Module IV

Forecasting

1 Introduction

Forecasting the future values of an observed time series is an important problem in many areas, including economics, production planning, sales forecasting and stock control.

Suppose we have an observed time series, $x_1, x_2, ..., x_N$. Then the problem is to estimate future values such as x_{N+h}, where the integer h is called the **lead time** or **forecasting horizon** –h for horizon.The forecast of x_{N+h} made at time N for h steps ahead will be denoted by $\hat{x}(N,h)$ or $\hat{x}_N(h)$.

A wide variety of different forecasting procedure is aviailable and it is important to realize that no single method is universally applicable. Rather, the analyst must choose the procedure that is most appropriate for a given set of conditions. Forecast are conditional statement about the future based on some specific assumptions. Thus forecast are not sacred[1] and the analyst should always be prepared to modify them as necessary in the light of any external information. For long-term forecasting, it can be helpful to produce a range of forecasts based on different set of assumptions so that alternative 'scenarios' can be explored.

1.1 Forecasting Methods

Forecasting methods may be broadly classified into three groups as follows:

1.1.1 Subjective

Forecasts can be made on a subjective basis using judgement, intuition, commercial knowledge and any other relevant information. Methods range widely from bold freehand extrapolation to the Delphi technique, in which a group of forecasters tries to obtain a consensus forecast with controlled feedback of other analysts' predictions and opinions as well as other relevant information. However, some subjective judgement is often used in a more statistical approach, for example, to choose an appropriate model and perhaps make adjustments to the resulting forecasts.

1.1.2 Univariate

Forecasts of a given variable are based on a model fitted only to present and past observations of a given time series, so that $\hat{x}_N(h)$ depends only on the values of $x_N, x_{N-1},,$ possibly augmented by a simple function of time, such as a global linear trend. This would mean, for example, that univariate forecasts of the future sales of a given product would be based entirely on past sales, and would not take account of other economic factors. Methods of this type are sometimes called naive or projection methods.

1.1.3 Multivariate

Forecasts of a given variable depend at least partly on values of one or more additional series, called **predictor** or **explanatory** variables[2]. For example, sales forecasts may depend on stocks and/or on economic indices. Models of this type are sometimes called causal models.
In practice, a forecasting procedure may involve a combination of the above approaches. For example, marketing forecasts are often made by combining statistical predictions with the subjective knowledge and insight of people involved in the market.

[1] religious

[2] They are also sometimes called *independent variable* but this terminology is misleading, as they are typically *not* independent of each other

1.2 Alternative way of classiying Forecasting

An alternative way of classifying forecasting methods is between an **automatic** approach requiring no human intervention, and a **non-automatic** approach requiring some subjective input from the forecaster.The latter applies to subjective methods and most multivariate methods. Most univariate methods can be made fully automatic but can also be used in a non-automatic form.

The choice of method depends on a variety of considerations, including:

- How the forecast is to be used.
- The type of time series (e.g. macroeconomic series or sales figures) and its properties (e.g. are trend and seasonality present?). Some series are very regular and hence 'very predictable', but others are not. As always, a time plot of the data is very helpful.
- How many past observations are available.
- The length of the forecasting horizon.
- The number of series to be forecast and the cost allowed per series.
- The skill and experience of the analyst. Analysts should select a method with which they feel 'happy' and for which relevant computer software is available. They should also consider the possibility of trying more than one method.

It is particularly important to clarify the objectives. This means finding out how a forecast will actually be used, and whether it may even influence the future. In a commercial environment, forecasting should be an integral part of the management process leading to what is sometimes called a *systems approach*.

Some forecasting procedures simply produce **point forecasts**. But these do not indicate the uncertainty associated with the estimation of future demand. Thus it is usually desirable to produce an **interval forecast**. Taking one more step away from a point forecast, it may be desirable to calculate the entire probability distribution of a future value of interest. This is called **density forecasting**.

Whatever forecasting method is used, some sort of forecast monitoring scheme is often advisable, particularly with large numbers of series, to ensure that forecast errors are not systematically positive or negative.

2 Univariate Procedure

In this section, we use the projection methods. For all the procedures, the first step is to plot the data as much useful information can often be obtained from a visual examination of the data, and this may help to suggest an appropriate forecasting procedure.

2.1 Extrapolation of trend curves

For long-term forecasting of non-seasonal data, it is often useful to fit a **trend curve** (or **growth curve**) to successive values and then extrapolate[3]. This approach is most often used when the data are yearly totals, and hence clearly non-seasonal. A variety of curves may be tried including polynomial, exponential, logistic and Gompertz curves. When the data are annual totals, at least 7 to 10 years of historical data are required to fit such curves. The method is worth considering for short annual series where fitting a complicated model to past data is unlikely to be worthwhile. Although primarily intended for longterm forecasting, it is inadvisable to make forecasts for a longer period ahead than about half the number of past years for which data are available.

