# FORECASTING THEORY

The ability to forecast future events accurately has been highly valued throughout time. Whether it is in business or in our private lives, forecasting future events helps us to plan for them adequately. We all make forecasts daily; we develop them from our experiences and knowledge about certain situations. The same is true in management or administrative situations. In business, industry, and government, decision makers must anticipate the future behavior of many variables before they can make decisions. Based on these forecasts, proper planning can take place. Forecasting can therefore be seen as a critical aid to planning effectively for the future.

In business organizations, forecasts are made in virtually every function and at every organizational level. For example, a bank manager might need to predict cash flows for the next quarter, a control engineer may wish to control future values of an output variable by manipulating a set of input variables, a company manager might need to forecast sales, and a production manager may need to estimate labor-hours required to meet a given production schedule. In all these scenarios, statements about the future are made based on the past and the assumption that the future will be similar to the past.

Although each forecasting situation is unique, certain general principles are common to almost all forecasting problems. A large range of forecasting methodologies vary in complexity, cost, and accuracy, allowing the forecaster great choice in model selection. Understanding the basic principles of forecasting and existing forecasting options is the first step in being able to generate good forecasts.

## FORECASTING FUNDAMENTALS

### Principles of Forecasting

One of the most basic principles of forecasting is that forecasts are rarely perfect. Forecasting future events involves uncertainty, and as such perfect prediction is almost impossible. Forecasters know that they must live with a certain amount of error. Our goal in forecasting is to generate *on the average* good forecasts over time and minimize forecast errors.

Another principle of forecasting is that forecasts are more accurate for groups or families of items rather than for individual items themselves. Because of pooling of variances, the behavior of group data can have very stable characteristics even when individual items in the group exhibit high degrees of randomness. Consequently, it is easier to obtain a high degree of accuracy when forecasting groups of items rather than individual items themselves.

Finally, forecasts are more accurate for shorter than longer time horizons. The shorter the time horizon of the forecast, the lower the uncertainty of the future. There is a certain amount of inertia inherent in the data, and dramatic pattern changes typically do not occur over the short run. As the time horizon increases, however, there is a much greater likelihood that a change in established patterns and relationships will occur. Therefore, forecasters cannot expect to have the same degree of forecast accuracy for long range forecasts as they do for shorter ranges.

### Classification of Forecasting Methods

Forecasting methods can be classified into two groups: *qualitative* and *quantitative methods.* Qualitative or judgmental forecasting methods are subjective in nature. They are based on intuition, personal knowledge, and experience and are educated guesses of forecasters or experts in the field. These forecasts can be generated very informally or follow a structured decision-making process. Because these forecasts are based upon individual opinions, they lack consistency, and different forecasters will typically generate different forecasts for the same situation. Although qualitative forecasting involves a nonrigorous approach, under certain circumstances these methodologies may be quite appropriate and the only method of choice.

Quantitative forecasting methods are approaches based on mathematical or statistical modeling. Based on mathematics, these models generate consistent forecasts that are reproducible by any forecaster. Three conditions are required for quantitative forecasting methods to be applied. First, information about the past must be available. Second, available information must be quantified in the form of data. Finally, we must be reasonably confident that past patterns will continue into the future. This last condition is known as the assumption of constancy and is an underlying premise of all quantitative models.

Quantitative forecasting models vary considerably, each having its own properties, accuracies, and costs that must be considered when choosing a specific method. Quantitative models can be divided into two major categories: *time-series* and *causal models.* The objective of time-series forecasting methods is to discover the pattern in the historical data series and extrapolate that pattern into the future. Causal models, on the other hand, assume that the factor to be forecast exhibits a cause-effect relationship with one or more independent variables. For example, sales = $f$(income, prices, advertising, competition). The purpose of the causal model is to discover the form of that relationship and to use it to forecast future values of the dependent variable.

Qualitative methods, sometimes called technological methods, do not require data in the same manner as quantitative methods do. The inputs required are mainly intuitive thinking, judgment, and accumulated knowledge, often developed by a number of specially trained people. Qualitative forecasting methods can be further divided into two groups. These are exploratory and normative methods (1). Exploratory methods start with the present and move toward the future in a heuristic manner considering all possibilities. Examples of exploratory methods include techniques such as Delphi, S-curves, analogies, and morphological research. Normative methods, on the other hand, start with future objectives and work backward to see if these can be achieved, considering all known constraints. Normative methods include such techniques as decision matrices, relevance trees, and system analysis.

Like their quantitative counterparts, qualitative techniques vary widely in cost, complexity, and value. They can be used separately but are often used in combination with each other or in conjunction with quantitative methods. In certain situations, such as formulating strategy, developing new products and technologies, and developing long-range plans, they are the only techniques possible because relevant data are unavailable.

#### Selecting a Forecasting Model

A number of factors influence the selection of a forecasting model. The first determining factor to consider is the type and amount of available data. Certain types of data are required for using quantitative forecasting models and, in the absence of these, qualitatively generated forecasts may be the only option. Also, different quantitative models require different amounts of data. The amount of data available may preclude the use of certain quantitiative models narrowing the pool of possible techniques.

Another important factor to consider in model selection is degree of accuracy required. Some situations require only crude forecasts, wheres others require great accuracy. Increasing accuracy, however, usually raises the costs of data acquisition, computer time, and personnel. A simpler but less accurate model may be preferred over a complex but highly accurate one, if the loss in accuracy is not critical and if there are substantial savings in cost. In general, it is best to use as simple a model as possible for the conditions present and data available. This is also known as the principle of parsimony, which says that, when deciding among alternative models, the simplest is preferable, all other things being equal.

A third factor to consider is the length of the forecast horizon. Forecasting methods vary in their appropriateness for different time horizons, and short-term versus long-term forecasting methods differ greatly. It is essential to select the correct forecasting model for the forecast horizon being used. For example, a manufacturer who is trying to forecast the sales of a product for the next 3 months is going to use a vastly different forecast than an electric utility trying to forecast demand for electricity over the next 25 years.

Finally, an important criterion in selecting an appropriate method is to consider the types of patterns present in the data so that the methods most appropriate to those patterns can be selected. Four basic types of data patterns can be distinguished:

1. Horizontal—A horizontal pattern exists when data values have no persistent upward or downward movement. An example of this would be a product whose sales do not increase or decrease over time. This type of pattern may not be uncommon for products in the mature stage of their life cycle or in a steady state environment.
2. Trend—When there is an increase or decrease in the data over time, we say that the data exhibit a trend pattern. The sales of many companies and products, as well as many business or economic indicators follow a trend pattern in their movement over time.
3. Seasonality—A seasonal pattern is any pattern that regularly repeats itself and is of constant length. This pattern exists when a series is influenced by seasonal factors, such as the quarter or month of the year or the day of the week. An example of this could be a retail operation with high sales during the months of November and December or a restaurant with peak sales on Fridays and Saturdays.
4. Cycles—When data are influenced by longer-term economic fluctuations such as those associated with the business cycle, we say that a cyclical pattern is present. The major distinction between a seasonal and a cyclical pattern is that a cyclical pattern varies in length and magnitude. Because of this, cyclical factors can be much more difficult to forecast than other patterns.

Any one of these patterns can be present in a time series. Also, many time series contain a combination of these patterns. Forecasting models differ based on their ability to forecast different data patterns. A critical issue in forecasting is to make sure that the model selected can forecast the patterns present in the data set.

