3
Time series modelling

In Chapter 1 we discussed briefly the relationship between the purely statistical approach to time series modelling and an approach which may be more appropriately thought of as economic modelling. In this chapter we discuss the purely statistical approach to time series modelling in more detail. Time series modelling is, of course, a discipline in its own right but our interest is more in the use which has been made of time series techniques in other branches of econometrics. So basic time series representations of data are often useful in modelling expectations (see Chapter 6); the cointegration analysis of Chapter 5 also grew out of this approach.

While good applied econometrics is much more than time series analysis the techniques of time series analysis are now widely seen to be a basic building block of econometrics.

3.1 Autoregressive time series models

Perhaps the simplest, purely statistical time series model is the first-order autoregression, or AR(1) process:

$$x_t = \rho x_{t-1} + \xi_t \tag{3.1}$$

where $|\rho| < 1$ and ξ_t is a white noise error process. Equation (3.1) states that the time series behaviour of x_t can be approximated by assuming that it is determined largely by its own value in the preceding period.

More generally, an nth order autoregressive process, or AR(n) can be written as

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} t \ldots + \rho_n x_{t-n} + \xi_t \quad (3.2)$$

So that n lags of x are deemed to be important in determining the time series behaviour of x_t.

The concept of stationarity was introduced in Chapter 1. In (3.1) we imposed the condition $|\rho| < 1$ specifically in order to guarantee stationarity of the AR(1) process. If we had $|\rho| > 1$, then x would tend to get bigger and bigger each period, in absolute value, and so would be an explosive series. We can write the AR(n) process (3.2) using the lag operator as

$$x_t(1 - \rho_1 L - \rho_2 L^2 \ldots - \rho_n L^n) = \xi_t$$

or

$$\Phi(L)x_t = \xi_t \quad (3.3)$$

The stationarity of an AR(n) process is guaranteed only if the n roots of the polynomial equation

$$\Phi(z) = 0 \quad (3.4)$$

(where z is a real variable) are greater than one in absolute value. For the AR(1) equation (3.1), this condition reduces to the roots of

$$(1 - \rho z) = 0$$

being greater than one in absolute value. If this is so, and if the first root is λ, then the condition is

$$|\lambda| = |1/\rho| > 1$$

which is the same as

$$|\rho| < 1.$$

A necessary, but not sufficient requirement for an AR(n) process to be stationary is that the sum of the n autoregressive coefficients should sum to less than unity:

$$\sum_{i=1}^{n} \rho_i < 1 \quad (3.5)$$

3.2 Moving average time series models

A second kind of pure time series model which is frequently applied is one where the stochastic process under consideration is postulated to be a moving average of current and lagged values of a white noise process. For example, a first-order moving average, MA(1), model would be written:

$$x_t = \xi_t + \theta\xi_{t-1} \quad (3.6)$$

where ξ is a white noise process. More generally, an MA(n) process would be written

$$x_t = (1 + \theta_1 L + \theta_2 L^2 + \ldots \theta_n L^n)\xi_t \quad (3.7)$$

or

$$x_t = \Theta(L)\xi_t \quad (3.8)$$

Because any MA(n) process is, by definition, an average of n stationary white noise terms, it follows that every moving average is stationary.

A property which is often discussed in relation to moving average processes is *invertibility*. Consider the MA(1) process (3.6), rewritten using the lag operator:

$$x_t = (1 + \theta L)\xi_t$$

Equivalently

$$x_t/(1 + \theta L) = \xi_t \quad (3.9)$$

if $|\theta| < 1$, then the left-hand side of (3.9) can be considered as the sum of an infinite geometric progression:

$$x_t(1 - \theta L + \theta^2 L^2 - \theta^3 L^3 + \ldots) = \xi_t \quad (3.10)$$

Alternatively, (3.9) could be derived by substituting repeatedly for lagged values of ξ_t in (3.6). Thus, the MA(1) process has been inverted into a high-order autoregressive process with geometrically declining weights. For the MA(1) model to be invertible, we require $|\theta| < 1$.

