Applied Time Series Analysis

1.Time Series and Their Features

Time series display a wide variety of features and an appreciation of these is essential for understanding both their properties and their evolution, including calculating future forecasts and, therefore, unknown values of xt at, say, times T 1 1; T 1 2; …; T 1 h, where h is referred to as the forecast horizon.

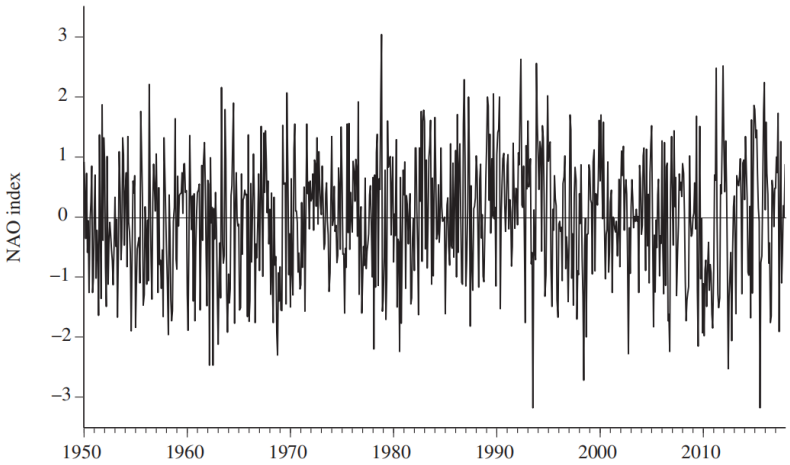

FIGURE 1.1 NAO index: monthly, January 1950–December 2017. *NAO*, North Atlantic Oscillation. *Data from Climate Prediction Center, NOAA Center for Weather and Climate Prediction.*

Such a conclusion may, however, be premature for there might well be internal correlations within the index that could be useful for identifying interesting periodic movements and for forecasting future values of the index. These are typically referred to as the autocorrelations between a current value, xt, and previous, or lagged, values, xt2k, for k 5 1; 2; …. The lag-k (sample) autocorrelation is defined as

$$r_k = \frac{\sum_{t=k+1}^{T} (x_t - \bar{x})(x_{t-k} - \bar{x})}{Ts^2} \tag{1.1}$$

1.6

Fig. 1.3 shows the temperature of a hospital ward taken every hour for several months during 2011 and 2012 (see Iddon et al., 2015, for more details and description of the data).

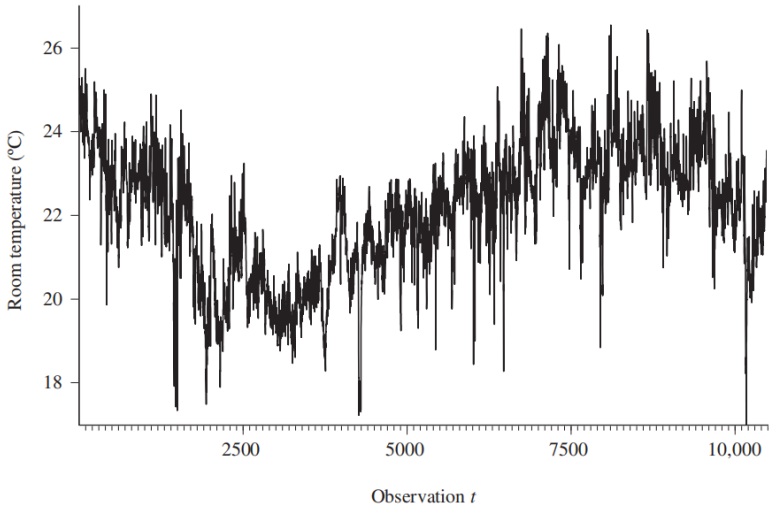

FIGURE 1.3 Hourly temperatures in °C of a ward in Bradford Royal Infirmary during 2011 and 2012. *Data from Iddon, C.R., Mills, T.C., Giridharan, R., Lomas, K.J., 2015. The influence of ward design on resilience to heat waves: an exploration using distributed lag models. Energy Build., 86, 573–588.*

TRENDS

1.10

Just as clearly, the exchange rate does not exhibit an overall trend throughout the observation period, this being informally thought of as a generally monotonic upward or downward movement, which would here imply either a perpetual appreciation or depreciation of the currency; a movement that could not happen in practice as it would, again, offer a one¬way bet to dealers in what is perhaps the most efficient of financial markets.

1.11

Trends, however, are to be found in many time series.3 Fig. 1.6 shows per capita wine and spirits consumption for the United Kingdom from 1950 to 2015. Both show positive trends; that for wine being stronger than that for spirits with, to a first approximation, both trends having reasonably constant slopes. The two series, thus, appear to exhibit linear trends.

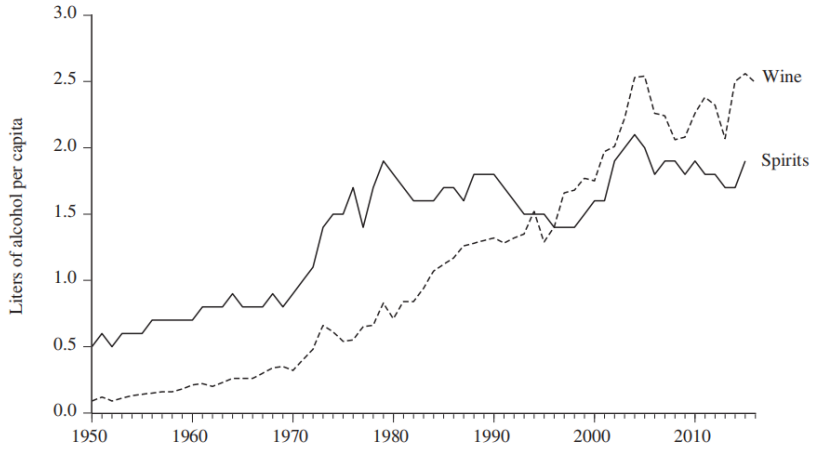

FIGURE 1.6 Annual consumption of wines and spirits in the United Kingdom: liters of alcohol per capita, 1950–2015. *Data from megafile_of_global_wine _data_1835_to_2016 _031117.xlxs. Available from: <www.adelaide.edu.au/wine-econ/databases>.*

VOLATILITY

1.14

A second condition of stationarity is that of constant variance. Fig. 1.9 shows the daily percentage change in the $ d exchange rate plotted in Fig. 1.5. Although the plot of the percentage changes is domi┐nated by the 8% point decline in sterling on June 24, 2016, after the announcement of the Brexit referendum result, the entire sequence of changes is characterized by periods of relative calm interspersed by bursts of volatility, so that the variance of the series changes continuously.

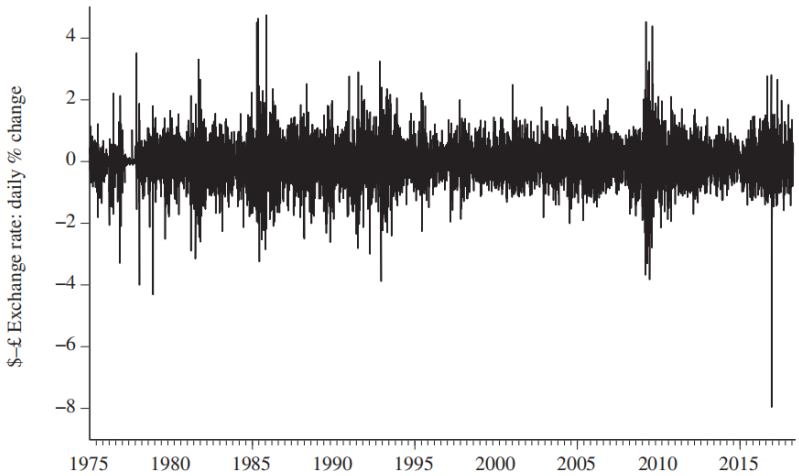

FIGURE 1.9 % Daily change in the $–£ exchange rate: January 1975–December 2017.

COMMON FEATURES

1.15

Two or more time series may contain common features. Consider Fig. 1.10 where the top panel shows long (the yield on 20-year gilts, R20) and short (the yield on 3-month Treasury bills, RS) UK interest rates monthly from 1952 to 2017. Both exhibit random walk nonstationarity, but appear to have a strong tendency to be "bound together" over time, as will be demon┐strated in Example 14.1. The bottom panel of Fig. 1.10 shows the "spread" between the rates, defined as S 5 R20 2 RS, which is stationary, albeit exhibiting persistent (highly positively autocorrelated) deviations from the sample mean of 1.2%. The nonstationarity in the individual series has, thus, been "annihilated" by taking the difference between them. The interest rates thus share a common trend and are said to cointegrate; a potential property of nonstationary time series that will be extensively developed in Chapter 14, Error Correction, Spurious Regressions, and Cointegration, and Chapter 15, VARs with Integrated Variables, VECMs, and Common Trends.

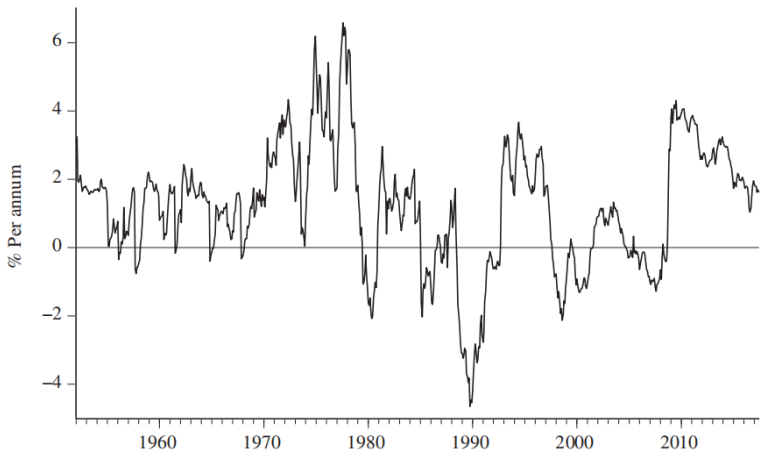

FIGURE 1.10 Long ($R20$) and short (RS) interest rates for the United Kingdom and their spread $S = R20 - RS$: monthly, January 1952–June 2017. *Data from Bank of England.*

2.Transforming Time Series

DISTRIBUTIONAL TRANSFORMATIONS

2.2

Many statistical procedures perform more effectively on data that are normally distributed, or at least are symmetric and not excessively kurtotic (fat-tailed), and where the mean and variance are approximately constant. Observed time series frequently require some form of transformation before they exhibit these distributional properties, for in their "raw" form they are often asymmetric. For example, if a series is only able to take positive (or at least nonnegative) values, then its distribution will usually be skewed to the right, because although there is a natural lower bound to the data, often

zero, no upper bound exists and the values are able to “stretch out,” possibly to infinity. In this case a simple and popular transformation is to take loga┐rithms, usually to the base e (natural logarithms).

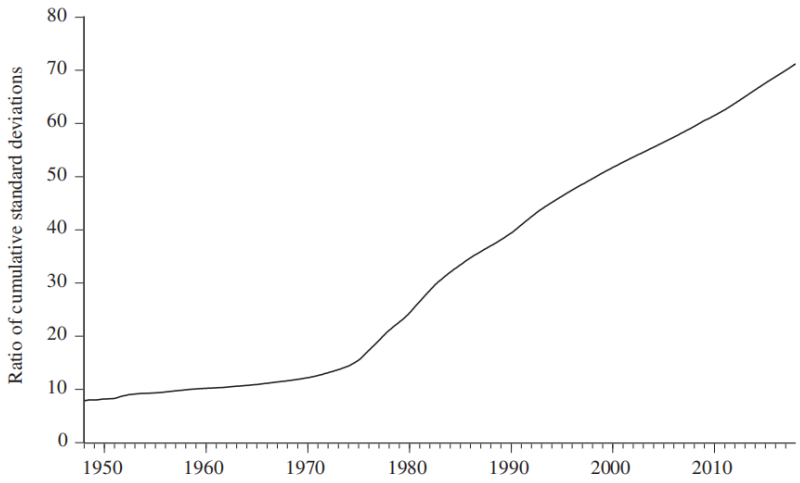

FIGURE 2.3 Ratio of cumulative standard deviations $s_i(\text{RPI})/s_i(\log \text{RPI})$.

also stabilizes the variance. Fig. 2.3 plots the ratio of cumulative standard

deviations, si

RPI ð Þ=si log RPI ð Þ, defined using (1.2) and (1.3) as:

$$s_i^2(x) = i^{-1}\sum_{t=1}^{i}(x_t - \overline{x}_i)^2 \quad \overline{x}_i = i^{-1}\sum_{t=1}^{i} x_t$$

It is also clear that for attaining approximate normality, the availability of a more general class of transformations would be useful. A class of power transformations that contains the logarithmic as a special case is that proposed by Box and Cox (1964) for positive x:3

$$f^{\text{BC}}(x_t, \lambda) = \begin{cases} (x_t^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log x_t & \lambda = 0 \end{cases} \tag{2.1}$$

The restriction to positive values that is required by the BoxCox transformation can be relaxed in several ways. A shift parameter may be introduced in (2.1) to handle situations where x may take negative values but is still bounded below, but this may lead to inferential problems when λ is estimated as in y2.6. Possible alternatives are the signed power transforma┐tion proposed by Bickel and Doksum (1981):

$$f^{\text{SP}}(x_t, \lambda) = \left(\text{sgn}(x_t)\left|x_t^{\lambda}\right| - 1\right)/\lambda \quad \lambda > 0 \tag{2.2}$$

$$f^{\mathrm{GP}}(x_t, \lambda) = \begin{cases} \left((x_t+1)^{\lambda} - 1\right)/\lambda & x_t \geq 0,\ \lambda \neq 0 \\ \log(x_t + 1) & x_t \geq 0,\ \lambda = 0 \\ -\left((-x_t+1)^{2-\lambda} - 1\right)/(2-\lambda) & x_t < 0,\ \lambda \neq 2 \\ -\log(-x_t + 1) & x_t < 0,\ \lambda \neq 2 \end{cases} \tag{2.3}$$

$$f^{\mathrm{IHS}}(x_t, \lambda) = \frac{\sinh^{-1}(\lambda x_t)}{\lambda} = \log \frac{\lambda x_t + \left(\lambda^2 x_t^2 + 1\right)^{1/2}}{\lambda} \quad \lambda > 0 \tag{2.4}$$

zero mean and constant variance. The ML estimator ^λ is then obtained by maximizing over λ the concentrated log-likelihood function:

$$\ell(\lambda) = C_f - \left(\frac{T}{2}\right) \sum_{t=1}^{T} \log \hat{a}_t^2 + D_f(x_t, \lambda) \tag{2.5}$$

STATIONARITY INDUCING TRANSFORMATIONS

2.9

A simple stationarity transformation is to take successive differences of a series

Some caution is required when taking higher-order differences. The second-differences

$$\nabla x_t = x_t - x_{t-1} = x_t - Bx_t = (1 - B)x_t \tag{2.6}$$

$$\nabla^2 x_t = (1-B)^2 x_t = \left(1 - 2B + B^2\right)x_t = x_t - 2x_{t-1} + x_{t-2} \tag{2.7}$$

When attention is focused on the percentage change in a price index, then these changes are typically referred to as the rate of inflation. There may be several different rates of inflation depending on the frequency of observation, say monthly, quarterly, or annually, and the span over which the inflation rate is calculated.

