An Introduction to Time-Series Modelling

"Forecasting is the art of saying what will happen, and then explaining why it didn't!" (Anonymous, quoted in Chatfield (1989, p.66)).

INTRODUCTION

The purpose of this article is to give the reader a brief introduction to the Box-Jenkins approach to time-series modelling. It is hoped that after reading this article the reader will be able to model his or her own time-series.

Unlike most econometric modelling, time-series analysis involves modelling a dependent variable solely in terms of the past history of itself. The main use of time-series modelling is to estimate future values of the series, or more preferably confidence intervals for the future values. Section 1 will outline two of the most basic concepts used in time-series analysis, namely the ideas of stationarity and the autocorrelation function (acf). Section 2 will introduce some of the basic models of the Box-Jenkins approach, and section 3 will summarize some of the main features of Box-Jenkins modelling. For illustrative purposes section 3 contains an example of a time-series model which I have developed myself. The time-series in question is *The Economist's* Metals Dollar Index for 1987, 1988 and 1989. Observations occur weekly, and for the purposes of determining the predictive power of my model I have omitted the last 8 observations. Further details of the index can be found in *The Economist* of 12 March 1988, 6 May 1989 and 6 January 1990.

1. BASIC CONCEPTS

The principle of *stationarity* is one of the most basic principles needed for time-series modelling. *A stationary time-series is one whose characteristics are invariant with respect to time.* More mathematically, the concept of stationarity is defined as follows. For a stochastic process to be stationary the following conditions must be satisfied for all values of t:

$$E(y_t) = \mu,$$
$$E[(y_t - \mu)^2] = \pi(0),$$
$$E[(y_t - \mu)(y_{t-k} - \mu)] = \pi(k),\ k=1,2,....$$

where μ, $\pi(0)$, and $\pi(k)$ are all constant. If any of these conditions are not satisfied then the characteristics of the process will tend to change with time. One can intuitively see that a stationary time-series is much easier to estimate and to forecast with, and Box-Jenkins modelling (which I shall introduce later) relies crucially on the time-series being stationary.

Most economic time-series are not stationary. Many involve an upward trend (prices, for example), so immediately the first condition required for stationarity is broken (i.e. the average of the process is increasing with time). Fortunately, however, there are methods which can sometimes be used to derive a stationary process from a non-stationary one which I shall mention later.

The *autocorrelation function* (acf) is obtained by plotting

$$P(k) = \pi(k)/\pi(0)$$

against k, k=1,2,....., where $\pi(k)$ is the autocovariance between y_t and y_{t-k} as defined before, and $\pi(0)$ is a scaling factor. In practice, of course, we only have estimates of the values of $\pi(k)$ and consequently we only have an estimate of the acf, usually called the *correlogram* or the *sample autocorrelation function.* If T is the number of observations and μ is the sample average then the correlogram is denoted by

$$r(k) = c(k)/c(0),\ k=1,2,...$$
$$\text{where } c(0) = T^{-1}\Sigma(y_t - \mu)^2$$
$$\text{and } c(k) = T^{-1}\Sigma(y_t - \mu)(y_{t-k} - \mu),\ k=1,2,.....$$

As we shall see later, the correlogram is one of the main tools used in trying to identify time-series models.

2. AR, MA, ARMA AND ARIMA PROCESSES

The processes most used in time-series modelling are *autoregressive integrated moving average* (ARIMA) processes. I will look at autoregressive and moving average processes separately, before combing them to get ARMA and ARIMA processes.

An autoregressive process of order p is written as

$$y_t = \varnothing_1 y_{t-1} + \ldots\ldots + \varnothing_p y_{t-p} + e_t,$$

where the e_t's are normally distributed random variables with mean zero and constant variance. More concisely,

$$\varnothing(L)y_t = e_t,$$
$$\text{where } \varnothing(L) = 1 - \varnothing_1 L - \varnothing_2 L^2 - \ldots\ldots - \varnothing_p L^p.$$

(L is known as the *lag operator* and is defined as

$$Ly_t = y_{t-1},$$

so obviously $L^j y_t = y_{t-j}$). This is usually denoted AR(p).

For example,

$$y_t = \varnothing y_{t-1} + e_t$$

is an AR(1) process. Since

$$y_t = \varnothing y_{t-1} + e_t$$
$$= \varnothing^t y_0 + \varnothing^{t-1} e_1 + \ldots. + \varnothing e_{t-1} + e_t,$$

then $E(y_t) = \varnothing^t y_0$

If $|\varnothing|<1$ and if the process started a long time ago (i.e. t is large), then

$$E(y_t) = E(y_{t-1}) = 0.$$

However, if $|\varnothing|>1$, then $E(y_t)$ grows exponentially, so the process is non-stationary. In fact if $\varnothing = 1$, the process is also non-stationary, so y_t is stationary if and only if $|\varnothing|<1$. Stationarity for higher order processes is harder to envisage, but it can be shown that for an AR(p) process to be stationary the roots of the polynomial equation

$$1 - \varnothing_1 L - \varnothing_2 L^2 - \ldots\ldots - \varnothing_p L^p = 0$$

must lie outside the unit circle. (This allows for complex roots). A proof of this for p = 2 is outlined in Harvey (1981, pp. 29-32). Notice that for the AR(1) process, P(k) decays geometrically as k gets large. The same is true for an AR(p) process, but for small k, P(k) depends very much on the values of $\varnothing_1, \ldots\ldots, \varnothing_p$ and little can be said in general about the first few values of the acf.

