Chapter 1: Basic Concepts of Forecasting

Types of forecasts

In time series forecasting, we seek to make statements about \( x_{n+h} \), the value the series will take at the future time period \( n+h \). The quantity \( h \) is called the lead time. Forecasts may be classified as short run, middle run or long run, according to the extent of the lead time. The precise meaning of these terms is based on personal judgement, and on the particular situation in question. In weather forecasting, the short run may be 24 hours ahead; in forecasting unemployment, the short run may be 10-15 months. In general, the further ahead one forecasts the less well one does, in that larger errors are likely to occur.

The appearance of plots of \( x_t \) versus \( t \), as well as the inherent complexities of the situations being studied, suggests that in many situations the series \( x_t \) may profitably be thought of as a sequence of random variables. Viewed from time \( n \), then, \( x_{n+h} \) is an unobserved random variable, and the most we can hope to know about \( x_{n+h} \) is its probability distribution function, or perhaps its conditional distribution given the available information, \( x_1, \ldots, x_n \). If we knew this distribution, then we could compute \( \text{Prob}(C < x_{n+h} < D) \) for any fixed numbers \( C \) and \( D \). Without assuming that the exact distribution is known, it may still be possible to construct confidence intervals (or, more precisely, prediction intervals) for \( x_{n+h} \). A prediction interval for \( x_{n+h} \) is an interval with random (i.e., data-dependent) endpoints \( (B \, \, A) \) such that

\[
\text{Prob}(B < x_{n+h} < A) = 1-\alpha.
\]

Here, \( \alpha \) is a freely chosen constant (usually we use \( \alpha=.05 \) or \( \alpha=.01 \)). The interpretation of the prediction interval is that if a large number of such intervals are constructed on different data sets, then about \( (1-\alpha)100\% \) of these intervals will contain \( x_{n+h} \). Thus, even though we can never be sure that the particular interval at hand will cover \( x_{n+h} \), we are willing to bet that it will. Further, if we declare that it will, then in the long run our declarations will prove to be correct \( (1-\alpha)100\% \) (i.e., "most") of the time.

Although an interval forecast (i.e., a prediction interval) should always be given if feasible, forecasters are often content with providing just a point forecast, a single guess for \( x_{n+h} \) that accurately
reflects the whole distribution of possible values. As an example of a point forecast, we might predict
that there will be 35 near misses by pairs of commercial airliners over the U.S. next year. If the (condi-
tional) distribution of \( x_{n+h} \) is known, then the optimal point forecast is the mean of this distribution. A
\textit{trace forecast} is a whole sequence of point forecasts for \( h=1,2,\ldots,H \). Trace forecasts represent
an attempt to forecast the whole future of the time series, not just the value at a particular future time.

\textbf{Information Sets}

The \textit{information set} available at time \( n \), denoted by \( \psi_n \), is simply the set of information to be
used for forecasting. For time series forecasting, the most commonly used information set is

\[ \psi_n : x_{n-j}, \quad j \geq 0. \]

Here, \( \psi_n \) consists of the observed present and past values of the series. It may be helpful to also use the
past and present values of other relevant series, leading to

\[ \psi_n : x_{n-j}, y_{n-j}, z_{n-j}, \text{ etc.} \quad j \geq 0. \]

In this course, however, we will assume that the information set is confined to the present and past
values of the series we wish to forecast.

\textbf{Cost Functions}

\textit{Cost functions} provide a criterion by which forecast methods can be compared. When a forecast
is in error, a cost will typically be incurred because the decision maker will not make the optimum
decision. Generally, the larger the magnitude of the error, the larger the cost. For example, suppose you
seek to forecast the number of cakes that people will want to buy tomorrow in a bakery. The number of
cakes prepared tonight (hopefully to be sold tomorrow) will be put equal to your forecast. Any of these
cakes not sold tomorrow will have to be thrown away. If each cake costs 70 cents to make and sells for
1 dollar, then the cost \( C(e) \) of making an error \( e \) is

\[ C(e) = \begin{cases} 
70e & e > 0 \\
30(-e) & e < 0 \\
0 & e = 0 
\end{cases}. \]
In the absence of such precise knowledge of the costs as given in the above example, it is con-
venient (especially for mathematical reasons) to adopt the squared error cost function, \( C(e) = e^2 \). Given a cost function \( C(e) \), the criterion for choosing between forecasting methods is: Choose the method which results in the lowest expected (or average) cost.

The advantage of the squared-error cost function is that it corresponds to a least–squares criterion and so allows the use of much standard statistical theory. The forecasting method which minimizes the average squared error (among all possible methods based on an information set \( \psi_n \)) is called the optimum least squares forecast. In this course, instead of considering all possible forecasting methods, we will restrict our attention to linear forecasts (i.e., forecasts which are linear functions of the data). For example, given a single-series information set \( \psi_n : x_{n-j}, j \geq 0 \), a linear forecast of \( x_{n+h} \) takes the form

\[
f_{n,h} = \sum_{j=0}^{m} \gamma_j x_{n-j}.
\]

If some appropriate \( m \) is chosen and the parameters \( \gamma_j \) are then chosen so that the forecast minimizes the average squared error, then the result is an optimal linear least-squares forecast. The corresponding \( h \)-step forecast error is

\[
e_{n,h} = x_{n+h} - f_{n,h}.
\]