CHAPTER 5

MODELS FOR NONSTATIONARY TIME SERIES

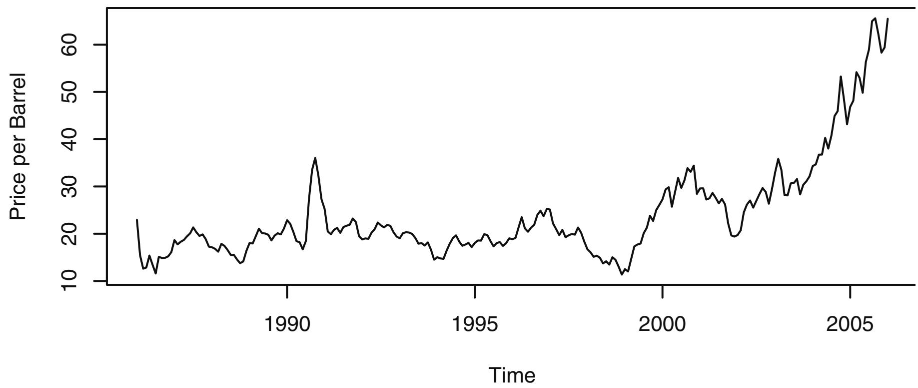
Any time series without a constant mean over time is nonstationary. Models of the form

$$Y_t = \mu_t + X_t$$

where μ_t is a nonconstant mean function and X_t is a zero-mean, stationary series, were considered in Chapter 3. As stated there, such models are reasonable only if there are good reasons for believing that the deterministic trend is appropriate "forever." That is, just because a segment of the series looks like it is increasing (or decreasing) approximately linearly, do we believe that the linearity is intrinsic to the process and will persist in the future? Frequently in applications, particularly in business and economics, we cannot legitimately assume a deterministic trend. Recall the random walk displayed in Exhibit 2.1, on page 14. The time series appears to have a strong upward trend that might be linear in time. However, also recall that the random walk process has a constant, zero mean and contains no deterministic trend at all.

As an example consider the monthly price of a barrel of crude oil from January 1986 through January 2006. Exhibit 5.1 displays the time series plot. The series displays considerable variation, especially since 2001, and a stationary model does not seem to be reasonable. We will discover in Chapters 6, 7, and 8 that no deterministic trend model works well for this series but one of the nonstationary models that have been described as containing stochastic trends does seem reasonable. This chapter discusses such models. Fortunately, as we shall see, many stochastic trends can be modeled with relatively few parameters.

Exhibit 5.1 Monthly Price of Oil: January 1986−January 2006




```
> win.graph(width=4.875,height=3,pointsize=8)
> data(oil.price)
> plot(oil.price, ylab='Price per Barrel',type='l')
```

5.1 Stationarity Through Differencing

Consider again the AR(1) model

$$Y_t = \phi Y_{t-1} + e_t \tag{5.1.1}$$

We have seen that assuming e_t is a true "innovation" (that is, e_t is uncorrelated with $Y_{t-1}, Y_{t-2},\ldots$), we must have $|\phi| < 1$. What can we say about solutions to Equation (5.1.1) if $|\phi| \geq 1$? Consider in particular the equation

$$Y_t = 3Y_{t-1} + e_t \tag{5.1.2}$$

Iterating into the past as we have done before yields

$$Y_t = e_t + 3e_{t-1} + 3^2 e_{t-2} + \cdots + 3^{t-1}e_1 + 3^t Y_0 \tag{5.1.3}$$

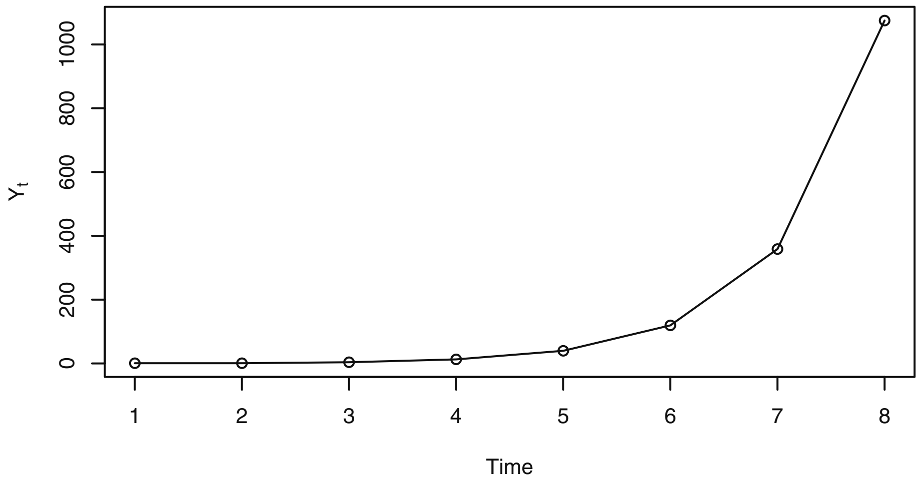
We see that the influence of distant past values of Y_t and e_t does not die out—indeed, the weights applied to Y_0 and e_1 grow exponentially large. In Exhibit 5.2, we show the values for a very short simulation of such a series. Here the white noise sequence was generated as standard normal variables and we used $Y_0 = 0$ as an initial condition.

Exhibit 5.2 Simulation of the Explosive "AR(1) Model" $Y_t = 3Y_{t-1} + e_t$

t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
e_t	0.63	−1.25	1.80	1.51	1.56	0.62	0.64	−0.98
Y_t	0.63	0.64	3.72	12.67	39.57	119.33	358.63	1074.91

Exhibit 5.3 shows the time series plot of this explosive AR(1) simulation.