In general, the MA(n) process (3.7) is invertible if the roots of the polynomial

$$\Theta(z) = 0 \quad (3.11)$$

are greater than one in absolute value.

It is often useful to consider, at least in theory, infinite-order moving average representations. Such a process would take the form:

$$x_t = \sum_{i=0}^{\infty} \theta_i \xi_{t-i} \quad (3.12)$$

If ξ is a white noise process with constant variance σ_ξ^2, then the variance of x_t is given by

$$x_t = \sigma_\xi^2 \sum_{i=1}^{\infty} \theta_i^2 \quad (3.13)$$

Clearly, although all finite-order moving average processes are stationary, in the infinite-order case it is necessary to place some restrictions on the moving average parameters in order to ensure that the process has finite variance. From (3.13), these restrictions are easily seen to be

$$\sum_{i=0}^{\infty} \theta_i^2 < \infty \tag{3.14}$$

Any model which can be written in the form (3.12) where (3.14) holds is said to be an indeterministic process. An alternative terminology, motivated by (3.14), is to say that it is an infinite-order, square-summable process. Note that any finite-order moving average process can be thought of as an indeterministic process with coefficients on lags beyond a certain point identically equal to zero.

3.3 ARMA and ARIMA process

Sometimes, as we shall discuss later in the chapter, it may be appropriate to model a time series as a combination of both autoregressive and moving average components. Such a process is termed, not unnaturally, an autoregressive moving average, or ARMA, process. An ARMA(p, q) process can be written

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} \ldots + \phi_p x_{t-p} + \xi_t + \theta_1 \xi_{t-1} + \ldots + \theta_q \xi_{t-q}$$

or, more generally:

$$\Phi(L) x_t = \Theta(L) \xi_t \tag{3.15}$$

The stationarity of an ARMA process depends entirely upon its autoregressive component, and requires that all the roots of

$$\Phi(z) = 0$$

lie outside the unit circle (i.e. greater than one in absolute value). Similarly, invertibility of an ARMA process requires that all the roots of the characteristic equation associated with the moving average polynominal, i.e.

$$\Theta(z) = 0$$

lie outside the unit circle.

In Chapter 5 we will discuss integrated processes. An integrated process of order d must be differenced d times before it has a stationary, invertible ARMA representation. If this ARMA representation is of order (p, q) then the original, undifferenced process is said to have an autoregressive integrated moving average representation of order (p, d, q), i.e. it is ARIMA(p, d, q). If a process, x say, has an ARIMA(p, d, q) representation, then $\Delta^d x_t$ has an ARMA(p, q) representation:

$$(1 - \phi_1 L - \ldots - \phi_p L^P) \Delta^d x_t = (1 + \theta_1 L + \ldots + d_q L^q) \xi_t \tag{3.16}$$

3.4 Wold's decomposition

An important result in time series analysis, known as Wold's decomposition theorem, states that any covariance stationary process, x say, can be represented as the sum of a deterministic component, $\bar{x}_t$ say, and an indeterministic component:

$$x_t = \bar{x}_t + \sum_{i=0}^{\infty} \theta_i \xi_{t-i} \tag{3.17}$$

Note that the deterministic component $\bar{x}_t$ need not be constant but it must be non-stochastic – it may for example be a trigonometric function of t (e.g. $\cos \lambda t$). Often, however, $\bar{x}_t$ is in practice thought of as a constant.

Wold's decomposition can be thought of as a central motivation for considering pure time series models. If we believe or know that a certain time series is covariance stationary, then Wold's decomposition tells us that it can be decomposed into deterministic and indeterministic components. We can then take out the deterministic component by assuming that it takes a particular form, such as a constant. This leaves a potentially infinite-order, square-summable component. If the coefficients of this component are very small beyond a certain lag, then it may be convenient to approximate the indeterministic component by a finite-order moving average process, Alternatively, it may be better to approximate the process by an ARMA process of finite orders.