#### A Framework of the Forecasting Process

Before we can study specific forecasting techniques, it is important to understand the general process used to develop a quantitative forecasting model and generate forecasts. There are certain procedural steps that must be followed regardless of which forecasting model is used. In general, developing and using a quantitative forecasting model consists of two major stages. The first stage is *model building,* where the forecasting model is selected based on historical data and available theory. The selected model then must be *fit* to the known data by carefully selecting parameters and initializing procedures. For example, these parameters may be selected through an estimation approach, such as least squares. Finally, in this stage, the forecaster must check the adequacy of the fitted model. This is done by applying the forecasting model to historical data and obtaining *fitted values. Fitted errors* that test the goodness of fit of the model are generated. Based on the fitted errors, the model could be found inadequate for a number of reasons, such as including inappropriate parameters or incorrectly specifying the functional relationship. If the forecasting model is found to be inadequate, it has to be respecified. This cycle of model specification, parameter estimation, and diagnostic checking is iterative and must be repeated until a satisfactory model is found.

The second stage in this framework is the *forecasting stage.* This is where the final model is used to obtain the forecasts. As data patterns change over time, the forecaster must make sure that the specified model and its parameters are adjusted accordingly. The adequacy of the forecasting model must be assessed continually by checking the forecasts against the new observations.

### Measuring Forecast Accuracy

One of the most important criteria for choosing a forecasting method is its accuracy. The model's accuracy can be assessed only if forecast performance is measured over time. The adequacy of parameters and models change over time as data change. In order to account for this and respond to the need for model change, we must track model performance. Measuring forecast accuracy also has another use. This is in the model development stage. Evaluating the accuracy of the model on the fitting data helps us to select a model for forecasting.

Many statistical measures can be used to evaluate forecast model performance. Unfortunately, there is little consensus among forecasters as to the best and most reliable forecast-error measures (2). Complicating this issue is that different error measures often provide conflicting results. Different forecast-error measures each have their shortcomings but provide unique information to the forecaster. Knowing when to rely on which measure can be highly beneficial for the forecaster.

Most forecast-error measures can be divided into two groups—*standard* and *relative* error measures (1). Some of the more common forecast-error measures in these categories follow, accompanied by specific suggestions with regard to their use.

If $X_t$ is the actual value for time period $t$ and $F_t$ is the forecast for the period $t$, the forecast error for that period can be computed as the difference between the actual and the forecast:

$$e_t = X_t - F_t$$

When evaluating performance for multiple observations, say $n$, there will be $n$ error terms. We can define the following *standard* forecast-error measures:

1. Mean Error:

$$\mathrm{ME} = \sum_{t=1}^{n} e_t/n$$

2. Mean Absolute Deviation:

$$\mathrm{MAD} = \sum_{t=1}^{n} |e_t|/n$$

3. Mean Square Error:

$$\mathrm{MSE} = \sum_{t=1}^{n} (e_t)^2/n$$

4. Root-Mean-Square Error:

$$\mathrm{RMSE} = \left[\sum_{t=1}^{n} (e_t)^2/n\right]^{1/2}$$

Next are some of the most common *relative* forecast error measures:

1. Mean Percentage Error:

$$\mathrm{MPE} = \sum_{t=1}^{n} \mathrm{PE}_t/n$$

where

$$\mathrm{PE} = [(X_t - F_t)/X_t](100)$$

2. Mean Absolute Percentage Error:

$$\mathrm{MAPE} = \sum_{t=1}^{n} |\mathrm{PE}_t|/n$$

### Standard Versus Relative Forecast-Error Measures

Standard error measures, such as mean error (ME) or mean square error (MSE), typically provide the error in the same units as the data. As such, the true magnitude of the error can be difficult to comprehend. For example, the forecast error of 50 units has a completely different meaning if the units are in dollars versus cartons. In addition, having the error in actual units of measurement makes it difficult to compare accuracies across time series or different time periods. In inventory control, for example, units of measure typically vary between series. Some series might be measured in dollars, whereas others are measured in pallets or boxes. When comparing accuracy between series, the results are not meaningful or the series with large numbers may dominate the comparison.

Relative-error measures, which are unit-free, do not have these problems. Because relative error measures are based on percentages, they are easy to understand. Also, relative-error measures make comparisons across different time series or different time intervals meaningful. However, these error measures are not without shortcomings. Because these measures are defined as a ratio, problems arise in the computation of values that are zero or close to zero. Mean absolute percentage error (MAPE) is one of the most popular of the relative-error measures.

### Error Measures Based on Absolute Values

Error measures that use absolute values, such as the mean absolute deviation (MAD) do not have the problem of errors of opposite signs canceling themselves out. For example, a low mean error may mislead the forecaster into thinking that the overall error is low, when in fact, high and low forecasts may be canceling each other out. This problem is avoided with absolute error measures. The typical shortcomings of these error measures is that they assume a symmetrical loss function. The forecaster is provided with the total magnitude of error but does not know the true bias or direction of that error.

When using error measures based on absolute values, it is also beneficial to compute an error measure of bias, such as mean error or mean percentage error (MPE). These error measures provide the direction of the error, which is a tendency of the model to over- or underforecast. It is very common for forecasters to have a biased forecast, particularly when qualitative forecasting methods are used. Frequently this may be in line with the organizational incentive system, such as being evaluated against a quota. Measuring the degree of bias is important because the forecast can then be adjusted for it. The two pieces of information, the error based on an absolute value as well as a measure of bias, work to complement each other and provide a more complete picture for the forecaster.

### Using Common Error Measures

**Mean Square Error.** Mean square error is an error measure that has particular benefits under certain circumstances. Squaring of error can be advantageous in certain situations as the errors are weighted based on magnitude. Larger errors are given greater weight than smaller errors, which can be quite beneficial in situations when the cost function increases with the square of the error. For example, in inventory control or production planning, larger errors can create costly problems. Overforecasting can lead to higher production and inventory levels. In inventory control, MSE is popular because it can be directly tied to the variability of the forecast errors. This is important for calculating safety stocks in order to cover the variability of demand during the lead time period. In general, this is a good error measure to use in situations when large errors are costly and decision making is very conservative (3).

The disadvantage of MSE is that it is inherently difficult to understand. Sometimes using the root-mean-square error (RMSE), which is simply the square root of MSE, may be preferred because the error is provided in the same units as the data. Like the MSE, the RMSE penalizes errors according to their magnitude. Also, because both MSE and RMSE are not unit-free, comparisons across series are difficult.

**Mean Absolute Deviation.** The mean absolute deviation is an error measure that provides the forecaster with the average total magnitude of error, regardless of sign. As indicated earlier, it is not unit-free, making comparisons across series difficult. Also, it assumes a symmetric loss function. A number of MAD properties can make it attractive for use. First, the following smoothing relationship can be used to approximate the values for MAD:

$$\mathrm{MAD}_t = \alpha|e_{t-1}| + (1-\alpha)\mathrm{MAD}_{t-1}$$

where $\alpha$ is a constant between 0 and 1. This relationship can provide computational advantages, such as requiring less historical data to be retained for each estimate. Also, through the use of $\alpha$, recent forecast performance can be emphasized more than past performance if the forecaster deems it most important.

Second, if forecast errors are normally distributed with a mean of 0, there is a simple relationship between the RMSE and MAD. Though this is only an approximation, it makes it easy to switch from one error measure to the other:

$$\mathrm{RMSE} = 0.8\,\mathrm{MAD}$$

**Mean Absolute Percentage Error.** The mean absolute percentage error is considered to be one of the most popular error measures among both practitioners and academicians. Makridakis (4) referred to it as "a relative measure that incorporates the best characteristics among the various accuracy criteria." MAPE provides the error in terms of percentages so that it is an easy measure to understand. MAPE is also dimensionless, allowing for comparison across different time series and time periods.

