T. C. Mills (1999): The Econometric Modelling of Financial Time Series, 2. laiton, CUP.

The resultant ARMA(p, q) process has the stationarity and invertibility conditions associated with the constituent AR(p) and MA(q) processes respectively. Its ACF will eventually follow the same pattern as that of an AR(p) process after $q-p$ initial values $\rho_1, \ldots, \rho_{q-p}$, while its PACF eventually (for $k > p-q$) behaves like that of an MA(q) process.

Throughout this development, we have assumed that the mean of the process, μ, is zero. Non-zero means are easily accommodated by replacing x_t with $x_t - \mu$ in (2.10), so that in the general case of an ARMA(p, q) process, we have

$$\phi(B)(x_t - \mu) = \theta(B)a_t$$

Noting that $\phi(B)\mu = (1 - \phi_1 - \ldots - \phi_p)\mu = \phi(1)\mu$, the model can equivalently be written as

$$\phi(B)x_t = \theta_0 + \theta(B)a_t$$

where $\theta_0 = \phi(1)\mu$ is a constant or intercept.

2.4 Linear stochastic processes

In this development of ARMA models, we have assumed that the innovations $\{a_t\}$ are uncorrelated and drawn from a fixed distribution with finite variance, and hence the sequence has been termed white noise, i.e., $a_t \sim WN(0, \sigma^2)$. If these innovations are also *independent*, then the sequence is termed *strict* white noise, denoted $a_t \sim SWN(0, \sigma^2)$, and a stationary process $\{x_t\}$ generated as a linear filter of strict white noise is said to be a linear process. It is possible, however, for a linear filter of a white noise process to result in a non-linear stationary process. The distinctions between white and strict white noise and between linear and non-linear stationary processes are extremely important when modelling financial time series and, as was alluded to in section 2.1.2, will be discussed in more detail in chapter 4.

2.5 ARMA model building

2.5.1 *Sample autocorrelation and partial autocorrelation functions*

An essential first step in fitting ARMA models to observed time series is to obtain estimates of the generally unknown parameters, μ, σ_x^2 and the ρ_k. With our stationarity and (implicit) ergodicity assumptions, μ and σ_x^2

can be estimated by the sample mean and sample variance, respectively, of the realisation $\{x_t\}_1^T$

$$\bar{x} = T^{-1}\sum_{t=1}^{T} x_t$$

$$s^2 = T^{-1}\sum_{t=1}^{T}(x_t - \bar{x})^2$$

An estimate of ρ_k is then given by the lag k *sample autocorrelation*

$$r_k = \frac{\sum_{t=k+1}^{T}(x_t - \bar{x})(x_{t-k} - \bar{x})}{Ts^2}, \quad k = 1, 2, \ldots$$

the set of r_ks defining the *sample autocorrelation function* (SACF).

For independent observations drawn from a fixed distribution with finite variance ($\rho_k = 0$, for all $k \neq 0$), the variance of r_k is approximately given by T^{-1} (see, for example, Box and Jenkins, 1976, chapter 2). If, as well, T is large, $\sqrt{T}r_k$ will be approximately standard normal, i.e., $\sqrt{T}r_k \overset{a}{\sim} N(0,1)$, so that an absolute value of r_k in excess of $2T^{-1/2}$ may be regarded as 'significantly' different from zero. More generally, if $\rho_k = 0$ for $k > q$, the variance of r_k, for $k > q$, is

$$V(r_k) = T^{-1}\left(1 + 2\rho_1^2 + \ldots + 2\rho_q^2\right)$$

Thus, by successively increasing the value of q and replacing the ρ_ks by their sample estimates, the variances of the sequence $r_1, r_2, \ldots, r_k$ can be estimated as $T^{-1}, T^{-1}(1 + 2r_1^2), \ldots, T^{-1}(1 + 2r_1^2 + \ldots + 2r_{k-1}^2)$ and, of course, these will be larger, for $k > 1$, than those calculated using the simple formula T^{-1}.