This can be illustrated by means of the following examples. Consider the following indeterministic process:

$$x_t = \xi_t(1 + L + 0.5L^2 + 0.25L^3 + 0.125L^4 + 0.0625L^5 + 0.03125L^6 + \ldots) \tag{3.18}$$

This is clearly a square-summable process since the coefficients are geometrically declining:

$$\sum_{i=0}^{\infty} \theta_i^2 = [1 + 1 + (0.5)^2 + (0.25)^2 + \ldots)$$
$$= 1/[1 - (0.5)^2] + 1 = 2\tfrac{1}{3} < \infty$$

The fact that the moving average coefficients are geometrically declining gives us a clue that (3.18) can be written as a stationary, invertible ARMA process of finite orders.

Consider the ARMA(1, 1) process:

$$x_t = 0.5x_{t-1} + \xi_t + 0.5\xi_{t-1} \tag{3.19}$$

This can be written

$$x_t (1 - 0.5L) = \xi_t(1 + 0.5L)$$

or

$$x_t = \xi_t(1 + 0.5L)(1 - 0.5L)^{-1}$$
$$= \xi_t(1 + 0.5L)(1 + 0.5L + 0.25L^2 + 0.125L^3 + \ldots)$$
$$= \xi_t(1 + L + 0.5L^2 + 0.25L^3 \ldots)$$

So that (3.18) and (3.19) are in fact equivalent. In practice, of course, there is no reason why the indeterministic component should have coefficients which decline in such an exact geometric pattern; but even if they do so only approximately, then an ARMA representation of finite orders may yield a sufficiently close approximation to the time series behaviour of the process.

As another example, consider the indeterministic process:

$$x_t = \xi_t(1 + 0.7L + 0.34L^2 + 0.068L^3 + 0.0136L^4$$
$$+ 0.00272L^5 + \ldots) \tag{3.20}$$

The reader should verify that, providing the coefficients beyond the fifth lag are very small, this process would be well approximated by an ARMA(1, 2) process of the form

$$x_t(1 - 0.2L) = \xi_t(1 + 0.5L + 0.2L^2) \tag{3.21}$$

3.5 Autocovariance and autocorrelation functions

The covariance between two random variables, w, z, is defined to be

$$\text{Cov}(w, z) = E\{(w - E[w])(z - E[z])\}$$

Thus, since any two elements of a stochastic process x_t, x_{t+i} say, are themselves random variables, we have:

$$\text{Cov}(x_t, x_{t+i}) = E\{(x_t - E[x_t])(x_{t+i} - E[x_{t+i}])\} \tag{3.22}$$

This is called the autocovariance function. The autocovariance function is an extremely useful tool in characterising the properties of stochastic processes. For example, in Chapter 1 we defined a white noise stochastic process to be one which had constant mean and finite variance and for which the autocovariance function was always zero.

Consider the first-order autoregressive process

$$x_{t+1} = \rho x_t + \xi_{t+1} \tag{3.23}$$

where $|\rho| < 1$ and where ξ_{t+1} is white noise and is, in particular, uncorrelated with x_t. From (3.23), substituting repeatedly for lagged x, we have:

$$x_{t+1} = \rho^t x_0 + (\rho^t \xi_1 + \rho^{t-1}\xi_2 + \ldots \xi_{t+1}) \tag{3.24a}$$

Since $|\rho| < 1$, ρ^t will be close to zero for large t. Thus, we have:

$$E(x_{t+1}) = E(\rho^t \xi_1 + \rho^{t-1}\xi_2 + \ldots \xi_{t+1}) = 0 \tag{3.24b}$$

Since $E(\xi_i) = 0$ for all i. Also,

$$\text{Var}(x_{t+1}) = E[\rho^{2t}\xi_1^2 + \rho^{2t-2}\xi_2^2 + \ldots \xi_{t+1}^2]$$
$$= \sigma_\xi^2(\rho^{2t} + \rho^{2t-2} + \ldots 1) \tag{3.25}$$

where σ_ξ^2 is the variance of ξ_{t+1}. For large t, the geometric progression (3.25) can be summed:

$$\text{Var}(x_{t+1}) = \sigma_\xi^2/(1 - \rho^2) \tag{3.26}$$

Now consider the autocovariance function for the autoregressive process:

$$\text{Cov}(x_t, x_{t+k}) = E(x_t x_{t+k})$$
$$= E[x_t(\rho x_{t+k-1} + \xi_{t+k})]$$
$$= \rho E(x_t x_{t+k-1}) + E(x_t \xi_{t+k})$$
$$= \rho E(x_t x_{t+k-1})$$
$$= \rho E[x_t(\rho x_{t+k-2} + \xi_{t+k-1})]$$
$$= \rho^2 E(x_t x_{t+k-2})$$

