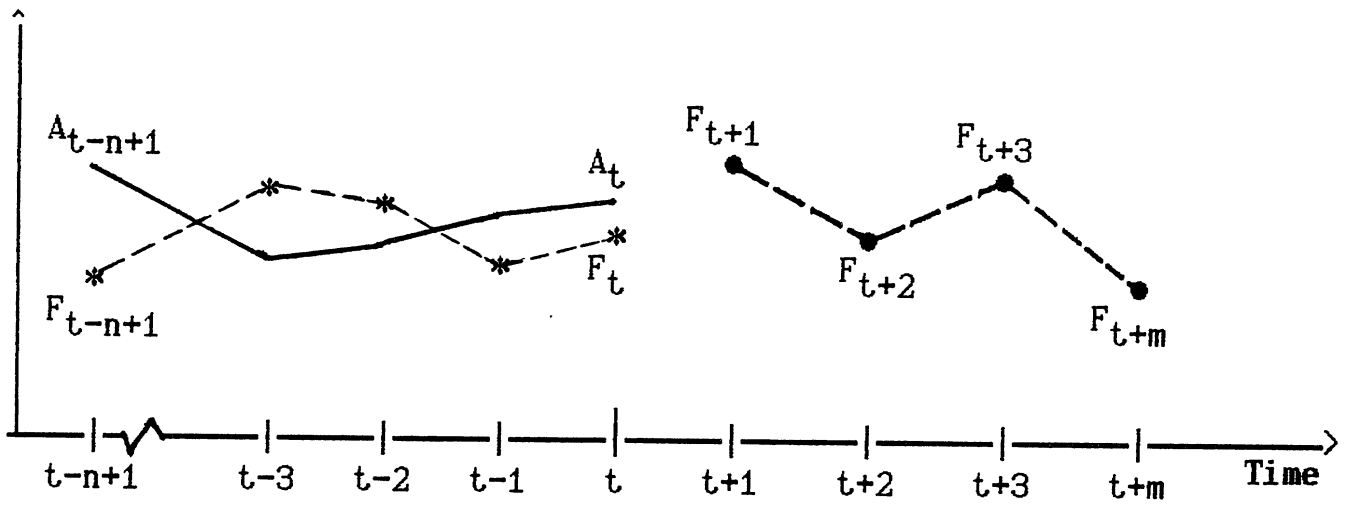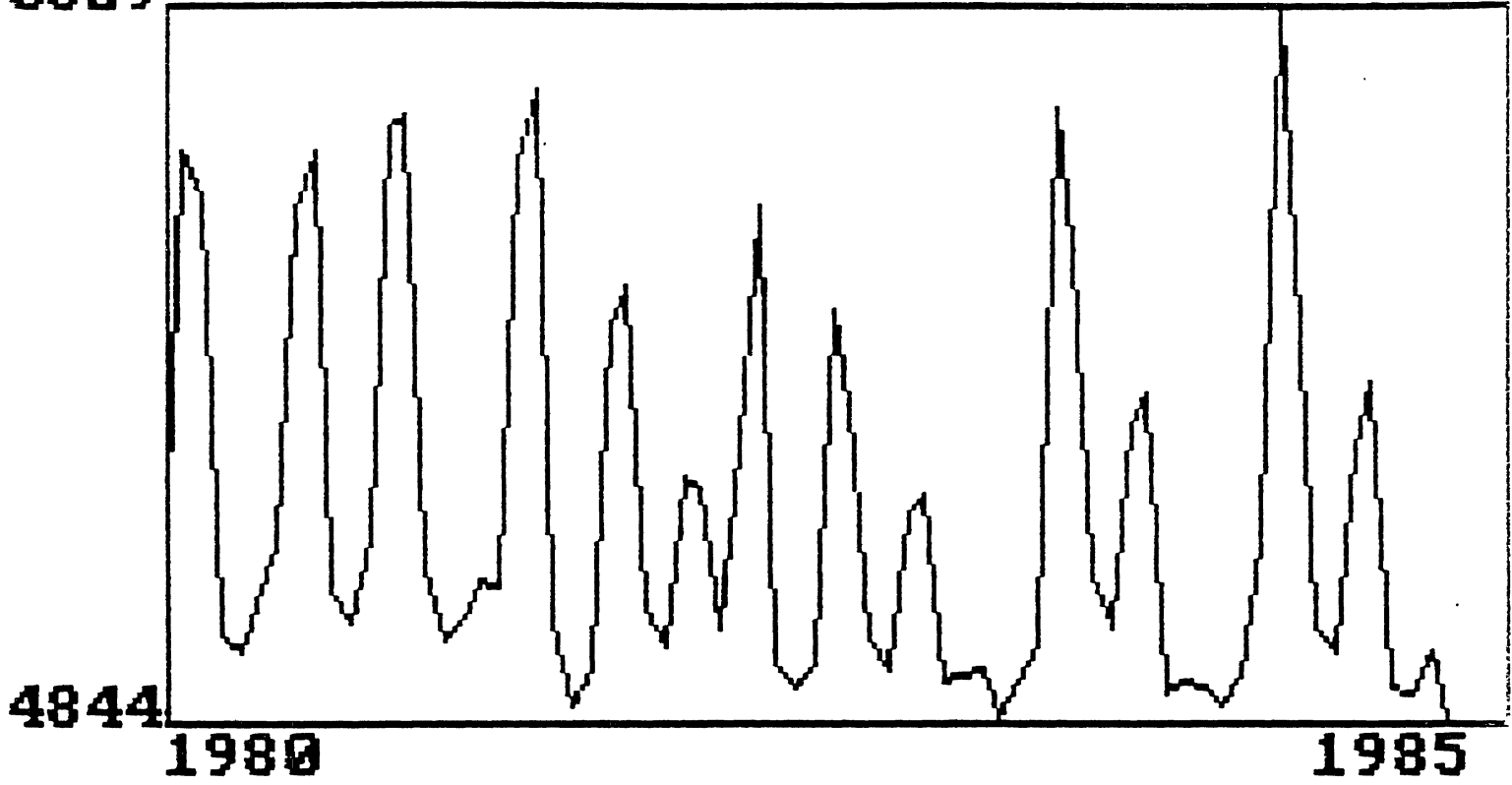
PRODUCTA