Exhibit 5.3 An Explosive "AR(1)" Series




```
> data(explode.s)
> plot(explode.s,ylab=expression(Y[t]),type='o')
```

The explosive behavior of such a model is also reflected in the model's variance and covariance functions. These are easily found to be

$$Var(Y_t) = \frac{1}{8}(9^t - 1)\sigma_e^2 \tag{5.1.4}$$

and

$$Cov(Y_t, Y_{t-k}) = \frac{3^k}{8}(9^{t-k} - 1)\sigma_e^2 \tag{5.1.5}$$

respectively. Notice that we have

$$Corr(Y_t, Y_{t-k}) = 3^k \sqrt{\frac{9^{t-k} - 1}{9^t - 1}} \approx 1 \qquad \text{for large } t \text{ and moderate } k$$

The same general exponential growth or explosive behavior will occur for any ϕ such that $|\phi| > 1$. A more reasonable type of nonstationarity obtains when $\phi = 1$. If $\phi = 1$, the AR(1) model equation is

$$Y_t = Y_t + e_t \tag{5.1.6}$$

This is the relationship satisfied by the random walk process of Chapter 2 (Equation (2.2.9) on page 12). Alternatively, we can rewrite this as

$$\nabla Y_t = e_t \tag{5.1.7}$$

where $\nabla Y_t = Y_t - Y_{t-1}$ is the **first difference** of Y_t. The random walk then is easily extended to a more general model whose first difference is some stationary process—not just white noise.

Several somewhat different sets of assumptions can lead to models whose first difference is a stationary process. Suppose

$$Y_t = M_t + X_t \tag{5.1.8}$$

where M_t is a series that is changing only slowly over time. Here M_t could be either deterministic or stochastic. If we assume that M_t is approximately constant over every two consecutive time points, we might estimate (predict) M_t at t by choosing β_0 so that

$$\sum_{j=0}^{1} (Y_{t-j} - \beta_{0,t})^2$$

is minimized. This clearly leads to

$$\hat{M}_t = \frac{1}{2}(Y_t + Y_{t-1})$$

and the "detrended" series at time t is then

$$Y_t - \hat{M}_t = Y_t - \frac{1}{2}(Y_t + Y_{t-1}) = \frac{1}{2}(Y_t - Y_{t-1}) = \frac{1}{2}\nabla Y_t$$

This is a constant multiple of the first difference, ∇Y_t.[†]

A second set of assumptions might be that M_t in Equation (5.1.8) is stochastic and changes slowly over time governed by a random walk model. Suppose, for example, that

$$Y_t = M_t + e_t \qquad \text{with} \qquad M_t = M_{t-1} + \varepsilon_t \tag{5.1.9}$$

where $\{e_t\}$ and $\{\varepsilon_t\}$ are independent white noise series. Then

$$\begin{aligned} \nabla Y_t &= \nabla M_t + \nabla e_t \\ &= \varepsilon_t + e_t - e_{t-1} \end{aligned}$$

which would have the autocorrelation function of an MA(1) series with

$$\rho_1 = -\{1/[2 + (\sigma_\varepsilon^2/\sigma_e^2)]\} \tag{5.1.10}$$

In either of these situations, we are led to the study of ∇Y_t as a stationary process.

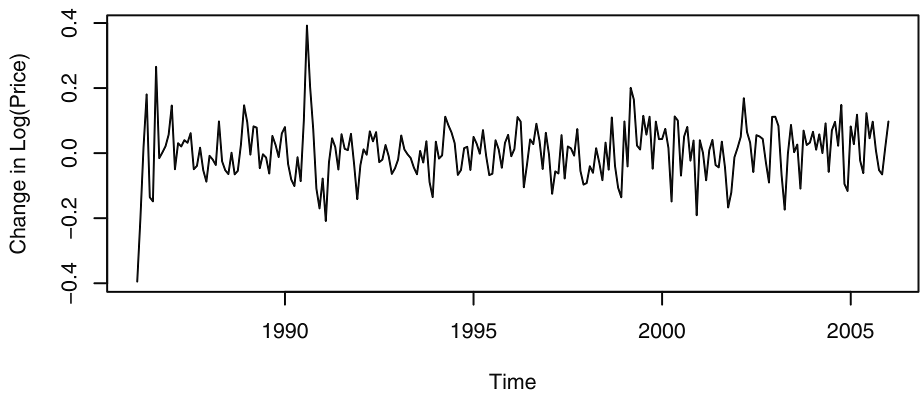
Returning to the oil price time series, Exhibit 5.4 displays the time series plot of the differences of logarithms of that series.[‡] The differenced series looks much more stationary when compared with the original time series shown in Exhibit 5.1, on page 88.

[†] A more complete labeling of this difference would be that it is a **first difference at lag 1**.

[‡] In Section 5.4 on page 98 we will see why logarithms are often a convenient transformation.

(We will also see later that there are outliers in this series that need to be considered to produce an adequate model.)

Exhibit 5.4 The Difference Series of the Logs of the Oil Price Time




```
> plot(diff(log(oil.price)),ylab='Change in Log(Price)',type='l')
```

We can also make assumptions that lead to stationary second-difference models. Again we assume that Equation (5.1.8) on page 90, holds, but now assume that M_t is linear in time over three consecutive time points. We can now estimate (predict) M_t at the middle time point t by choosing $\beta_{0,t}$ and $\beta_{1,t}$ to minimize

$$\sum_{j=-1}^{1}(Y_{t-j}-(\beta_{0,t}+j\beta_{1,t}))^2$$

The solution yields

$$\hat{M}_t = \frac{1}{3}(Y_{t+1}+Y_t+Y_{t-1})$$

and thus the detrended series is

$$\begin{aligned} Y_t-\hat{M}_t &= Y_t-\left(\frac{Y_{t+1}+Y_t+Y_{t-1}}{3}\right) \\ &= \left(-\frac{1}{3}\right)(Y_{t+1}-2Y_t+Y_{t-1}) \\ &= \left(-\frac{1}{3}\right)\nabla(\nabla Y_{t+1}) \\ &= \left(-\frac{1}{3}\right)\nabla^2(Y_{t+1}) \end{aligned}$$

a constant multiple of the centered **second difference** of Y_t. Notice that we have differenced twice, but both differences are at lag 1.