Continuing in this fashion it is easily seen:

$$\text{Cov}(x_t, x_{t+k}) = \rho^k \text{Var}(x_t)$$
$$= \rho^k \sigma_\xi^2/(1 - \rho^2) \tag{3.27}$$

If we write

$$\text{Var}(x_t) = \sigma_x^2$$

and

$$\text{Cov}(x_t, x_{t+k}) = \gamma_k$$

then (3.27) becomes

$$\gamma_k = \rho^k \sigma_x^2 \tag{3.28}$$

Moreover, it is clear that, for all k,

$$\gamma_k = \gamma_{-k}$$

so that it is only necessary to consider non-negative k. The sequence $\{\gamma_k\}_{k=0}^{\infty}$ is the autocovariance function. Note that $\gamma_0 = \sigma_x^2$.

Note that the autocovariance function is not expressed as a pure number – its units of measurement are dependent upon the underlying units of measurement of the x process. Thus, the size of γ_k would differ, for example, if x were measured in pounds rather than pence, or dollars rather than cents (by a factor of 10000). In order to circumvent this problem, it is sometimes more convenient to consider the *autocorrelation function* rather than the autocovariance function, since the former is expressed in terms of pure numbers, and is independent of the units of measurement of the underlying process. The autocorrelation function is obtained simply by dividing each of the γ_k by $\gamma_0 = \sigma_x^2$, thereby cancelling out the units of measurement:

$$\rho_k = \gamma_k/\gamma_0, \; k = 0, 1, 2, \ldots \tag{3.29}$$

Note that $\rho_0 = 1$, by definition.

For the autoregressive, AR(1) process considered above, for example, we have:

$$\rho_k = \gamma_k/\gamma_0 = \rho^k \sigma_x^2/\sigma_x^2 = \rho^k \tag{3.30}$$

As another example, consider the first-order moving average, MA(1) model:

$$y_t = \xi_t + \theta \xi_{t-1} \tag{3.31}$$

where ξ_t is again white noise. We have:

$$\begin{aligned}\gamma_0 &= E[(\xi_t + \theta\xi_{t-1})(\xi_t + \theta\xi_{t-1})] \\ &= E(\xi_t^2) + \theta^2 E(\xi_{t-1}^2) + 2\theta E(\xi_t \xi_{t-1}) \\ &= (1 + \theta^2)\sigma_\xi^2\end{aligned} \tag{3.32}$$

Similarly

$$\begin{aligned}\gamma_1 &= E[(\xi_t + \theta\xi_{t-1})(\xi_{t-1} + \theta\xi_{t-2})] \\ &= E(\xi_t \xi_{t-1}) + \theta E(\xi_{t-1}^2) + \theta^2 E(\xi_{t-1}\xi_{t-2}) \\ &= \theta\sigma_\xi^2\end{aligned} \tag{3.33}$$

Since ξ_t is serially uncorrelated, it also easily seen that

$$\gamma_k = 0, \; k > 1 \tag{3.34}$$

Thus, the autocorrelation function for an MA(1) process is given by:

$$\rho_0 = 1, \; \rho_1 = \theta/(1 + \theta^2), \; \rho_k = 0, \; k > 1 \tag{3.35}$$

3.6 The correlogram

In general, when a researcher is analysing a time series, he or she will have to estimate the autocovariance function by using the sample moments. This estimate is termed the *correlogram*. A commonly used estimator for ρ_k is

$$\gamma_k = c_k/c_0$$

where

$$c_k = \frac{1}{T}\sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \text{ for } x = 0, 1, 2 \ldots \tag{3.36}$$

and where

$$\bar{x} = \frac{1}{T}\sum_{t=1}^{T} x_t \tag{3.37}$$

and where T is the sample size. Sometimes, especially where the sample size is relatively small, T is replaced in the denominator in (3.36) by $(T - k)$ to correct for lost degrees of freedom.

