Time-series forecasting

As we go through life, everyone makes forecasts all the time, often without realising it. Sadly these forecasts are often (very) inaccurate. **Chris Chatfield** looks at the chequered history of forecasting and asks how we might do it better using time-series data, and what statistical techniques and models might help us.

We all make forecasts all the time. Some are trivial (shall I take an umbrella today?) while others are more substantial (shall I buy this new share issue?). In business, companies rely on forecasts of sales and of the economy in order to plan production. The government relies on forecasts to plan fiscal policy, the provision of new facilities and so on. Many forecasts are made using subjective judgement, but this article addresses the more statistical issues involved in time-series forecasting.

A brief history

People have been making forecasts since time immemorial—and falling flat on their faces. Apocryphal stories include: the Yale professor of economics who, in 1929, said "Stock prices have reached what looks like a permanently high plateau" shortly before the big crash; Lord Rutherford who, after splitting the atom, said "Anyone who expects a source of power from the transformations of these atoms is talking moonshine"; the founder of IBM who, in 1947, said "I think there is a world market for about five computers"; and Thomas Edison who, in 1880, said "The phonograph is not of any commercial value". These examples demonstrate that basing predictions on past events can be a recipe for disaster. Yet that is exactly how forecasts are usually made.

So, are there ways in which uncertainty can be reduced and accuracy improved? Fortunately there are, though the perceptive forecaster will still realise that the future may not be like the past and that the longer the forecasting horizon, the greater the uncertainty.

Time-series forecasting is a fairly recent phenomenon. Before about 1960, linear regression on time and fitting a constant seasonal pattern were essentially the only statistical techniques employed. Then came exponential smoothing (ES). The latter technique originated in work by Robert Brown for the US Navy in the Second World War. A method of smoothing seasonal data was independently developed by Charles Holt in work for the US Office of Naval Research and the resulting combined method for forecasting data showing both trend and seasonality is customarily called the Holt-Winters method. As early as 1960, John Muth showed that simple ES was optimal for a model called "random walk plus noise", and various other versions of ES have been found to be optimal for a variety of other models, including models with non-constant variance.

The key idea in ES, which has parallels in many other time-series problems, is that the updated forecast is a linear combination of the previous forecast and the latest observation. Thus, although the updated forecast involves all earlier observations via the previous forecast, the only quantities that need to be stored and used are the latest forecast and the latest observation. ES is a simple example of a more general updating procedure called the Kalman filter that is widely used elsewhere, especially in signal processing.

In the mid-1960s, George Box and Gwilym Jenkins began work on a class of models called auto-regressive integrated moving average (ARIMA) models and the forecasting approach that was optimal therefore. Their 1970 book, now in its third edition,[1] was hugely influential, not only for its coverage of ARIMA modelling, but also for some more general statistical ideas (e.g. those on how to carry out iterative model-building). The book is still worth reading. However, despite claims made at the time by some enthusiasts, it is now generally recognised that ARIMA modelling is a valuable addition to the forecaster's toolkit rather than a universal panacea for all

forecasting problems. Indeed, this last statement is true for all the many new forecasting methods that have been proposed over the years, often accompanied by grandiose, but subsequently unsubstantiated, claims.

In the 1980s and 1990s a broad class of state-space models became popular and was reinvented in different application areas with names such as structural models, unobserved component models and dynamic linear models. In recent years we have continued to hear about a variety of new alternative methods and models, some of which allow for non-linearity and changing variance. One particularly noteworthy example was the application of neural nets to time series. Neural nets can be thought of as a non-linear form of regression, and have the theoretical advantage that they are a "universal approximator" to any given function. However, in practice they behave like most other methods, meaning that they sometimes do well and sometimes do poorly. Moreover, they can be tricky to apply and, as with all complicated methods, it is easier to go wrong with them than with simpler methods. Models that concentrate on modelling changes in variance, such as ARCH and GARCH (Generalised Auto-Regressive Conditionally Heteroscedastic!) models, are also generally classified as non-linear. They are unsuitable for the novice and I will not attempt to explain them here. Other classes of non-linear model have also been used for forecasting, but are typically (much) more complicated than linear methods such as ES and, in my experience, seldom yield genuine out-of-sample forecasts that are sufficiently improved to justify the extra effort.