Alternatively, we might assume that

$$Y_t = M_t + e_t, \quad \text{where} \quad M_t = M_{t-1} + W_t \quad \text{and} \quad W_t = W_{t-1} + \varepsilon_t \tag{5.1.11}$$

with $\{e_t\}$ and $\{\varepsilon_t\}$ independent white noise time series. Here the stochastic trend M_t is such that its "rate of change," ∇M_t, is changing slowly over time. Then

$$\nabla Y_t = \nabla M_t + \nabla e_t = W_t + \nabla e_t$$

and

$$\begin{aligned}\nabla^2 Y_t &= \nabla W_t + \nabla^2 e_t \\ &= \varepsilon_t + (e_t - e_{t-1}) - (e_{t-1} - e_{t-2}) \\ &= \varepsilon_t + e_t - 2e_{t-1} + e_{t-2}\end{aligned}$$

which has the autocorrelation function of an MA(2) process. The important point is that the second difference of the nonstationary process $\{Y_t\}$ is stationary. This leads us to the general definition of the important integrated autoregressive moving average time series models.

5.2 ARIMA Models

A time series $\{Y_t\}$ is said to follow an **integrated autoregressive moving average** model if the dth difference $W_t = \nabla^d Y_t$ is a stationary ARMA process. If $\{W_t\}$ follows an ARMA(p,q) model, we say that $\{Y_t\}$ is an ARIMA(p,d,q) process. Fortunately, for practical purposes, we can usually take $d = 1$ or at most 2.

Consider then an ARIMA(p,1,q) process. With $W_t = Y_t - Y_{t-1}$, we have

$$\begin{aligned}W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \cdots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \\ - \cdots - \theta_q e_{t-q}\end{aligned} \tag{5.2.1}$$

or, in terms of the observed series,

$$\begin{aligned}Y_t - Y_{t-1} &= \phi_1(Y_{t-1} - Y_{t-2}) + \phi_2(Y_{t-2} - Y_{t-3}) + \cdots + \phi_p(Y_{t-p} - Y_{t-p-1}) \\ &\quad + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}\end{aligned}$$

which we may rewrite as

$$\begin{aligned}Y_t &= (1 + \phi_1)Y_{t-1} + (\phi_2 - \phi_1)Y_{t-2} + (\phi_3 - \phi_2)Y_{t-3} + \cdots \\ &\quad + (\phi_p - \phi_{p-1})Y_{t-p} - \phi_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}\end{aligned} \tag{5.2.2}$$

We call this the **difference equation form** of the model. Notice that it appears to be an ARMA(p + 1,q) process. However, the characteristic polynomial satisfies

$$\begin{aligned}1 - (1 + \phi_1)x - (\phi_2 - \phi_1)x^2 - (\phi_3 - \phi_2)x^3 - \cdots - (\phi_p - \phi_{p-1})x^p + \phi_p x^{p+1} \\ = (1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p)(1 - x)\end{aligned}$$

which can be easily checked. This factorization clearly shows the root at $x = 1$, which implies nonstationarity. The remaining roots, however, are the roots of the characteristic polynomial of the *stationary* process ∇Y_t.

Explicit representations of the observed series in terms of either W_t or the white noise series underlying W_t are more difficult than in the stationary case. Since nonstationary processes are not in statistical equilibrium, we cannot assume that they go infinitely into the past or that they start at $t = -\infty$. However, we can and shall assume that they start at some time point $t = -m$, say, where $-m$ is earlier than time $t = 1$, at which point we first observed the series. For convenience, we take $Y_t = 0$ for $t < -m$. The difference equation $Y_t - Y_{t-1} = W_t$ can be solved by summing both sides from $t = -m$ to $t = t$ to get the representation

$$Y_t = \sum_{j=-m}^{t} W_j \tag{5.2.3}$$

for the ARIMA(p,1,q) process.

The ARIMA(p,2,q) process can be dealt with similarly by summing twice to get the representations

$$\begin{aligned} Y_t &= \sum_{j=-m}^{t} \sum_{i=-m}^{j} W_i \\ &= \sum_{j=0}^{t+m} (j+1) W_{t-j} \end{aligned} \tag{5.2.4}$$

These representations have limited use but can be used to investigate the covariance properties of ARIMA models and also to express Y_t in terms of the white noise series $\{e_t\}$. We defer the calculations until we evaluate specific cases.

If the process contains no autoregressive terms, we call it an integrated moving average and abbreviate the name to IMA(d,q). If no moving average terms are present, we denote the model as ARI(p,d). We first consider in detail the important IMA(1,1) model.

The IMA(1,1) Model

The simple IMA(1,1) model satisfactorily represents numerous time series, especially those arising in economics and business. In difference equation form, the model is

$$Y_t = Y_{t-1} + e_t - \theta e_{t-1} \tag{5.2.5}$$

To write Y_t explicitly as a function of present and past noise values, we use Equation (5.2.3) and the fact that $W_t = e_t - \theta e_{t-1}$ in this case. After a little rearrangement, we can write

$$Y_t = e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1} \tag{5.2.6}$$

Notice that in contrast to our stationary ARMA models, the weights on the white noise terms *do not die out* as we go into the past. Since we are assuming that $-m < 1$ and $0 < t$, we may usefully think of Y_t as mostly an equally weighted accumulation of a large number of white noise values.

From Equation (5.2.6), we can easily derive variances and correlations. We have

$$Var(Y_t) = [1 + \theta^2 + (1-\theta)^2(t+m)]\sigma_e^2 \tag{5.2.7}$$

and

$$\begin{aligned} Corr(Y_t, Y_{t-k}) &= \frac{1-\theta+\theta^2+(1-\theta)^2(t+m-k)}{[Var(Y_t)Var(Y_{t-k})]^{1/2}} \\ &\approx \sqrt{\frac{t+m-k}{t+m}} \\ &\approx 1 \qquad \text{for large } m \text{ and moderate } k \end{aligned} \tag{5.2.8}$$

We see that as t increases, $Var(Y_t)$ increases and could be quite large. Also, the correlation between Y_t and Y_{t-k} will be strongly positive for many lags $k = 1, 2, \dots$.