[3] to predict by projecting past experience or known data

A drawback to the use of trend curves is that there is no logical basis for choosing among the different curves except by goodness-of-fit. Unfortunately it is often the case that one can find several curves that fit a given set of data almost equally well but which, when projected forward, give widely different forecasts.

2.2 Simple Exponential Smoothing

Exponential smoothing (ES) is the name given to the general class of forecasting procedure that rely [4] on simple updating equations to calculate forecast.The most basic form, is called **simple exponential smoothing** (SES), but this should only be used for non-seasonal time series showing no systematic trend.

Of course many time series that arise in practice do contain a trend or seasonal pattern, but these effect can be measured and removed to produce a stationary series for SES is appropriate. Alternatively, more complicated version of ES are available to cope [5] with trend and seasonality. Thus adaptations of ES are useful for many types of time series.

Given a non-seasonal time series, say $x_1, x_2, ..., x_N$, with no systematic trend, it is natural to forecast x_{N+1} by means of a weighted sum of the past observations:

$$\hat{x}_N(1) = c_0 x_N + c_1 x_{N-1} + c_2 x_{N-2} + ... \tag{1}$$

where the $\{c_i\}$ are weights.It seems sensible to give more weight to recent observations and less weight to observations further in past. An intuitively appealing set of weights are *geometric* weights, which decrease by a constant ratio for every unit increase in the lag. In order that weight sum to one, we take

$$c_i = \alpha(1-\alpha)^i \qquad i = 0, 1, ...$$

where α is a constant such that $0 < \alpha < 1$. Then Equation (1) becomes

$$\hat{x}_N(1) = \alpha x_N + \alpha(1-\alpha)x_{N-1} + \alpha(1-\alpha)^2 x_{N-2} + ... \tag{2}$$

Equation(2) implies an infinite number of past observations, but in practice there will be only a finite number. Thus Equation (2) is rewritten in **recurrence** form as

$$\begin{aligned} \hat{x}_N(1) &= \alpha x_N + (1-\alpha)[\alpha x_{N-1} + \alpha(1-\alpha)x_{N-2} + ...] \\ &= \alpha x_N + (1-\alpha)\hat{x}_{N-1}(1) \end{aligned} \tag{3}$$

If we set $\hat{x}_1(1) = x_1$, then equation (3) is can be used recursively[6] to compute forecast. Equation (3) can also reduce the amount of arithmetic involved, since forecast can easily be updated using only the latest observation and previous forecast.

The Procedure defined by Equation (3) is called simple exponential smoothing. The adjective 'exponential' arises from the fact that geometric weights lie on an exponential curve, but the procedure could equally well have been called geometric smoothing. Equation (3) is sometimes rewritten in the equivalent **error-correction** form

$$\begin{aligned} \hat{x}_N(1) &= \alpha[x_N - \hat{x}_{N-1}(1)] + \hat{x}_{N-1}(1) \\ &= \alpha e_N + \hat{x}_{N-1}(1) \end{aligned} \tag{4}$$

where $e_N = x_N - \hat{x}_{N-1}(1)$ is the prediction error at time N. Equation (3) and (4) give identical forecast, and it is a matter of practical convenience as to which one should be used.

Simple exponential smoothing is optimal if underlying model for time series is given by

$$X_t = \mu + \alpha \sum_{j<t} Z_j + Z_t \tag{5}$$

[4] to be dependent

[5] deal effectively with something difficult

[6] procedure that can repaet itself indefinitely

where $\{Z_t\}$ deontes a purely random process. This infinite-order moving average (MA) process is non-stationary, but the differences $(X_{t+1} - x_t)$ form a stationary first-order MA process. Thus X_t is an autoregressive integrated moving average (ARIMA) process of order $(0, 1, 1)$.In fact, there are many other models for which SES is optimal. This helps to explain why SES appears to be such a robust[7] method.

The value of the smoothing constant α depends on the properties of the given time series. Values between 0.1 and 0.3 are commanly used and produce a forecast that depends on a large number of past observations. Values close to one are used rather less often and give forecast that depends much more on recent observations. When $\alpha = 1$, the forecast is equal to the most recent observation.

The value of α may be estimated from past data by a similar procedure that used for estimating the parameters of an MA process. In more detail, for a given value of α, calculate

$$
\begin{aligned}
\hat{x}_1(1) &= x_1 \\
e_2 &= x_2 - \hat{x}_1(1) \\
\hat{x}_2(1) &= \alpha e_2 + \hat{x}_1(1) \\
e_3 &= x_3 - \hat{x}_2(1)
\end{aligned}
$$

and so until

$$e_N = x_N - \hat{x}_{N-1}(1)$$

and then compute $\sum_{i=2}^{N} e_i^2$.Repeat this procedure for values of α between 0 and 1, and select the value that minimize $\sum e_i^2$.

