Chapter 10 Autocorrelation in Time Series Data

The methods of regression and smoothing discussed in previous chapters are incomplete in that they do not take full advantage of the internal structure of time series data. In business and economic time series data, the data are usually correlated and the methods of regression and smoothing do take advantage of this additional information about the internal structure of time series data.

In this chapter we shall examine the internal correlation of time series data and introduce a new set of time series model, often called Box-Jenkins models, which can exploited this additional information. Thus, this chapter is in two parts: the first part is a discussion of autocorrelation in time series data (sections 1 — 4) and the second part is an introduction to the Box-Jenkins models which take advantage of the autocorrelation of time series data (sections 5 — 10).

Chapter 10

Stationary Time Series

With any set of time series data the first stage in forecasting is the stage of Data Collection and Analysis. Given a time series:

Table 10.1

t	Y_t
1	Y_1
2	Y_2
3	Y_3
4	Y_4
5	Y_5
$\vdots$	$\vdots$

We can think of these observations, Y_1, Y_2, Y_3, Y_4, Y_5,$\ldots$ as one possible realization of a stochastic process. That is, if it were possible to replay this time series over again, we would have another set of observations, Y_1, Y_2, Y_3, Y_4, Y_5, $\ldots$ another realization, and another, and so on.

Table 10.2

t		Y_t		
1	Y_1	Y_1	Y_1	$\cdots$
2	Y_2	Y_2	Y_2	$\cdots$
3	Y_3	Y_3	Y_3	$\cdots$
4	Y_4	Y_4	Y_4	$\cdots$
5	Y_5	Y_5	Y_5	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Thus, for each particular period, *t*, there is a probability distribution for all the Y_t*'s* associated with it. For each period *t*, there is a mean, μ_{Y_t}, and variance, $\sigma^2_{Y_t}$.

Table 10.3

t		Y_t				
1	Y_1	Y_1	Y_1	$\cdots$	μ_{Y_1}	$\sigma^2_{Y_1}$
2	Y_2	Y_2	Y_2	$\cdots$	μ_{Y_2}	$\sigma^2_{Y_2}$
3	Y_3	Y_3	Y_3	$\cdots$	μ_{Y_3}	$\sigma^2_{Y_3}$
4	Y_4	Y_4	Y_4	$\cdots$	μ_{Y_4}	$\sigma^2_{Y_4}$
5	Y_5	Y_5	Y_5	$\cdots$	μ_{Y_5}	$\sigma^2_{Y_5}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$

It would then be possible to determine all covariances between different periods. For example, the covariance between the set of observations at period 2, $\{\, Y_2 \,\}$, and observations at period 4, $\{\, Y_4 \,\}$.

$$COV(Y_2, Y_4).$$

The variance-covariance matrix for the first 5 periods

	Y_1	Y_2	Y_3	Y_4	Y_5
Y_1	$\sigma^2_{Y_1}$	$COV(Y_1,Y_2)$	$COV(Y_1,Y_3)$	$COV(Y_1,Y_4)$	$COV(Y_1,Y_5)$
Y_2	$COV(Y_1,Y_2)$	$\sigma^2_{Y_2}$	$COV(Y_2,Y_3)$	$COV(Y_2,Y_4)$	$COV(Y_2,Y_5)$
Y_3	$COV(Y_1,Y_3)$	$COV(Y_2,Y_3)$	$\sigma^2_{Y_3}$	$COV(Y_3,Y_4)$	$COV(Y_3,Y_5)$
Y_4	$COV(Y_1,Y_4)$	$COV(Y_2,Y_4)$	$COV(Y_3,Y_4)$	$\sigma^2_{Y_4}$	$COV(Y_4,Y_5)$
Y_5	$COV(Y_1,Y_5)$	$COV(Y_2,Y_5)$	$COV(Y_3,Y_5)$	$COV(Y_4,Y_5)$	$\sigma^2_{Y_5}$

Considering the 5 means, μ_{Y_1}, μ_{Y_2}, μ_{Y_3}, μ_{Y_4} μ_{Y_5} and the 10 variances-covariances, we have 15 possible parameters to be estimated based on the first 5 observations.

Clearly, it would be much simpler if all the means were equal and all the variances were equal; i. e.

$$\mu_{Y_1} = \mu_{Y_2} = \mu_{Y_3} = \mu_{Y_4} = \mu_{Y_5} = \mu_Y$$

$$\sigma_{Y_1} = \sigma_{Y_2} = \sigma_{Y_3} = \sigma_{Y_4} = \sigma_{Y_5} = \sigma_Y$$

Indeed, we define a time series as ***stationary*** if all the means are equal over time and all the variances are equal over time. Thus, by stationary we mean a series which is in "equilibrium" -- there is no trend in the series, there is no changing variance, it is "horizontal" data.