The IMA(2,2) Model

The assumptions of Equation (5.1.11) led to an IMA(2,2) model. In difference equation form, we have

$$\nabla^2 Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

or

$$Y_t = 2Y_{t-1} - Y_{t-2} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} \tag{5.2.9}$$

The representation of Equation (5.2.4) may be used to express Y_t in terms of $e_t, e_{t-1}, \dots$. After some tedious algebra, we find that

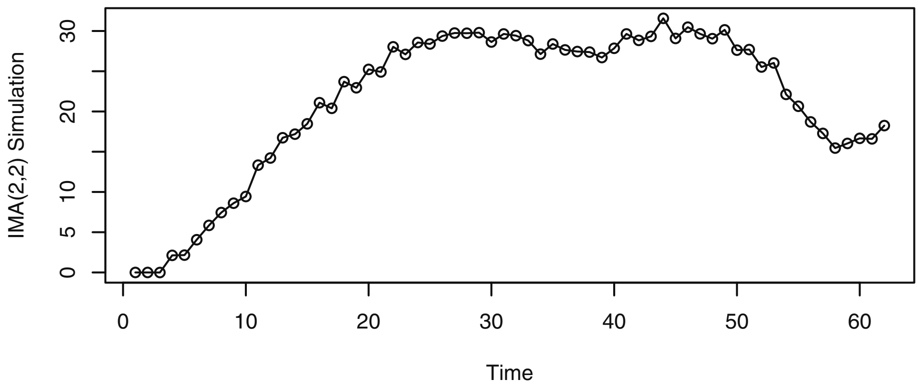
$$\begin{aligned} Y_t = e_t + \sum_{j=1}^{t+m} \psi_j e_{t-j} - [(t+m+1)\theta_1 + (t+m)\theta_2]e_{-m-1} \\ -(t+m+1)\theta_2 e_{-m-2} \end{aligned} \tag{5.2.10}$$

where $\psi_j = 1 + \theta_2 + (1 - \theta_1 - \theta_2)j$ for $j = 1, 2, 3, \dots, t + m$. Once more we see that the ψ-weights do not die out but form a linear function of j.

Again, variances and correlations for Y_t can be obtained from the representation given in Equation (5.2.10), but the calculations are tedious. We shall simply note that the variance of Y_t increases rapidly with t and again $Corr(Y_t, Y_{t-k})$ is nearly 1 for all moderate k.

The results of a simulation of an IMA(2,2) process are displayed in Exhibit 5.5. Notice the smooth change in the process values (and the unimportance of the zero-mean function). The increasing variance and the strong, positive neighboring correlations dominate the appearance of the time series plot.

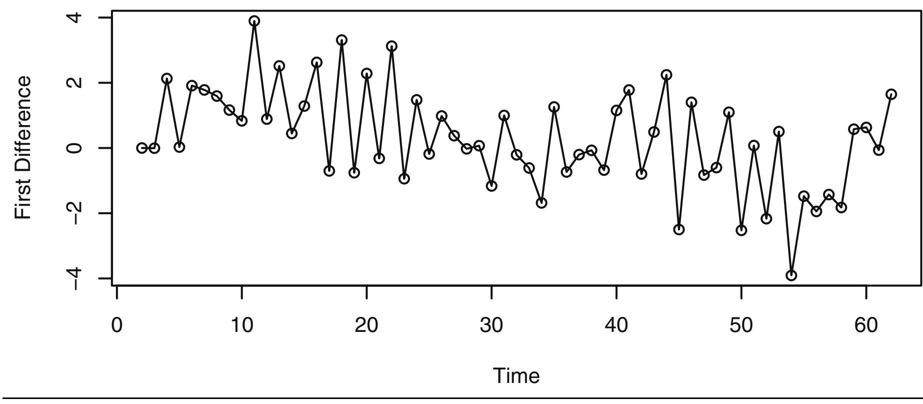
Exhibit 5.5 Simulation of an IMA(2,2) Series with $\theta_1 = 1$ and $\theta_2 = -0.6$




```
> data(ima22.s)
> plot(ima22.s,ylab='IMA(2,2) Simulation',type='o')
```

Exhibit 5.6 shows the time series plot of the first difference of the simulated series. This series is also nonstationary, as it is governed by an IMA(1,2) model.

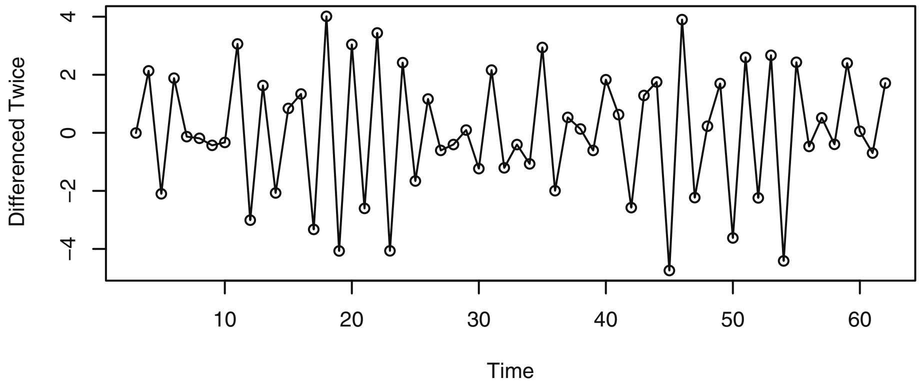
Exhibit 5.6 First Difference of the Simulated IMA(2,2) Series




```
> plot(diff(ima22.s),ylab='First Difference',type='o')
```

Finally, the second differences of the simulated IMA(2,2) series values are plotted in Exhibit 5.7. These values arise from a stationary MA(2) model with $\theta_1 = 1$ and $\theta_2 = -0.6$. From Equation (4.2.3) on page 63, the theoretical autocorrelations for this model are $\rho_1 = -0.678$ and $\rho_2 = 0.254$. These correlation values seem to be reflected in the appearance of the time series plot.