Under the null hypothesis that the x process consists of independent drawings from identical populations, it can be shown that, for large T, c_k will be approximately normally distributed with mean zero and variance $1/T$ under weak conditions:

$$c_k \sim N(0, 1/T) \tag{3.38}$$

Thus, the 95% confidence interval for c_k is given approximately by

$$c_k - 2/\sqrt{T}, \; c_k + 2/\sqrt{T} \tag{3.39}$$

Hence, if this interval does not contain zero, the null hypothesis

$\rho_k = 0$ can be rejected at a nominal significance level of approximately 5%.

If the series is not white noise, then the appropriate formula for the large-sample variance of c_k is

$$\frac{1 + 2(\rho_1^2 + \rho_2^2 + \ldots \rho_n^2)}{T} > \frac{1}{T} \text{ for } n > 0 \qquad (3.40)$$

where n is such that $\rho_k \neq 0$ for $k \leq n$ and $\rho_k = 0$ for $k > n$. Thus, the fact that $1/T$ is only an approximation to the sample autocorrelation variance suggests that, even when the sample autocorrelations are apparently insignificant when using $1/T$ as an estimate of the variance, one should still look for apparent regularities in the *shape of* the correlogram.

3.7 The partial autocorrelation function

The correlogram is useful for identifying a pure moving average model, since there will tend to be a cut-off of significant points on the correlogram after the appropriate lag depth. For autoregressive or mixed processes, however, the order of the autoregressive component may be harder to determine from the correlogram. For this reason, it is usual to use a complementary procedure which involves plotting the estimated coefficient of x_{t-k}, from an OLS estimate of an AR(k) model for x_t, against k. If the observations are generated by an AR(p) process, then the theoretical partial autocorrelations are zero at lags beyond p. Since any invertible MA process can be represented as an AR process with geometrically declining coefficients, the partial autocorrelation function for an MA process should decay slowly.

The identification of the orders of a mixed model may be more difficult to determine, and a good deal of skill must be exercised.

3.8 Common factors

A simple approach to econometric modelling involves deliberate over-parameterisation – more than enough variables and lags are included in a fitted equation, with the objective of eliminating those with poor explanatory power (i.e. insignificant coefficients). In the context of pure time series modelling, such deliberate over-parameterisation will often prove disastrous, because of the presence of

common factors. As an example, suppose that the 'true' model is an ARMA(1, 1):

$$x_t = \phi x_{t-1} + \xi_t + \theta \xi_{t-1} \qquad (3.41)$$

Multiplying the right-hand side of (3.41) by $(1 + \gamma L)/(1 + \gamma L) = 1$:

$$x_t = \frac{(1 + \gamma L)}{(1 + \gamma L)} (\phi x_{t-1} + \xi_t + \theta \xi_{t-1}) \qquad (3.42)$$

$$x_t = (\phi - \gamma) x_{t-1} + \gamma \phi x_{t-2} + \xi_t + (\theta + \gamma) \xi_{t-1} + \gamma \theta \xi_{t-2} \qquad (3.43)$$

(3.42) is in the form of an ARMA(2, 2):

$$x_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \xi_t + \lambda_1 \xi_{t-1} + \lambda_1 \xi_{t-2} \qquad (3.44)$$

with

$$\beta_1 = (\phi - \gamma),\ \beta_2 = \gamma\phi,\ \lambda_1 = (\theta + \gamma),\ \lambda_2 = \gamma\theta$$

Thus, if an ARMA(1, 1) model such as (3.41) is correct, then an ARMA(2, 2) model such as (3.44) will fit well. Moreover, this will be true for *any* value of γ. Thus, (3.44) is in fact unidentified.

This example therefore illustrates the pitfall in over-parameterisation of a pure time series model. A much more subtle approach to model selection is required.

3.9 Model selection: the Box–Jenkins approach

Box and Jenkins (1976) suggest a three-stage approach to pure time series modelling, the three stages being *identification, estimation* and *diagnostic checking*.