In contrast to an autoregressive process, a moving average process relates y_t to previous values of the error term e_t. More formally, an MA(q) process is written as

$$y_t = \beta_1 e_{t-1} + \ldots.. + \beta_q e_{t-q} + e_t,$$

or more concisely

$$y_t = \beta(L)e_t,$$

where $\beta(L) = 1 + \beta_1 L + \ldots.. + \beta_q L^q$. It is easy to see that a finite moving average process is always stationary, since y_t is uncorrelated with y_{t-k} for k>q. Thus the autocorrelation function of an MA(q) process will suddenly drop to zero for k>q.

Combining autoregressive and moving average processes we get processes of the form

$$y_t = \varnothing_1 y_{t-1} + \ldots.+ \varnothing_p y_{t-p} + e_t + \beta_1 e_{t-1} + \ldots. + \beta_q e_{t-q},$$
$$\text{or } \varnothing(L)y_t = \beta(L)e_t.$$

These are called ARMA(p,q) processes. For the process to be stationary we only require that the autoregressive part be stationary. In terms of the autocorrelation function, the only thing that can be said in general is that for k>q the acf is going to behave exactly as the acf of the autoregressive part of the process, i.e. it decays geometrically towards 0 for k>q.

I hinted when talking about stationarity that some non-stationary processes can be made stationary. Consider the following non-stationary process:

$$y_t = y_{t-1} + e_t.$$

Although the process is non-stationary, we see that if we let

$$z_t = y_t - y_{t-1}$$
then $z_t = e_t$,

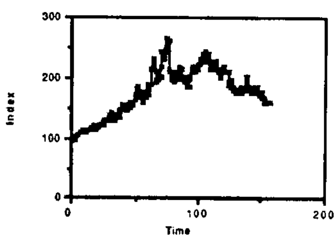
so z_t (the *differenced* process) is stationary. More generally it is sometimes possible to difference a non-stationary process d times to derive a stationary process. Such a process is denoted ARIMA(p,d,q), an autoregressive *integrated* moving average process. Such a process can be written as

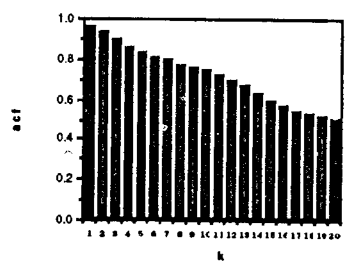
$$\text{Ø}(L)\Delta^d y_t = \beta(L)e_t,$$
where $\Delta y_t = y_t - y_{t-1}$, $\Delta^d y_t = \Delta^{d-1} y_t - \Delta^{d-1} y_{t-1}$.

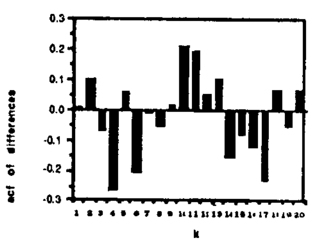
MODEL BUILDING AND ESTIMATION

Suppose we want to fit an ARIMA(p,d,q) model to a data set $\{y_t\}$. Box & Jenkins (1970) suggest a methodology for finding the best such ARIMA process. They suggest first making a tentative guess as to the values of p, d and q, then estimating the parameters, and finally subjecting the model to diagnostic tests to see if there is a significant divergence between the estimated model and the actual data.

The first stage is to make a guess as to what the values of p, d and q might be. The way that this is done is to examine the acf, or rather its approximation the correlogram, the characteristics of which should give us some hints about what values p, d and q might take on. The determination of d is probably the easiest part. For an ARMA process to be stationary the correlogram should be close to the zero for large k. If this isn't the case, the data should be differenced as many times as necessary until we think the model is stationary. Consider the Metals Dollar Index which I introduced earlier. Figure 1 shows the time-series plot of the data. The data appears to be non-stationary, and the correlogram bears this out (see figure 2 - the correlogram is not decaying towards zero). Thus differencing is required. Figure 3 shows the correlogram of the differenced data. Obviously differencing has produced a stationary model. Thus we can conclude that d=1.







To determine p and q is much harder. Box & Jenkins admit that this stage of their methodology is more like an art than a science. It is very hard to determine what p and q might be even from the acf, never mind the correlogram. Usually, though, we can narrow the possible models down to two or three (especially if p and q are low, which will usually be the case), and then we can use diagnostic tests to determine which is best. Pindyck & Rubinfeld (1976) show that for c(k) to be significantly different from zero its absolute value must be greater than $2/n^{1/2}$ (in the case of our example, approximately 0.164). One of the few values of k that attains this value is k=4. Similarly, the partial autocorrelation function (which is similar to the acf, except the characteristics of the AR and the MA parts of the process are reversed) suggest that k=4 is important (although I haven't provided a diagram). Thus I am going to suggest that the best model to fit the data is either ARIMA(4,1,0), ARIMA(0,1,4) or ARIMA(4,1,4). Further analysis may isolate one model.