Exhibit 5.7 Second Difference of the Simulated IMA(2,2) Series




```
> plot(diff(ima22.s,difference=2),ylab='Differenced
   Twice',type='o')
```

The ARI(1,1) Model

The ARI(1,1) process will satisfy

$$Y_t - Y_{t-1} = \phi(Y_{t-1} - Y_{t-2}) + e_t \tag{5.2.11}$$

or

$$Y_t = (1+\phi)Y_{t-1} - \phi Y_{t-2} + e_t \tag{5.2.12}$$

where $|\phi| < 1$.[†]

To find the ψ-weights in this case, we shall use a technique that will generalize to arbitrary ARIMA models. It can be shown that the ψ-weights can be obtained by equating like powers of x in the identity:

$$\begin{aligned}(1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi^p x^p)(1-x)^d(1 + \psi_1 x + \psi_2 x^2 + \psi_3 x^3 + \cdots) \\ = (1 - \theta_1 x - \theta_2 x^2 - \theta_3 x^3 - \cdots - \theta_q x^q)\end{aligned} \tag{5.2.13}$$

In our case, this relationship reduces to

$$(1 - \phi x)(1 - x)(1 + \psi_1 x + \psi_2 x^2 + \psi_3 x^3 + \cdots) = 1$$

or

$$[1 - (1+\phi)x + \phi x^2](1 + \psi_1 x + \psi_2 x^2 + \psi_3 x^3 + \cdots) = 1$$

Equating like powers of x on both sides, we obtain

[†] Notice that this looks like a special AR(2) model. However, one of the roots of the corresponding AR(2) characteristic polynomial is 1, and this is not allowed in stationary AR(2) models.

$$-(1+\phi)+\psi_1 = 0$$
$$\phi-(1+\phi)\psi_1+\psi_2 = 0$$

and, in general,

$$\psi_k = (1+\phi)\psi_{k-1}-\phi\psi_{k-2} \qquad \text{for } k \geq 2 \tag{5.2.14}$$

with $\psi_o = 1$ and $\psi_1 = 1 + \phi$. This recursion with starting values allows us to compute as many ψ-weights as necessary. It can also be shown that in this case an explicit solution to the recursion is given as

$$\psi_k = \frac{1-\phi^{k+1}}{1-\phi} \quad \text{for } k \geq 1 \tag{5.2.15}$$

(It is easy, for example, to show that this expression satisfies Equation (5.2.14).

5.3 Constant Terms in ARIMA Models

For an ARIMA(p,d,q) model, $\nabla^d Y_t = W_t$ is a stationary ARMA(p,q) process. Our standard assumption is that stationary models have a zero mean; that is, we are actually working with deviations from the constant mean. A nonzero constant mean, μ, in a stationary ARMA model $\{W_t\}$ can be accommodated in either of two ways. We can assume that

$$W_t-\mu = \phi_1(W_{t-1}-\mu)+\phi_2(W_{t-2}-\mu)+\cdots+\phi_p(W_{t-p}-\mu)$$
$$+e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}-\cdots-\theta_q e_{t-q}$$

Alternatively, we can introduce a constant term θ_0 into the model as follows:

$$W_t = \theta_0+\phi_1 W_{t-1}+\phi_2 W_{t-2}+\cdots+\phi_p W_{t-p}$$
$$+e_t-\theta_1 e_{t-1}-\theta_2 e_{t-2}-\cdots-\theta_q e_{t-q}$$

Taking expected values on both sides of the latter expression, we find that

$$\mu = \theta_0+(\phi_1+\phi_2+\cdots+\phi_p)\mu$$

so that

$$\mu = \frac{\theta_0}{1-\phi_1-\phi_2-\cdots-\phi_p} \tag{5.3.16}$$

or, conversely, that

$$\theta_0 = \mu(1-\phi_1-\phi_2-\cdots-\phi_p) \tag{5.3.17}$$

Since the alternative representations are equivalent, we shall use whichever parameterization is convenient.

What will be the effect of a nonzero mean for W_t on the undifferenced series Y_t? Consider the IMA(1,1) case with a constant term. We have

$$Y_t = Y_{t-1} + \theta_0 + e_t - \theta e_{t-1}$$

or

$$W_t = \theta_0 + e_t - \theta e_{t-1}$$

Either by substituting into Equation (5.2.3) on page 93 or by iterating into the past, we find that

$$\begin{aligned} Y_t = e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1} \\ + (t+m+1)\theta_0 \end{aligned} \tag{5.3.18}$$

Comparing this with Equation (5.2.6), we see that we have an added *linear deterministic time trend* $(t + m + 1)\theta_0$ with slope θ_0.

An equivalent representation of the process would then be

$$Y_t = Y_t' + \beta_0 + \beta_1 t$$

where Y_t' is an IMA(1,1) series with $E(\nabla Y_t') = 0$ and $E(\nabla Y_t) = \beta_1$.

For a general ARIMA(p,d,q) model where $E(\nabla^d Y_t) \neq 0$, it can be argued that $Y_t = Y_t' + \mu_t$, where μ_t is a deterministic polynomial of degree d and Y_t' is ARIMA(p,d,q) with $EY_t' = 0$. With $d = 2$ and $\theta_0 \neq 0$, a quadratic trend would be implied.

5.4 Other Transformations

We have seen how differencing can be a useful transformation for achieving stationarity. However, the logarithm transformation is also a useful method in certain circumstances. We frequently encounter series where increased dispersion seems to be associated with higher levels of the series—the higher the level of the series, the more variation there is around that level and conversely.

Specifically, suppose that $Y_t > 0$ for all t and that

$$E(Y_t) = \mu_t \qquad \text{and} \qquad \sqrt{Var(Y_t)} = \mu_t \sigma \tag{5.4.1}$$

Then

$$E[\log(Y_t)] \approx \log(\mu_t) \qquad \text{and} \qquad Var(\log(Y_t)) \approx \sigma^2 \tag{5.4.2}$$

These results follow from taking expected values and variances of both sides of the (Taylor) expansion

$$\log(Y_t) \approx \log(\mu_t) + \frac{Y_t - \mu_t}{\mu_t}$$

In words, if the standard deviation of the series is proportional to the level of the series, then transforming to logarithms will produce a series with approximately constant variance over time. Also, if the level of the series is changing roughly exponentially, the log-transformed series will exhibit a linear time trend. Thus, we might then want to take first differences. An alternative set of assumptions leading to differences of logged data follows.