The above remarks presuppose that a univariate approach is being used, so that forecasts of a given variable depend only on past values of the same variable. In many situations, it would appear that a multivariate model is appropriate in that it seems reasonable to expect that relevant explanatory variables may be used to improve forecasts. Following advances in computer software, various multivariate forecasting models have been developed, especially over the last 15 years or so. As for non-linear models, the empirical accuracy of the resulting forecasts can best be described as "mixed" when compared with those from simpler univariate methods.

The novice forecaster also needs to beware of claims made for new methods. They will typically be tried on some nice data, suitably chosen to show up the advantages of the new method. Sadly, when applied by different analysts to different data they may not do so well. Alternatively they may turn out to be mathematically equivalent to an existing method. This happened recently with a new, algebraically complicated, method that was published alongside claims that it gave good empirical results. After publication it was subjected to close scrutiny and shown to be equivalent to a form of ES that is already known to give good forecasts on average.

Yet another cause for concern is that empirical results for a new method are often found to be unfair, in that the method uses information about the future that is unknown at the time of forecasting. This unfair use of future data may be done consciously (so it is "cheating") or unconsciously. Thus, when comparing accuracy it is essential that all forecasts should be genuinely out-of-sample (called ex-ante forecasts by economists). In other words, the method or model should be fitted to the first part of the data before making forecasts of the second part, called the "hold-out" sample, without using any information from the hold-out period. Analysts do not always do this!

"Time-series forecasting flies in the face of all other statistical work"

It follows that there should be transparency in any empirical study or forecasting competition, so that empirical claims can be checked and forecast comparisons replicated.

Preliminaries

A time series is a collection of observations made sequentially through time. Examples include air temperature measured hourly, share prices measured daily and carbon monoxide levels measured annually. In time-series forecasting, the general approach is to look at past data, fit an appropriate model and then project future values using the model.

The first general point to make is that time-series forecasting flies in the face of all other statistical work. When students first learn techniques like regression they are customarily taught that it is unwise to use a fitted model to make predictions outside the range of values over which the model has been fitted. In other words, extrapolation is generally unwise. Yet, in time-series forecasting, that is exactly what the analyst has to do. Although often unavoidable, this does mean that time-series forecasting is potentially disastrous when the future is not like the past.

As in any statistical exercise, it is important to start by getting background information, clarifying objectives, checking the quality and quantity of the data, and so on. Make sure that you understand the context and have all relevant information before you start. Do not be afraid to ask questions. Although they may seem obvious, these preliminaries are often glossed over, which may result in the analyst not realising why a forecast is desired or how it is going to be used. The quality and properties of the data must also be assessed. Occasionally, for example, the time between observations (the sampling rate) is changed so that, for example, a series collected annually becomes monthly. This will naturally cause havoc with any model-fitting procedure and can easily trip up the unwary.

The time plot

In order to fit an appropriate model to a given time series, the first and most important step is to plot the data against time. This gives a graph, called a time plot, which will show the main properties of the data, provided that it is drawn carefully. In particular, the scales need to be chosen so that any past changes are not too shallow or too steep. Furthermore, common sense suggests that a clear title should be given and that the axes should be carefully labelled. I have found trial and error important in getting a clear graph, even with the good graphics software available today. Figure 1 shows an example of a time plot for some monthly sales data over a 7-year period. The series looks reasonably smooth, with a clear upward trend (annual sales are larger towards the end of the series) and clear seasonality (sales are generally higher in December but lower in January and February).