The *sample partial autocorrelation function* (SPACF) is usually calculated by fitting autoregressive models of increasing order: the estimate of the last coefficient in each model is the sample partial autocorrelation, $\hat{\phi}_{kk}$. If the data follow an AR(p) process, then for lags greater than p the variance of $\hat{\phi}_{kk}$ is approximately T^{-1}, so that $\sqrt{T}\hat{\phi}_{kk} \overset{a}{\sim} N(0,1)$.

2.5.2 *Model-building procedures*

Given the r_k and $\hat{\phi}_{kk}$, with their respective standard errors, the approach to ARMA model building proposed by Box and Jenkins (1976) is essentially to match the behaviour of the SACF and SPACF of a particular time series with that of various theoretical ACFs and PACFs, picking the best match (or set of matches), estimating the unknown model parameters (the ϕ_is, θ_is and σ^2), and checking the residuals from the fitted models for any possible misspecifications.

Another popular method is to select a set of models based on prior considerations of maximum possible settings of p and q, estimate each possible model and select that model which minimises a chosen selection criterion based on goodness of fit considerations. Details of these model building procedures, and their various modifications, may be found in many texts, e.g., Mills (1990, chapter 8), and hence will not be discussed in detail: rather, they will be illustrated by way of a sequence of examples.

Example 2.1 Are the returns on the S&P 500 a fair game?

An important and often analysed financial series is the real return on the annual Standard & Poor (S&P) 500 stock index for the US. Annual observations from 1872 to 1995 are plotted in figure 2.6 and its SACF up to $k = 12$ is given in table 2.1. It is seen that the series appears to be stationary around a constant mean, estimated to be 3.08 per cent. This is confirmed by the SACF and a comparison of each of the r_k with their corresponding standard errors, computed using equation

Table 2.1. *SACF of real S&P 500 returns and accompanying statistics*

k	r_k	$s.e.(r_k)$	$Q(k)$
1	0.043	0.093	0.24 [0.62]
2	−0.169	0.093	3.89 [0.14]
3	0.108	0.093	5.40 [0.14]
4	−0.057	0.094	5.83 [0.21]
5	−0.117	0.094	7.61 [0.18]
6	0.030	0.094	7.73 [0.26]
7	0.096	0.094	8.96 [0.25]
8	−0.076	0.096	9.74 [0.28]
9	−0.000	0.097	9.74 [0.37]
10	0.086	0.097	10.76 [0.38]
11	−0.038	0.099	10.96 [0.45]
12	−0.148	0.099	14.00 [0.30]

Note: Figures in [..] give $P(\chi^2_k > Q(k))$.

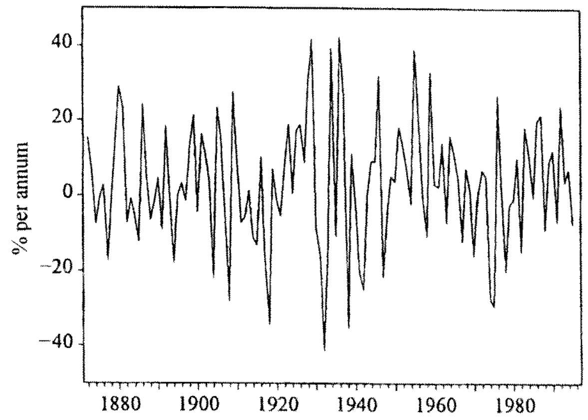

Figure 2.6 Real S&P 500 returns (annual 1872–1995)

(2.10), shows that none are individually significantly different from zero, thus suggesting that the series is, in fact, white noise.