There is a useful relationship between the rate of change of a variable and its logarithm that is often worth bearing in mind, namely:

$$\frac{x_t - x_{t-1}}{x_{t-1}} = \frac{x_t}{x_{t-1}} - 1 \approx \log \frac{x_t}{x_{t-1}} = \log x_t - \log x_{t-1} = \nabla \log x_t \tag{2.8}$$

DECOMPOSING A TIME SERIES AND SMOOTHING

TRANSFORMATIONS

2.15

It is often the case that the long-run behavior of a time series is of particular interest and attention is then focused on isolating these "perma¬ nent" movements from shorter-run, more "transitory," fluctuations, that is, by separating the observations through a decomposition, generally of the form "data 5 fit 1 residual." Because such a decomposition is more than likely going to lead to a smooth series, this might be better thought of as "data 5 smooth 1 rough," terminology borrowed from Tukey (1977). Tukey himself favored running or moving medians to do this, but moving averages (MAs) have become by far the most popular approach to smoothing a time series.

$$\mathrm{WMA}_t(2n+1) = \sum_{i=-n}^{n} \omega_i x_{t-i} \qquad (2.9)$$

Weighted MAs lie behind many of the trend filters that have been pro¬ posed over the years and which will be introduced in Chapter 8, Unobserved Component Models, Signal Extraction, and Filters: see Mills (2011, chapter 10) for a historical discussion.

$$X_t = T_t + S_t + I_t \qquad (2.11)$$

$$X_t = T_t \times S_t \times I_t \qquad (2.12)$$

$$X_t^{\mathrm{SA,A}} = X_t - S_t = T_t + I_t \qquad (2.13)$$

$$X_t^{\mathrm{SA,M}} = \frac{X_t}{S_t} = T_t \times I_t \qquad (2.14)$$

The dominant features of the beer sales series shown in Fig. 1.4 were the downward trend in the second half of the observation period and the prominent seasonal pattern of sales.

The irregular is now calculated "by residual" as:

$$I_t = X_t - T_t - S_t$$

3.ARMA Models for Stationary Time Series

Specifying the complete form of the probability distribution, however, will typically be too ambitious a task, so attention is usually concentrated on the first and second moments; the T means:

$$E(x_1), E(x_2), \ldots, E(x_T)$$

Perhaps the most important simplifying assumption has already been introduced in Chapter 1, Time Series and Their Features, that of stationarity, which, as we have seen, requires the process to be in a state of "statistical equilibrium." A stochastic process is said to be strictly stationary if its prop¬erties are unaffected by a change of time origin, that is, the joint probability distribution at any set of times t1; t2; ... ; tm must be the same as the joint probability distribution at t11k; t21k; ... ; tm1k, where k is an arbitrary shift in time. For m 5 1, strict stationarity implies that the marginal probability dis¬tributions at t1; t2; ... do not depend on time, which in turn implies that as long as E x2t ,N (which is part of a finite second moment assumption) both the mean and variance of xt must be constant, so that:

$$E(x_1) = E(x_2) = \cdots = E(x_T) = \mu$$

$$V(x_1) = V(x_2) = \cdots = V(x_T) = \sigma_x^2$$

$$\rho_k = \frac{Cov(x_t, x_{t-k})}{(V(x_t)V(x_{t-k}))^{1/2}} = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_x^2}$$

The set of assumptions that the mean and variance of xt are both constant and the autocovariances and autocorrelations depend only on the lag k is known as weak or covariance stationarity.

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

WOLD'S DECOMPOSITION AND AUTOCORRELATION

3.6

A fundamental theorem in time series analysis, known as Wold's decom¬position, states that every weakly stationary, purely nondeterministic, stochas¬tic process xt 2 μ can be written as a linear combination (or linear filter) of a sequence of uncorrelated random variables.2 "Purely nondeterministic" means that any deterministic components have been subtracted from xt 2 μ. Such components are those that can be perfectly predicted from past values of them¬selves and examples commonly found are a (constant) mean, as is implied by writing

the process as xt 2 μ, periodic sequences (e.g., sine and cosine func¬tions), and polynomial or exponential sequences in t. This linear filter representation is given by:

$$x_t - \mu = a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \psi_0 = 1 \qquad (3.2)$$

Although Eq. (3.2) may appear complicated, many realistic models result from specific choices for the ψ-weights. Taking μ 5 0 without loss of generality, choosing ψj 5 φj allows (3.2) to be written as:

$$\begin{aligned} x_t &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots \\ &= a_t + \phi(a_{t-1} + \phi a_{t-2} + \cdots) \\ &= \phi x_{t-1} + a_t \end{aligned}$$

$$x_t - \phi x_{t-1} = a_t \qquad (3.3)$$

The lag operator B introduced in y2.10 allows (possibly infinite) lag expressions to be written in a concise way. For example, by using this opera¬tor the AR(1) process can be written as:

$$(1 - \phi B)x_t = a_t$$

$$\begin{aligned} x_t &= (1-\phi B)^{-1} a_t = \left(1 + \phi B + \phi^2 B^2 + \cdots\right) a_t \\ &= a_t + \phi a_{t-1} + \phi^2 a_{t-2} + \cdots \end{aligned} \qquad (3.4)$$

The ACF of an AR(1) process may now be deduced. Multiplying both sides of (3.3) by xt2k, k . 0, and taking expectations yields:

$$\gamma_k - \phi \gamma_{k-1} = E(a_t x_{t-k}). \qquad (3.5)$$

FIRST-ORDER MOVING AVERAGE PROCESSES

3.12

Now consider the model obtained by choosing ψ1 5 2 θ and ψj 5 0, j $ 2, in (3.2):

$$x_t = a_t - \theta a_{t-1} \qquad (3.6)$$

Since any MA model consists of a finite number of ψ-weights, all MA models are stationary. To obtain a converging autoregressive representa┐tion, however, the restriction θ , 1 must be imposed. This restriction is known as the invertibility condition and implies that the process can be writ┐ten in terms of an infinite autoregressive representation:tationary Time Series

$$x_t = \pi_1 x_{t-1} + \pi_2 x_{t-2} + \cdots + a_t$$

GENERAL AR AND MA PROCESSES

3.16

Extensions to the AR(1) and MA(1) models are immediate. The gen┐eral autoregressive model of order p (AR(p)) can be written as:

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \cdots - \phi_p x_{t-p} = a_t$$

The stationarity conditions required for convergence of the ψ-weights are that the roots of the characteristic equation:

$$\phi(B) = (1 - g_1 B)(1 - g_2 B)\cdots\left(1 - g_p B\right) = 0$$

The ACF of an AR(2) process is given by:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}$$

for k $ 2 with starting values ρ0 5 1 and ρ1 5 φ1= 1 2 φ2 ┌ └ The behavior of this ACF for four combinations of φ1; φ2 is shown in Fig. 3.3 If g1 and g2 are real (cases (A) and (C)), the ACF is a mixture of two damped exponen┐tials. Depending on their sign, the autocorrelations can also damp out in an oscillatory manner. If the roots are complex (cases (B) and (D)), then the ACF follows a damped sine wave. Fig. 3.4 shows plots of generated time series from these four AR(2) processes, in each case with atBNID 0; 25 ð Þ. Depending on the signs of the real roots, the series may be either smooth or jagged, while complex roots tend to induce “periodic-type” behavior.

The kth partial autocorrelation is the coefficient φkk in the AR(k) process: 6963}

$$x_t = \phi_{k1} x_{t-1} + \phi_{k2} x_{t-2} + \cdots + \phi_{kk} x_{t-k} + a_t \qquad (3.8)$$

and measures the additional correlation between xt and xt2k after adjustments have been made for the intervening lags.

In general, the φkk can be obtained from the Yule

┐ Walker equations that correspond to (3.8). These are given by the set shown in Eq. (3.7) with p 5 k and φi 5 φii, and solving for the last coefficient φkk using Cramer's Rule leads to:

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

We may also entertain combinations of autoregressive and moving average models. For example, consider the natural combination of the AR(1)

and MA(1) models, known as the first-order autoregressive-moving average, or ARMA(1,1), process:

$$x_t - \phi x_{t-1} = a_t - \theta a_{t-1} \tag{3.9}$$

$$(1 - \phi B)x_t = (1 - \theta B)a_t$$

$$x_t = \psi(B)a_t = \left(\sum_{i=0}^{\infty} \phi^i B^i\right)(1 - \theta B)a_t = a_t + (\phi - \theta)\sum_{i=1}^{\infty} \phi^{i-1} a_{t-i} \tag{3.10}$$

More general ARMA models are obtained by combining AR(p) and MA(q) processes:

$$x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p} = a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

$$\left(1 - \phi_1 B - \cdots - \phi_p B^p\right)x_t = \left(1 - \theta_1 - \cdots - \theta_q B^q\right)a_t \tag{3.11}$$

An essential first step in fitting ARMA models to observed time series is to obtain estimates of the generally unknown parameters μ, σ2x , and the ρk. With the stationarity and (implicit) ergodicity assumptions, μ and σ2x can be estimated by the sample mean and sample variance, respectively, of the reali┐zation x1; x2; … ; xT , that is, by Eqs. (1.2) and (1.3). An estimate of ρk is then provided by the lag k sample autocorrelation given by Eq. (1.1), which, because of its importance, is reproduced here:

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

Consider a time series generated as independent observations drawn from a fixed distribution with finite variance (i.e., ρk 5 0 for all k 6¼ 0). Such a series is said to be independent and identically distributed or i.i.d. For such a series the variance of rk is approximately given by T21. If T is large as well, ffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffi T p rk will be approximately standard normal, so that rk Ba N 0; T21 ┌ └ , implying that an absolute value of rk in excess of 2=ffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffiffi T p may be regarded as “significantly” different from zero at the 5% significance level. Moregener┐ally, if ρk 5 0 for k . q, the variance of rk, for k . q, is:

$$V(r_k) = T^{-1}\left(1 + 2\rho_1^2 + \cdots + 2\rho_q^2\right). \qquad (3.12)$$

$$Q(k) = T(T+2)\sum_{i=1}^{k} (T-i)^{-1} r_i^2 \overset{a}{\sim} \chi^2(k) \qquad (3.13)$$

4.ARIMA Mode ls for Nons3.14

NONSTATIONARITY

4.1

The autoregressive-moving average (ARMA) class of models relies on the assumption that the underlying process is weakly stationary, which restricts the mean and variance to be constant and requires the autocovar┐iances to depend only on the time lag. As we have seen, however, many time series are certainly not stationary, for they tend to exhibit time┐changing means and/or variances.

To deal with such nonstationarity, we begin by characterizing a time series as the sum of a nonconstant mean level plus a random error component:

$$x_t = \mu_t + \varepsilon_t \tag{4.1}$$

$$x_t = \mu_t + \varepsilon_t = \sum_{j=0}^{d} \beta_j t^j + \psi(B)a_t \tag{4.2}$$

Trends of this type can be removed by a simple transformation.

Consider the linear trend obtained by setting d 5 1, where, for simplicity, the error component is assumed to be a white noise sequence:

$$x_t = \beta_0 + \beta_1 t + a_t \tag{4.3}$$

and subtracting this from (4.3) itself yields

$$x_t - x_{t-1} = \beta_1 + a_t - a_{t-1} \tag{4.4}$$

An alternative way of generating a nonstationary mean level is to employ ARMA models whose autoregressive parameters do not satisfy sta- tionarity conditions. For example, consider the AR(1) process:

$$x_t = \phi x_{t-1} + a_t \tag{4.5}$$

$$x_t = x_0\phi^t + \sum_{i=0}^{t} \phi^i a_{t-i} \tag{4.6}$$

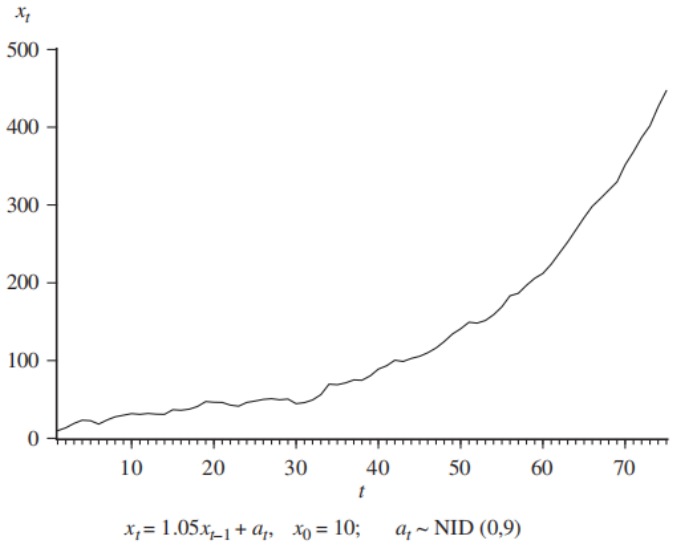

FIGURE 4.2 A simulated explosive AR(1) model.

essentially follows an exponential curve with the generating ats playing almost no further part. The same behavior would be observed if additional autoregressive and moving average terms were added to the model, as long as the stationarity conditions are violated.

ARIMA PROCESSES

4.6

As we can see from (4.6), the solution to (4.5) is explosive if φ . 1 but stationary if φ , 1. The case φ 5 1 produces a process that is neatly balanced between the two. If xt is generated by the model:

$$x_t = x_{t-1} + a_t \tag{4.7}$$

$$x_t = x_{t-1} + \theta_0 + a_t \tag{4.8}$$

$$\gamma_{k,t} = Cov(x_t, x_{t-k}) = (t-k)\sigma^2 \quad k \geq 0$$

$$\rho_{k,t} = \frac{\gamma_{k,t}}{\sqrt{\gamma_{0,t}\gamma_{0,t-k}}} = \frac{t-k}{\sqrt{t(t-k)}} = \sqrt{\frac{t-k}{t}}$$

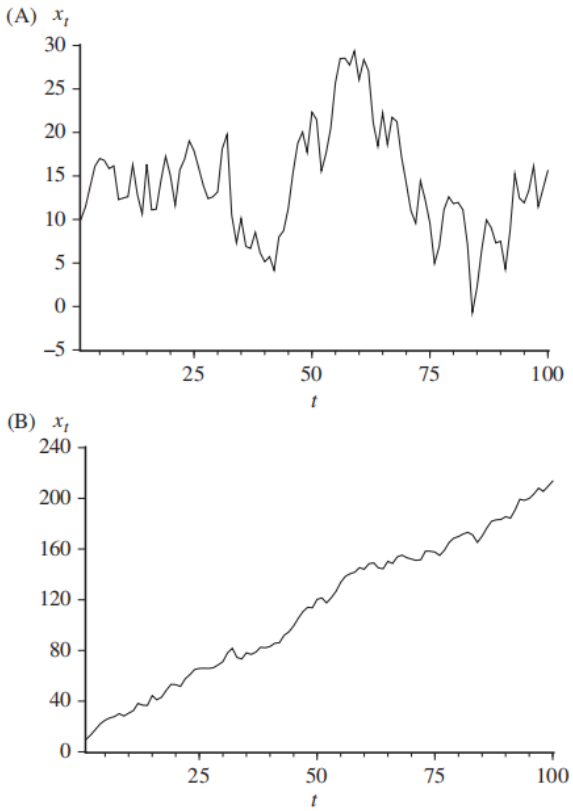

FIGURE 4.3 Simulated random walks. (A) $x_t = x_{t-1} + a_t$, $x_0 = 10$; $a_t \sim$ NID(0,9) and (B) $x_t = 2 + x_{t-1} + a_t$, $x_0 = 10$; $a_t \sim$ NID(0,9).

The random walk is an example of a class of nonstationary models known as integrated processes. Eq. (4.8) can be written as:

$$\nabla x_t = \theta_0 + a_t$$

and so first-differencing xt leads to a stationary model, in this case the white noise process at. Generally, a series may need first-differencing d times to attain stationarity, and the series so obtained may itself be autocorrelated. If this autocorrelation is modeled by an ARMA(p,q) process, then the model for the original series is of the form:

$$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t \qquad (4.9)$$

It will usually be the case that the order of integration d or, equiva¬lently, the degree of differencing, will be 0, 1 or, occasionally, 2 (recall the examples in yy2.92.10). It will continue to be the case that the autocorrela¬tions of an ARIMA process will be close to 1 for all nonlarge k. For exam¬ple, consider the (stationary) ARMA(1,1) process:

$$x_t - \phi x_{t-1} = a_t - \theta a_{t-1}$$

as previously noted, the model corresponds to assuming that rdxt

can be represented by a stationary and invertible ARMA process.