2.3 Box-Jenkins forecasting procedure

AR, MA and ARMA models have been around for many years and are associated, in particular, with early work by G.U.Yule and H.O.Wold. A major contribution of Box and Jenkins has been to provide a general strategy for time-series forecasting, which emphasizes the importance of identifying an appropriate model in an iterative way. Furthermore, Box and Jenkins showed how the use of differencing can extend ARMA models to ARIMA models and hence cope with non-stationary series. In addition, Box and Jenkins show how to incorporate seasonal terms into seasonal ARIMA (SARIMA) models. Because of all these fundamental contributions, ARIMA models are often referred to as Box-Jenkins models.

In brief, the main stages in setting up a Box-Jenkins forecasting model are as follows:

Model identification Examine the data to see which member of the class of ARIMA processes appears to be most appropriate.

Estimation Estimate the parameters of the chosen model.

Diagnostic checking Examine the residuals from the fitted model to see if it is adequate.

Consideration of alternative models if necessary If the first model appears to be inadequate for some reason, then alternative ARIMA models may be tried until a satisfactory model is found. When such a model has been found, it is usually relatively straightforward to calculate forecasts as conditional expectations.

We now consider these stages in more detail. In order to identify an appropriate ARIMA model, the first step in the Box-Jenkins procedure is to difference the data until they are stationary. This

[7] powerful

is achieved by examining the correlograms of various differenced series until one is found that comes down to zero 'fairly quickly' and from which any seasonal cyclic effect has been largely removed, although there could still be some 'spikes' at the seasonal lags $s, 2s$, and so on, where s is the number of observations per year. For non-seasonal data, first-order differencing is usually sufficient to attain stationarity.

For monthly data (of period 12), the operator $\nabla\nabla_{12}$ is often used if the seasonal effect is additive, while the operator ∇_{12}^2 may be used if the seasonal effect is multiplicative. Sometimes the operator ∇_{12} by itself will be sufficient. Over-differencing should be avoided. For a seasonal period of length s, the operator ∇_s may be used, and, in particular, for quarterly data we may use ∇_4. The differenced series will be denoted by $\{w_t; t = 1, 2, , N-c\}$, where c terms are 'lost' by differencing. For example, if the operator $\nabla\nabla_{12}$ is used, then $c = 13$.

For non-seasonal data, an ARMA model can now be fitted to w_t. If the data are seasonal, Box and Jenkins (1970) generalized the ARIMA model to deal with seasonality, and defined a general multiplicative seasonal ARIMA (SARIMA) model as

$$\phi_p(B)\Phi_P(B^s)W_t = \theta_q(B)\Theta_Q(B^s)Z_t \tag{6}$$

Where B denotes the backward shift operator, $\phi_p, \Phi_P, \theta_q, \Theta_Q$ are polynomials of order p, P, q, Q respectively, and $Z - t$ denotes a purely random process and

$$W_t = \nabla^d\nabla^D X_t \tag{7}$$

denotes the differenced series. If the integer D is not zero, then seasonal differencing is involved. The above model is called a SARIMA model of order $(p, d, q) \times (P, D, Q)_s$.

Plausible values of p, P, q, Q are selected by examining the correlogram and the partial autocorrelation function (a.c.f.) of the differenced series w_t. values of p, P, q and Q need to be assessed by looking at the ac.f. and partial ac.f. of the differenced series and choosing a SARIMA model whose ac.f. and partial ac.f. are of similar form. Also, values of P and Q are selected primarily by examining the values of r_k at $k = 12, 24, ,$ when the seasonal period is given by $s = 12$. If, for example, r_{12} is 'large' but r_{24} is 'small', this suggests one seasonal moving average term, so we would take $P = 0, Q = 1$ as this SARIMA model has an ac.f. of similar form. Box et al. (1994) list the autocovariance functions of various SARIMA models.

Having tentatively identified what appears to be a reasonable ARMA *or* SARIMA model, least squares estimates of the model parameters may be obtained by minimizing the residual sum of squares. In the case of seasonal series, it is advisable to estimate initial values of a_t and w_t by backforecasting rather than setting them equal to zero.

For both seasonal and non-seasonal data, the adequacy of the fitted model should be checked by what Box and Jenkins call 'diagnostic checking'. This essentially consists of examining the residuals from the fitted model to see whether there is any evidence of non-randomness. The correlogram of the residuals is calculated and we can then see how many coefficients are significantly different from zero.