### Other Useful Error Measures

**Theil's U Statistic.** One useful way of evaluating forecast performance is to compare accuracy against a baseline forecast. A forecasting technique that commonly serves as a baseline is the Naive model or random walk, which is nothing more than last period's actual serving as next period's forecast. The idea is that a chosen forecasting model must perform better than Naive in order to justify its use. The accuracy of multiple forecasting procedures can be compared with this baseline.

One statistic that performs an automatic comparison against the Naive model and, much like MSE, considers the disproportionate cost of large errors is Theil's $U$ statistic. The statistic allows a relative comparison of formal forecasting methods with Naive and also squares the errors involved so that large errors are given much more weight than small errors. Theil's $U$ statistic can be difficult to understand intuitively, and readers who are interested in the mathematical definition are referred to Ref. 1 and 5. For practicing forecasters, the interpretation of the value of this statistic is significant because it falls into easily interpreted ranges. A Theil's $U$ statistic equal to 1 means that the forecasting model being evaluated is equal in performance to the Naive model. A Theil's $U$ statistic greater than 1 indicates that the Naive model produces better results than your model. Finally, a Theil's $U$ less than 1 indicates that the forecasting model evaluated is providing better forecasts than Naive. Most statistical and forecasting software packages provide Theil's $U$ statistic, and the easy range of interpretation makes this statistic quite valuable.

## SMOOTHING FORECASTING MODELS

The first forecasting models to be discussed belong to a category known as smoothing models. Smoothing models are based on a simple weighing or smoothing of past observations in a time series in order to obtain a forecast of the future. Through the process of averaging of historical values, random errors are averaged in order to provide a "smooth" forecast. Smoothing models are one of the most popular groups of quantitative forecasting models, finding numerous applications, particularly for short to medium range forecasting.

### The Mean

The simplest smoothing model available is the mean, or the simple average. Given a data set covering $N$ time periods, $X_1, X_2, \ldots, X_n$, the forecast for next time period $t + 1$ is given as

$$F_{t+1} = \sum_{i=1}^{T} X_i/T$$

This model is useful only for horizontal data patterns. As the mean becomes based on a larger and larger historical data set, forecasts become more stable. One of the advantages of this model is that only two historical pieces of information need to be carried, the mean itself and the number of observations the mean was based on.

### Simple Moving Average

When using the mean to forecast, one way to control the influence of past data is to specify at the outset how many observations will be included in the mean. This process is described by the term *moving average* because as each new observation becomes available, the oldest observation is dropped, and a new average is computed. The number of observations in the average is kept constant and includes the most recent observations. Like the simple mean, this model is good only for forecasting horizontal, nonseasonal data and is not able to forecast data with trend or seasonality.

Using a moving average for forecasting is quite simple. Given $M$ data points and a decision to use $T$ observations for each average, the simple moving average is computed as follows:

| Time | Forecast |
|---|---|
| $T$ | $F_{t+1} = \sum_{i=1}^{T} X_i/T$ |
| $T+1$ | $F_{t+2} = \sum_{i=2}^{T+1} X_i/T$ |
| $T+2$ | $F_{t+3} = \sum_{i=3}^{T+2} X_i/T$ |

The decision on how many periods to include in the moving average is important, and several conflicting effects need to be considered. In general, the greater the number of observations in the moving average, the greater the smoothing on the random elements. However, if there is a change in data pattern, such as a trend, the larger the number of observations in the moving average, the more the forecast will lag this pattern.

### Exponential Smoothing Models

This section describes a class of models called exponential smoothing models. These models are characterized by exponentially decreasing weights placed on progressively older observations. They are based on the premise that the importance of past data diminishes as the past becomes more distant.

Exponential smoothing models are the most used of all forecasting techniques and are an integral part of many computerized forecasting software programs. They are widely used for forecasting in practice, particularly in production and inventory control environments. There are many reasons for their widespread use. First, these models have been shown to produce accurate forecasts under many conditions (6). Second, model formulation is relatively easy, and the user can understand how the model works. Finally, little computation is required to use the model, and computer storage requirements are quite small.

**Single Exponential Smoothing.** The simplest case of exponential smoothing models is single exponential smoothing (SES). Forecasts using SES are generated as follows:

$$F_{t+1} = \alpha X_t + (1-\alpha)F_t \tag{1}$$

where $F_{t+1}$ and $F_t$ are next period's and this period's forecasts, respectively. $X_t$ is this period's actual observation, and $\alpha$ is a smoothing constant that can theoretically vary between 0 and 1. Selection of $\alpha$, which is discussed later, is a critical component to generating good forecasts. The implication of exponential smoothing can be seen if Eq. (1) is expanded to include past components:

$$\begin{aligned} F_{t+1} &= \alpha X_t + (1-\alpha)[\alpha X_{t-1} + (1-\alpha)F_{t-1}] \\ &= \alpha X_t + \alpha(1-\alpha)X_{t-1} + (1-\alpha)^2 F_{t-1} \\ &= \alpha X_t + \alpha(1-\alpha)X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} \\ &\quad + \alpha(1-\alpha)^3 X_{t-3} + \cdots + \alpha(1-\alpha)^{N-1} X_{t-(N-1)} \end{aligned} \tag{2}$$

An alternative way of writing Eq. (2) follows:

$$\begin{aligned} F_{t+1} &= F_t + \alpha(X_t - F_t) \\ F_{t+1} &= F_t + \alpha e_t \end{aligned} \tag{3}$$

where $e_t$ is the forecast error for period $t$. This provides another interpretation of SES. It can be seen that the forecast provided through SES is simply the old forecast plus an adjustment for the error that occurred in the last forecast. When $\alpha$ is close to 1, the new forecast includes a large adjustment for the error. The opposite is true when $\alpha$ is close to 0. The new forecast will include very little adjustment. These equations demonstrate that SES has a built-in self-adjusting mechanism using the basic principle of negative feedback. The past forecast error is used to correct the next forecast in a direction opposite to that of the error, the same principle used to adjust thermostats and automatic pilots.

Equations (1) and (3) also demonstrate that the best this forecasting model can do is to develop the next forecast from some percentage of error. As such, SES is appropriate only for horizontal, nonseasonal data and is not appropriate for data containing trend because the forecasts will always lag the trended data.

**Selection of the Smoothing Constant $\alpha$.** As indicated earlier, the proper selection of $\alpha$ is a critical component to generating good forecasts with exponential smoothing. High values of $\alpha$ will generate responsive forecasts but will not offer much data smoothing. On the other hand, low $\alpha$ values will not allow the model to respond rapidly to changes in data pattern.

There are a number of ways to select $\alpha$. A common approach is to select $\alpha$ in such a way so that some criteria, such as MSE, is minimized over the initialization set in the fitting stage of model development (7). Another approach is to use what is known as adaptive-response-rate single exponential smoothing (ARRSES), which allows $\alpha$ to change as changes in the data pattern occur (8). This adaptive approach allows $\alpha$ to change automatically based on the distribution of past errors, making $\alpha$ more responsive or stable, based on the pattern in the data. The basic equation for exponential smoothing is the same, except that $\alpha$ is replaced by $\alpha_t$:

$$F_{t+1} = \alpha_t X_t + (1-\alpha_t)F_t \tag{4}$$

where

$$\alpha_{t+1} = |E_t/M_t| \quad (5)$$

$$E_t = \beta e_t + (1-\beta)E_{t-1} \quad (6)$$

$$M_t = \beta|e_t| + (1-\beta)M_{t-1} \quad (7)$$

$$e_t = X_t - F_t \quad (8)$$

where both $\alpha$ and $\beta$ are parameters between 0 and 1.