We can construct a 'portmanteau' statistic based on the complete set of r_ks. On the hypothesis that $x_t \sim WN(\mu, \sigma^2)$, Box and Pierce (1970) show that the statistic

$$Q^*(k) = T\sum_{i=1}^{k} r_i^2$$

is asymptotically distributed as χ^2 with k degrees of freedom, i.e., $Q^*(k) \overset{a}{\sim} \chi^2_k$. Unfortunately, simulations have shown that, even for quite large samples, the true significance levels of $Q^*(k)$ could be much smaller than those given by this asymptotic theory, so that the probability of incorrectly rejecting the null hypothesis will be smaller than any chosen significance level. Ljung and Box (1978) argue that a better approximation is obtained when the modified statistic

$$Q(k) = T(T+2)\sum_{i=1}^{k}(T-i)^{-1}r_i^2 \overset{a}{\sim} \chi^2_k$$

is used. $Q(k)$ statistics, with accompanying marginal significance levels of rejecting the null, are also reported in table 2.1 for $k = 1, \ldots, 12$, and they confirm that there is no evidence against the null hypothesis that returns are white noise. Real returns on the S&P 500 would therefore appear to be consistent with the fair game model in which the expected return is constant, being 3.08 per cent per annum.

Example 2.2 Modelling the UK interest rate spread,

As we shall see in chapter 8, the 'spread', the difference between long and short interest rates, is an important variable in testing the expectations hypothesis of the term structure of interest rates. Figure 2.7 shows the spread between 20 year UK gilts and 91 day Treasury bills using monthly observations for the period 1952 to 1995 ($T = 526$), while table 2.2 reports the SACF and SPACF up to $k = 12$, with accompanying standard errors. (The spread may be derived from the interest rate series $R20$ and RS given in the data appendix.)

The spread is seen to be considerably smoother than one would expect if it was a realisation from a white-noise process, and this is confirmed by the SACF, all of whose values are positive and significant (the accompanying portmanteau statistic is $Q(12) = 3701$!). The SPACF has both $\hat{\phi}_{11}$ and $\hat{\phi}_{22}$ significant, thus identifying an AR(2) process. Fitting such a model to the series by ordinary least squares (OLS) regression yields

$$x_t = \underset{(0.023)}{0.045} + \underset{(0.043)}{1.182}\, x_{t-1} - \underset{(0.043)}{0.219}\, x_{t-2} + \hat{a}_t, \quad \hat{\sigma} = 0.448$$

Figures in parentheses are standard errors and the intercept implies a fitted mean of $\hat{\mu} = \hat{\theta}_0/(1 - \hat{\phi}_1 - \hat{\phi}_2) = 1.204$, with standard error 0.529. Since $\hat{\phi}_1 + \hat{\phi}_2 = 0.963$, $-\hat{\phi}_1 + \hat{\phi}_2 = -1.402$ and $\hat{\phi}_2 = -0.219$, the stationarity conditions associated with an AR(2) process are satisfied but,

Table 2.2. *SACF and SPACF of the UK spread*

k	r_k	$s.e.(r_k)$	$\hat{\phi}_{kk}$	$s.e.(\hat{\phi}_{kk})$
1	0.969	0.044	0.969	0.044
2	0.927	0.075	−0.217	0.044
3	0.884	0.094	0.011	0.044
4	0.844	0.109	0.028	0.044
5	0.803	0.121	−0.057	0.044
6	0.761	0.131	−0.041	0.044
7	0.719	0.139	−0.007	0.044
8	0.678	0.146	−0.004	0.044
9	0.643	0.152	0.057	0.044
10	0.613	0.157	0.037	0.044
11	0.586	0.162	0.008	0.044
12	0.560	0.166	−0.020	0.044