At the *identification* stage, a tentative ARIMA model is specified that may approximate the data-generating process for the given sample, through examination of the correlogram and the partial autocorrelation functions. This stage is discussed in more detail below.

Once a model has been tentatively identified, the next stage is to *estimate* its parameters. The estimation stage is also discussed further in a following section.

Once the tentative model has been estimated, a set of estimated residuals are automatically generated. For example, for an AR(1) model we would have the estimated residuals:

$$\hat{\xi}_t = x_t - \hat{\phi} x_{t-1}$$

(where a circumflex denotes a fitted value), while for an ARIMA(0, 1, 1), we have:

$$\hat{\xi} = \Delta\hat{\xi}_{t-1}$$

(where one would normally set $\hat{\xi}_0 = 0$). If the fitted model is correct, then this residual series should be approximately white noise. One test of the adequacy of the model thus includes testing for the whiteness of the fitted reiduals using diagnostic checks such as the Box–Pierce or Ljung–Box portmanteau statistics (see Chapter 4).

If the estimated parameters of the fitted model are significantly different from zero and the fitted residuals appear to be approximate white noise, then the fitted ARIMA model may be held to be adequate. If the model fails on either of these counts, then the identification stage should be returned to.

The Box–Jenkins approach to model selection thus involves three stages: identification, estimation and diagnostic checking. The procedure is summarised by the flow chart presented as Figure 3.1.

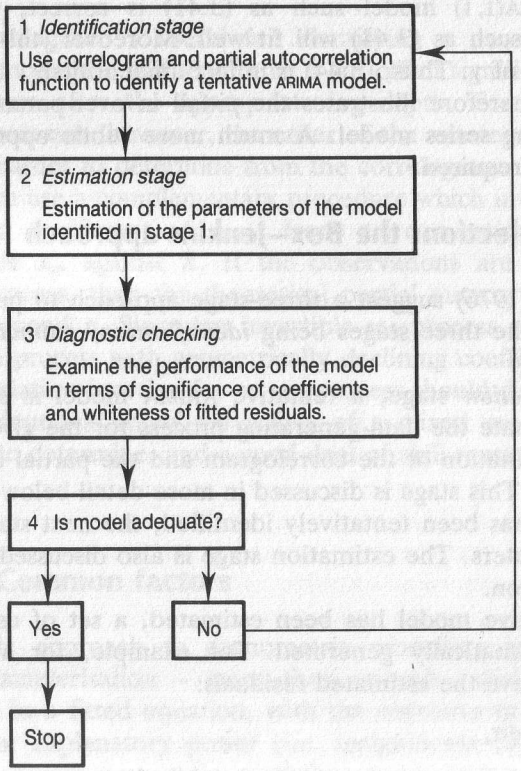

Figure 3.1 Flow diagram for the Box-Jenkins model selection procedure.

3.10 Model identification

The theoretical autocorrelation function for an AR(1) process was shown above to be given by

$$\rho_k = \phi^k \tag{3.45}$$

where ϕ is the first-order autoregressive coefficient. Stationarity of the process requires

$$|\phi| < 1$$

Thus, non-stationarity of a first-order process will be revealed by a correlogram which shows no sign of decay in the absolute magnitude of the estimated autocorrelations. This is, in fact true more generally: if the estimated autocorrelations to not die out or show signs of decay, the data must be transformed to induce stationarity. A common stationarity-inducing transformation with economic data is to take logarithms and then to first difference once. This will induce stationarity, for example, if the data exhibits a fairly consistent average growth rate.

Once apparent stationarity has been achieved, the next step is to identify the orders of the ARIMA process. For a pure moving average process of order q, MA(q), the correlogram will tend to show estimated autocorrelations which are significantly different from zero only up to lag q, while the partial autocorrelation function will tend to taper off. For a pure autoregressive process of order p, the estimated partial autocorrelations will tend to be insignificantly different from zero beyond lag p while the correlogram will show the estimated autocorrelations tapering off.