Percentage Changes and Logarithms

Suppose Y_t tends to have relatively stable percentage changes from one time period to the next. Specifically, assume that

$$Y_t = (1 + X_t)Y_{t-1}$$

where $100X_t$ is the percentage change (possibly negative) from Y_{t-1} to Y_t. Then

$$\begin{aligned}\log(Y_t) - \log(Y_{t-1}) &= \log\left(\frac{Y_t}{Y_{t-1}}\right)\\ &= \log(1 + X_t)\end{aligned}$$

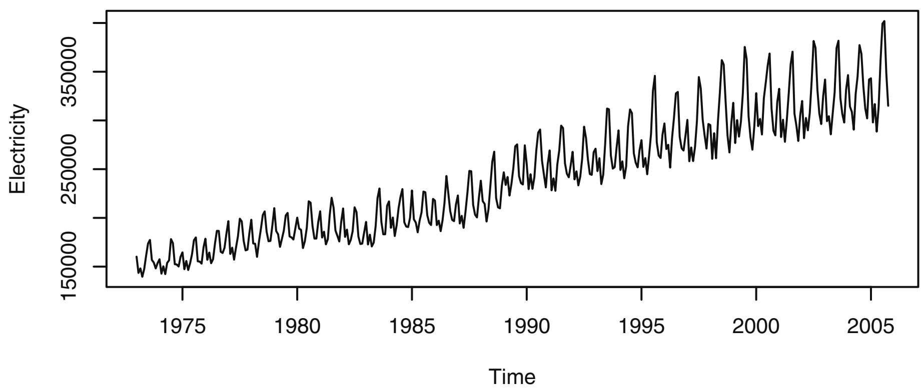
If X_t is restricted to, say, $|X_t| < 0.2$ (that is, the percentage changes are at most ±20%), then, to a good approximation, $\log(1+X_t) \approx X_t$. Consequently,

$$\nabla[\log(Y_t)] \approx X_t \tag{5.4.3}$$

will be relatively stable and perhaps well-modeled by a stationary process. Notice that we take logs first and then compute first differences—the order does matter. In financial literature, the differences of the (natural) logarithms are usually called **returns**.

As an example, consider the time series shown in Exhibit 5.8. This series gives the total monthly electricity generated in the United States in millions of kilowatt-hours. The higher values display considerably more variation than the lower values.

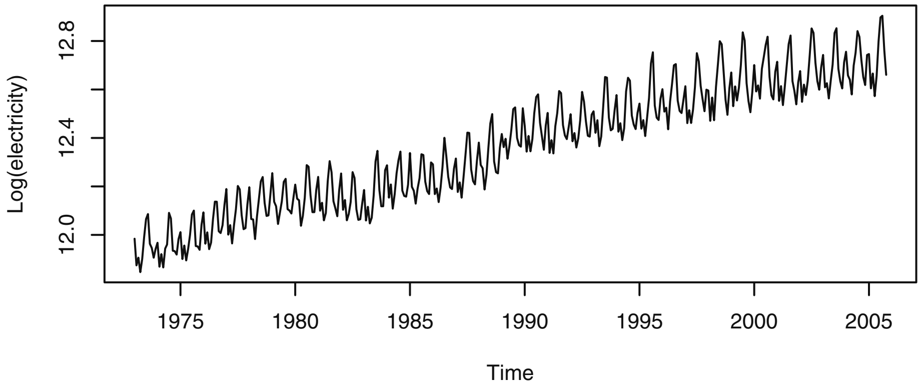
Exhibit 5.8 U.S. Electricity Generated by Month




```
> data(electricity); plot(electricity)
```

Exhibit 5.9 displays the time series plot of the logarithms of the electricity values. Notice how the amount of variation around the upward trend is now much more uniform across high and low values of the series.

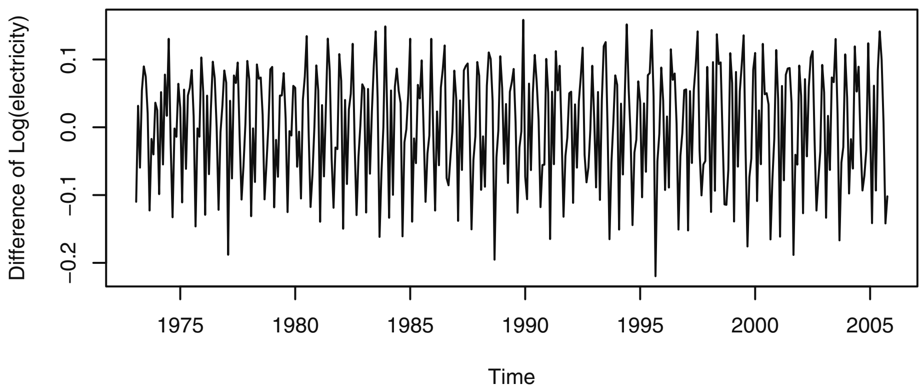
Exhibit 5.9 Time Series Plot of Logarithms of Electricity Values




```
> plot(log(electricity),ylab='Log(electricity)')
```

The differences of the logarithms of the electricity values are displayed in Exhibit 5.10. On the basis of this plot, we might well consider a stationary model as appropriate.

Exhibit 5.10 Difference of Logarithms for Electricity Time Series




```
> plot(diff(log(electricity)),
   ylab='Difference of Log(electricity)')
```

Power Transformations

A flexible family of transformations, the **power transformations**, was introduced by Box and Cox (1964). For a given value of the parameter λ, the transformation is defined by

$$g(x) = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & \text{for } \lambda \neq 0 \\ \log x & \text{for } \lambda = 0 \end{cases} \tag{5.4.4}$$

The term x^{λ} is the important part of the first expression, but subtracting 1 and dividing by λ makes $g(x)$ change smoothly as λ approaches zero. In fact, a calculus argument[†] shows that as $\lambda \to 0$, $(x^{\lambda} - 1)/\lambda \to \log(x)$. Notice that $\lambda = ½$ produces a square root transformation useful with Poisson-like data, and $\lambda = -1$ corresponds to a reciprocal transformation.