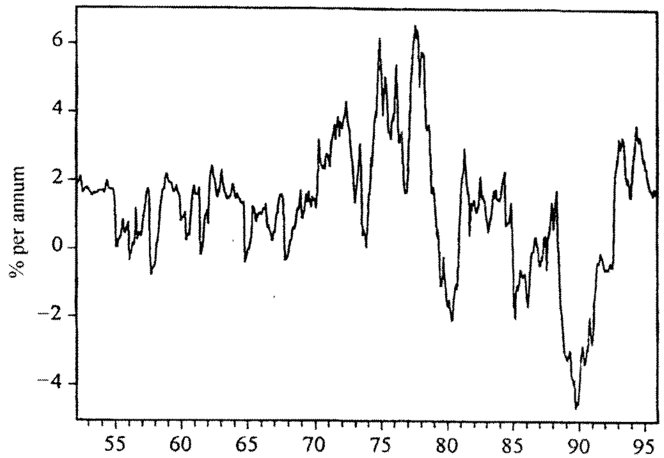

Figure 2.7 UK interest rate spread (monthly 1952.03–1995.12)

although $\hat{\phi}_2$ is negative, $\hat{\phi}_1^2 + 4\hat{\phi}_2 = 0.522$, so that the roots are real, being $\hat{g}_1 = 0.95$ and $\hat{g}_2 = 0.23$. The closeness of $\hat{g}_1$ to unity will be discussed further later in the chapter.

Having fitted an AR(2) process, it is now necessary to check whether such a model is adequate. As a 'diagnostic check', we may examine the properties of the residuals $\hat{a}_t$. Since these are estimates of a_t, they should mimic its behaviour, i.e., they should behave as white noise. The portmanteau statistics Q^* and Q can be used for this purpose, although the degrees of freedom attached to them must be amended: if an ARMA(p, q) process is fitted, they are reduced to $k - p - q$. With $k = 12$, our residuals yield the value $Q(12) = 6.62$, which is now asymptotically distributed as χ^2_{10} and hence gives no evidence of model inadequacy.

An alternative approach to assessing model adequacy is to overfit. For example, we might consider fitting an AR(3) process or, perhaps, an ARMA(2,1) to the series. These yield the following pair of models (methods of estimating MA processes are discussed in, for example, Hamilton, 1994, chapter 5. We use here maximum likelihood (ML))

$$x_t = \underset{(0.023)}{0.044} + \underset{(0.044)}{1.185}\, x_{t-1} - \underset{(0.067)}{0.235}\, x_{t-2} + \underset{(0.044)}{0.013}\, x_{t-3} + \hat{a}_t,$$

$$\hat{\sigma} = 0.449$$

$$x_t = \underset{(0.025)}{0.046} + \underset{(0.196)}{1.137}\, x_{t-1} - \underset{(0.191)}{0.175}\, x_{t-2} + \hat{a}_t + \underset{(0.199)}{0.048}\, \hat{a}_{t-1},$$

$$\hat{\sigma} = 0.449$$

In both models, the additional parameter is insignificant, thus confirming the adequacy of our original choice of an AR(2) process.

Other methods of testing model adequacy are available. In particular, we may construct formal tests based on the Lagrange Multiplier (LM) principle: see Godfrey (1979), with Mills (1990, chapter 8.8) providing textbook discussion.

Example 2.3 Modelling returns on the FTA All Share index

The broadest-based stock index in the UK is the *Financial Times-Actuaries (FTA) All Share*. Table 2.3 reports the SACF and SPACF (up to $k = 12$) of its nominal return calculated using equation (1.2) from monthly observations from 1965 to 1995 ($T = 371$). The portmanteau statistic is $Q(12) = 26.4$, with a marginal significance level of 0.009, and both r_k and $\hat{\phi}_{kk}$ at lags $k = 1$ and 3 are greater than two standard errors. This suggests that the series is best modelled by some ARMA process of reasonably low order, although a number of models could be consistent with the behaviour shown by the SACF and SPACF.