If neither the correlogram nor the partial autocorrelation function show a definite cut off, then a mixed process is suggested. In seeking to identify the orders of the moving average and autoregressive parts, it then may be useful to think of the correlogram and partial autocorrelation functions of the pure MA and AR processes being superimposed upon one another. For example, if both the correlogram and the partial autocorrelation function show signs of a slow tapering off, then an ARMA(1, 1) process may be identified. Similarly if the correlogram has two spikes at lags one and two and then exponential decay, while the partial autocorrelation function shows either exponential decay or damped sinewave behaviour, then an ARMA(1, 2) may be identified. Table 3.1 lists other possible combinations for low-order ARMA processes.

Table 3.1 Correlogram and partial autocorrelation patterns for low-order ARMA models

Correlogram pattern	Partial autocorrelation pattern	Underlying model
Single spike at lag 1	Exponential decay or damped sinewave	MA(1)
Exponential decay or damped sinewave	Single spike at lag 1	AR(1)
Spike at lag 1 followed by exponential decay or damped sinewave	Spike at lag 1 followed by exponential decay or damped sinewave	ARMA(1, 1)
Spikes at first two lags followed by exponential decay or damped sinewave	Spike at lag 1 followed by exponential decay or damped sinewave	ARMA(1, 2)
Spike at lag 1 followed by exponential decay or damped sinewave	Spikes at first two lags followed by exponential decay or damped sinewave	ARMA(2, 1)
Spikes at first two lags followed by exponential decay or damped sinewave	Spikes at first two lags followed by exponential decay or damped sinewave	ARMA(2, 2)

In general however, identifying mixed processes involves a fair degree of trial and error, and this is why the estimation and diagnostic checking stages are important to see if the tentatively identified model 'flies'.

3.11 Estimation

Estimation of pure time series models can be carried out by maximum likelihood methods, which were discussed in general in Chapter 2.

For a pure autoregressive AR(p), process, maximum likelihood estimation is in fact little different from an ordinary least squares regression applied to x_t on p lags, $x_{t-1}, \ldots, x_{t-p}$.

For moving average and mixed processes, non-linear techniques must be used. In nearly all cases, maximum likelihood estimation can be approximated closely by minimising a sum of squares function. This is because, as we showed in Chapter 2, the likelihood function can always be broken down into a form involving only squared one-step-ahead prediction errors. But the one-step-ahead prediction errors are in fact the residuals of the model.

Consider, for example, the AR(1) model:

$$x_t = \phi x_{t-1} + \xi_t$$

Conditional on information (i.e. observations on x) at time $t-1$, the model predicts x_t as

$$\hat{x}_t = \phi x_{t-1}$$

Thus the prediction error is

$$x_t - \hat{x}_t = \hat{\xi}_t$$

Thus, for any given value of ϕ, the prediction error is just the fitted residual. One could then form the sum of squared residuals function:

$$S(\phi) = \sum \hat{\xi}_t^2$$

and ϕ could be chosen by minimising it – which is, of course, exactly what OLS does.

For an ARMA(p, q) model, the residuals can be generated recursively by an equation of the form:

$$\hat{\xi}_t = x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} - \theta_1 \hat{\xi}_{t-1} - \ldots - \theta_q \hat{\xi}_{t-q}$$

with $\hat{\xi}_p = \hat{\xi}_{p-1} = \hat{\xi}_{p-q+1} = 0$. The conditional sum of squares function would then be:

$$S(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q) = \sum_{t=p+1}^{T} \hat{\xi}_t^2$$

In general, the covariance matrix of the estimates can also be obtained from an approximation of the information matrix and applying standard maximum likelihood properties (see Chapter 2).

3.12 Conclusion

We have discussed how data may be categorised and described purely in terms of their time series properties; this is an important and useful technique which has a wide range of applications. In Chapter 4 we will discuss dynamic structural models; the time series representation is often a useful benchmark against which to measure a structural model. In Chapter 5 we will discuss cointegration; this analysis rests heavily on recognising the importance for structural modelling of the univariate properties of the data we are dealing with. Chapter 6 considers rational expectations; here again time series modelling is used widely as a way of capturing the expectations formation procedure for variables which are not of central interest to the model at hand.