Alternatively, for d $ 1, (4.10) can be inverted to give:

$$x_t = S^d w_t \qquad (4.11)$$

This type of nonstationary behavior is often referred to as homogenous

nonstationarity, and it is important to discuss why this form of nonstationar┐

ity is felt to be useful when describing the behavior of time series from

many fields. Consider again the first-order autoregressive process (4.2). A

basic characteristic of the AR(1) model is that, for both φ| | | , 1 and φ . 1,

the "local" behavior of a series generated from the model is heavily depen┐

dent on the level of xt. In the former case local behavior will always be dom┐

inated by an affinity to the mean, while in the latter the series will

eventually increase rapidly with t. For many time series, however, local

behavior appears to be roughly independent of level, and this is what we

mean by homogenous nonstationarity.

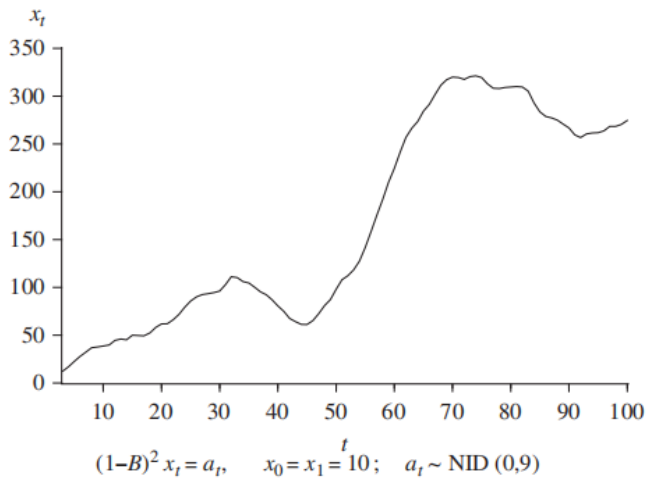

FIGURE 4.4 A simulated "second-difference" model.

In general, if a constant is included in the model for dth differences,

then a deterministic polynomial trend of degree d is automatically allowed

for. Equivalently, if θ0 is taken to be nonzero, then:

$$E(w_t) = E\left(\nabla^d x_t\right) = \mu_w = \frac{\theta_0}{\left(1 - \phi_1 - \phi_2 - \cdots - \phi_p\right)}$$

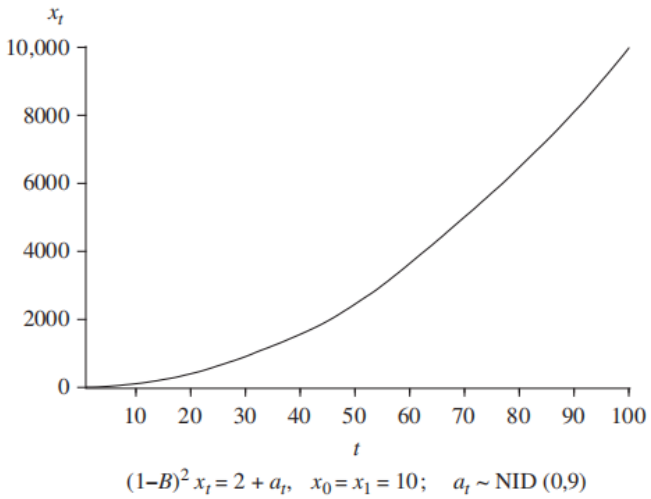

FIGURE 4.5 A simulated "second-difference with drift" model.

5. Unit Roots, Difference and Trend Stationarity,and Fractional Differencing

DETERMINING THE ORDER OF INTEGRATION OF A

TIME SERIES

5.1

As we have shown in yy4.54.13, the order of integration, d, is a cru┐cial determinant of the properties exhibited by a time series. If we restrict ourselves to the most common values of zero and one for d, so that xt is either I 0ð Þ or I 1ð Þ, then it is useful to bring together the properties of these two processes. If xt is I 0ð Þ, which we will sometimes denote xtBI 0ð Þ even though such a notation has been used previously to denote the distributional characteristics of a series, then, if we assume for convenience that xt has zero mean;

\1. the variance of xt is finite and does not depend on t;

\2. the innovation at has only a temporary effect on the value of xt;

\3. the expected length of time between crossings of x 5 0 is finite, so that xt fluctuates around its mean of zero;

\4. the autocorrelations, ρk, decrease steadily in magnitude for large enough k, so that their sum is finite.

If, on the other hand, xtBI 1ð Þ with x0 5 0, then;

\1. the variance of xt goes to infinity as t goes to infinity;

\2. an innovation at has a permanent effect on the value of xt because xt is the sum of all previous innovations: recall from y4.10 that xt 5 r21at 5 Sat 5 Pt21 i50 at2i;

\3. the expected time between crossings of

x 5 0 is infinite;

\4. the autocorrelations ρk-1 for all k as t goes to infinity.

5.2

As was shown in Chapter 2, Transforming Time Series, the fact that a time series is nonstationary is often self-evident from a plot of the series. Determining the actual form of nonstationarity, however, is not so easy from just a visual inspection and, consequently, an examination of the SACFs for alternative differences of the series may be required.

To see why this may be so, recall from y3.17 that a stationary AR(p) pro¬ cess requires that all roots gi in

$$\phi(B) = (1 - g_1 B)(1 - g_2 B)\ldots(1 - g_p B)$$

$$\rho_k = A_1 g_1^k + A_2 g_2^k + \cdots + A_p g_p^k$$

$$A_1 g_1^k = A_1(1-\delta)^k = A\left(1 - \delta k + \delta^2 k^2 - \ldots\right) \cong A_1(1 - \delta k)$$

Sole reliance on the SACF can sometimes lead to problems of overdif¬

ferencing. Although further differences of a stationary series will themselves be stationary, overdifferencing can lead to serious difficulties. Consider the stationary MA(1) process xt 5 1 2 θB ð Þat. The first-difference of this is

$$\begin{aligned}\nabla x_t &= (1-B)(1-\theta B)a_t \\ &= \left(1-(1+\theta)B+\theta B^2\right)a_t \\ &= \left(1-\theta_1 B-\theta_2 B^2\right)a_t.\end{aligned}$$

TESTING FOR A UNIT ROOT

5.4

Given the importance of choosing the correct order of differencing, we should have available a formal testing procedure to determine d. To intro┐ duce the issues involved in developing such a procedure, we begin by con┐ sidering the simplest case, that of the zero mean AR(1) process:

$$x_t = \phi x_{t-1} + a_t \quad t = 1, 2, \ldots, T \tag{5.1}$$

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} x_{t-1}x_t}{\sum_{t=1}^{T} x_t^2}$$

$$t_\phi = \frac{\hat{\phi}_T - 1}{\hat{\sigma}_{\hat{\phi}_T}} = \frac{\hat{\phi}_T - 1}{\left(s_T^2 / \sum_{t=1}^{T} x_{t-1}^2\right)^{1/2}} \tag{5.2}$$

$$\hat{\sigma_{\phi_T}} = \left(\frac{s_T^2}{\sum_{t=1}^{T} x_{t-1}^2}\right)^{1/2}$$

Unfortunately, the distribution of tφ does not have the usual limiting standard normal distribution when φ 5 1. Rather, its distribution is as shown in Fig. 5.1, where it is called the τ-distribution in recognition of its nonnorm┐ ality. The test statistic (5.2) is renamed τ, rather than tφ, and is often known as the Dickey ┌ Fuller test, as indeed is the distribution.1 Fig. 5.1 shows that the limiting distribution of τ is approximately stan┐ dard normal, but shifted to the left by roughly 0.3: the large T 5%, 2.5%, and 1% critical

values for τ are 21.95, 22.23, and 22.58, rather than the stan┐dard normal critical values of 21.65, 21.96, and 22.33.

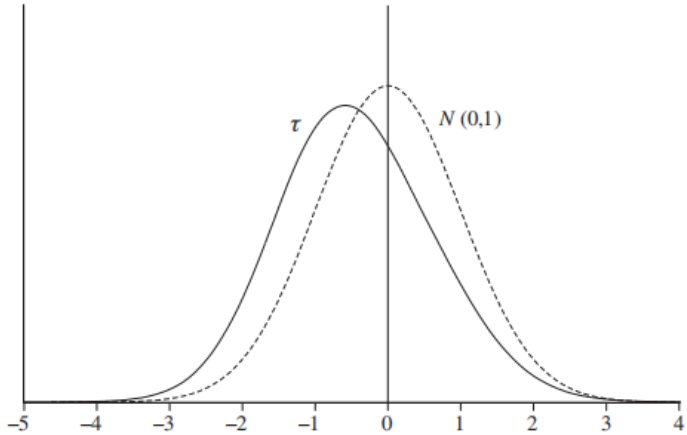

FIGURE 5.1 Limiting distribution of τ.

This case has the merit of being simple, but is not particularly realis┐tic, for it implies that the alternative to a driftless random walk is a sta┐tionary AR(1) process about a zero mean, which would rule out series that can take only positive values, of which there are many (most economic and financial time series, for example). A more sensible alternative would be for the AR(1) process to fluctuate about a nonzero mean, so that we have the model:

$$x_t = \theta_0 + \phi x_{t-1} + a_t \quad t = 1, 2, \ldots, T \tag{5.3}$$

A further generalization is to allow the innovations to be autocorre┐lated. Suppose that xt is generated by the AR(p) process:

$$\left(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p\right) x_t = \theta_0 + a_t$$

$$x_t = \theta_0 + \sum_{i=1}^{p} \phi_i x_{t-i} + a_t \tag{5.4}$$

$$\phi = \sum_{i=1}^{p} \phi_i$$

$$\delta_i = -\sum_{j=i+1}^{p-1} \phi_j \quad i = 1, 2, \ldots, p-1$$

$$x_t = \theta_0 + \phi x_{t-1} + \sum_{i=1}^{k} \delta_i \nabla x_{t-i} + a_t \qquad (5.5)$$

The analysis of yy5.4 5.7 has implicitly assumed that the AR order p is known, so that we are certain that xt is generated by a pth order autore┐gression. If the generating process is an ARMA(p,q), then the τμ statistic obtained from estimating the model:

$$x_t = \theta_0 + \phi x_{t-1} + \sum_{i=1}^{p} \delta_i \nabla x_{t-i} + a_t - \sum_{j=1}^{q} \theta_j a_{t-j}$$

TREND VERSUS DIFFERENCE STATIONARITY

5.9

In the unit root testing strategy outlined previously, the implicit null hypothesis is that the series is generated as a driftless random walk with, possibly, autocorrelated innovations. In popular terminology introduced by Nelson and Plosser (1982), xt is said to be difference stationary (DS),

$$\nabla x_t = \varepsilon_t$$

$$\nabla x_t = \theta + \varepsilon_t \qquad (5.7)$$

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t \qquad (5.8)$$

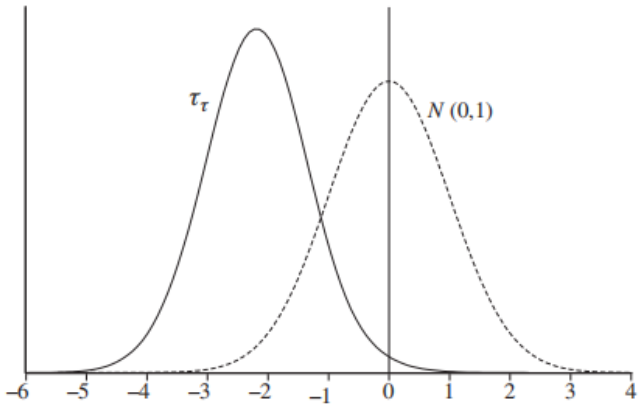

FIGURE 5.3 Limiting distribution of τ_τ.

with a unit root and drift [model (5.7)]. Indeed, rejection of a null hypothesis

of a unit root is unlikely using this statistic if the series is stationary around a

linear trend and becomes impossible as the sample size increases, i.e., the test

is inconsistent, results that were first announced by Perron (1988).5

5.10

A test of (5.7) against (5.8) is, however, straightforward to carry out

by using an extension of the testing methodology previously discussed: the

ADF regression (5.5) is simply extended by the inclusion of the time trend t

as an additional regressor,

$$x_t = \beta_0 + \beta_1 t + \phi x_{t-1} + \sum_{i=1}^{k} \delta_i \nabla x_{t-i} + a_t \qquad (5.9)$$

$$\tau_\tau = \frac{\hat{\phi}_T - 1}{se\left(\hat{\phi}_T\right)}$$

TESTING FOR MORE THAN ONE UNIT ROOT

5.11

This development of unit root tests has been predicated on the

assumption that xt contains at most one unit root, so that it is either I(0) or I 1ð Þ

. If the null hypothesis of a unit root is not rejected, then it may be nec¬

essary to test whether the series contains a second unit root—in other words

whether xt is I 2ð Þ and, thus, needs differencing twice to induce stationarity.

OTHER APPROACHES TO TESTING FOR A UNIT ROOT

5.12

An alternative unit root test to the ADF for dealing with autocorrela¬

tion in at, which also allows for heterogeneity of variance, has been proposed

by Phillips and Perron (1988). Rather than including extra lags of rxt to

ensure that the errors of (5.4) are indeed white noise, the idea here is to esti¬

mate an "unaugmented" model—(5.3), say—and to modify the test statistics

so that the effects of any autocorrelation are accounted for. This will enable

the same DF limiting distributions and, hence, critical values to be used.

Under a specific set of conditions placed upon at, known as weak depen¬

dency, which are described in detail by Phillips (1987), the τμ statistic

obtained from the estimation of (5.3) is modified to

$$Z(\tau_\mu) = \tau_\mu(\hat{\sigma}_0/\hat{\sigma}_\ell) - \frac{1}{2}(\hat{\sigma}_\ell^2 - \hat{\sigma}_0^2)/\Sigma_\ell \qquad (5.10)$$

$$\begin{aligned} \hat{\sigma}_0^2 &= T^{-1}\sum_{t=1}^{T}\hat{a}_t^2 \\ \hat{\sigma}_\ell^2 &= \hat{\sigma}_0^2 + 2T^{-1}\sum_{j=1}^{\ell}w_j(\ell)\left(\sum_{t=j+1}^{T}\hat{a}_t\hat{a}_{t-j}\right) \\ \Sigma_\ell^2 &= T^{-2}\hat{\sigma}_\ell^2\sum_{t=2}^{T}(x_{t-1}-\bar{x}_{-1})^2 \quad \bar{x}_{-1} = (T-1)^{-1}\sum_{t=1}^{T-1}x_t \end{aligned} \qquad (5.11)$$

σ2‘ is a consistent estimator of the long-run variance and employs a window or kernel function wj ‘Þ to weight the sample autocovariances appearing in the formula. This ensures that the estimator remains positive, with ‘ acting as a truncation lag, much like k in the ADF regression. A range of kernel functions are available, such as the “triangular” set of lag weights wj ‘ð Þ ‘ 2 j= ‘ 1 1 ð Þ. Z τμ ┐ └ is often referred to as the Phillips ┌ Perron (PP) non-parametric unit root test. Z τμ ┐ └ has the same limiting distribution as τμ, so that the latter’s critical values may again be used. If xt has zero mean, the adjusted statistic, Z τð Þ, is as in (5.10) with x21 removed and has the same limiting distribution as τ. If a time trend is included then a further adjustment is required to enable the statistic, now denoted Z ττ ð Þ, to have the limiting ττ distribution (Mills and Markellos, 2008, page 87

5.13

Many alternative unit root tests have been developed since the initial ADF and PP tests were introduced. A recurring theme of unit root testing is the low power and severe size distortion inherent in many tests: see, espe┐cially, the review by Haldrup and Jansson (2006). For example, the PP tests suffer severe size distortions when there are moving average errors with a large negative root and, although their ADF counterparts are better behaved in this respect, the problem is not negligible even here. Moreover, many tests

have low power when the largest autoregressive root is close to, but never┐theless less than, unity.