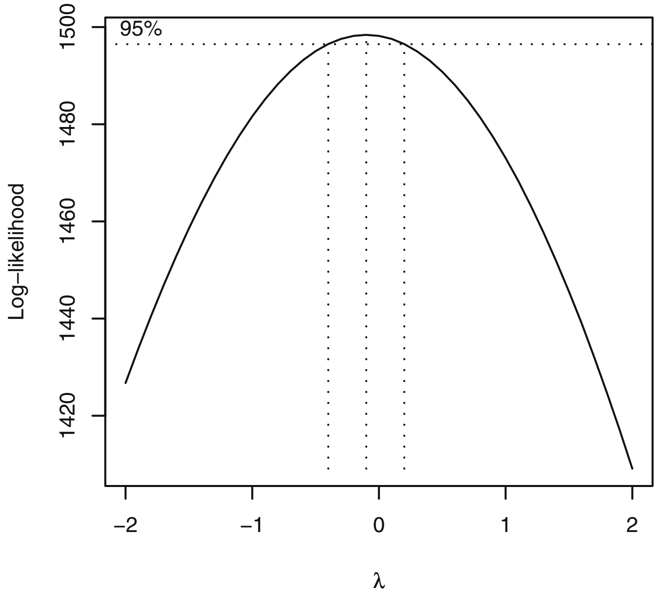
The power transformation applies only to positive data values. If some of the values are negative or zero, a positive constant may be added to all of the values to make them all positive before doing the power transformation. The shift is often determined subjectively. For example, for nonnegative catch data in biology, the occurrence of zeros is often dealt with by adding a constant equal to the smallest positive data value to all of the data values. An alternative approach consists of using transformations applicable to any data—positive or not. A drawback of this alternative approach is that interpretations of such transformations are often less straightforward than the interpretations of the power transformations. See Yeo and Johnson (2000) and the references contained therein.

We can consider λ as an additional parameter in the model to be estimated from the observed data. However, precise estimation of λ is usually not warranted. Evaluation of a range of transformations based on a grid of λ values, say ± 1, $\pm 1/2$, $\pm 1/3$, $\pm 1/4$, and 0, will usually suffice and may have some intuitive meaning.

Software allows us to consider a range of lambda values and calculate a log-likelihood value for each lambda value based on a normal likelihood function. A plot of these values is shown in Exhibit 5.11 for the electricity data. The 95% confidence interval for λ contains the value of $\lambda = 0$ quite near its center and strongly suggests a logarithmic transformation ($\lambda = 0$) for these data.

[†] Exercise (5.17) asks you to verify this.

Exhibit 5.11 Log-likelihood versus Lambda




```
> BoxCox.ar(electricity)
```

5.5 Summary

This chapter introduced the concept of differencing to induce stationarity on certain nonstationary processes. This led to the important integrated autoregressive moving average models (ARIMA). The properties of these models were then thoroughly explored. Other transformations, namely percentage changes and logarithms, were then considered. More generally, power transformations or Box-Cox transformations were introduced as useful transformations to stationarity and often normality.

EXERCISES

5.1 Identify the following as specific ARIMA models. That is, what are p, d, and q and what are the values of the parameters (the ϕ's and θ's)?

(a) $Y_t = Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1}$.

(b) $Y_t = 2Y_{t-1} - Y_{t-2} + e_t$.

(c) $Y_t = 0.5Y_{t-1} - 0.5Y_{t-2} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$.

5.2 For each of the ARIMA models below, give the values for $E(\nabla Y_t)$ and $Var(\nabla Y_t)$.

(a) $Y_t = 3 + Y_{t-1} + e_t - 0.75e_{t-1}$.

(b) $Y_t = 10 + 1.25Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1}$.

(c) $Y_t = 5 + 2Y_{t-1} - 1.7Y_{t-2} + 0.7Y_{t-3} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$.

5.3 Suppose that $\{Y_t\}$ is generated according to $Y_t = e_t + ce_{t-1} + ce_{t-2} + ce_{t-3} + \cdots + ce_0$ for $t > 0$.

(a) Find the mean and covariance functions for $\{Y_t\}$. Is $\{Y_t\}$ stationary?

(b) Find the mean and covariance functions for $\{\nabla Y_t\}$. Is $\{\nabla Y_t\}$ stationary?

(c) Identify $\{Y_t\}$ as a specific ARIMA process.

5.4 Suppose that $Y_t = A + Bt + X_t$, where $\{X_t\}$ is a random walk. First suppose that A and B are constants.

(a) Is $\{Y_t\}$ stationary?

(b) Is $\{\nabla Y_t\}$ stationary?

Now suppose that A and B are random variables that are independent of the random walk $\{X_t\}$.

(c) Is $\{Y_t\}$ stationary?

(d) Is $\{\nabla Y_t\}$ stationary?

5.5 Using the simulated white noise values in Exhibit 5.2, on page 88, verify the values shown for the explosive process Y_t.

5.6 Consider a stationary process $\{Y_t\}$. Show that if $\rho_1 < ½$, ∇Y_t has a larger variance than does Y_t.

5.7 Consider two models:

A: $Y_t = 0.9Y_{t-1} + 0.09Y_{t-2} + e_t$.

B: $Y_t = Y_{t-1} + e_t - 0.1e_{t-1}$.

(a) Identify each as a specific ARIMA model. That is, what are p, d, and q and what are the values of the parameters, ϕ's and θ's?

(b) In what ways are the two models different?

(c) In what ways are the two models similar? (Compare ψ-weights and π-weights.)

5.8 Consider a nonstationary "AR(1)" process defined as a solution to Equation (5.1.2) on page 88, with $|\phi| > 1$.

(a) Derive an equation similar to Equation (5.1.3) on page 88, for this more general case. Use $Y_0 = 0$ as an initial condition.

(b) Derive an equation similar to Equation (5.1.4) on page 89, for this more general case.

(c) Derive an equation similar to Equation (5.1.5) on page 89, for this more general case.

(d) Is it true that for any $|\phi| > 1$, $Corr(Y_t, Y_{t-k}) \approx 1$ for large t and moderate k?

5.9 Verify Equation (5.1.10) on page 90.

5.10 Nonstationary ARIMA series can be simulated by first simulating the corresponding stationary ARMA series and then "integrating" it (really partially summing it). Use statistical software to simulate a variety of IMA(1,1) and IMA(2,2) series with a variety of parameter values. Note any stochastic "trends" in the simulated series.

5.11 The data file winnebago contains monthly unit sales of recreational vehicles (RVs) from Winnebago, Inc., from November 1966 through February 1972.

(a) Display and interpret the time series plot for these data.

(b) Now take natural logarithms of the monthly sales figures and display the time series plot of the transformed values. Describe the effect of the logarithms on the behavior of the series.