0529

4844

1980 **1985**

FITTED FORECAST PRODUCTA

Figure 4

FITTED FORECAST PRODUCTA

REFERENCES

1. Armstrong, J. S., Long-Range Forecasting: From Crystal Ball to Computer, John Wiley, Second Edition, 1985.

2. Beaumont, C.; Mahmoud, E. and McGee, V., Microcomputer Forecasting Software: A Survey. Journal of Forecasting, vol. 4, pp. 305-311.

3. Fildes, Robert, Forecasting: The Issues, The Handbook of Forecasting: A Manager's Guide, Edited by Makridakis, S. and Wheelwright, S. C. New York : John Wiley and Sons, 1982, pp. 83-104.

4. Fildes, R. and Fitzgerald,D., The Use of Information in Balance of Payments Forecasting, paper presented at the First International Symposium on Forecasting, Quebec, Canada, May 1981.

5. Gardenfors, P., On the Information Provided by Forecasting Models, Technological Forecasting and Social Change, 16, 1980, pp. 351-361.

6. Gardner, E. S., Exponential Smoothing: The State of the Art, Journal of Forecasting, Vol. 4, No. 1, 1985, pp. 1-28.

7. Gardner, E. S., Automatic Monitoring of Forecast Errors, Journal of Forecasting, Vol. 2, No. 1, 1983, pp. 1-21.

8. Gardner, E. S. and Dannenbring, Forecasting with Exponential Smoothing: Some Guidelines for Model Selection, Decision Sciences, Vol. 11, 1980, pp. 370-383.

9. Granger, C. W. J., Prediction With a Generalized Cost of Error Function: Operations Research Quarterly, Vol. 20, No. 2, 1969, pp. 199-207.