A related issue is that unlike many hypothesis testing situations, the power of tests of the unit root hypothesis against stationary alternatives depends less on the number of observations per se and more on the span of the data (i.e., the length of the observation period). For a given number of observations, power has been found to be highest when the span is longest; conversely, for a given span, additional observations obtained using data sampled more frequently lead to only a marginal increase in power, the increase becoming negligible as the sampling interval is decreased. Hence, a

series containing fewer annual observations over an extended time period will often lead to unit root tests having higher power than those computed from a series containing more observations over a shorter period.

5.14

Several subsequent tests have explicitly concentrated on improving power and reducing size distortion. Many of these are based on generalized least squares (GLS) “detrending” prior to calculating a test statistic. The DF¬ GLS and point optimal unit root tests of Elliott, Rothenberg, and Stock (ERS, 1996) were the initial pair of tests based on this approach, both employing the quasi-differences

$$d(x_t|\alpha) = \begin{cases} x_t & \text{if } t = 1 \\ x_t - \alpha x_{t-1} & \text{if } t > 1 \end{cases}$$

$$x_t^d = \phi x_{t-1}^d + \sum_{i=1}^{k} \nabla x_{t-i}^d + a_t$$

$$P_\tau = \frac{S(\overline{\phi}) - \overline{\phi} S(1)}{\hat{\sigma}_\ell^2}$$

Ng and Perron (2001) construct four further tests that are based on the GLS-detrended data xdt . It is useful to define the term

$$\kappa = T^{-2} \sum_{t=1}^{T-1} \left(x_t^d\right)^2$$

whereupon the test statistics are defined as

$$\mathrm{MZ}^d(\phi) = \frac{T^{-1}\left(x_T^d\right) - \hat{\sigma}_\ell^2}{2\kappa}$$

$$\mathrm{MSB}^d = \left(\frac{\kappa}{\hat{\sigma}_\ell^2}\right)^{1/2}$$

$$\mathrm{MZ}_t^d = \mathrm{MZ}^d(\phi) \times \mathrm{MSB}^d$$

$$\mathrm{MP}_T^d(k) = \frac{\bar{c}_k^2\kappa - (\bar{c}_k - k)T^{-1}\left(x_T^d\right)^2}{\hat{\sigma}_\ell^2}$$

ESTIMATING TRENDS ROBUSTLY

5.17

Consider again the linear trend model (5.8): xt 5 β0 1 β1t 1 εt. As

we have seen, correct specification of the trend is crucially important for

unit root and stationarity testing. As was pointed out in y5.9, incorrectly

excluding a linear trend renders the τμ statistic inconsistent, while it is also

the case that unnecessarily including a trend vastly reduces the power of

the ττ test, with similar problems affecting the KPSS stationarity statistics

ημ and ητ .

Often, however, the trend parameter β1 is of direct interest, especially

when ascertaining whether a trend is present β1 60 ð Þ or not β1 5 0 ð Þ. This may be assessed by either constructing a direct test of the no trend hypothe¬sis β1 5 0 or by forming a confidence interval for β1. Such tests rely on

whether εt, and hence, xt, is either I 0ð Þ or I 1ð Þ, but this can only be estab¬

lished after a unit root or stationarity test has been performed—yet the prop¬

erties of these latter tests rely, in turn, on whether a trend has been correctly

included or not! This circularity of reasoning has prompted the development

of trend function testing procedures that are robust, in the sense that, at least

asymptotically, inference on the trend function is unaffected as to whether εt

is I 0ð Þ or I 1ð Þ. 5.18 To develop robust tests of trend, we start with the simplest case

in which εt 5 ρεt21 1 at, where εt is I 0ð Þ if ρ⌟⌟⌟⌟ , 1 and I 1ð Þ if

5 1. We then wish to test H0:β1 5 β01 against the alternative H1β1 6¼ β0\1. If εt is known to be I 0ð Þ then an optimal test of H0 against H1 is given by the "slope" t-ratio

$$z_0 = \frac{\hat{\beta}_1 - \beta_1^0}{s_0} \qquad s_0 = \sqrt{\frac{\hat{\sigma}_\varepsilon^2}{\sum_{t=1}^{T}(t-\bar{t})^2}} \tag{5.12}$$

$$\nabla x_t = \beta_1 + \nu_t \quad t = 2, \ldots, T \tag{5.13}$$

where $\nu_t = \nabla \varepsilon_t$:

$$z_1 = \frac{\tilde{\beta}_1 - \beta_1^0}{s_1} \qquad s_1 = \sqrt{\frac{\tilde{\sigma}_\nu^2}{T-1}}$$

Here

$$\tilde{\beta}_1 = (T-1)\sum_{t=2}^{T} \nabla x_t = (T-1)(x_T - x_1)$$

What if it is not known whether εt is I 0ð Þ or I ð Þ? Harvey, Leybourne, and Taylor (HLT, 2007) show that a weighted average of z0 and z1, say

$$z_\lambda = (1 - \lambda(U,S))z_0 + \lambda(U,S)z_1 \tag{5.14}$$

where U is a standard unit root test statistic, S is a standard trend-stationarity

test statistic and

$$\lambda = \exp\left(-\kappa\left(\frac{U}{S}\right)^2\right) \tag{5.15}$$

$$\hat{\beta}_{1,\lambda} \pm c_{\alpha/2}\frac{s_0 s_1}{(1-\lambda(U,S))s_1 + \lambda(U,S)s_0} \tag{5.16}$$

HLT suggest using the DF-GLS statistic for U and the KPSS ητ statis┐

tic for S, along with setting the constant in (5.15) to κ 5 0:00025. In a typical case in which there is autocorrelation, the augmented versions of these tests should be employed.

5.22

The HLT approach has the advantage that it may be computed using only statistics that are readily available, but no claims can be made for its optimality and other tests may have better size and power properties. Two alternative approaches have been proposed by Bunzel and Vogelsang (2005) and Perron and Yabu (2009), but while the latter approach appears to have some good statistical properties, its resulting trend estimate is rather more complicated to obtain.

FRACTIONAL DIFFERENCING AND LONG MEMORY

5.23

Our analysis has so far only considered cases where the order of differencing, d, is either zero, one, or possibly two. Concentrating on the first

two cases, if xtBI

1

ð Þ

then its ACF declines linearly, whereas if xtBI

0

ð Þ

its

ACF exhibits an exponential decline, so that observations far apart may be assumed to be independent, or at least nearly so. Many empirically observed

time series, however, although appearing to satisfy the assumption of stationarity (perhaps after differencing), nevertheless seem to exhibit some dependence between distant observations that, although small, is by no means negligible. This may be termed long range persistence or dependence, although the term long memory is now popular.6

Such series have particularly been found in hydrology, where the long-range persistence of river flows is known as the Hurst effect (see, e.g., Mandelbrot and Wallis, 1969; Hosking, 1984), but many financial time series also exhibit similar characteristics of extremely long persistence. This may be characterized as a tendency for large values to be followed by further large values of the same sign, in such a way that the observations appear to go through a succession of "cycles," including long cycles whose length is comparable to the total sample size.

5.24

The class of ARIMA processes may be extended to model this type of long-range persistence by relaxing the restriction to just integer values of d, so allowing fractional differencing within the class of AR-fractionally integrated-MA (ARFIMA) processes. This notion of fractional differencing/integration seems to have been proposed independently by Granger and Joyeux (1980) and Hosking (1981) and is made operational by considering the binomial series expansion of rd for any real d . 2 1:

$$\begin{aligned}\nabla^d &= (1-B)^d = \sum_{k=0}^{\infty} \frac{d!}{(d-k)!k!}(-B)^k \\ &= 1 - dB + \frac{d(d-1)}{2!}B^2 - \frac{d(d-1)(d-2)}{3!} + \cdots\end{aligned} \tag{5.17}$$

$$f(\omega) = 1 + 2\sum_{k=1}^{\infty} \phi^k \cos k\omega = \frac{\sigma_a^2}{2\pi} \frac{1-\phi^2}{1+\phi^2-2\phi\cos\omega} \tag{5.21}$$

$$f_x(\omega) = |1-z|^{-2d} f_y(\omega) = \left(2\sin|\omega/2|\right)^{-2d} f_y(\omega) \tag{5.22}$$

TESTING FOR FRACTIONAL DIFFERENCING

5.28

The "classic" approach to detecting the presence of long memory in a

time series is to use the range over standard deviation or rescaled range

(R=S) statistic. This was originally developed by Hurst (1951) when studying

river discharges and a revised form was later proposed in an economic context

by Mandelbrot (1972). It is defined as the range of partial sums of deviations

of a time series from its mean, rescaled by its standard deviation, i.e.,

$$R_0 = \hat{\sigma}_0^{-1}\left[\max_{1\le i\le T}\sum_{t=1}^{i}(x_t-\bar{x}) - \min_{1\le i\le T}\sum_{t=1}^{i}(x_t-\bar{x})\right] \quad \hat{\sigma}_0^2 = T^{-1}\sum_{t=1}^{T}(x_t-\bar{x})^2 \tag{5.23}$$

The first term in brackets is the maximum of the partial sums of the first

i deviations of xt from the sample mean. Since the sum of all T deviations of

the xts from their mean is zero, this maximum is always nonnegative. The

second term is the minimum of the same sequence of partial sums, and hence

is always nonpositive. The difference between the two quantities, called the

"range" for obvious reasons, is therefore always nonnegative, so that R0 $ 0.

5.29

Although it has long been established that the R=S statistic is certainly

able to detect long-range dependence, it is nevertheless sensitive to short-run

influences. Consequently, any incompatibility between the data and the pre¬

dicted behavior of the R=S statistic under the null of no long run dependence

need not come from long memory, but may merely be a symptom of short¬

run autocorrelation.

The R=S statistic was, thus, modified by Lo (1991), who incorporated

short-run dependence into the estimator of the standard deviation, replacing

(5.23) with

$$R_q = \hat{\sigma}_q^{-1}\left[\max_{1\le i\le T}\sum_{t=1}^{i}(x_t-\bar{x}) - \min_{1\le i\le T}\sum_{t=1}^{i}(x_t-\bar{x})\right] \qquad (5.24)$$

5.31

An obvious approach to testing for fractional differencing is to con¬

struct tests against the null of either d 5 1 or d 5 0. ADF and nonparametric

tests of d 5 1 and KPSS tests of d 5 0 are consistent against fractional d

alternatives, but have the drawback that rejection of the respective nulls can¬

not be taken as evidence of the presence of fractional d.

Extensions of the Dickey

┌ Fuller testing approach have, thus, been pro¬ posed, which evaluate the null hypothesis δ 5 0 in the model rd1δxt 5 at. Breitung and Hassler (2002), building upon the Lagrange Multiplier (LM) approach of Agiakloglou and Newbold (1994), show that a simple test of this null is the t-statistic testing φ 5 0 from the regression

$$\nabla^d x_t = \phi x_{t-1}^{*} + a_t \qquad (5.25)$$

$$\varepsilon_t = \phi\varepsilon_{t-1}^{*} + \sum_{i=1}^{p}\gamma_i\nabla^d x_{t-i} + a_t \qquad (5.26)$$

ESTIMATING THE FRACTIONAL DIFFERENCING PARAMETER

5.32

A drawback of the FD-F procedure is that, if d1 is not known a priori,

as it is in the standard Dickey┌ Fuller case, then a consistent estimate must

be provided. A variety of estimators have been suggested, many of which

involve quite complex calculations. Perhaps the simplest is suggested by R=S

analysis and is

$$\tilde{d} = \frac{\log R_0}{\log T} - 0.5$$

$$\log I(\omega_j) = a - d \log 4 \sin^2\left(\frac{\omega}{2}\right) \qquad (5.27)$$

6.Breaking and Nonlinear Trends

BREAKING TREND MODELS

6.1

The trend stationary (TS) versus difference stationary (DS) dichotomy

and associated testing procedure outlined in yy5.9┌ 5.10 is both simple and

straightforward to implement, but is it necessarily realistic? Could the TS

alternative of a "global" linear trend be too simplistic in some situations,

thus indicating that a more sophisticated trend function might be warranted?

Often a more plausible candidate for a trend is a linear function that "breaks"

at one or more points in time.

There are several ways in which a trend may break. Assume, for simplic┐

ity, that there is a single break at a known point in time Tbc 1 , Tbc , T ┌ └ ,

with the superscript "c" denoting the "correct" break date, a distinction that

will become important in due course. The simplest breaking trend model is

the “level shift” in which the level of xt shifts from μ0 to μ1 5 μ0 1 μ at Tbc.

This may be parameterized as

$$x_t = \mu_0 + (\mu_1 - \mu_0)\mathrm{DU}_t^c + \beta_0 t + \varepsilon_t = \mu_0 + \mu\mathrm{DU}_t^c + \beta_0 t + \varepsilon_t \qquad (6.1)$$

$$x_t = \mu_0 + \beta_0 t + (\beta_1 - \beta_0)\mathrm{DT}_t^c + \varepsilon_t = \mu_0 + \beta_0 t + \beta\mathrm{DT}_t^c + \varepsilon_t \qquad (6.2)$$

$$\begin{aligned} x_t &= \mu_0 + (\mu_1 - \mu_0)\mathrm{DU}_t^c + \beta_0 t + (\beta_1 - \beta_0)\mathrm{DT}_t^c + \varepsilon_t \\ &= \mu_0 + \mu\mathrm{DU}_t^c + \beta_0 t + \beta\mathrm{DT}_t^c + \varepsilon_t \end{aligned} \qquad (6.3)$$

$$\nabla x_t = \beta_0 + \mu\nabla\mathrm{DU}_t^c + \varepsilon_t^* = \beta_0 + \mu\mathrm{D}(\mathrm{TB}^c)_t + \varepsilon_t^* \qquad (6.4)$$

$$\nabla x_t = \beta_0 + \beta\nabla\mathrm{DT}_t^c + \varepsilon_t^* = \beta_0 + \beta\mathrm{DU}_t^c + \varepsilon_t^* \qquad (6.5)$$

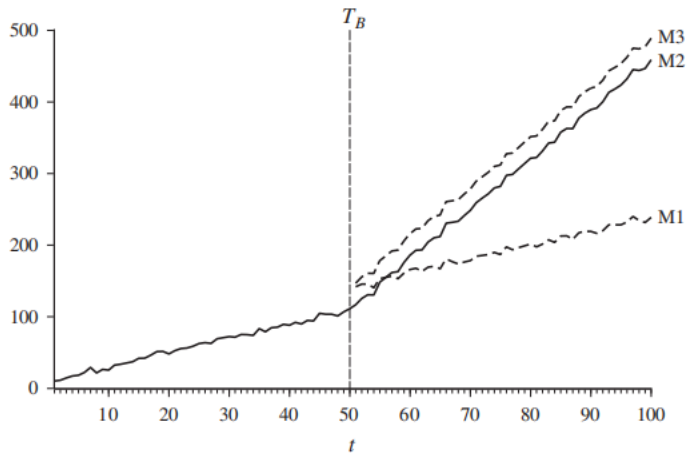

FIGURE 6.1 Examples of TS breaking trend functions.