Equation (5) shows that $\alpha$ is made equal to the absolute value of the ratio of smoothed error $E_t$ over the smoothed absolute error $M_t$. $E_t$ and $M_t$ are obtained through Eqs. (6) and (7), and the error is defined by Eq. (8). Through the distributions of past errors, $\alpha$ is automatically adjusted from period to period. The reader is referred to Ref. 8 for a description of this process.

**Holt's Two-Parameter Model.** Holt's two-parameter model, also known as linear exponential smoothing, is one of many models applicable for forecasting data with a trend pattern (9). As noted earlier, horizontal models will generate forecasts that will lag trended data. Trend models have some mechanisms that allows for tracking of trend and adjusting the level of the forecast to compensate for the trend. Holt's model does this through the development of a separate trend equation that is added to the basic smoothing equation to generate the final forecast. The series of equations follows:

1. Overall smoothing

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + b_{t-1}) \quad (9)$$

2. Trend smoothing

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad (10)$$

3. Forecast

$$F_{t+m} = S_t + b_t m \quad (11)$$

In Eq. (9), the last smoothed value $S_{t-1}$ is directly adjusted for last period's trend $b_{t-1}$ to generate next period's trend $S_t$. This is the technique that helps bring up the value of $S_t$ to the level of trend and eliminate any lagging. The level of trend is updated over time through Eq. (10), where the trend is expressed as the difference between the last two smoothed values. The form of Eq. (10) is the basic single smoothing equation applied to trend. Much like $\alpha$, the coefficient $\gamma$ is used to smooth out the randomness in the trend. Finally, Eq. (11) is used to generate forecasts. The trend $b_t$ is multiplied by $m$, the number of periods ahead to be forecast, and added to the base value $S_t$.

**Winters' Three-Parameter Trend and Seasonality Model.** As indicated earlier in this chapter, a critical part of forecasting is to match the forecasting model to the characteristic patterns of the time series being forecast. If the data are horizontal and nonseasonal, then models such as the mean, moving averages, or SES would be the models of choice. If the data have a trend present, then Holt's linear model or any one of a number of trend models (1) could be selected.

Winters' model is just one of several models that is appropriate for seasonal data. It is based on three smoothing equations—one for stationarity of the series, one for trend, and one for seasonality. The equations of this model are similar to Holt's model, with the addition of an equation to deal with seasonality. The model is described as follows:

1. Overall smoothing

$$S_t = \alpha X_t/I_{t-L} + (1-\alpha)(S_{t-1} + b_{t-1}) \quad (12)$$

2. Trend smoothing

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad (13)$$

3. Seasonal smoothing

$$I_t = \beta X_t/S_t + (1-\beta)I_{t-L} \quad (14)$$

4. Forecast

$$F_{t+m} = (S_t + b_t m)I_{t-L+m} \quad (15)$$

Equations (12)–(15) are similar to Holt's equations, with a few exceptions. Here, $L$ is the length of seasonality, such as the number of months or quarters in a year and $I_t$ is the corresponding seasonal adjustment factor. As in Holt's model, the trend component is given by $b_t$, and the forecast for $m$ periods ahead is $F_{t+m}$. Equation (14) is the seasonal smoothing equation that is comparable to a seasonal index that is found as a ratio of the current values of the series $X_t$, divided by the current single smoothed value for the series $S_t$. When $X_t$ is larger than $S_t$, the ratio is greater than 1. The opposite is true when $X_t$ is smaller than $S_t$, when the ratio will be less than 1. It is important to understand that $S_t$ is a smoothed value of the series that does not include seasonality. The data values $X_t$, on the other hand, do contain seasonality, which is why they are deseasonalized in Eq. (12). $X_t$ also contains randomness, which Eq. (14) smooths out through $\beta$, allowing us to weight the newly computed seasonal factor with the most recent seasonal number.

As with other smoothing models, one of the problems in using Winters' method is to determine the values of parameters $\alpha$, $\beta$, and $\gamma$. The approach for determining these values is the same as for selecting parameters for other smoothing procedures. Trial and error on historical data is one approach that can be used. Another option is to use a nonlinear optimization algorithm to give optimal parameter values that minimize MSE or MAPE. The reader is referred to Ref. 10 for more information on this method.

## AUTOREGRESSIVE/MOVING AVERAGE FORECASTING MODELS

Autoregressive/moving average (ARMA) models are another category of forecasting models that are in many ways similar to smoothing methods in that they are based on historical time-series analysis. However, ARMA models have a unique approach to identifying the patterns in historical time series and extrapolating those into the future. These models are fairly complex, which has, in many cases, hindered their widespread use. Nevertheless, ARMA models have a strong theoretical and statistical appeal. Over the years, many useful guidelines for the use have been developed; the guidelines have made using these models much easier (11–13). Autoregressive/moving average models are actually a combination of two separate models: autoregressive models and moving average models.

### Autoregressive Models

The general class of autoregressive (AR) models take on the following form:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \phi_3 X_{t-3} + \cdots + \phi_p X_{t-p} + \epsilon_t \tag{16}$$

The forecast is formed from the time-lagged values of the dependent variable; therefore, we have the name autoregression. The general AR($p$) equation can take on a number of forms depending upon the order of $p$. When $p = 1$, it is a first-order AR model or AR(1). The first step in using an AR model is to identify its order $p$, which specifies the number of terms to be included in the model. This is achieved through an examination of the autocorrelation coefficients.

Application of the autoregressive equation also requires estimates for the values of the autoregressive parameters. The method of *adaptive filtering* can be applied to an AR model to estimate parameter values. Through this procedure, parameter values are estimated with a nonlinear least-squares approach using the method of steepest descent to minimize MSE. This method starts with an initial set of $\Phi_i$ values and proceeds to adjust them based on the following equation:

$$\begin{aligned} \phi'_{it} &= \phi_{i,t-1} + 2Ke_t X_{t-i} \\ i &= 1, 2, \ldots, p \\ t &= p+1, p+2, \ldots, n \end{aligned} \tag{17}$$

where $\Phi'_{it}$ is the new adapted parameter, $\Phi_{it-1}$ is the old parameter, and $K$ is the learning constant that controls the speed of adaptation. As before, $e_t$ and $X_{t-i}$ are the residual error and time-series value at period $t - i$, respectively. The method of adaptive filtering allows the parameters to adjust over time in a similar manner that ARRES adjusts $\alpha$ over time in exponential smoothing.

### Moving Average Models

AR($p$) models cannot always isolate all patterns, particularly when $p$ is fairly small. Another type of model, called a *moving average* (MA) model can be used in this case to either substitute or supplement an AR($p$) model. In contrast to AR models, which express $X_t$ as a linear function of $p$ actual values of $X_t$, MA models use a linear combination of past errors to generate a forecast. The general MA model is

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \tag{18}$$

Even though this equation is called a moving average model in the literature, it has no relationship to the moving average models discussed earlier. As with AR($p$) models, the issue of parameter selection is important and; the method of adaptive filtering can be used to find optimal parameters for an MA($q$) model.

### Mixed Autoregressive Moving Average Models

AR($p$) and MA($q$) models can be mixed together in the same equation to form an autoregressive moving average model. ARMA models are defined by the order $p$ and $q$, which is shown in the following equation:

$$\begin{aligned} X_t = {} & \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} \\ & + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \end{aligned} \tag{19}$$

An ARMA(1,1) model is

$$X_t = \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1}$$

ARMA models are quite comprehensive in nature, and their performance is generally superior to that obtained by using an AR or MA model separately. Because of their accuracy, ARMA models have been used widely in practice. As with separate AR and MA models, optimizing parameter values using the steepest descent method can be applied to mixed ARMA models. The adaptive filtering procedure discussed earlier can also be applied to mixed ARMA models. Here it is referred to as *generalized adaptive filtering* (14).