If a given time series exhibits trend or seasonality (or both) the forecasting method must be chosen so as to take this into account. Many government agencies use a filtering approach called X-11, or the later version X-12, which deals with trend and seasonality as well as handling other data anomalies, such as missing values. Filtering methods are simply ways of applying a series of linear smoothing operations and no modelling is involved (except implicitly). In the Box-Jenkins approach trend and seasonal variation are removed from the data by a sequence of differencing operators, which are a special type of linear filter. An ARMA model is then fitted to the differenced series. This is not very helpful when it is of prime importance to understand and model the trend and seasonal variation. In contrast, state-space models (or structural models) include specific terms to model these important effects explicitly.

Although they sound like simple phenomena, it is my experience that trend and seasonality can often be quite hard to model and understand. This is partly because both effects may be changing through time. Many years ago analysts would often model a trend by fitting a linear regression line to the data. This is now generally seen as inadequate and analysts prefer to model a "local trend" rather than a fixed (global) trend. A technique like ES is ideal for updating the local estimate of trend. Seasonal effects may also change through time and it is necessary to distinguish between different types of seasonality, such as additive seasonality (which is constant from year to year) and multiplicative seasonality.

The time plot should also reveal any other important features of the data, such as a discontinuity, a step change in trend or the pres-

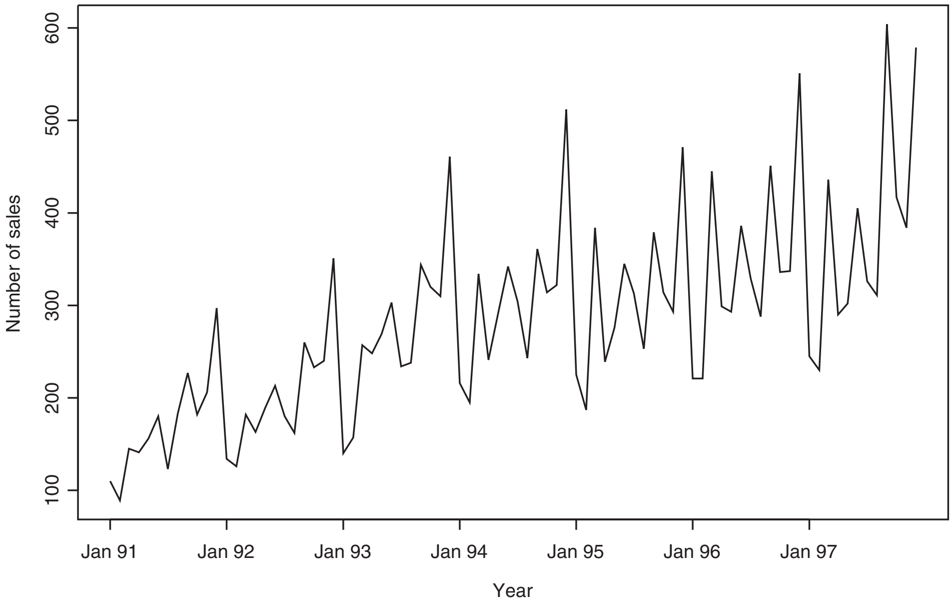

Figure 1. An example of a time plot showing monthly sales data over a 7-year period

ence of any outlying observations. At this stage, it may prove necessary to "clean" the data and consider transformations, for example to reduce skewness.

Modelling and forecasting

Having established the main properties of a dataset, the next stage is to choose an appropriate forecasting method. Sometimes this is done on a fairly ad hoc basis, especially when applying an automatic method to a large number of series, as in stock control. For example, if the variation is dominated by trend and seasonal, then it may be sensible to use the Holt-Winters version of ES. Here the modelling is implicit rather than explicit. Even so, some care is needed, for example to decide whether seasonality should be assumed additive or multiplicative.