BREAKING TRENDS AND UNIT ROOT TESTS

6.3

How can we distinguish between TS breaking trends and breaking DS processes? Clearly unit root tests should be applicable, but what is the influ┐ence of breaking trends upon such tests? Perron (1989, 1990: see also Perron and Vogelsang, 1993) was the first to consider the impact of breaking trends

and shifting levels on unit root tests, showing that standard tests of the type discussed in Chapter 5, Unit Roots, Difference and Trend Stationarity, and Fractional Differencing, are not consistent against TS alternatives when the trend function contains a shift in slope. Here the estimate of the largest auto¬

regressive root is biased toward unity and, in fact, the unit root null becomes impossible to reject, even asymptotically. Although the tests are consistent against a shift in the intercept of the trend function, their power is neverthe¬

less reduced considerably because the limiting value of the estimated autoregressive root is inflated above its true value.

$$\tilde{x}_t^i = \tilde{\phi}^i \tilde{x}_{t-1}^i + \sum_{j=0}^{k} \gamma_j \mathrm{D}(\mathrm{TB}^c)_{t-j} \sum_{j=1}^{k} \delta_j \nabla \tilde{x}_{t-j}^i \quad i = A,\ C \tag{6.7}$$

$$\tilde{x}_t^B = \tilde{\phi}^i \tilde{x}_{t-1}^B + \sum_{j=1}^{k} \delta_j \nabla \tilde{x}_{t-j}^B + a_t \tag{6.8}$$

One way to incorporate such a gradual change into the trend function is to suppose that xt responds to a trend shock in the same way as it reacts to any other shock. Recalling the ARMA specification for εt made in y6.2, viz., φBð Þεt 5 θ Bð Þat, this would imply that ψ ð Þ 5 φ BÞ21θ Bð Þ, which would be

analogous to an “innovation outlier” (IO) model. With this specification, tests

for the presence of a unit root can be performed using a direct extension of the ADF regression framework to incorporate dummy variables as appropriate:

$$x_t = \mu^A + \theta^A \mathrm{DU}_t^c + \beta^A t + d^A \mathrm{D}(\mathrm{TB}^c)_t + \phi^A x_{t-1} + \sum_{i=1}^{k} \delta_i \nabla x_{t-i} + a_t \tag{6.9}$$

$$x_t = \mu^B + \theta^B \mathrm{DU}_t^c + \beta^B t + \gamma^B \mathrm{DT}_t^c + \phi^B x_{t-1} + \sum_{i=1}^{k} \delta_i \nabla x_{t-i} + a_t \tag{6.10}$$

$$x_t = \mu^C + \theta^C \mathrm{DU}_t^c + \beta^C t + \gamma^C \mathrm{DT}_t^c + d^C \mathrm{D}(\mathrm{TB}^c)_t + \phi^C x_{t-1} + \sum_{i=1}^{k} \delta_i \nabla x_{t-i} + a_t \tag{6.11}$$

On the face of it, the finding that the S&P 500 index may be characterized as a breaking TS process looks odd, as it would suggest that (the logarithms) of stock prices evolve as transitory deviations from a linear trend (albeit with a single break at the Great Crash) and, hence, would allow traders to bet successfully on a reversal of price whenever the index moved too far away from the trend line, in sharp contrast to the findings for the UK All Share Index in Example 5.4.

Perron (1989) argued, however, that the rejection of the null hypothesis of a unit root, conditional on the possibility of shifts in the underlying trend function at known dates, does not imply that the series can necessarily be modeled as stationary fluctuations around a completely deterministic trend function. To do this, Perron invoked the general statistical principle that a rejection of a null hypothesis does not imply the acceptance of any specific alternative hypothesis. What Perron had in mind was the class of maintained hypotheses that could be parameterized as

$$\begin{aligned} x_t &= \eta_t + \varepsilon_t \qquad \eta_t = \mu_t + \beta_t t \\ \nabla\mu_t &= \upsilon(B)v_t \quad \nabla\beta_t = \omega(B)w_t \end{aligned} \tag{6.12}$$

The intuitive idea behind the model (6.12) is that the coefficients of the trend function are determined by long-term "fundamentals" which rarely change. The exogeneity assumption about the changes in the trend function

is then a device that allows us to take these rare shocks out of the noise and

into the trend without having to model specifically the stochastic behavior of

μt and βt. Perron's framework is then to test whether the noise εt is an inte¬

grated process or not by removing those events that occur at dates where

nonzero values of vt and wt are believed to have occurred and to model these

as part of the trend function.

UNIT ROOTS TESTS WHEN THE BREAK DATE IS UNKNOWN

6.10

The procedure set out in yy6.3┌6.5 is only valid when the break date

is known independently of the data, for if a systematic search for a break is

carried out then the limiting distributions of the tests are no longer appropri¬

ate. Problems also occur if an incorrect break date is selected exogenously,

with the tests then suffering size distortions and loss of power.

Consequently, several approaches have been developed that treat the occur¬

rence of the break date as unknown and needing to be estimated: see, for

example, Zivot and Andrews (1992), Perron (1997), and Vogelsang and Perron

(1998). Thus, suppose now that the correct break date Tbc is unknown. Clearly,

if this is the case then the models of yy6.3┌6.5 are not able to be used until

some break date, say ^T, is selected, since none of the dummy variables that these models require can be defined until this selection has been made.

6.11

Two data-dependent methods for choosing ^Tb have been considered,

both of which involve estimating the appropriate detrended AO regression,

(6.7) or (6.8), or IO regression (6.9┌6.11), for all possible break dates. The

first method chooses ^Tb as the break date that is most likely to reject the unit

root hypothesis, which is the date for which the t-statistic for testing φ 5 1 is

minimized (i.e., is most negative). The second approach involves choosing ^T

b as the break date for which some statistic that tests the significance of the break parameters is maximized.

This is equivalent to minimizing the residual sum of squares across all possible

regressions, albeit after some preliminary trimming has been performed, that is,

if only break fractions τ 5 Tb=T between 0 , τmin; τmax , 1 are considered.

ROBUST TESTS FOR A BREAKING TREND

6.13

Of course, the broken trend will typically be of interest in itself, and

so it is natural for the robust trend analysis of yy5.17 ┌ 5.22 to have been

extended to cover such specifications, most notably by Harvey, Leybourne,

and Taylor (HLT, 2009). If the break date is known to be at Tbc with break

fraction τc then, focusing on the segmented trend model (B), the HLT

method is extended by focusing on autocorrelation corrected t-tests of β 5 0

in (6.2) and (6.5), which we denote as t0 τc ð Þ and t1 τc ð Þ. A weighted average of these two statistics is again considered,

$$t_\lambda = \lambda(S_0(\tau^c), S_1(\tau^c)) \times \left|t_0(\tau^c)\right| + (1 - \lambda(S_0(\tau^c), S_1(\tau^c))) \times \left|t_1(\tau^c)\right| \qquad (6.13)$$

CONFIDENCE INTERVALS FOR THE BREAK DATE AND

MULTIPLE BREAKS

6.16

When the break date is estimated it is often useful to be able to pro┐

vide a confidence interval for the unknown Tbc. Perron and Zhu (2005) show

that for the segmented trend model (B) and I 1ð Þ errors

$$\sqrt{T}(\hat{\tau} - \tau^c) \xrightarrow{d} N\left(0, 2\sigma^2/15\beta^2\right)$$

The limiting distributions for the break date do not depend on the auto¬

correlation structure of the errors, only requiring an estimate of the error var¬

iance σ2. When the errors are I 1ð Þ the limiting distribution is invariant to the

location of the break, whereas for I 0ð Þ errors, the limiting distribution

depends on the location of the break in such a way that the variance is smal¬

ler the closer the break is to the middle of the sample. In both cases the vari¬

ance decreases as the shift in slope increases.

For model (C) the limiting distributions for the break date are no longer

normal but are complicated functions of nuisance parameters and, thus, can

only be simulated, so that no simple results are available.

6.17

In theory, all the procedures available when there is only one break,

may be extended to the case of multiple breaks, but, in practice, when there

are multiple breaks at unknown times, only the sequential procedure of

Kejriwal and Perron (2010), which requires specialized programming, is cur¬

rently available.

NONLINEAR TRENDS

6.18

A breaking linear trend may be interpreted as a form of nonlinear

trend and their use begs the question of why not model nonlinear trends

explicitly, particularly when the shift in the trend evolves smoothly over a

sequence of observations rather than occurring instantaneously with a sharp break. While many types of deterministic nonlinear trends could be specified, the logistic smooth transition (LSTR) and exponential smooth transition (ESTR) have proved to be popular. The LSTR function may be defined as

$$S_t(\gamma, m) = (1+\exp(-\gamma(t-mT)))^{-1} \tag{6.14}$$

while the ESTR takes the form

$$S_t(\gamma, m) = 1 - \exp\left(-\gamma(t-mT)^2\right) \tag{6.15}$$

Analogous to (6.1)–(6.3), three alternative smooth transition trend models may then be specified as

$$x_t = \mu_0 + \mu S_t(\gamma, m) + \varepsilon_t \tag{6.16}$$

$$x_t = \mu_0 + \beta_0 t + \mu S_t(\gamma, m) + \varepsilon_t \tag{6.17}$$

$$x_t = \mu_0 + \beta_0 t + \mu S_t(\gamma, m) + \beta t S_t(\gamma, m) + \varepsilon_t \tag{6.18}$$

$$x_t = \mu + \beta t + \sum_{f=1}^{n} \gamma_{1f} \sin\left(\frac{2\pi f t}{T}\right) + \sum_{f=1}^{n} \gamma_{2f} \cos\left(\frac{2\pi f t}{T}\right) + \varepsilon_t \tag{6.19}$$

$$\nabla x_t = \mu + \beta t + \phi x_{t-1} + \sum_{f=1}^{n} \gamma_{1f} \sin\left(\frac{2\pi f t}{T}\right) + \sum_{f=1}^{n} \gamma_{2f} \cos\left(\frac{2\pi f t}{T}\right) + \sum_{i=1}^{k} \delta_i \nabla x_{t-i} + \varepsilon_t \tag{6.21}$$

7.An Introduction to Forecasting With Univariate Models

FORECASTING WITH AUTOREGRESSIVE-INTEGRATED MOVING AVERAGE (ARIMA) MODELS

7.1

An important feature of the univariate models introduced in previous chapters is their ability to provide forecasts of future values of the observed series. There are two aspects to forecasting: the provision of a forecast for a future value of the series and the provision of a forecast error that can be

attached to this point forecast. This forecast error may then be used to con¬ struct forecast intervals to provide an indication of the precision these fore¬ casts are likely to possess. The setup is, thus, analogous to the classic statistical problem of estimating an unknown parameter of a model and pro¬ viding a confidence interval for that parameter.

What is often not realized when forecasting is that the type of model used to construct point and interval forecasts will necessarily determine the properties of these forecasts. Consequently, forecasting from an incorrect or misspecified model may lead to forecasts that are inaccurate and which incorrectly measure the precision that may be attached to them.1

7.2

To formalize the forecasting problem, suppose we have a realization x12d; x22d; ...; xT ð Þ from a general ARIMA (p,d,q) process

$$\phi(B)\nabla^d x_t = \theta_0 + \theta(B)a_t \tag{7.1}$$

$$f_{T,h} = E(\alpha_1 x_{T+h-1} + \alpha_2 x_{T+h-2} + \cdots + \alpha_{p+d} x_{T+h-p-d} + \theta_0 + a_{T+h} - \theta_1 a_{T+h-1} - \cdots - \theta_q a_{T+h-q} | x_T, x_{T-1}, \ldots). \tag{7.2}$$

$$E\left(x_{T+j} | x_T, x_{T-1}, \ldots\right) = \begin{cases} x_{T+j}, & j \leq 0 \\ f_{T,j}, & j > 0 \end{cases},$$

$$E\left(a_{T+j} | x_T, x_{T-1}, \ldots\right) = \begin{cases} a_{T+j}, & j \leq 0 \\ 0, & j > 0 \end{cases},$$

The h-step ahead forecast error for origin T, defined in y 7.2, may be expressed as

$$e_{T,h} = x_{T+h} - f_{T,h} = a_{T+h} + \psi_1 a_{T+h-1} + \cdots + \psi_{h-1} a_{T+1} \tag{7.3}$$

$$V\left(e_{T,h}\right) = \sigma^2\left(1 + \psi_1^2 + \psi_2^2 + \cdots + \psi_{h-1}^2\right) \tag{7.4}$$

FORECASTING A TREND STATIONARY PROCESS

7.11

Let us now consider the trend stationary (TS) process

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t \quad \phi(B)\varepsilon_t = \theta(B)a_t \tag{7.5}$$

8. Unobserved Component Models, Signal Extraction, and Filters

UNOBSERVED COMPONENT MODELS

8.1

A difference stationary, that is, I(1), time series may always be decomposed into a stochastic nonstationary trend, or signal, component and a stationary noise, or irregular, component:

$$x_t = z_t + u_t \tag{8.1}$$

$$\nabla x_t = \mu + v_t + u_t - u_{t-1} \tag{8.2}$$

$$\rho_1 = -\frac{\sigma_u^2}{\sigma_u^2 + 2\sigma_v^2} \tag{8.3}$$

$$\nabla x_t = \mu + e_t - \theta e_{t-1} \tag{8.4}$$

Models of the form of (8.1) are known as unobserved component (UC) models, a more general formulation for the components being:

$$\begin{aligned} \nabla z_t &= \mu + \gamma(B)v_t \\ u_t &= \lambda(B)a_t \end{aligned} \tag{8.5}$$

$$\nabla x_t = \mu + \theta(B)e_t \tag{8.6}$$

$$\sigma_e^2 \frac{\theta(B)\theta(B^{-1})}{(1-B)(1-B^{-1})} = \sigma_v^2 \frac{\gamma(B)\gamma(B^{-1})}{(1-B)(1-B^{-1})} + \sigma_a^2 \lambda(B)\lambda(B^{-1}) \tag{8.7}$$

The assumption that the trend component, zt, follows a random walk is not as restrictive as it may at first seem. Consider the Wold decomposition for rxt:

$$\nabla x_t = \mu + \psi(B)e_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j} \tag{8.8}$$

$$\nabla z_t = \mu + \left(\sum_{j=0}^{\infty} \psi_j\right) e_t = \mu + \psi(1)e_t \tag{8.9}$$

$$\nabla z_t = \mu + (1-\theta)e_t \tag{8.10}$$

$$u_t = \theta e_t \tag{8.11}$$

In a more general context, it is possible for a time series xt with Wold decomposition (8.8) to be written as (8.1) with zt being a random walk and ut being stationary and where the innovations of the two components are cor¬ related to an arbitrary degree. However, only the Beveridge¬ Nelson decom¬ position is guaranteed to exist.

SIGNAL EXTRACTION

8.8

Given a UC model of the form of (8.1) and models for zt and ut, it is often useful to provide estimates of these two unobserved components, a pro¬ cedure that is known as signal extraction.

As noted earlier, if we are given only a realization of

xt and its model, that is, (8.6), then component models for zt and ut are in general unidentified.