### The Box-Jenkins Method

George Box and Gwilym Jenkins (13) have studied ARMA models extensively, and their names have frequently been used synonymously with general ARMA processes. Box and Jenkins (13) have put together the relevant information required to understand ARMA processes in a comprehensive manner. Their methodology consists of the following four steps. The first step is *model identification* and involves identifying a tentative model by using autocorrelations and partial autocorrelations. After a model has been identified, the second step is *estimation of model parameters.* The third step is *diagnostic checking* where an evaluation is made of the adequacy of the model identified. Finally, the last step is the actual *forecasting.* This methodology is iterative in that the cycle of model identification, parameter estimation, and diagnostic checking is repeated until a satisfactory model is identified (see References 11–13 for more details).

## PROBABILISTIC FORECASTING

Any forecast of future events can be viewed as a hypothesis or conjecture about the future. As such, a forecast always contains some degree of uncertainty. Many forecasts appear deterministic to their users when in fact they are highly conditional, based on the historical data and the underlying assumptions used to generate them. Forecasts are therefore much more accurately described through parametrized distributions rather than by fixed statements. Probabilistic forecasting is a methodology that allows us to generate a forecast in this manner. Although more complex, these techniques allow us to more accurately capture and represent forecasting problems.

### Bayesian Forecasting and Dynamic Models

Bayesian statistics is the foundation of probabilistic forecasting and is based on the premise that all uncertainties are represented and measured by probabilities. Based on the laws of probability, the Bayesian paradigm provides rules for managing these uncertainties. These laws of probability can be applied to produce probabilistic inferences about any quantity of interest. In forecasting, the quantities of interest may be future values of a time series or values of variables used to model the time series. Bayesian forecasting allows us to model forecast information as probability distributions that represent uncertainty. Forecasts are then derived from such models as predictive probability distributions. Throughout this process, keep in mind that these distributions represent uncertain knowledge and that all probabilities are subjective beliefs of the forecaster or modeler responsible for providing forecast information.

To illustrate how relationships can be modeled through Bayesian processes, assume that the output variable $Y$ and input variable $X$ are related through the following general form:

$$Y = X\theta + \epsilon \tag{20}$$

where $\theta$ and $\epsilon$ represent an uncertain parameter and uncertain random error term, respectively. The forecaster's beliefs about the parameter $\theta$ can be expressed through a probability distribution $P(\theta)$.

However, Eq. 20 does not account for the dynamic nature of processes that occur over time, and we say that the form is only *locally* appropriate. As time passes, $\theta$ may take on different values, or the form defining the process may even change. A methodology that allows us to change processes because of the passage of time is referred to as dynamic modeling. The most common class of dynamic models are dynamic linear models (DLMs) (15,16).

To illustrate how dynamic models work, we can assume that at any given time, a dynamic model $M$ consists of possible models M and that the forecaster's uncertainty is described through a prior distribution $P(\mathrm{M})$, $(\mathrm{M} \in M)$. In producing a forecast for output $Y$, at any time $t$, each member model M provides a conditional forecast in terms of a probability distribution $P(Y|\mathrm{M})$, where M directly relates to the parametrization $\theta$. The forecast from the dynamic model M can then be defined as the following marginal probability distribution:

$$P(Y) = \int_{\mathrm{M}\in M} P(Y)/(\mathrm{M})\, dP(\mathrm{M}) \tag{21}$$

## Structuring Dynamic Models

The Bayesian methodology and dynamic modeling allow for changes in model form to take place as new information becomes available over time. Modeling forecasting problems using these methodologies first involves defining the sequential model and structuring parametric model forms. Next, probabilistic representation of information about parameters is necessary. Forecasts are then derived as probability distributions. As time evolves, new information relevant to forecasting the future is received and may be used in revising the forecaster's views. This revision can be at the quantitative level, the model form level, or even the conceptual level of the general model structure. This sequential approach generates statements about future values of a time series conditional on existing information.

This process can mathematically be described as follows. Assume that $Y_t$ denotes the $t$th value of a series. At $t = 0$, which is the current time, we can assume that the initial information set available to the forecaster is denoted by $D_0$. The primary objective in forecasting is to calculate the forecast distribution for $(Y_t|D_0)$ when $t > 0$. As time evolves, statements at any time $t$ about the future are conditional on the existing information set at that unique time $D_t$. To generalize, forecasting ahead to any time $s > t$ involves calculating the forecast distribution for $(Y_s|D_t)$ where $D_t$ includes both the previous information set $D_{t-1}$ and the observation $Y_t$, namely $[D_t = \{Y_t, D_{t-1}\}]$.

A parametric model can then be used to formulate the beliefs of the forecaster as

$$P(Y_t|\theta_t, D_{t-1})$$

where $\theta_t$ is a defining parameter vector at time $t$. Information relevant to forecasting the future is summarized through parameter $\theta_t$ and used in forming forecast distributions. The sequential revising of the state of knowledge about such parameters over time creates the learning process of the dynamic model. This transfer of information through time occurs through a prior distribution $P(\theta_t|D_{t-1})$ and posterior distribution $P(\theta_t|D_{t;})$. At time $t$, prior to actual observation of $Y_t$, the historical information $D_{t+1}$ is summarized through a prior distribution $P(\theta_t|D_{t;-1})$.

The following joint distribution can be used to describe the relationship of these parameters and observations:

$$P(Y_t, \theta_t|D_{t-1}) = P(Y_t|\theta_t, D_{t-1})P(\theta_t|D_{t-1}) \tag{22}$$

Finally, the desired forecast can be developed from this as

$$P(Y_t|D_{t-1}) = \int P(Y_t, \theta_t|D_{t-1})\, d\theta_t \tag{23}$$

Inferences about the future $Y_t$ are made by summarizing information contained in the forecast distribution.

## Types of Dynamic Models

**The First-Order Polynomial Model.** The general class of dynamic linear models can be exemplified by two simple model structures. The first DLM is the first-order polynomial model. For any time $t$, this model can be described as follows:

$$Y_t = \mu_t + v_t \tag{24}$$

where $v_t \sim N[0, V_t]$. The level of the series at time is given as $\mu_t$, and $v_t$ is the random error or noise about the underlying level. This system can be modeled as changing through time using a random walk:

$$\mu_t = \mu_{t-1} + w_t \tag{25}$$

where $w_t \sim N[0, W_t]$ and represents random changes in level between time $t-1$ and $t$. Initial information available to the forecaster is assumed as

$$(\mu_0|D_0) \sim N[m_0, C_0]$$

This last formulation is a probabilistic representation of the forecaster's beliefs about the level at time $t = 0$ given available information $D_0$. The mean $m_0$ and variance $C_0$ are estimates of the level and a measure of the uncertainty about the mean. The only new information becoming available at any time is the value of the time-series observation so that $D_t = \{Y_t, D_{t-1}\}$. In this formulation, the error sequences $v_t$ and $w_t$ are assumed to be independent over time, mutally independent, as well as independent of $(\mu_0|D_0)$.

The components of this model are represented as distributions, which are sequentially updated over time as new information becomes available. Although simple, this model type has found wide application in short-term forecasting, such as forecasting product demand and inventory levels.