In such circumstances, there are a variety of selection criteria that may be used to choose an appropriate model, of which perhaps the most popular is Akaike's (1974) Information Criteria, defined as

$$AIC(p, q) = \log \hat{\sigma}^2 + 2(p + q)T^{-1}$$

although a criterion that has better theoretical properties is Schwarz's (1978)

Table 2.3. *SACF and SPACF of* FTA All Share *nominal returns*

k	r_k	$s.e.(r_k)$	$\hat{\phi}_{kk}$	$s.e.(\hat{\phi}_{kk})$
1	0.153	0.052	0.153	0.052
2	−0.068	0.053	−0.094	0.052
3	0.109	0.053	0.139	0.052
4	0.093	0.054	0.046	0.052
5	−0.066	0.054	−0.072	0.052
6	−0.017	0.054	−0.006	0.052
7	0.056	0.054	0.030	0.052
8	−0.022	0.054	−0.030	0.052
9	0.101	0.055	0.138	0.052
10	0.044	0.055	−0.017	0.052
11	−0.025	0.055	−0.014	0.052
12	0.019	0.055	0.017	0.052

$$BIC(p, q) = \log \hat{\sigma}^2 + (p + q)T^{-1} \log T$$

A number of other criteria have been proposed, but all are structured in terms of the estimated error variance $\hat{\sigma}^2$ plus a penalty adjustment involving the number of estimated parameters, and it is in the extent of this penalty that the criteria differ. For more discussion about these, and other, selection criteria, see Judge *et al.* (1985, chapter 7.5).

The criteria are used in the following way. Upper bounds, say p_{max} and q_{max}, are set for the orders of $\phi(B)$ and $\theta(B)$, and with $\bar{p} = \{0, 1, \ldots, p_{max}\}$ and $\bar{q} = \{0, 1, \ldots, q_{max}\}$, orders p_1 and q_1 are selected such that, for example

$$AIC(p_1, q_1) = \min AIC(p, q), \quad p \in \bar{p}, q \in \bar{q}$$

with parallel strategies obviously being employed in conjunction with *BIC* or any other criterion. One possible difficulty with the application of this strategy is that no specific guidelines on how to determine $\bar{p}$ and $\bar{q}$ seem to be available, although they are tacitly assumed to be sufficiently large for the range of models to contain the 'true' model, which we may denote as having orders (p_0, q_0) and which, of course, will not necessarily be the same as (p_1, q_1), the orders chosen by the criterion under consideration.

Given these alternative criteria, are there reasons for preferring one to another? If the true orders (p_0, q_0) are contained in the set $(p, q), p \in \bar{p}, q \in \bar{q}$, then for all criteria, $p_1 \geq p_0$ and $q_1 \geq q_0$, almost surely, as $T \to \infty$. However, *BIC* is *strongly consistent* in that it determines the true model asymptotically, whereas for *AIC* an overparameterised model will emerge no matter how long the available realisation. Of course, such properties are not necessarily guaranteed in finite samples, as we find below.

Given the behaviour of the SACF and SPACF of our returns series, we set $\bar{p} = \bar{q} = 3$ and table 2.4 shows the resulting *AIC* and *BIC* values. *AIC* selects the orders (2,2), i.e., an ARMA(2,2) process, while *BIC* selects the orders (0,1), so that an MA(1) process is chosen. The two estimated models are

$$\begin{aligned} x_t = \underset{(0.10)}{1.57} - \underset{(0.059)}{1.054}\, x_{t-1} - \underset{(0.056)}{0.822}\, x_{t-2} + \hat{a}_t + \underset{(0.049)}{1.204}\, \hat{a}_{t-1} \\ + \underset{(0.044)}{0.895}\, \hat{a}_{t-2}, \quad \hat{\sigma} = 5.89 \end{aligned}$$

$$x_t = \underset{(0.04)}{0.55} + \hat{a}_t + \underset{(0.051)}{0.195}\, \hat{a}_{t-1}, \quad \hat{\sigma} = 5.98$$

Table 2.4. *Model selection criteria for nominal returns*

	q / p	0	1	2	3
AIC	0	−5.605	−5.629	−5.632	−5.633
	1	−5.621	−5.631	−5.626	−5.629
	2	−5.622	−5.624	−5.649	−5.647
	3	−5.634	−5.629	−5.629	−5.646
BIC	0	−5.594	−5.608	−5.601	−5.590
	1	−5.600	−5.599	−5.584	−5.576
	2	−5.591	−5.582	−5.596	−5.583
	3	−5.591	−5.576	−5.565	−5.571