If xt follows the ARIMA(0,1,1) process

$$\nabla x_t = (1 - \theta B)e_t \qquad (8.16)$$

$$\nabla z_t = (1 - \Theta B)v_t \qquad (8.17)$$

9.Seasonality and Exponential Smoothing

SEASONAL PATTERNS IN TIME SERIES

9.1

In yy2.162.17 we introduced the idea of seasonal patterns appearing

in time series observed at frequencies greater than annual, typically monthly or quarterly. The presence of seasonality is often immediately apparent from a plot of the series (recall the quarterly United Kingdom beer sales series of Fig. 1.4), but it will also manifest itself in the sample autocorrelation func┐ tion (SACF) of the appropriately differenced data. Fig. 9.1 shows the SACF of the first differences of beer sales, which is dominated by a pronounced seasonal pattern, and a similar effect is seen in Fig. 9.2, which shows the SACF for the square root of monthly England and Wales rainfall (recall

2.3

┌ 2.6 and Fig. 2.4). Clearly a seasonal pattern is a predictable feature of these series and is, therefore, susceptible to either modeling explicitly or to being removed by a suitable seasonal adjustment procedure

MODELING DETERMINISTIC SEASONALITY

9.2

A simple model for seasonality was alluded to in yy2.62.7, which is

to use a “seasonal mean” model in which there is a different mean for each

season, that is, the model for xt is:

$$x_t = \sum_{i=1}^{m} \alpha_i s_{i,t} + \varepsilon_t \qquad (9.1)$$

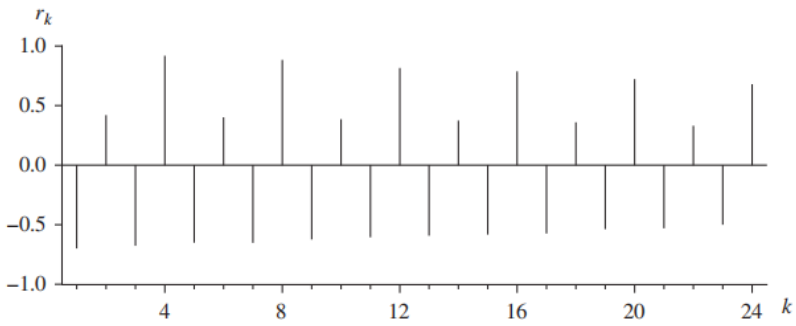

FIGURE 9.1 SACF of the first-difference of quarterly UK beer sales, 1997–2017.

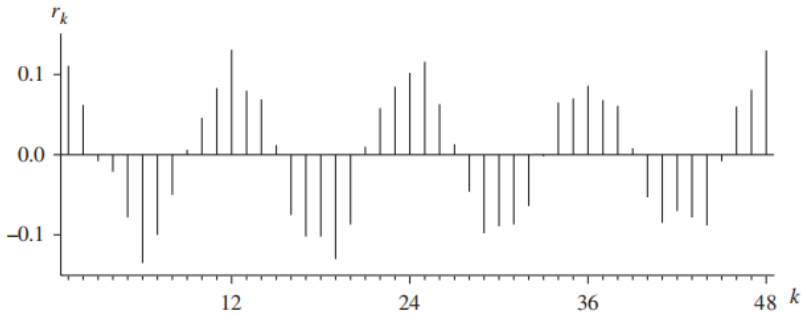

FIGURE 9.2 SACF of the square root of England and Wales monthly rainfall, 1875–2017.

MODELING STOCHASTIC SEASONALITY

9.3

It would, however, be imprudent to rule out the possibility of an evolv¬

ing seasonal pattern: in other words, the presence of stochastic seasonality.

As in the modeling of stochastic trends, ARIMA processes have been found

to do an excellent job in modeling stochastic seasonality, albeit in an

extended form to that developed in previous chapters.

9.4

An important consideration when attempting to model a seasonal time series with an ARIMA model is to determine what sort of process will best match the SACFs and PACFs that characterize the data. Concentrating on the beer sales series, we have already noted the seasonal pattern in the SACF for rxt shown in Fig. 9.1. In considering the SACF further, we note that the seasonality manifests itself in large positive autocorrelations at the seasonal lags

(4k; k $ 1) being flanked by negative autocorrelations at the "satellites"

The slow decline of these seasonal autocorrelations is indicative of seasonal nonstationarity and, analogous to the analysis of "nonseasonal nonstationarity," this may be dealt with by seasonal differencing, that is, by using the

r4 5 1 2 B4 operator in conjunction with the usual r operator. Fig. 9.3 shows the SACF of

rr4 transformed beer sales and this is now clearly stationary and, thus, potentially amenable to ARIMA identification.

9.5

In general, if we have a seasonal period of m then the seasonal differencing operator may be denoted as rm. The nonseasonal and seasonal differencing operators may then be applied d and D times, respectively, so that a seasonal ARIMA model may take the general form

$$\nabla^d \nabla_m^D \phi(B) x_t = \theta(B) a_t \tag{9.2}$$

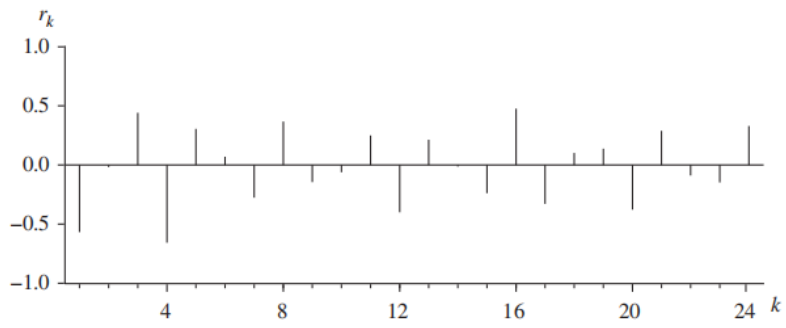

FIGURE 9.3 SACF of $\nabla\nabla^4$ transformed beer sales.

Appropriate forms of the θ polynomials can then, at least in

principle, be obtained by the usual methods of identification and/or model selection. Unfortunately, two difficulties are typically encountered. First, the PACFs of seasonal models are difficult both to derive and to interpret, so that conventional identification is usually based solely on the behavior of the appropriate SACF. Second, since the θ polynomials need to account for the seasonal autocorrelation, at least one of them must be of min┐imum order m. This often means that the number of models which need to be considered in model selection procedures can become prohibitively large.

The Q4 observations may then be linked by a model of the form:

$$\Phi(B^m)\nabla_m^D x_t = \Theta(B^m)\alpha_t \qquad (9.3)$$

On the assumption that the model is of the form of (9.6), the variances for the estimated sample autocorrelations at lags higher than m 1 1 are given by:

$$V(r_k) = T^{-1}\left(1 + 2\left(r_1^2 + r_{m-1}^2 + r_m^2 + r_{m+1}^2\right)\right) \quad k > m+1 \qquad (9.7)$$

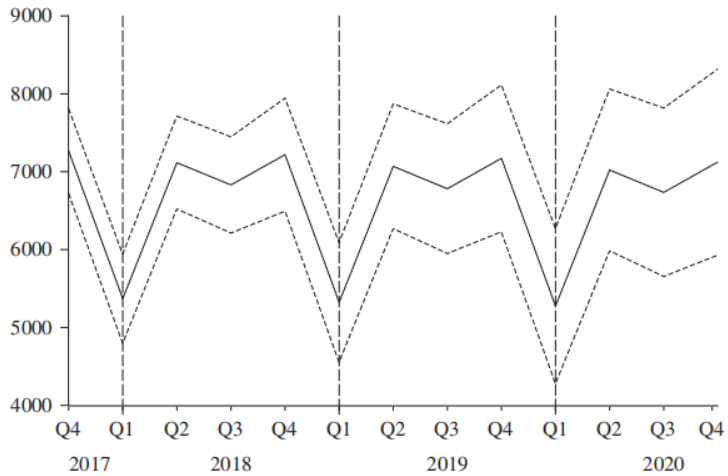

FIGURE 9.4 Airline model forecasts of beer sales out to 2020Q4 accompanied by two-standard error bounds.

The deterministic and stochastic seasonal models, (9.1) and (9.5), may be combined to form, on setting d 5 D 5 1 for both simplicity and because

these are the settings that are typically found,

$$x_t = \sum_{i=1}^{m} \alpha_i s_{i,t} + \frac{\theta_q(B)\Theta_Q(B^m)}{\phi_p(B)\Phi_P(B^m)\nabla\nabla_m} a_t \qquad (9.8)$$

SEASONAL ADJUSTMENT

9.13

In y2.16 we introduced a decomposition of an observed time series into trend, seasonal, and irregular (or noise) components, focusing attention on estimating the seasonal component and then eliminating it to provide a seasonally adjusted series. Extending the notation introduced in (8.1), this implicit UC decomposition can be written as

$$x_t = z_t + s_t + u_t \qquad (9.9)$$

EXPONENTIAL SMOOTHING

9.15

Returning to the two-component UC model, (8.1), where xt 5 zt 1 ut, then a simple model for the signal or "level" zt is to assume that its current value is an exponentially weighted moving average of current and past observations of xt:

$$\begin{aligned} z_t &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \cdots = \alpha \textstyle\sum_{j=0}^{\infty} (1-\alpha)^j x_{t-j} \\ &= \alpha\big(1 + (1-\alpha)B + (1-\alpha)^2 B^2 + \cdots + (1-\alpha)^j B^j + \cdots\big) x_t \end{aligned}$$

Simple exponential smoothing is, therefore, a suitable forecasting proce- dure for a series in which a trend is absent. To capture a linear trend, the approach may be generalized by extending (9.11) to include a trend component,

$$\begin{aligned} z_t &= \alpha x_t + (1-\alpha)(z_{t-1} + \tau_{t-1}) \\ &= z_{t-1} + \tau_{t-1} + \alpha e_t \end{aligned} \qquad (9.12)$$

where the error correction is now et 5 xt 2 zt21 2 τt21, and defining a second updating equation for the trend τt:

$$\begin{aligned}\tau_t &= \beta(z_t - z_{t-1}) + (1-\beta)\tau_{t-1} \\ &= \tau_{t-1} + \alpha\beta e_t\end{aligned} \qquad (9.13)$$

10.Volatility and Generalized AutoregressiveConditional Heteroskedastic Processes

VOLATILITY

10.1

Following initial research on portfolio theory during the 1950s, volatility became an extremely important concept in finance, appearing regularly in models of, for example, asset pricing and risk management. Although there are various definitions of volatility, in the context of a time series it is generally taken to be a period in the evolution of the series that is associated with high variability or, equivalently, high variance. This was prompted by the observation that many time series, not just financial returns, appear to be

characterized by alternating periods of relative tranquility in which variability is low and relative volatility where variability is considerably higher.

10.2

Much of the initial interest in volatility had to do with it not being directly observable, and several alternative measures were consequently developed to approximate it empirically.1 In the early 1980s, it was proposed that volatility should be embedded within a formal stochastic model for the observed time series. This was prompted by the fact that, although some series appeared to be serially uncorrelated, they were certainly not indepen-

dent through time. They, thus, had the potential to exhibit rich dynamics in their higher moments, these often being accompanied by interesting non¬ Gaussian distributional properties. Under such circumstances, attention should be focused on the characteristics of the higher moments of the series, rather than just on modeling the conditional mean.

10.3

A straightforward way of doing this is to allow the variance (or typi¬ cally, the conditional variance) of the process generating the series xt to change either continuously or at certain discrete points in time. Although a stationary process must have a constant variance, certain conditional var¬ iances can change, so that although the unconditional variance V xt ð Þ may be constant for all t, the conditional variance V x

ð tjxt21; xt22; …Þ, which depends on the realization of xt, is able to alter from observation to observation.

10.4

A stochastic model having time-varying conditional variances may be defined by supposing that xt is generated by the product process:

$$x_t = \mu + \sigma_t U_t \qquad (10.1)$$

AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC PROCESSES

10.5

Up until this point we have said nothing about how the conditional variances σ2

t might be generated. We now consider the case where they are a function of past values of xt:

$$\sigma_t^2 = f(x_{t-1}) = \alpha_0 + \alpha_1 (x_{t-1} - \mu)^2 \qquad (10.2)$$

a model that has proved extremely popular for describing financial time

series, these conditions simply require that all three parameters are nonnegative. The equivalent form of the GARCH(p,q) process is

$$\varepsilon_t^2 = \alpha_0 + (\alpha(B) + \beta(B))\varepsilon_{t-1}^2 + \nu_t - \beta(B)\nu_{t-1} \qquad (10.3)$$

TESTING FOR THE PRESENCE OF ARCH ERRORS

10.12

Let us suppose that an ARMA model for xt has been estimated, from which the residuals et have been obtained. The presence of ARCH may lead to serious model misspecification if it is ignored. As with all forms of heteroskedasticity (i.e., nonconstant error variance), analysis assuming its absence will result in inappropriate parameter standard errors, these typically being too small. For example, ignoring ARCH will lead to the identification of ARMA models that tend to be overparameterized, as parameters that should be set to zero will show up as significant.

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 U_{t-1}^2 + \beta_1\sigma_{t-1}^2 \qquad (10.4)$$

An early alternative was to model conditional standard deviations rather than variances (Schwert, 1989):

$$\sigma_t = \alpha_0 + \alpha_1|\varepsilon_{t-1}| + \beta_1\sigma_{t-1} = \alpha_0 + \alpha_1\sigma_{t-1}|U_{t-1}| + \beta_1\sigma_{t-1} \qquad (10.5)$$

An asymmetric response to shocks is made explicit in the exponential GARCH (EGARCH) model of Nelson (1991):

$$\log(\sigma_t^2) = \alpha_0 + \alpha_1 g\left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) + \beta_1 \log(\sigma_{t-1}^2) \qquad (10.6)$$

FORECASTING FROM AN ARMA-GARCH MODEL

10.20

Suppose we have the ARMA(P,Q)-GARCH(p,q) model of y10.11:

$$x_t = \Phi_1 x_{t-1} + \cdots + \Phi_P x_{t-P} + \Theta_0 + \varepsilon_t - \Theta_1 \varepsilon_{t-1} - \cdots - \Theta_Q \varepsilon_{t-Q} \qquad (10.7)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_p \varepsilon_{t-p}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_q \sigma_{t-q}^2 \qquad (10.8)$$

11.Nonlinear Stochastic Processes

MARTINGALES, RANDOM WALKS, AND NONLINEARITY

11.1

In y10.2 a distinction was drawn between serial uncorrelatedness

and independence. Although this distinction lies at the heart of GARCH

modeling, it is also of more general importance, manifesting itself in the con¬

cept of a martingale; a stochastic process that is a mathematical model of

“fair play.”1 A martingale may be defined as a stochastic process xt having

the following properties:2

$$E(x_t - x_s | x_s, x_{s-1}, \ldots) = 0, \quad s < t, \qquad (11.1)$$

NONLINEAR STOCHASTIC MODELS

11.7

As discussed in y3.6, Wold’s decomposition theorem allows us to rep¬

resent every weakly stationary, purely nondeterministic, stochastic process as

a linear combination of a sequence of uncorrelated random variables, as in

(3.2). A stochastic process can then be considered nonlinear if it does not

satisfy the assumptions underlying the decomposition, for example, if the

representation is:

$$x_t - \mu = f(a_t, a_{t-1}, a_{t-2}, \ldots) \qquad (11.2)$$

BILINEAR MODELS

11.9

An important class of nonlinear model is the bilinear, which takes the general form

$$\phi(B)(x_t - \mu) = \theta(B)\varepsilon_t + \sum_{i=1}^{R}\sum_{j=1}^{S}\gamma_{ij}x_{t-i}\varepsilon_{t-j} \qquad (11.3)$$

Charemza et al. (2005) discuss nonstationary generalizations of bilinear models that allow for unit roots. For example, they consider the following simple model:

$$x_t = (a + b\varepsilon_{t-1})x_{t-1} + \varepsilon_t \qquad (11.4)$$

$$\nabla x_t = bx_{t-1}\varepsilon_{t-1} + \varepsilon_t \qquad (11.5)$$

THRESHOLD AND SMOOTH TRANSITION AUTOREGRESSIONS

11.18

A popular class of nonlinear model is the self-exciting threshold autoregressive (SETAR) process, which allows for asymmetry by defining a set of piecewise autoregressive models whose switch points, or "thresholds," are generally unknown (see Tong and Lim, 1980; Tong, 1990; Tera¨svirta, 2006):

$$x_t = \sum_{j=1}^{r}\left(\phi_{j,1}x_{t-1} + \cdots + \phi_{j,p}x_{t-p} + a_{j,t}\right)\mathbf{1}\left(c_{j-1} < x_{t-d} \le c_j\right) \qquad (11.7)$$

The SETAR formulation requires that the shift from one regime to another is immediate. Allowing the shift to be smooth is accomplished by defining the exponential autoregressive (EAR) process:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + G(\gamma, x_{t-d})\left(\varphi_1 x_{t-1} + \cdots + \varphi_p x_{t-p}\right) + a_t \qquad (11.8)$$

$$G_E(\gamma, c, x_{t-d}) = \exp\left(-\gamma\left(x_{t-d} - c\right)^2\right), \quad \gamma > 0 \qquad (11.9)$$

$$G_L(\gamma, c, x_{t-d}) = (1+\exp(-\gamma(x_{t-d}-c)))^{-1}, \quad \gamma > 0 \qquad (11.10)$$

11.20

Since the SETAR is a piecewise linear model it can be estimated by a variant of OLS. Tera¨svirta (2006) provides details and discusses the weak stationarity and ergodicity conditions required for consistent estimation. Both ESTAR and LSTAR models can be estimated by nonlinear least squares (NLS) and ML techniques, although the properties of the ML estimator generally remain unknown. Numerical problems may arise, however, when the transition parameter γ is large, since then the transition is rapid and accurate estimation of this parameter requires many observations to lie in a small neighborhood of the location parameter c. Convergence of the optimization algorithm may be further exacerbated if γ is of a much higher order of magnitude than the other parameters. Tera ¨svirta (1994) suggests that, when the transition is known to be quick, γ can be fixed at an appropriately large value rather than being estimated imprecisely.