**The Dynamic Regression Model.** The second general DLM can be applied in the context of regression modeling where we are concerned with quantitatively modeling relationships between variables, such as that existing between two time series. If we assume that time series $X_t(t = 1, 2, \ldots, n)$ is observed contemporaneously with $Y_t$, in regression modeling we typically focus on the extent to which changes in the mean $\mu_t$ of $Y_t$ can be explained through $X_t$. $Y_t$ is generally referred to as the dependent or response variable and $X_t$, as the independent variable or regressor. The mean response $\mu_t$ is then related to the regressor variable through a mean response function $\mu_t = f(X_t, X_{t-1}, \ldots)$.

This function can be modeled as a simple linear model of the following form:

$$\mu_t = \alpha + \beta X_t \tag{26}$$

with defining parameters $\alpha$ and $\beta$. However, we say that this linear model is only adequate *locally* but not *globally* because it may not describe the change in the preceding relationships as time evolves and $X_t$ varies. This flexibility is provided by allowing for the probability of time variation in the coefficients, namely

$$\mu_t = \alpha_t + \beta_t X_t \tag{27}$$

The formulation of Eq. (27) allows for the model to have different defining parameters at different points in time. The variation of parameters through time can be modeled through random walk-type evolutions such as

$$\alpha_t = \alpha_{t-1} + w_1 \tag{28}$$

$$\beta_t = \beta_{t-1} + w_2 \tag{29}$$

where $w_1$ and $w_2$ are zero-mean error terms. Again, the components of this model are distributions updated sequentially over time. This basic linear model can be further expanded to include a multiple regression DLM.

**Vector Modeling.** The general DLM can be expanded to a multivariate DLM for a time series of vector observations $\boldsymbol{Y}_t$ where $\boldsymbol{Y}_t$ is an $(r \times 1)$ column vector. According to West and Harrison (15), the multivariate DLM is characterized by a quadruple:

$$\{\boldsymbol{F}, \boldsymbol{G}, \boldsymbol{V}, \boldsymbol{W}\}_t = \{\boldsymbol{F}_t, \boldsymbol{G}_t, \boldsymbol{V}_t, \boldsymbol{W}_t\}$$

for each time $t$, where

1. $\boldsymbol{F}_t$ is a known $(n \times r)$ dynamic regression matrix.
2. $\boldsymbol{G}_t$ is a known $(n \times n)$ state evolution matrix.
3. $\boldsymbol{V}_t$ is a known $(r \times r)$ observational variance matrix.
4. $\boldsymbol{W}_t$ is a known $(n \times n)$ evolution variance matrix.

$\boldsymbol{Y}_t$ is related to the $(n \times 1)$ parameter vector $\theta_t$ through a model that is defined by this quadruple. The parameter vector $\theta_t$ is sequentially specified through time in the following manner:

$$(\boldsymbol{Y}_t|\boldsymbol{\theta}_t) \sim N[\boldsymbol{F}_t'\boldsymbol{\theta}_t, \boldsymbol{V}_t]$$
$$(\boldsymbol{\theta}_t|\boldsymbol{\theta}_{t-1}) \sim N[\boldsymbol{G}_t\boldsymbol{\theta}_{t-1}, \boldsymbol{W}_t]$$

As in the scalar case, these equations are conditional on the information set available prior to time $t$, namely $D_{t-1}$. This model can be further specified through the following set of equations:

$$\boldsymbol{Y}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t + \boldsymbol{v}_t, \quad \text{where } \boldsymbol{v}_t \sim N[0, \boldsymbol{V}_t] \tag{30}$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{w}_t, \quad \text{where } \boldsymbol{w}_t \sim N[0, \boldsymbol{W}_t] \tag{31}$$

Equation (30) is the observation equation that defines the sampling distribution of $\boldsymbol{Y}_t$ and is conditional on the quantity $\boldsymbol{\theta}_t$. $\boldsymbol{F}_t$ is a regression matrix of known values of independent variables, and $\boldsymbol{\theta}_t$ is the dynamic vector of regression parameters known as the state vector or system vector of the model. At time period $t$, the mean response is $\boldsymbol{\mu}_t = \boldsymbol{F}_t'\boldsymbol{\theta}_t$ or the expected value of $\boldsymbol{Y}_t$, which defines the level of the series at time $t$. As in the scalar case, the term $\boldsymbol{v}_t$ is the observational error at time $t$.

Equation (30) is the evolutionary equation enabling the evolution of the state vector through a one-step Markov process. Through this equation, the distribution of $\boldsymbol{\theta}_t$ is determined solely based on $\boldsymbol{\theta}_{t-1}$ and the known values of $\boldsymbol{G}_t$ and $\boldsymbol{w}_t$ and is determined independently of values of the state vector and data prior to time $t - 1$. The transition of $\boldsymbol{\theta}_t$ over time is enabled through the use of the evolution transfer matrix $\boldsymbol{G}_t$. Finally, the term $\boldsymbol{w}_t$ is the evolution error, with the known evolution variance $\boldsymbol{W}_t$.

## DYNAMIC MODELING WITH BELIEF NETWORKS

Probabilistic dependencies and nonlinearities, which are characteristic of many real-world problems, are difficult to model with classical time-series methodologies. An approach to forecasting and decision making that has shown success along these lines is the use of graphical models of decision theory known as influence diagrams or belief networks. Interest and use of belief networks has attracted decision modelers and forecasters, as well as designers of knowledge-based systems. These models come from research in artificial intelligence and decision analysis and are the basis of diagnostic systems for many real-world applications (17,18). From a theoretical perspective, they combine graph theory, probability theory, and decision theory. Many techniques for probabilistic inference in belief networks and for their specification have been developed as a response to their increased use (19–25).

Belief networks are graphical representations of probabilistic dependencies among domain variables. A belief network consists of a directed acyclic graph (DAG) and a set of conditional probability functions that model the conditional interdependence in multivariate systems. The nodes of the DAG represent the variables of the belief network. The directed arcs in the DAG represent explicit dependencies between the variables. Let $X_1, \ldots, X_n$ represent the nodes of the DAG and let $\pi(X_i)$ denote the set of parents of each variable $X_i$ in the DAG. Then for each variable $X_i$ in the belief network, we can specify a conditional probability function as

$$P[X_i|\pi(X_i)]$$

The full joint probability distribution is then given as (26,27)

$$P[X_1,\ldots,X_n] = \prod_{i=1}^{n} P[X_i|\pi(X_i)] \tag{32}$$

According to Dagum et al. (28), probabilistic inference in belief networks entails the computation of an inference probability that is $P[\boldsymbol{X}= x|\boldsymbol{E} = \boldsymbol{e}]$ for any given set of nodes $\boldsymbol{X}$ instantiated to value $x$ and conditioned on observation nodes $\boldsymbol{E}$ instantiated to value $e$. Even though this probabilistic inference can be difficult for large and complex belief networks, there are inference approximation procedures that can provide estimates of posterior probabilities.

In developing belief networks, two tasks are required. The first is identification of the dependency structure of the model, representing the set of causal relationships between domain variables. Here probability distributions are used to infer relationships and causality between domain variables. This is in contrast to classical time-series models, AR models, dynamic linear models, or transfer-function modes, which use cross correlations between the variables to construct the model. The second task in belief-network development is specification of the conditional probabilities. These are typically derived using maximum-likelihood estimates from time-series data.

### Additive Belief-Network Models

Using belief networks for forecasting can pose some difficulties. The main disadvantage has to do with large storage and computational requirements that occur with complex problems such as those containing multivariate time series with multiple lagged dependencies. The need to overcome these problems encountered in large belief-network applications has lead to the development of additive beliefs networks models (28,29). Additive belief-network models belong to a more general class of additive models that approximate multivariate functions by sums of univariate functions. As such, additive belief network models can reduce the specification of a large contingency table into the specification of a few small tables, substantially improving the efficiency of computation.