Although these models appear quite different, they are, in fact, similar in two respects. The estimate of the mean return implied by the ARMA(2,2) model is 0.55, the same as that obtained directly from the MA(1) model, while the sum of the weights of the respective AR(∞) representations are 0.93 and 0.84 respectively. The short-run dynamics are rather different, however. For the ARMA(2,2) model the initial weights are $\pi_1 = -0.150$, $\pi_2 = 0.108$, $\pi_3 = 0.005$, $\pi_4 = -0.102$; while for the MA(1) they are $\pi_1 = -0.195$, $\pi_2 = 0.038$, $\pi_3 = -0.007$, $\pi_4 = 0.001$.

There is, however, one fundamental difference between the two models: the MA(1) does not produce an acceptable fit to the returns series, for it has a $Q(12)$ value of 20.9, with a marginal significance level of 0.035. The ARMA(2,2) model, on the other hand, has a $Q(12)$ value of only 8.42.

Thus, although theoretically the *BIC* has advantages over the *AIC*, it would seem that the latter selects the model that is preferable on more general grounds. However, we should observe that, for both criteria, there are other models that yield criterion values very close to that of the model selected. Using this idea of being 'close to', Poskitt and Tremayne (1987) introduce the concept of a *model portfolio*. Models are compared to the selected (p_1, q_1) process by way of the statistic, using *AIC* for illustration

$$\Re = \exp\left[-\frac{1}{2}T\{AIC(p_1, q_1) - AIC(p, q)\}\right]$$

Although $\Re$ has no physical meaning, its value may be used to 'grade the decisiveness of the evidence' against a particular model. Poskitt and Tremayne (1987) suggest that a value of $\Re$ less than $\sqrt{10}$ may be thought of as being a close competitor to (p_1, q_1), with the set of closely competing models being taken as the model portfolio.

Using this concept, with $\sqrt{10}$ taken as an approximate upper bound, *no* models are found to be close to (0,1) for the *BIC*. For the *AIC*, however, a model portfolio containing (2,2), (2,3) and (3,3) is obtained.

All these models have similar fits and, although it is difficult to compare them using the estimated AR and MA polynomials, their 'closeness' can be seen by looking at the roots of the characteristic equations associated with the $\phi(B)$ and $\theta(B)$ polynomials. The (2,2) model has complex AR roots $-0.53 \pm 0.74i$ and MA roots of $-0.60 \pm 0.73i$. The (3,3) model has AR roots of 0.82 and $-0.58 \pm 0.75i$ and MA roots 0.76 and $-0.64 \pm 0.75i$: the real roots are close to each other and hence 'cancel out' to leave the (2,2) model. The (0,1) model has a real MA root of -0.20, while the (2,3) model has a real MA root of -0.30 and complex AR roots of $-0.49 \pm 0.77i$ and MA roots of $-0.54 \pm 0.81i$. 'Cancelling out' these complex roots yields the (0,1) model.

We should also note the similarity of the complex AR and MA roots in the higher-order models. This could lead to problems of parameter redundancy, with roots again approximately cancelling out. From this perspective, the (2,2) model selected by the *AIC* may be thought of as providing a trade-off between the parsimonious, but inadequate, (0,1) model selected by *BIC* and the other, more profligately parameterised, models contained in the *AIC* portfolio.

2.6 Non-stationary processes and ARIMA models

The class of ARMA models developed in the previous sections of this chapter relies on the assumption that the underlying process is weakly stationary, thus implying that the mean, variance and autocovariances of the process are invariant under time translations. As we have seen, this restricts the mean and variance to be constant and requires the autocovariances to depend only on the time lag. Many financial time series, however, are certainly not stationary and, in particular, have a tendency to exhibit time-changing means and/or variances.

2.6.1 *Non-stationarity in variance*

We begin by assuming that a time series can be decomposed into a *non-stochastic* mean level and a random error component