MARKOV-SWITCHING MODELS

11.21

Yet another way of introducing asymmetry is to consider "regime switching" models. Hamilton (1989, 1990), Engle and Hamilton (1990), and Lam (1990) all propose variants of a switching-regime Markov model, which can be regarded as a nonlinear extension of an ARMA process that can accommodate complicated dynamics, such as asymmetry and condi¬ tional heteroskedasticity. The setup is that of the UC model of y8.1, i.e., Eq. (8.1), where zt now evolves as a two-state Markov process:

$$z_t = \alpha_0 + \alpha_1 S_t \tag{11.11}$$

NEURAL NETWORKS

11.24

Neural networks (NNs) refer to a broad class of nonparametric models which have gained a good deal of popularity in recent years across a wide range of disciplines, including computer science, psychology, biology, linguistics, and pattern recognition (for a textbook treatment, see, for example, Haykin, 1999). These models originate from research in the cognitive sciences on emulating the structure and behavior of the human brain.

One of the most common types of NN is the multi-layered perceptron (MLP), which can be used for nonparametric regression and classification. These models are organized in three basic layers: the input layer of independent variables, the output layer of dependent variables, and one or more hidden layers in-between. An activation function regulates the dependencies between the elements of each layer. A univariate autoregressive MLP model with a single hidden layer can be represented as:

$$x_t = \sum_{i=1}^{p} \phi_i x_{t-i} + \sum_{j=1}^{q} \beta_j G\left(\sum_{i=1}^{p} \varphi_i x_{t-i}\right) + \varepsilon_t \tag{11.13}$$

NONLINEAR DYNAMICS AND CHAOS

11.28

So far, all the processes introduced in this chapter have the common aim of modeling stochastic nonlinearities in time series. This would seem the natural approach to take when dealing with stochastic time series processes, but a literature has also developed that considers the question of

whether such series could have been generated, at least in part, by nonlinear deterministic laws of motion.

11.29

Research in the general area of nonlinear dynamics is concerned with the behavior of deterministic and stochastic nonlinear systems. Both applied and theoretical research has flourished over the past four decades across a variety of disciplines and an extensive overview of the research on nonlinear dynamics, albeit with a bias toward the natural sciences, is given by Hilborn (1997). The meaning of the term "nonlinear dynamics" seems to vary considerably across scientific disciplines and eras. For example, a popular interpretation, since the early 1980s, associates nonlinear dynamics with deterministic nonlinear systems and a specific dynamic behavior called chaos, although this term has itself been given several different interpretations.

TESTING FOR NONLINEARITY

11.36

As the previous sections have demonstrated, there have been a wide variety of nonlinear processes proposed for modeling time series. We have, for example, compared ARCH and bilinear models, and in so doing have discussed LM tests for each. Nevertheless, given the range of alternative nonlinear models, it is not surprising that other tests for nonlinearity have also been proposed. Since the form of the departure from linearity is often difficult to specify a priori, many tests are diagnostic in nature, i.e., a clear alternative to the null hypothesis of linearity is not specified. This, of course, leads to difficulties in discriminating between the possible causes of

any “nonlinear misspecification” that might be uncovered by such tests

FORECASTING WITH NONLINEAR MODELS

11.51

One-step ahead forecasting from nonlinear models is straightforward,

but multistep forecasting may be complicated. If we have the simple nonlin¬

ear model, xt 5 g x

FORECASTING WITH NONLINEAR MODELS

11.51

One-step ahead forecasting from nonlinear models is straightforward,

but multistep forecasting may be complicated. If we have the simple nonlin¬

ear model, xt 5 g x

FORECASTING WITH NONLINEAR MODELS

11.51

One-step ahead forecasting from nonlinear models is straightforward,

but multistep forecasting may be complicated. If we have the simple nonlin¬

ear model, xt 5 g x

FORECASTING WITH NONLINEAR MODELS

11.51

One-step ahead forecasting from nonlinear models is straightforward,

but multistep forecasting may be complicated. If we have the simple nonlin¬

ear model, xt 5 g x then the one-step ahead forecast is straightfor¬ wardly given by fT

. However, the two-step ahead forecast is then the one-step ahead forecast is straightfor¬

wardly given by fT. However, the two-step ahead forecast is then the one-step ahead forecast is straightfor¬

wardly given by fT. However, the two-step ahead forecast is

$$f_{T,2} = E\left(x_{T+2}\middle|x_T\right) = E(g(g(x_{T+1}) + \varepsilon_{T+1})) = \int_{\varepsilon} g(g(x_{T+1}) + \varepsilon_{T+1})dF(\varepsilon) \tag{11.15}$$

12.Transfer Functions and Autoregressive DistributedLag Modeling

TRANSFER FUNCTION-NOISE MODELS

12.1

The models that have been developed so far in this book have all been univariate, so that the current value of a time series depends, linearly or otherwise, only on past values of itself and, perhaps, a deterministic function of time. While univariate models are important in themselves, they also play a key role in providing a "baseline" to which multivariate models may be compared. We shall analyze several multivariate models over the next chapters, but our development begins with the simplest. This is the single-input transfer function-noise model, in which an endogenous, or output, variable yt is related to a single input, or exogenous, variable xt through the dynamic model

$$y_t = \upsilon(B)x_t + n_t \tag{12.1}$$

$$\upsilon(B) = \frac{\omega(B)B^b}{\delta(B)} \tag{12.2}$$

$$y_t = \frac{\omega(B)}{\delta(B)}x_{t-b} + \frac{\theta(B)}{\phi(B)}a_t \tag{12.3}$$

If identifying a univariate ARMA model is often considered to be an "art form," then identifying a transfer function in this way is even more so and, if there are multiple inputs, can become increasingly difficult, for the model is now:

$$y_t = \sum_{j=1}^{M} \upsilon_j(B)x_{j,t} + n_t = \sum_{j=1}^{M} \frac{\omega_j(B)B^{b_j}}{\delta_j(B)}x_{j,t} + \frac{\theta(B)}{\phi(B)}a_t \tag{12.4}$$

AUTOREGRESSIVE DISTRIBUTED LAG MODELS

12.8

Nevertheless, it would clearly be useful if an automatic model selec¬ tion procedure could be developed. This has not been done for the multiple input model (12.4), but if a restricted form of it is specified then such a pro¬ cedure becomes feasible. This restricted form is known as the autoregressive distributed lag, or ARDL, model and is obtained by placing the following restrictions on (12.4):

$$\phi(B)y_t = \beta_0 + \sum_{j=1}^{M} \beta_j(B)x_{j,t} + a_t \qquad (12.5)$$

This is known as the ARDL p;s1; ... ;sM

ð Þ

model and restricts all the

autoregressive lag polynomials to be the same and excludes a moving average noise component, although this exclusion is not essential. These restrictions reduce the noise component to white noise through constraining the dynamics and enables (12.5) to be estimated by OLS, so that on selecting a maximum lag order of, say, m, goodness-of-fit statistics, such as informa¬ tion criteria, can be used to select the appropriate specification.

13.Vector Autoregressions and Granger Causality

MULTIVARIATE DYNAMIC REGRESSION MODELS

13.1

In a natural extension to the ARDL model of the previous chapter,

suppose that there are now two endogenous variables, y1;t and y2;t, that may

both be related to an exogenous variable xt and its lags as well as to lags of each other. In the simplest case, such a model would be:

$$\begin{aligned} y_{1,t} &= c_1 + a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + b_{10}x_t + b_{11}x_{t-1} + u_{1,t} \\ y_{2,t} &= c_2 + a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + b_{20}x_t + b_{21}x_{t-1} + u_{2,t} \end{aligned} \tag{13.1}$$

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \sum_{i=0}^{q} \mathbf{B}_i \mathbf{x}_{t-i} + \mathbf{u}_t \tag{13.2}$$

VECTOR AUTOREGRESSIONS

13.4

Suppose the model (13.2) does not contain any exogenous variables, so that all the Bi matrices are zero, and that there are p lags of the endogenous variables in every equation:

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^{p} \mathbf{A}_i \mathbf{y}_{t-i} + \mathbf{u}_t \tag{13.3}$$

GRANGER CAUSALITY

13.5

In the VAR (13.3) the presence of nonzero off-diagonal elements in the Ai matrices, ars;i s, implies that there are dynamic relationships between the variables, otherwise the model would collapse to a set of n univariate AR processes. The presence of such dynamic relationships is known as Granger (-Sims) causality.4 The variable ys does not Granger-cause the variable yr if ars;i 5 0 for all i 5 1; 2; ... ; p. If, on the other hand, there is at least one ars;i 6¼ 0 then ys is said to Granger-cause yr because if that is the case then past values of ys are useful in forecasting the current value of yr: Granger causality is, thus, a criterion of "forecastability." If yr also Granger-causes ys, the pair of variables are said to exhibit feedback.

DETERMINING THE LAG ORDER OF A VECTOR

AUTOREGRESSION

13.8

To enable the VAR to become operational the lag order p, which will typically be unknown, needs to be determined empirically.

$$LR(p,m) = (T - np)\log\left(\frac{\left|\hat{\Omega}_m\right|}{\left|\hat{\Omega}_p\right|}\right) \sim \chi^2_{n^2(p-m)} \qquad (13.4)$$

VARIANCE DECOMPOSITIONS AND INNOVATION ACCOUNTING

13.10

While the estimated coefficients of a VAR(1) are relatively easy to interpret, this quickly becomes problematic for higher order VARs because not only do the number of coefficients increase rapidly (each additional lag introduces a further n2 coefficients), but many of these coefficients will be imprecisely estimated and highly intercorrelated, so becoming statistically insignificant. This can be seen in the estimated VAR(2) of Example 13.1,

$$\mathbf{y}_t = \mathbf{A}^{-1}(B)\mathbf{u}_t = \mathbf{\Psi}(B)\mathbf{u}_t = \mathbf{u}_t + \sum_{i=1}^{\infty} \mathbf{\Psi}_i \mathbf{u}_{t-i} \qquad (13.5)$$

STRUCTURAL VECTOR AUTOREGRESSIONS

13.15

The “noninvariance property” of VARs has generated much detailed analysis and criticism of the variance decomposition methodology, mainly focusing on the inability of VARs to be regarded as “structural” in the tradi- tional econometric sense, so that shocks cannot be uniquely identified with a specific variable unless prior identifying assumptions are made, without

which the computed impulse response functions and variance decompositions

would be invalid. The triangular "recursive" structure of S has been criti┐

cized for being atheoretical, and has led to the development of other sets of

identifying restrictions that are based more explicitly on theoretical consid┐

erations using the structural VAR (SVAR) approach: see Cooley and LeRoy

(1985); Blanchard (1989); and Blanchard and Quah (1989).

13.16

The Cholesky decomposition of y13.12 can be written as ut 5 Svt

with SS0 5 Ωp and E 5 In. A more general formulation is:

$$\mathbf{BB}' = \mathbf{A}\boldsymbol{\Omega}_p\mathbf{A}' \qquad (13.8)$$

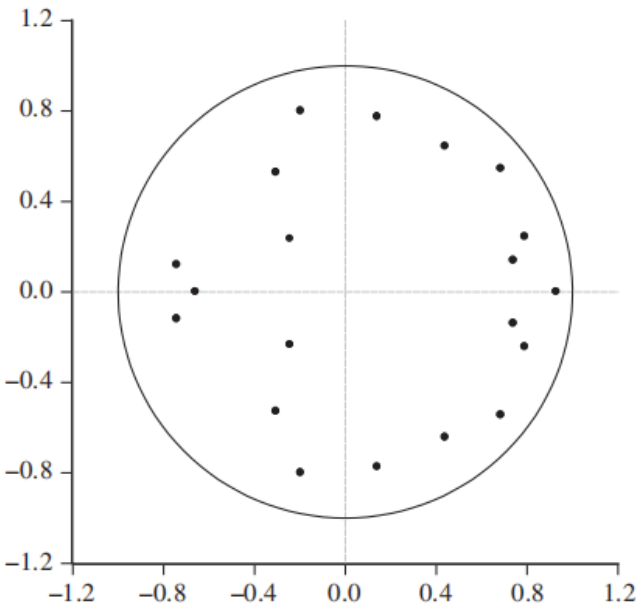

FIGURE 13.6 Roots of the characteristic equation associated with the VAR(4) fit.

14.Error Correction, Spurious Regressions, and Cointegration

THE ERROR CORRECTION FORM OF AN AUTOREGRESSIVE

DISTRIBUTED LAG MODEL

14.1

The simplest case of the ARDL (autoregressive distributed lag) model

introduced in y12.7 is the ARDL 115.Vector Autoregressions With Integrated Variables,Vector Error Correction Models, and Common Trends

$$y_t = \beta_0 + \phi y_{t-1} + \beta_{1,0} x_t + \beta_{1,1} x_{t-1} + a_t \qquad (14.1)$$

SPURIOUS REGRESSIONS

14.2

As in Chapter 12, Transfer Functions and Autoregressive Distributed

Lag Modeling, it has been implicitly assumed that all the variables entering

the ARDL/ECM are stationary, so that any nonstationary series have been

appropriately differenced beforehand. What would happen if nonstationary

variables were not prior differenced but were entered as levels? That there is

a theoretical issue is clearly apparent. The standard proof of the consistency

of OLS when there are stochastic regressors, as there are here, relies on the

assumption that the probability limit of T21X0X, where X is a matrix con¬

taining the data on the explanatory variables, tends to a fixed matrix; i.e., the

matrix of expectations of sums of squares and cross-products of the data

tends to a matrix of constants (see, for example, Mills, 2013a, yy6.3└6.4). In

other words, as the sample size T increases, the sample moments of the data

settle down to their population values. For there to be fixed population

moments to which these sample moments converge, the data must be station¬

ary. If it was not, then, as in the case of integrated series, the data may dis¬

play a tendency to increase in magnitude over time, so that there are no

fixed values in the matrix of expectations of sums of squares and cross¬ products of these data.

