ARIMA processes

Rob J Hyndman   25 May 2001

ARIMA processes are mathematical models used for forecasting. ARIMA is an acronym for *AutoRegressive, Integrated, Moving Average*. Each of these phrases describes a different part of the mathematical model.

ARIMA processes have been studied extensively and are a major part of time series analysis. They were popularized by George Box and Gwilym Jenkins in the early 1970s; as a result, ARIMA processes are sometimes known as Box-Jenkins models. Box and Jenkins (1970) effectively put together in a comprehensive manner the relevant information required to understand and use ARIMA processes.

The ARIMA approach to forecasting is based on the following ideas:

1 The forecasts are based on linear functions of the sample observations;

2 The aim is to find the simplest models that provide an adequate description of the observed data. This is sometimes known as the principle of *parsimony*.

Each ARIMA process has three parts: the autoregressive (or AR) part; the integrated (or I) part; and the moving average (or MA) part. The models are often written in shorthand as ARIMA(p,d,q) where p describes the AR part, d describes the integrated part and q describes the MA part.

AR: This part of the model describes how each observation is a function of the previous p observations. For example, if $p = 1$, then each observation is a function of only one previous observation. That is,

$$Y_t = c + \phi_1 Y_{t-1} + e_t$$

where Y_t represents the observed value at time t, Y_{t-1} represents the previous observed value at time $t - 1$, e_t represents some random error and c and ϕ_1 are both constants. Other observed values of the series can be included in the right-hand side of the equation if $p > 1$:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t.$$

I: This part of the model determines whether the observed values are modelled directly, or whether the *differences* between consecutive observations are modelled instead. If $d = 0$, the observations are modelled directly. If $d = 1$, the differences between consecutive observations are modelled. If $d = 2$, the differences of the differences are modelled. In practice, d is rarely more than 2.

MA: This part of the model describes how each observation is a function of the previous q errors. For example, if $q = 1$, then each observation is a function

of only one previous error. That is,

$$Y_t = c + \theta_1 e_{t-1} + e_t.$$

Here e_t represents the random error at time t and e_{t-1} represents the previous random error at time $t-1$. Other errors can be included in the right-hand side of the equation if $q > 1$.

Combining these three parts gives the diverse range of ARIMA models.

There are also ARIMA processes designed to handle seasonal time series, and vector ARIMA processes designed to model multivariate time series. Other variations allow the inclusion of explanatory variables.

ARIMA processes have been a popular method of forecasting because they have a well-developed mathematical structure from which it is possible to calculate various model features such as prediction intervals. These are a very important feature of forecasting as they enable forecast uncertainty to be quantified.

See entry for Box-Jenkins modelling.

References

BOX, G.E.P. and G.M. JENKINS (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.

MAKRIDAKIS, S., S.C. WHEELWRIGHT, and R.J. HYNDMAN (1998) *Forecasting: methods and applications*, New York: John Wiley & Sons.

PANKRATZ, A. (1983) *Forecasting with univariate Box–Jenkins models: concepts and cases*, New York: John Wiley & Sons.