(c) Calculate the fractional relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them with the differences of (natural) logarithms, $\nabla\log(Y_t) = \log(Y_t) - \log(Y_{t-1})$. How do they compare for smaller values and for larger values?

5.12 The data file SP contains quarterly Standard & Poor's Composite Index stock price values from the first quarter of 1936 through the fourth quarter of 1977.

(a) Display and interpret the time series plot for these data.

(b) Now take natural logarithms of the quarterly values and display and the time series plot of the transformed values. Describe the effect of the logarithms on the behavior of the series.

(c) Calculate the (fractional) relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them to the differences of (natural) logarithms, $\nabla\log(Y_t)$. How do they compare for smaller values and for larger values?

5.13 The data file airpass contains international airline passenger monthly totals (in thousands) flown from January 1960 through December 1971. This is a classic time series analyzed in Box and Jenkins (1976).

(a) Display and interpret the time series plot for these data.

(b) Now take natural logarithms of the monthly values and display and the time series plot of the transformed values. Describe the effect of the logarithms on the behavior of the series.

(c) Calculate the (fractional) relative changes, $(Y_t - Y_{t-1})/Y_{t-1}$, and compare them to the differences of (natural) logarithms, $\nabla\log(Y_t)$. How do they compare for smaller values and for larger values?

5.14 Consider the annual rainfall data for Los Angeles shown in Exhibit 1.1, on page 2. The quantile-quantile normal plot of these data, shown in Exhibit 3.17, on page 50, convinced us that the data were not normal. The data are in the file larain.

(a) Use software to produce a plot similar to Exhibit 5.11, on page 102, and determine the "best" value of λ for a power transformation of the data.

(b) Display a quantile-quantile plot of the transformed data. Are they more normal?

(c) Produce a time series plot of the transformed values.

(d) Use the transformed values to display a plot of Y_t versus Y_{t-1} as in Exhibit 1.2, on page 2. Should we expect the transformation to change the dependence or lack of dependence in the series?

5.15 Quarterly earnings per share for the Johnson & Johnson Company are given in the data file named JJ. The data cover the years from 1960 through 1980.

(a) Display a time series plot of the data. Interpret the interesting features in the plot.

(b) Use software to produce a plot similar to Exhibit 5.11, on page 102, and determine the "best" value of λ for a power transformation of these data.

(c) Display a time series plot of the transformed values. Does this plot suggest that a stationary model might be appropriate?

(d) Display a time series plot of the differences of the transformed values. Does this plot suggest that a stationary model might be appropriate for the differences?

5.16 The file named gold contains the daily price of gold (in dollars per troy ounce) for the 252 trading days of year 2005.

(a) Display the time series plot of these data. Interpret the plot.

(b) Display the time series plot of the differences of the logarithms of these data. Interpret this plot.

(c) Calculate and display the sample ACF for the differences of the logarithms of these data and argue that the logarithms appear to follow a random walk model.

(d) Display the differences of logs in a histogram and interpret.

(e) Display the differences of logs in a quantile-quantile normal plot and interpret.

5.17 Use calculus to show that, for any fixed $x > 0$, as $\lambda \to 0$, $(x^\lambda - 1)/\lambda \to \log x$.

Appendix D: The Backshift Operator

Many other books and much of the time series literature use what is called the **backshift operator** to express and manipulate ARIMA models. The backshift operator, denoted B, operates on the time index of a series and shifts time back one time unit to form a new series.[†] In particular,

$$BY_t = Y_{t-1}$$

The backshift operator is linear since for any constants a, b, and c and series Y_t and X_t, it is easy to see that

$$B(aY_t + bX_t + c) = aBY_t + bBX_t + c$$

Consider now the MA(1) model. In terms of B, we can write

$$\begin{aligned} Y_t &= e_t - \theta e_{t-1} = e_t - \theta B e_t = (1 - \theta B) e_t \\ &= \theta(B) e_t \end{aligned}$$

where $\theta(B)$ is the MA characteristic polynomial "evaluated" at B.

Since BY_t is itself a time series, it is meaningful to consider BBY_t. But clearly $BBY_t = BY_{t-1} = Y_{t-2}$, and we can write

$$B^2 Y_t = Y_{t-2}$$

More generally, we have

$$B^m Y_t = Y_{t-m}$$

for any positive integer m. For a general MA(q) model, we can then write

$$\begin{aligned} Y_t &= e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q} \\ &= e_t - \theta_1 B e_t - \theta_2 B^2 e_t - \cdots - \theta_q B^q e_t \\ &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) e_t \end{aligned}$$

or

$$Y_t = \theta(B) e_t$$

where, again, $\theta(B)$ is the MA characteristic polynomial evaluated at B.

For autoregressive models AR(p), we first move all of the terms involving Y to the left-hand side

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

and then write

$$Y_t - \phi_1 B Y_t - \phi_2 B^2 Y_t - \cdots - \phi_p B^p Y_t = e_t$$

or

[†] Sometimes B is called a **Lag operator**.

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) Y_t = e_t$$

which can be expressed as

$$\phi(B) Y_t = e_t$$

where $\phi(B)$ is the AR characteristic polynomial evaluated at B.

Combining the two, the general ARMA(p,q) model may be written compactly as

$$\phi(B) Y_t = \theta(B) e_t$$

Differencing can also be conveniently expressed in terms of B. We have

$$\begin{aligned} \nabla Y_t &= Y_t - Y_{t-1} = Y_t - BY_t \\ &= (1-B) Y_t \end{aligned}$$

with second differences given by

$$\nabla^2 Y_t = (1-B)^2 Y_t$$

Effectively, $\nabla = 1 - B$ and $\nabla^2 = (1 - B)^2$.

The general ARIMA(p,d,q) model is expressed concisely as

$$\phi(B)(1-B)^d Y_t = \theta(B) e_t$$

In the literature, one must carefully distinguish from the context the use of B as a backshift operator and its use as an ordinary real (or complex) variable. For example, the stationarity condition is frequently given by stating that the roots of $\phi(B) = 0$ must be greater than 1 in absolute value or, equivalently, must lie outside the unit circle in the complex plane. Here B is to be treated as a dummy variable in an equation rather than as the backshift operator.