When a satisfactory model is found, forecasts may readily be computed. The minimum mean square error forecast of X_{N+h} at time N is the conditional expectation of X_{N+h} at time N,

$$\hat{x}_N(h) = E(X_{N+h}|X_N, X_{N-1}, ...).$$

In evaluating this conditional expectation, we use the fact that the 'best' forecast of all future Zs is simply zero (or more formally that the conditional expectation of Z_{N+h}, given data up to time N, is zero for all $h > 0$). Box et al. describe general approaches to computing forecasts.

Point forecasts are usually computed most easily directly from model equation, which Box et al.(1994) call the difference equation form. Assuming that the model equation is known exactly, then $\hat{x}_N(h)$ is obtained from the model equation by replacing *(i)* future values of Z by zero, *(ii)* future values of X by their conditional expectation and *(iii)* present and past values of X and Z by their observed values.

As an example, consider the SARIMA $(1,0,0)\times(0,1,1)_{12}$ model

$$\phi_1(B)W_t = \Theta_1(B^{12})Z_t$$
$$(1-\alpha B)W_t = (1+\Theta B^{12})Z_t$$

where $W_t = \nabla_{12}X_t$ and Θ is a constant perameter(rather than a function as in (6)). It may help to write this out in terms of the original observed variable X_t as

$$X_t = X_{t-12} + \alpha(X_{t-1} - X_{t-13}) + Z_t + \theta Z_{t-12}$$

Then we find

$$\hat{x}_N(1) = x_{N-11} + \alpha(x_N - x_{N-12}) + \theta z_{N-11}$$
$$\hat{x}_N(2) = x_{N-10} + \alpha(\hat{x}_N(1) - x_{N-11}) + \theta z_{N-10}$$

Although some packages have been written to carry out ARIMA modelling and forecasting in an automatic way, the Box-Jenkins procedure is primarily intended for a non-automatic approach where the analyst uses subjective judgement to select an appropriate model from the large family of ARIMA models according to the properties of the individual series being analysed. Thus, although the procedure is more versatile than many competitors, it is also more complicated and considerable experience is required to identify an appropriate ARIMA model. Unfortunately, the analyst may find several different models, which fit the data equally well but give rather different forecasts, while sometimes it is difficult to find any sensible model. Of course, an inexperienced analyst will sometimes choose a 'silly' model. Another drawback is that the method requires several years of data (e.g. at least 50 observations for monthly seasonal data).

2.4 Stepwise Autoregression

Granger and Newbold (1986) describe a procedure called stepwise autoregression, which can be regarded as a subset of the Box-Jenkins procedure. It has the advantage of being fully automatic and relies on the fact that AR models are much easier to fit than MA or ARMA models even though an AR model may require extra parameters to give as good a representation of the data. The first step is to take first differences of the data to allow for non-stationarity in the mean. Then a maximum possible lag, say p, is chosen and the best AR model with just one lagged variable at a lag between 1 and p, is found, namely

$$W_t = \mu + \alpha_k^{(1)} W_{t-k} + e_t^{(1)} \tag{8}$$

where $W_t = X_t - X_{t-1}, 1 \leq k \leq p, \alpha_k^{(1)}$ is the autoregression coefficient at lag k when fitting one lagged variable only, and $e_t^{(1)}$ is the corresponding error term. Then the best AR model with $2, 3, ,$ lagged variables is found. The procedure is terminated when the reduction in the sum of squared residuals at the j^{th} stage is less than some preassigned quantity. Thus an integrated AR model is fitted, which is a special case of the Box-Jenkins ARIMA class. Granger and Newbold suggest choosing p=13 for quarterly data and p=25 for monthly data.

2.5 Other Methods

Several other foreeasting proeedures have been proposed. Brown (1963) has suggested a technique called general exponential smoothing which consists of fitting polynomial, sinusoidal or exponential functions to the data and finding appropriate updating formula. One special case of this is double exponential smoothing which is applicable to series containing a linear trend. Note that

Brown suggests fitting by discounted least squares, in which more weight is given to recent observations. Generally speaking, Brown's method compares unfavourably with other univariate forecasting procedures, especially on seasonal data (see Reid, 1975).

Harrison (1965) has proposed a modification of seasonal exponential smoothing, which consists essentially of performing a Fourier analysis of the seasonal factors and replacing them by smoothed factors. Parzen's ARARMA approach (Parzen, 1982; Meade and Smith, 1985) relies on fitting an AR model to remove the trend (rather than just differencing the trend away) before fitting an ARMA model. An apparently new method, called the theta method, gave promising results (Makridakis and Hibon, 2000), but subsequent research has shown that it is actually equivalent to a form of exponential smoothing.

There are two general forecasting methods, called Bayesian forecasting (West and Harrison, 1997) and structural modelling (Harvey, 1989), which rely on updating model parameters by a technique called Kalman filtering.