10. Mahmoud, Essam, Accuracy in Forecasting: A Survey, Journal of Forecasting, Vol. 3, No. 3, 1984, pp. 139-159.

11. Mahmoud, Essam, Short-term Forecasting: Matching
    Techniques to Tasks. An Integrated Framework and Empirical
    Investigation, Ph.D. Dissertation, State University of
    New York at Buffalo, June 1982.

12. Mahmoud, Essam and Pegels, C. Alternative Methods for
    Ranking Time Series Forecasting Models, (working paper),
    1984.

13. Mahmoud, Essam; Goyal, Suresh K.; and Shalchi, H.,
    Loss-cost Functions for Measuring the Accuracy of
    Sales Forecasting Methods, (working paper), 1985.

14. Mahmoud, Essam; Rice, Gillian; McGee, Victor and Beaumont,
    Chris, Mainframe Specific Purpose Forecasting : A Survey.
    Journal of Forecasting, Vol. 5, 1986.

15. Makridakis, Sypros, Empirical Evidence versus Personal
    Experience, Journal of Forecasting, Vol. 2, No. 3,
    1983, pp. 295-311.

16. Makridakis, Sypros, The Art and Science of Forecasting: An
    Assessment and Future Directions, International Journal
    of Forecasting, Forthcoming, 1986.

17. Makridakis, S. and Hibon, M., Accuracy of Forecasting:
    An Empirical Investigation, J. R. Statistical Soc.,
    Part 2, Vol 142, 1979, pp. 97-145.

18. Makridakis, S.; Wheelwright, S.; and McGee, V.E.,
    Forecasting: Methods and Applications, Second Edition,
    New York: John Wiley and Sons, 1983.

19. Makridakis, S.; Andersen, A.; Carbone,R.; Fildes, R.;
    Hibon, M.; Lewandowski, R.; Parzen, E. and Winkler, R.,
    The Accuracy of Extrapolation (Time Series) Methods:
    Results of Forecasting Computation, Journal of
    Forecasting, Vol 1, 1982, pp. 111-153.

20. Makridakis, S. and Wheelwright, S. C., Forecasting: Framework and Overview, Forecasting TIMS Studies in Management Science, Vol 12, S. Makridakis and S. C. Wheelwright, eds. North-Holland Publishing Company, 1979.

21 Makridakis, S. and Winkler, R., Averages of Forecasting: Some Empirical Results, Management Science, Vol. 29, No. 9, 1983, pp. 987-996.

22. Moriarty, Mark M., Design Features of Forecasting Systems Involving Management Judgements, Journal of Marketing Research, Vol. 22, November 1985, pp. 353-364.

23. Nelson, C. R., The First Order Moving Average Process Identification, Estimation and Prediction, Journal of Econometrics, Vol. 2, 1974, pp. 121-141.

24. Rice, Gillian and Mahmoud, Essam, Forecasting and the Database: An Analysis of Databases for International Business, Journal of Forecasting, Vol. 4, 1985, pp.89-97.

25. Steece, Bert, Evaluation of Forecasts, The Handbook of Forecasting: A Manager's Guide, edited by Spyros Makridakis and Steven C. Wheelwright, New York: John Wiley and Sons, 1982, pp. 457-468.

26. Theil,H., Applied Economic Forecasting, Amsterdam: North-Holland Publishing Company, 1971.

27. Tiao, G. C., Box, G. E. P. and Hamming, W. J., Analysis of Los Angeles Photochemical Smog Data: A Statistical Overview, Journal of the Air Pollution Control Association, Vol. 25, 1975.

28. Winkler, R. L. and Makridakis, S., The Combination of Forecasts, JRSS, Series A.

29. Wright, David J., Evaluation of Forecasting Methods for Decision Support, The International Journal of Forecasting, Forthcoming, 1986.

30. Wheelwright, S. and Makridakis, S., Forecasting Methods for Management, 4th edition, New York: John Wiley and Sons, 1985.

31. Zarnowitz, V., The Accuracy of Individual and Group Forecasts From Business Outlook Surveys, Journal of Forecasting, Vol. 3, 1984, pp. 10-27.