Additive belief-network models possess the same properties as other separable models where the joint effect of a set of causes $X_1, \ldots, X_p$ on the dependent variable $Y$ can be expressed in terms of the effects of each individual cause. Here we can assume that for each cause $X_i$, there exists an off state in which $X_i$ has no bearing on the value of $Y$. If these distinguished states are denoted by $s_i^*$, the conditional probabilities $P[Y|X_i, X_{j\neq i} = x_j^*]$for $i = 1, \ldots, p$, represent the isolated effects of each $X_i$ on $Y$. The joint effects of the causes on the dependent variables can then be expressed in terms of sets of causes $\boldsymbol{X}_i$, $i = 1, \ldots, k$, that partition the set $\{X_1, \ldots, X_p\}$. The isolated effects of each set $\boldsymbol{X}_i$ on $Y$ is represented by the conditional probability $P[Y|\boldsymbol{X}_i, \boldsymbol{X}_{j\neq i} = \boldsymbol{x}_j^*]$.

If we let $y^*$ denote the off state of the variable $Y$, an additive belief-network model is a separable model that satisfies

$$\begin{aligned} P[Y = y|X_1,\ldots,X_p] &\\ &= \sum_{i=1}^{k} \phi_i P[Y = y|\boldsymbol{X}_i, \boldsymbol{X}_{j\neq i} = \boldsymbol{x}_j^*] \quad \text{if } y \neq y* \\ &= 1 - \sum_{y'\neq y^*} P[Y = y'|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_k] \quad \text{if } y = y^* \end{aligned} \tag{33}$$

The parameters $\phi_i \geq 0$, for $i = 1, \ldots, k$, must satisfy

$$\sum_{i=1}^{k} \phi_i P[Y|\boldsymbol{X}_i, \boldsymbol{X}_{j\neq i} = x_j^*] \leq 1 \tag{34}$$

Dagum et al. (28) show that to specify the conditional probabilities $P[Y|X_1, \ldots, X_p]$ of an additive belief-network model, like with other separable models, only the conditional probabilities of the $k$ isolated effects need to be specified. For example, the size of the conditional probability table for a binary-valued belief network is reduced from $2^{p+1}$ to $\sum_{i=1}^{k} 2^{|Xi|+1}$. In addition to this reduction in the size of the conditional probability table, additive models improve the efficiency of the belief-network influence algorithm.

### Temporal Belief-Network Models

Modeling dynamic domains temporally is possible with dynamic network models (DNMs), which are based on the integration of Bayesian time-series analysis with belief-network representation and inference techniques (30). DNMs can be used to structure forecasting models capable of capturing explicit domain dependencies. DNMs have all the capabilities inherent in belief networks and are therefore well suited for domains with categorical variables. The causal relationships between these variables and their dependencies are represented through the graphical structure of the DNM.

DNMs are additive belief-network models with variables indexed by time. The conditional probabilities of the model can be expressed using the same additive decomposition as with additive belief networks. However, after each new observation, the parameters of the decomposition are reestimated. As with belief networks, the first step in constructing a DNM is to identify the dependencies among domain variables in the model. These are then used to specify the directed acyclic graph of the model. For example, assuming that a single variable $Y_t$ is dependent on the set of variables $X_{t-i} = \{X_{1,t-i}, \ldots, X_{m,t-i}\}$, we can specify the explicit dependencies between the domain variables. Next, the conditional probability for the DAG of the node $Y_i$ is specified. This conditional probability can be specified using the same additive decomposition illustrated in Eq. (33):

$$\begin{aligned} P[Y_t = y|X_t,\ldots,X_{t-k}] &\\ &= \sum \phi_{ti} P[Y_t = y|\boldsymbol{X}_{t-i}, \boldsymbol{X}_{t-j,j\neq i} = \boldsymbol{x}j_{t-j}^*] \quad \text{if } y \neq y^* \\ &= 1 - \sum_{y'\neq y^*} P[Y_t = y'|\boldsymbol{X}_t,\ldots,\boldsymbol{X}_{t-k}] \quad \text{if } y = y^* \end{aligned} \tag{35}$$

Either expert assessment or maximum likelihood estimates can be used to estimate the conditional probabilities $P[Y_t|\boldsymbol{X}_{t-i}, \boldsymbol{X}_{t-j,j\neq i} = \boldsymbol{x}^*_{t-j}]$.

Through reestimation of parameters $\phi_{t-1}, \ldots, \phi_{t-k}$, the additive decomposition provides a means of updating the conditional probabilities with new information. Dagum et al. (28) show that forecasting using a DNM reduces to probabilistic inference in the forecast model, which yields probability distributions for the forecast nodes.

**Application of Belief Networks.** A number of applications of probabilistic reasoning about change over time and temporal reasoning using belief networks and influence diagrams have been provided. They have found great applicability in modeling situations where modelers need to coordinate hard data with data available only from expert judgment. Belief networks have been used in many diagnostic reasoning systems to assign probabilities to alternative hypotheses, such as about a patient's health or about a source of failure in complex machinery. Real-world applications of forecasting with belief networks have included forecasting crude-oil prices (31,32) and predicting outcome in critically ill patients (28).

## ORGANIZATIONAL FORECASTING

The vast majority of this article has been dedicated toward quantitative or statistical models. As this article has demonstrated, many statistical forecasting procedures have been developed and tested. They certainly provide forecasters with great technique choice and better guidelines for use than ever before (33). Despite these advances in the field of formal forecasting, surveys of forecasting practices in business consistently show that qualitative forecasting methods continue to be used in practice more frequently than statistical methods (34–37). Also, these surveys show that the more sophisticated statistical methods are used less than the simpler methods. When quantitative methods are used, they are frequently adjusted by practitioners to include "inside knowledge" (36).

There are a number of reasons for the heavy reliance on qualitative methods. Practitioners may view a mathematical model as a "black box" that is not fully understood; with this attitude users may be reluctant to use mathematical models (38). Also, practitioners may believe that qualitative methods provide a certain advantage because they allow the incorporation of outside information exogenous to the model (39,40). Both qualitative and quantitative methodologies have their advantages and shortcoming as already discussed.

### Qualitative Versus Quantitative Forecasting Models

The high reliance of business practitioners on qualitative forecasting methods is often disheartening to academicians. Qualitative methods are based on judgment and are highly subjective. As such, they are subject to numerous shortcomings. A large portion of the forecasting literature has pointed out the information processing limitations and many biases inherent in human decision making (41,42). Many of these limitations are obvious, such as limited attention span, limited processing ability, and short-term memory. Human decision making is also subject to many biases, such as inconsistency, selective perception, illusory correlation, and overoptimism (41). Because of problems in human decision making, qualitative forecasts are never consistent and are prone to forecast errors. Also, because of limited information processing ability of humans, qualitative forecasts cannot consider the volume of information possible with quantitative models.

Quantitative forecasting models, on the other hand, are always consistent. This means that, for the same set of data, the same model will always generate the same forecast. Also, these models can process large amounts of information. However, quantitative forecasts will only be as good as the quality of the data upon which they are based.

Given the differences between qualitative and quantitative forecasting models, it is clear that both methodologies have a place in the forecasting process. However, each of these models should be used for different purposes based on their strengths and weaknesses. Because quantitative models are always consistent and can incorporate much information, they should be the primary tool for forecasting. However, there are exceptions to using quantitative models. The first is when meaningful data are unavailable, as is often the case for forecasting demand for a new product, for long-range strategic forecasting, or in new technology fields where qualitative forecasts are the only alternative. Another reason is when practitioners have certain inside knowledge of their environment, which may be difficult or overly costly to incorporate into a forecasting model. Examples may include knowledge of planned advertising campaigns by a competitor; a change in management, which may call for a change in policy; changes in general purchasing patterns of customers; or something as simple as the weather, which may delay a shipment of goods. There is ample evidence to suggest that judgmental forecasts are successful in such circumstances (43).