If a more formal modelling approach is indicated for a particular time series then the analyst will usually begin by looking at a function called the autocorrelation function, perhaps after removing trend and seasonal variation. An autocorrelation coefficient measures the correlation between observations at a given time lag apart. That is, it is the correlation of a variable with itself at a particular lag—hence the term *auto*correlation. It is often found that autocorrelations are high at small lags and also at the lag corresponding to seasonality (e.g. lag 12 for monthly data). Given this and other relevant information, a suitable class of models is chosen and fitted to the data using available software.

Having fitted a model, it is essential to look at the residuals, which, in time-series analysis, are the one-step-ahead forecast errors. They should be plotted against time to see if any patterns emerge, and their autocorrelation function can also be computed. If necessary, alternative models can be tried in an iterative way. Once a satisfactory model has been found, it is usually clear how it can be used to make forecasts.

Time-series modelling is a large topic which cannot really be covered here, and remains something of an art even with the help of modern software packages.

If a multivariate model is envisaged the analyst will need to look at the properties of the explanatory variables as well as those of the variable to be forecasted. The analyst will also need to look at quantities, called cross-correlations, that assess the covariation of two series. However, in my experience, multivariate model-fitting and forecasting is fraught with problems and is not for the faint-hearted. Forecasts of the variable of interest may in any case require forecasts of explanatory variables, and multivariate forecast accuracy is often little better than that given by simpler univariate forecasts, except perhaps when there is a "leading indicator", meaning that the effect of one explanatory variable takes several time periods to manifest itself on the response variable.

More generally, it has to be said that, whatever method is used in practice, empirical results show that out-of-sample forecast accuracy is often disappointing when compared with within-sample fit. There are many reasons for this, the main one being that the relevant model may be wrongly identified or may be changing through time. It is a pity that the possible presence of model uncertainty has generally received rather little attention from statisticians, and this is of particular concern for time-series forecasting.

Many forecasting methods are primarily intended to produce a forecast as a single value, called a point forecast. However, this gives no indication of the likely uncertainty and it is usually preferable to produce what are called interval forecasts. As the name suggests, an interval forecast covers a range of values that are likely to cover the future value with a prescribed probability, and I would encourage more forecasters to produce them.

A question commonly asked of forecasters is which method is best, but there is no simple answer, given the wide variety of forecasting problems. The first response should be to ask what is meant by "best", as minimising the average out-of-sample squared forecast error is not necessarily the only criterion that may apply. More generally, the context is crucial and the choice of method must be matched to the particular practical situation. The choice will depend *inter alia* on the properties of the series, the number of data points, the length of the forecasting horizon, the number of series to be forecasted, and so on. Often a simple univariate method is as good as anything, though causal methods should be considered when there is good domain knowledge, especially when the explanatory variables are of the leading indicator type.

Further reading

There are many good books on time-series analysis and forecasting. The classic intermediate text[2] by Clive Granger and Paul Newbold still provides a good general coverage of forecasting topics, while my own book[3] provides more up-to-date general coverage with emphasis on model uncertainty and prediction intervals. The introductory text[4] by Francis Diebold is aimed more at business and finance students.

References

1. Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994) *Time Series Analysis, Forecasting and Control*, 3rd edn. Englewood Cliffs, N. J.: Prentice-Hall.
2. Granger, C. W. J. and Newbold, P. (1986) *Forecasting Economic Time Series*, 2nd edn. New York: Academic Press.
3. Chatfield, C. (2001) *Time-Series Forecasting*. Boca Raton, Florida: Chapman & Hall/CRC Press.
4. Diebold, F. X. (2004) *Elements of Forecasting*, 3rd edn. Cincinnati: South-Western.

Chris Chatfield is Reader in Statistics at Bath University, an Honorary Fellow of the International Institute of Forecasters and the author of 5 textbooks. His interests include all forms of practical problem-solving as well as time-series forecasting (not to mention golf, now that he is working part-time!).