Having tentatively suggested values for p,d and q the next stage is to estimate the values of the parameters. It is usual to assume that the white noise errors e are independently and identically distributed. Unfortunately, estimation of the parameters of an ARMA process involve non-linear maximum likelihood techniques. I do not propose to discuss the details here, but the interested reader is referred to Harvey (1981, pp.124-130) or to Pindyck & Rubinfeld (1976, pp.481-489). From a practical point of view, however, most computer statistical packages will carry out the estimation automatically, so the non-specialist reader need not worry about the details. For my model, the following results were produced:

For the ARIMA(4,1,0) model,

$$E(y_t) = 0.0027y_{t-1} + 0.1318y_{t-2} - 0.0652y_{t-3} - 0.2665y_{t-4},$$

for the ARIMA(0,1,4) model,

$$E(y_t) = 0.0452e_{t-1} + 0.0679e_{t-2} - 0.0307e_{t-3} - 0.2936e_{t-4},$$

and for the ARIMA(4,1,4) model,

$$E(y_t) = -0.3722y_{t-1} + 0.4017y_{t-2} + 0.0269y_{t-3} - 0.2021y_{t-4} + 0.4272e_{t-1} - 0.2969e_{t-2} - 0.0706e_{t-3} - 0.1370y_{t-4}.$$

Sometimes we can eliminate a model at this stage if we discover that it is non-stationary. However, all of the models above are stationary (a result which I won't prove).

Since we have assumed that the residuals of the true process are white noise (i.e. distributed normally and independently of each other) then it seems logical that we should use this assumption to test the model. The best way to do this is to use the Box-Pierce test. Denote the correlogram of the residuals by r_k

$$\text{i.e. } r_k = (\Sigma e_t e_{t-k})/(\Sigma e_t^2)$$

where e_t are the estimated residuals. If the model is correctly specified, then for large k the residual autocorrelations r_k are themselves uncorrelated, normally distributed random variables with mean 0 and variance T. Thus the statistic

$$G = T\Sigma r_k^2$$

is approximately chi-squared distributed with K-p-q degrees of freedom, and so by subjecting G to a chi-squared test we can decide whether to accept the model or not.

The computer package which I used for my model (namely MINITAB) produces Box-Pierce statistics for K=12, 24, 26, 48. Unfortunately this test failed to eliminate any of the models (for example, for K=48, G=49.9 for the ARIMA(4,1,0) model. The 95% confidence interval for a chi-squared distribution with 44 degrees of freedom is approximately (29.4,53.3), which G easily falls within), so to evaluate how well each of the models work, compare the predicted future values with the actual values for each model from the table below. Once again, MINITAB produces forecasts of the future values, so there is no point going into the theory behind the forecasting.

Actual Value	ARIMA(4,1,0)	ARIMA(0,1,4)	ARIMA(4,1,4)
174.0	175.1	175.6	175.7
173.0	175.2	175.3	175.9
166.6	177.1	177.7	177.9
165.6	177.7	178.1	177.7
160.8	178.4	178.1	178.8
159.6	178.3	178.1	178.3
158.4	177.9	178.1	178.5
160.5	177.6	178.1	178.3

As a crude measure of how each model performed, the ARIMA(4,1,0) model was out by an average of 7.6% for the 8 observations, the ARIMA(0,1,4) by 7.8% and the ARIMA (4,1,4) by 7.9%. From these measures the ARIMA model is the best, but there is very little to distinguish between the three of them. Although I haven't included them, the actual values all fall withing the 95% confidence intervals for all three models. Notice that the predictions for the first two observations were quite good in each case, but not quite so good after that.

This approach hasn't isolated a single model for my time-series, but rather has given us three potential ones. Remember, time-series modelling is used mainly for short-term forecasting, so having three potential models doesn't pose any real problems - we could just take an average of the three forecasts. The model that I have used isn't very susceptible to time-series modelling and was used purely for illustrative purposes; something like ice-cream sales for the past three years would be a lot more striking. Despite this my model has thrown up some potentially useful forecasts, although I wouldn't advise any reader to use them for arbitrage purposes!

CONCLUSION

The purpose of this article was to give the reader a flavour of the intricacies of time-series modelling. Computer packages such as MINITAB or SPSSX take much of the drudgery out of time-series analysis. The reader who is interested in exploring the theory in more detail is referred to Pindyck & Rubinfeld (1976) for a fairly readable introduction. More advanced material is to be found in Harvey (1981) and Chatfield (1989), while the original work on the subject is found in Box & Jenkins (1970).

Karl Murphy

Bibliography

Box, G.E.P. & Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, 1970 (revised edition 1976).

Chatfield, C., *The Analysis of Time Series*, 1989 (4th edition).

Harvey, A.C., *Time Series Models*, 1981.

Pindyck, R.S. & Rubinfeld, D.L., *Econometric Models and Economic Forecasts*, 1976.