14.4

Granger and Newbold considered the following data generation pro¬ cess (DGP):

$$y_t = \phi y_{t-1} + u_t \quad u_t \sim \text{i.i.d.}\left(0, \sigma_u^2\right) \tag{14.5a}$$

$$x_t = \phi^* x_{t-1} + v_t \quad v_t \sim \text{i.i.d.}\left(0, \sigma_v^2\right) \tag{14.5b}$$

ERROR CORRECTION AND COINTEGRATION

14.16

An equivalent way of expressing the spurious nature of (14.6) is to note that the error, εt 5 yt 2 β0 2 β1xt, may, under the DGP (14.5), be regarded as a linear combination of I(1) processes and should therefore be I(1) as well, thus invalidating all least squares regression theory. While this would appear, on the face of it, to be an eminently sensible argument

More generally, if ytBI dð Þ and xtBI dð Þ then the linear combination

$$e_t = y_t - ax_t \tag{14.8}$$

TESTING FOR COINTEGRATION

14.22

Given the crucial role that cointegration plays in regression models with integrated variables, it is clearly important to test for its presence. Tests may be based on the residuals from the cointegrating regression, i.e.,

$$\hat{e}_t = y_t - \hat{\beta}_0 - \hat{\beta}_1 x_{1,t} - \cdots - \hat{\beta}_M x_{M,t} \tag{14.9}$$

$$\nabla y_t = \alpha_0 + \alpha_1 t - \phi(1) y_{t-1} + \sum_{j=1}^{M} \beta_j(1) x_{j,t-1} + \phi^*(B) \nabla y_{t-1} + \sum_{j=0}^{M} \gamma_j(B) \nabla x_{j,t} + a_t \tag{14.10}$$

ESTIMATING COINTEGRATING REGRESSIONS

14.25

Having found that yt cointegrates with x1;t; ... ; xM;t, the parameters

in the cointegrating regression

$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_M x_{M,t} + e_t \tag{14.11}$$

$$y_t = \beta_0 + \sum_{j=1}^{M} \beta_{j,t} x_{j,t} + \sum_{i=1}^{p} \gamma_i \nabla y_{t-i} + \sum_{j=1}^{M} \sum_{i=-p_1}^{p_2} \delta_{j,i} \nabla x_{j,t-i} + e_t \tag{14.12}$$

15.Vector Autoregressions WithIntegrated Variables, Vector Error Correction Models, and Common Trends

VECTOR AUTOREGRESSIONS WITH INTEGRATED VARIABLES

15.1

Having analyzed the impact of I 1ð Þ variables, and hence the possibil┐

ity of cointegration, on single-equation autoregressive distributed lag models

in Chapter 14, Error Correction, Spurious Regressions, and Cointegration,

the implications of allowing vector autoregressions to contain I 1ð Þ

variables clearly require discussing. Consider, then, the n-variable VAR

pð Þ of y13.11,

$$\boldsymbol{A}(B)\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{u}_t \tag{15.1}$$

$$\nabla \boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}(\mathbf{B}) \nabla \boldsymbol{y}_{t-1} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \boldsymbol{u}_t$$

$$\nabla \boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}(B) \nabla \boldsymbol{y}_{t-1} + \boldsymbol{u}_t \tag{15.3}$$

VECTOR AUTOREGRESSIONS WITH COINTEGRATED

VARIABLES

15.4

The condition A 5 In implies that

$$|\boldsymbol{\Pi}| = \left|\boldsymbol{A}_1 + \cdots + \boldsymbol{A}_p - \boldsymbol{I}_n\right| = 0 \qquad (15.4)$$

$$\nabla \boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}(B)\nabla \boldsymbol{y}_{t-1} + \boldsymbol{\beta}\boldsymbol{\alpha}'\boldsymbol{y}_{t-1} + \boldsymbol{u}_t \qquad (15.5)$$

Consequently, if yt is cointegrated with cointegrating rank r, then it

can be represented as the vector error correction model (VECM)

$$\nabla \boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}(B)\nabla \boldsymbol{y}_{t-1} + \boldsymbol{\beta}\boldsymbol{e}_{t-1} + \boldsymbol{u}_t \qquad (15.6)$$

ESTIMATION OF VECTOR ERROR CORRECTION MODELS AND

TESTS OF COINTEGRATING RANK

15.8

Estimation of the VECM (15.5) is nonstandard because the α and β

matrices enter in nonlinear fashion as the product βα0 . Without going into

unnecessary technical details, ML estimates are obtained in the following

way. Consider again (15.5) but now written as:

$$\nabla \boldsymbol{y}_t = \boldsymbol{c} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i \nabla \boldsymbol{y}_{t-i} + \boldsymbol{\beta}\boldsymbol{\alpha}'\boldsymbol{y}_{t-1} + \boldsymbol{u}_t \qquad (15.7)$$

More precisely, this procedure maximizes the likelihood of (15.7) by

treating it as a generalized eigenvalue problem and solving a set of equations

of the form:

$$\left(\lambda_i \boldsymbol{S}_{11} - \boldsymbol{S}_{10}\boldsymbol{S}_{00}^{-1}\boldsymbol{S}_{01}\right)\boldsymbol{v}_i = 0 \quad i = 1, \ldots, n \qquad (15.8)$$

This procedure can be straightforwardly adapted when a trend is

included in (15.7) and when various restrictions are placed upon the intercept

and trend coefficients. This involves adjusting the first- and second-step

regressions to accommodate these alterations. Consider again the levels VAR

(15.1) with a linear trend included:

$$\boldsymbol{A}(B)\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{d}t + \boldsymbol{u}_t \qquad (15.9)$$

The error correction in Example 15.2 has been normalized by setting

α2 5 1. With one cointegrating vector (r 5 1), imposing one restriction is

sufficient to identify the cointegrating vector.

If, for the moment, the identifying restrictions are imposed only on the α

matrix, if they are linear, and if there are no cross-cointegrating vector

restrictions, then these restrictions can be written for the ith cointegrating

vector as Riαi 5 ai, where Ri and ai are an r 3 n matrix and an r 3 1 vector,

respectively. A necessary and sufficient condition for α to be uniquely iden¬

tified is that the rank of each Riαi is r, while the necessary condition is that

there must be r restrictions placed on each of the r cointegrating vectors.1

Note that the identification of α, and hence Π, is achieved solely through

restrictions on α itself. Long-run relationships cannot be identified through

restrictions on the short-run dynamics: consequently, the Φi coefficients in

(15.6) may be estimated freely.

Following Johansen and Juselius (1994), a “structural VECM” may

be written as

$$\mathbf{\Gamma}_0 \nabla \mathbf{y}_t = \sum_{i=1}^{p-1} \mathbf{\Gamma}_i \nabla \mathbf{y}_{t-i} + \mathbf{\Theta} \boldsymbol{\alpha}' \mathbf{y}_{t-1} + \boldsymbol{\upsilon}_t \qquad (15.11)$$

CAUSALITY TESTING IN VECTOR ERROR CORRECTION

MODELS

15.18

Consider a “fully partitioned” form of the marginal VECM (15.10a,b):

IMPULSE RESPONSE ASYMPTOTICS IN NONSTATIONARY

the various impulse responses of the VAR are computed from the sequence of matrices

VECTOR ERROR CORRECTION MODEL-X MODELS

15.22

A straightforward extension of the CVAR/VECM model is to include

a vector of I exogenous variables, wt say, which may enter each equation:

$$\nabla \mathbf{y}_t = \boldsymbol{c} + \boldsymbol{d}t + \sum_{i=1}^{p-1} \mathbf{\Phi}_i \nabla \mathbf{y}_{t-i} + \beta \boldsymbol{\alpha}' \mathbf{y}_{t-1} + \mathbf{\Lambda} \boldsymbol{w}_t + \boldsymbol{u}_t \qquad (15.12)$$

$$\boldsymbol{C}(B)\mathbf{\Pi}(B) = \nabla \boldsymbol{I}_n \qquad (15.13)$$

$$\begin{aligned} \mathbf{y}_t &= \mathbf{y}_0 + \boldsymbol{b}_0 t + \boldsymbol{b}_1 \tfrac{t(t+1)}{2} + \boldsymbol{C}(B) \sum_{s=1}^{t} \boldsymbol{u}_t \\ &= \mathbf{y}_0 + \boldsymbol{b}_0 t + \boldsymbol{b}_1 \tfrac{t(t+1)}{2} + \left(\boldsymbol{C}(1) + \boldsymbol{C}^* \nabla\right) \sum_{s=1}^{t} \boldsymbol{u}_t \\ &= \mathbf{y}_0 + \boldsymbol{b}_0 t + \boldsymbol{b}_1 \tfrac{t(t+1)}{2} + \boldsymbol{C}(1)\boldsymbol{s}_t + \boldsymbol{C}^*(B)(\boldsymbol{u}_t - \boldsymbol{u}_0) \\ &= \mathbf{y}_0^* + \boldsymbol{b}_0 t + \boldsymbol{b}_1 \tfrac{t(t+1)}{2} + \boldsymbol{C}(1)\boldsymbol{s}_t + \boldsymbol{C}^*(B)\boldsymbol{u}_t \end{aligned} \qquad (15.14)$$

16.Compositional and Count Time Series

CONSTRAINED TIME SERIES

16.1

In previous chapters we considered time series that generally have no restrictions placed upon them apart from when they have a natural lower bound, this often being zero. There are, however, some series, or groups of series, that are bound by further constraints. When modeling such series, a "good" model should be unable to predict values which violate the known constraints, that is, the model should be "forecast coherent." Two examples of these types of series are considered in this chapter: (1) compositional time series in which a group of series are defined as shares of a whole, so that they must be positive fractions that sum to unity; and (2) "count" time series that can only take on positive, and typically low, integer values.

MODELING COMPOSITIONAL DATA

16.2

A compositional data set is one in which the T observations on

D 5 d 1 1 variables, written in matrix form as

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,D} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,D} \\ \vdots & \vdots & & \vdots \\ x_{T,1} & x_{T,2} & \cdots & x_{T,D} \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_D \end{bmatrix} \qquad (16.1)$$

There are several difficulties encountered when analyzing X within the

simplex sample space, these being a consequence of the summation condition

(16.2) rendering standard covariance and correlation analysis invalid.

Aitchison (1982) proposed mapping X from Sd to the d-dimensional real space

within Rd and then examining the statistical properties of the transformed data

Rd. Several transformations have been proposed for doing this, the most popular being the

additive log-ratio transformation which is defined as

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_d \end{bmatrix} = a_d\left(\boldsymbol{X}^{(d)}\right) = \left[\log\left(\frac{\boldsymbol{x}_1}{\boldsymbol{x}_D}\right) \quad \log\left(\frac{\boldsymbol{x}_2}{\boldsymbol{x}_D}\right) \quad \cdots \quad \log\left(\frac{\boldsymbol{x}_d}{\boldsymbol{x}_D}\right) \right] \tag{16.3}$$

FORECASTING COMPOSITIONAL TIME SERIES

16.7

Let us now denote the tth rows of X and Y as Xt and Yt; respectively,

and let us assume that an h-step ahead forecast of Yt1h, which may not yet be observed, is available. This may be denoted Yt with covariance matrix

. Since Yt is multivariate normal, such forecasts may have been

obtained from a wide variety of multivariate models; for example, Brunsdon and Smith (1998) consider modeling Yt as a vector ARMA process and other regression frameworks are available in which covariates and trends may be introduced (see Mills, 2010).

$$\boldsymbol{X}_t(h) = a_d^{-1}(\boldsymbol{Y}_t(h)) \tag{16.4}$$

$$\left(\boldsymbol{Y}_t(h) - \log\left(\frac{\boldsymbol{X}_{t+h}^{(d)}}{\boldsymbol{X}_{D,t+h}}\right)\right)' \boldsymbol{\Sigma}_t^{-1}(h) \left(\boldsymbol{Y}_t(h) - \log\left(\frac{\boldsymbol{X}_{t+h}^{(d)}}{\boldsymbol{X}_{D,t+h}}\right)\right) \leq \chi_\alpha^2(d) \tag{16.5}$$

TIME SERIES MODELS FOR COUNTS: THE IN-AR(1)

BENCHMARK MODEL

16.8

Time series of small numbers of counts arise in various fields and typ¬

ically consist of integer values, usually including zero, with a sample mean perhaps no higher than 10, making it inappropriate to treat the data as if it were continuous. We will focus here on the so-called integer-valued ARMA (IN-ARMA) models that provide an interesting class of discrete valued pro¬

cesses that are able to specify not only the dependence structure of the series

of counts, but also enable a choice to be made between a wide class of (dis-

crete) marginal distributions.

The "benchmark" IN-AR(1) process is defined by the difference

equation:

$$x_t = a \circ x_{t-1} + w_t \tag{16.6}$$

OTHER INTEGER-VALUED ARMA PROCESSES

16.12

The IN-MA(1) process is defined as

$$x_t = w_t + b \circ w_{t-1}$$

where $0 \leq b < 1$ and the thinning operation is defined analogously to (16.7) as

$$b \circ w_{t-1} = \sum_{i=1}^{w_{t-1}} y_{i,t-1}$$

The autocorrelation function (ACF) of x_t is now

$$\rho_k = \begin{cases} \dfrac{b\sigma_w^2}{b(1-b)\mu_w + (1+b^2)\sigma_w^2} & \text{for} \quad k = 1 \\ 0 & \text{for} \quad k > 1 \end{cases} \tag{16.8}$$

17.State Space Models

FORMULATING STATE SPACE MODELS

17.1

Many time series models can be cast in state space form (SSF), and

this enables a unified framework of analysis to be presented within which,

for example, the differences and similarities of the alternative models may

be assessed.

The state space model for a univariate time series xt consists of both a measurement equation (alternatively known as the signal or observation equation) and a transition equation (alternatively state equation: see, e.g., Harvey, 1989, Chapters 3 and 4; Hamilton, 1994, Chapter 13; or Durbin and Koopman, 2012, for full textbook treatments). Although there are various specifications of the SSF, a popular version has the measurement equation taking the form:

$$x_t = z_t'\alpha_t + d_t + \varepsilon_t \quad t = 1, 2, \ldots, T \tag{17.1}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{T}_t\boldsymbol{\alpha}_{t-1} + \boldsymbol{c}_t + \boldsymbol{R}_t\boldsymbol{\eta}_t \tag{17.2}$$

THE KALMAN FILTER

17.8

Once a model has been put into state space form, several important algorithms may be applied. Central to these is the Kalman (└ Bucy) filter. The Kalman filter is a recursive procedure for computing the optimal estimate of the state vector at time t, based on the information available at that time, which consists of all the observations up to and including xt.1 The system matrices, together with the initial values a0 and P0, are assumed to be known for all t and so do not need to be included explicitly in the information set.

$$\boldsymbol{a}_{t|t-1} = \boldsymbol{T}_t\boldsymbol{a}_{t-1} + \boldsymbol{c}_t \tag{17.4}$$

$$\boldsymbol{P}_{t|t-1} = \boldsymbol{T}_t\boldsymbol{P}_{t-1}\boldsymbol{T}_t' + \boldsymbol{R}_t\boldsymbol{Q}_t\boldsymbol{R}_t' \tag{17.5}$$

ML ESTIMATION AND THE PREDICTION ERROR

DECOMPOSITION

17.12

The system matrices may depend on a set of unknown parameters,

as with the ARMA process whose SSF was given in y17.3. The parameters

may be denoted by an n 3 1 vector ψ and will be referred to as the

hyperparameters of the SSF. These hyperparameters may be estimated by

ML, the classical theory of which is based on the T observations x1; … ; xT

being i.i.d. This allows the joint density function of the observations to be

written as:

$$\mathcal{L}(\boldsymbol{x}:\psi)=\prod_{t=1}^{T}p(x_t) \tag{17.10}$$

$$\mathcal{L}(\boldsymbol{x}:\psi)=\prod_{t=1}^{T}p(x_t|\boldsymbol{x}_{t-1}) \tag{17.11}$$

$$\log\mathcal{L}=\ell=-\frac{T}{2}\log 2\pi-\frac{1}{2}\sum_{t=1}^{T}f_t-\frac{1}{2}\sum_{t=1}^{T}\nu_t^2/f_t \tag{17.12}$$

PREDICTION AND SMOOTHING

17.15

Applying the Kalman filter to (17.1) and (17.2) yields aT , the

MMSE of αT , based on all T observations. In addition, it gives:

$$\boldsymbol{a}_{T+1|T}=\boldsymbol{T}_{T+1}\boldsymbol{a}_T+\boldsymbol{c}_{T+1} \tag{17.13}$$