## CONCLUSION

This article covered a wide array of forecasting concepts and methodologies. There are numerous forecasting models, and today's access to computer software provides the forecaster with options never before available. However, it should be evident that the process of forecasting involves much more than merely applying a forecasting model to historical data. The forecaster must understand the context in which the generated forecast will be used, accuracies required, data available for modeling, complexities required, as well as costs involved. Decisions need to be made regarding the types of forecasting methodologies that need to be considered and whether judgmental inputs through qualitative forecasts are needed. Much thought needs to go into analyzing historical data, selecting the correct model and parameters, and monitoring forecast performance over time. Understanding these processes is the first step toward generating good forecasts.

## BIBLIOGRAPHY

1. S. Makridakis, S. Wheelwright, and V. McGee, *Forecasting: Methods and Applications,* 2nd ed., New York: Wiley, 1983.
2. J. S. Armstrong and F. Collopy, Error measures for generalizations about forecasting methods: Empirical comparisons with discussion, *Int. J. Forecasting,* **8**: 69–80, 1992.
3. B. E. Flores, A pragmatic view of accuracy measurement in forecasting, *OMEGA,* **14** (2): 93–98, 1986.

4. S. Makridakis, Accuracy measures: Theoretical and practical concerns, *Int. J. Forecasting,* **9** (4): 527–529, 1993.
5. H. Theil, *Applied Economic Forecasting,* Amsterdam: North-Holland Publishing, 1996.
6. S. Makridakis, A. Andersen, R. Carbone, R. Fildes, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. Winkler, The accuracy of extrapolation (time series) methods: Results of a forecasting competition, *J. Forecasting,* **1**: 111–153, 1982.
7. D. J. Dalrymple and B. E. King, Selecting parameters for short-term forecasting techniques, *Decision Sci.,* **12**: 661–669, 1981.
8. D. W. Trigg and D. H. Leach, Exponential smoothing with an adaptive response rate, *Operational Res. Quart.,* **18**: 53–59, 1976.
9. C. C. Holt, Forecasting seasonal and trends by exponentially weighted moving averages, *Office of Naval Research, Research Memorandum,* No. 52, 1957.
10. P. R. Winters, Forecasting sales by exponentially weighted moving averages, *Management Sci.,* **6**: 324–342, 1960.
11. S. Makridakis and S. Wheelwright, *Forecasting Methods and Applications,* New York: Wiley, 1978.
12. B. L. Bowerman and R. T. O'Connell, *Time Series Forecasting,* 3rd ed., Boston: Duxbury Press, 1993.
13. G. E. P. Box, G. M. Jenkins, and G. C. Reinsel, *Time Series Analysis, Forecasting and Control,* San Francisco: Holden-Day, 1994.
14. S. Wheelwright and S. Makridakis, An examination of the use of adaptive filtering in forecasting, *Operational Res. Quart.,* **24** (1): 55–64, 1973.
15. M. West and J. Harrison, *Bayesian Forecasting and Dynamic Models,* New York: Springer-Verlag, 1989.
16. A. H. Jazwinski, *Stochastic Processes and Filtering Theory,* New York: Academic Press, 1970.
17. E. J. Horvitz, J. S. Breese, and M. Henrion, Decision theory in expert systems and artificial intelligence, *J. Approximate Reasoning,* **2**: 247–302, 1988.
18. E. Charniak, Bayesian networks without tears, *The AI Magazine,* **12** (4): 50–63, 1991.
19. S. Lauritzen and D. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, *J. Roy. Statistical Soc. B,* **50** (19): 157–224, 1988.
20. R. Chavez and G. Cooper, A randomized approximation algorithm for probabilistic inference on Bayesian belief networks, *Networks,* **20**: 661–685, 1990.
21. P. Dagum and R. M. Chavez, Approximating probabilistic inference in Bayesian belief networks, *IEEE Trans. Pattern Anal. Mach. Intell.,* **15** (3): 246–255, 1993.
22. P. Dagum and E. Horvitz, A Bayesian analysis of simulation algorithms for inference in belief networks, *Networks,* **23**: 499–516, 1993.
23. G. Cooper and E. Heskovits, A Bayesian method for the induction of probabilistic networks from data, *Machine Learning,* **9**: 309–347, 1992.
24. J. Pearl and T. Verma, A statistical semantics for causation, *Statistics and Computing,* **2**: 91–95, 1992.
25. I. Matzkevich and B. Abramson, Decision analytic networks in artificial intelligence, *Management Sci.,* **41** (1): 1–22, 1995.
26. J. Peark, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* San Mateo, CA: Morgan Kaufmann, 1991.
27. R. Neapolitan, *Probabilistic Reasoning in Expert Systems,* New York: Wiley, 1990.
28. P. Dagum, A. Galper, E. Horvitz, and A. Seiver, Uncertain reasoning and forecasting, *Int. J. Forecasting,* **11** (1): 73–87, 1995.
29. P. Dagum and A. Galper, Additive belief network models, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence,* Association for Uncertainty in Artificial Intelligence, Washington, DC, pp. 91–98, 1993.
30. P. Dagum and A. Galper, Forecasting sleep apnea with dynamic network models, *Proceedings of the Ninth Conference on Uncertainty in Artificial Intelligence,* Association for Uncertainty in Artificial Intelligence, Washington, DC, pp. 64–71, 1993.
31. B. Abramson and A. J. Finizza, Using belief networks to forecast oil prices, *Int. J. Forecasting,* **7**: 299–316, 1991.
32. B. Abramson and A. J. Finizza, Probabilistic forecasts from probabilistic models: A case study in the oil market, *Int. J. Forecasting,* **11**: 63–72, 1995.
33. J. S. Armstrong, *Long-range Forecasting: From Crystal Ball to Computer,* New York: Wiley, 1995.
34. D. J. Dalrymple, Sales forecasting practices-results from a United States survey, *Int. J. Forecasting,* **4** (3): 51–59, 1987.
35. J. Mentzer and J. Cox, Familiarity, application and performance of sales forecasting techniques, *J. Forecasting,* **3** (1): 27–36, 1984.
36. N. R. Sanders and K. B. Manrodt, Forecasting practices in US corporations: Survey results, *Interfaces,* **24** (2): 91–100, 1994.
37. J. T. Mentzer and K. B. Kahn, Forecasting technique familiarity, satisfaction, usage and application, *J. Forecasting,* **14** (5): 465–476, 1995.
38. B. Kleinmuntz, Why we still use our heads instead of formulas: Towards an integrative approach, *Psychological Bulletin,* **107**: 296–310, 1990.
39. R. Edmundson, M. Lawrence, and M. O'Connor, The use of non-time series information in sales forecasting: A case study, *J. Forecasting,* **7**: 201–211, 1988.
40. N. R. Sanders and L. P. Ritzman, The need for contextual and technical knowledge in judgmental forecasting, *J. Behavioral Decision Making,* **5**: 39–52, 1992.
41. R. M. Hogarth, *Judgment and Choice,* New York: Wiley, 1987.
42. R. Webby and M. O'Connor, Judgmental and statistical time series forecasting: A review of the literature, *Int. J. Forecasting,* **12**: 91–118, 1996.
43. P. Goodwin and G. Wright, Improving judgmental time series forecasting: A review of the guidance provided by research, *Int. J. Forecasting,* **9**: 147–161, 1993.

NADA R. SANDERS
Wright State University