Stationarity in Time Series Data

Formally, a time series which is in "equilibrium' is termed a "stationary time series."

This means that times series satisfies three conditions.

(1) the series has constant mean:
μ_Y is constant over all time.

Constant Mean

$$E(Y_t) = \mu_Y \qquad \textbf{10.1}$$

(2) the series has constant variance:
$\sigma^2_{Y_t}$ is constant over time.

Constant Variance

$$\gamma_0 = Var(Y_t) \qquad \textbf{10.2}$$

(γ_0, "gamma zero," or "gamma nought," replacing $\sigma^2_{Y_t}$, for reasons that will become clear shortly.)

(3) the covariances within the series only depend on lag:

This means that the covariance between observations is only dependent on the number of periods separating the observations. We shall explain condition (3) in more detail now in section 2.

Autocovariance in Time Series Data

For example, $COV(Y_1,Y_2)$ is the covariance between the observations at period 1 and the observations at period 2, they are separated by 1 period. Another covariance, $COV(Y_4,Y_5)$, is a covariance between the observations of period 4 and the observations of period 5, they too are separated by 1 period. In a stationary process,

$$COV(Y_1, Y_2) = COV(Y_4, Y_5).$$

Thus, in our example of the five periods of observations, if the process is stationary, then

$$COV(Y_{1,}Y_2) = COV(Y_{2,}Y_3) = COV(Y_{3,}Y_4) = COV(Y_{4,}Y_5)$$

because they all are observations separated by 1 period.

We denote this by ("gamma 1")

$$\gamma_1 = COV(Y_{1,}Y_2) = COV(Y_{2,}Y_3) = COV(Y_{3,}Y_4) = COV(Y_{4,}Y_5)$$

Because we are constructing a covariance between a series and itself, between Y_t and Y_{t-1}, this is called the ***autocovariance.*** Because the autocovariance is is based on observations 1 period apart, this is called the ***autocovariance at lag 1.***

We formally denote this as

Autocovariance at lag 1

$$\boldsymbol{\gamma_1 = COV(Y_t, Y_{t-1})} \qquad \textbf{10.3}$$

Similarly, in a stationary process

$$\gamma_2 = COV(Y_1, Y_3) = COV(Y_2, Y_4) = COV(Y_3, Y_5)$$

These are covariances of observations separated by two periods.

$$\gamma_2 = \textit{autocovariance at lag 2}$$

Autocovariance at lag 2

$$\boldsymbol{\gamma_2 = COV(Y_t, Y_{t-2})} \qquad \textbf{10.4}$$

Then, in a stationary process

$$\gamma_3 = COV(Y_1, Y_4) = COV(Y_2, Y_5)$$

These are covariances of observations separated by three periods.

$$\gamma_3 = \textit{autocovariance at lag 3}$$

Autocovariance at lag 3

$$\boldsymbol{\gamma_3 = COV(Y_t, Y_{t-3})} \qquad \textbf{10.5}$$

And so on.

Notice, that the autocovariance at lag 0 is

Autocovariance at lag 0

$$\boldsymbol{\gamma_0 = COV(Y_t, Y_{t-0})} \qquad \textbf{\textit{10.6}}$$

$$\gamma_0 = COV(Y_t, Y_t)$$
$$\gamma_0 = VAR(Y_t), \text{ the variance of } Y_t$$

Autocovariance at lag 0

$$\boldsymbol{\gamma_0 = \sigma^2_{Y_t}, \textbf{ the variance of } Y_t}$$

The variance-covariance matrix of this stationary process is thus

$$\Sigma = \begin{matrix} \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 \\ \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 & \gamma_3 \\ \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 & \gamma_2 \\ \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_4 & \gamma_3 & \gamma_2 & \gamma_1 & \gamma_0 \end{matrix}$$

In our example, we have reduced the matrix to one mean and one variance, μ_Y and σ^2_Y, and the autocovariances at lags 1 through 4.

Our example listed only 5 periods of observations. In theory a time series Y_t is an infinite process so that $\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_k, \ldots$ exist for all lags.

γ_0

$$\gamma_0 = COV(Y_t, Y_t) = VAR(Y_t) \qquad \textbf{10.7}$$

γ_1

$$\gamma_1 = COV(Y_t, Y_{t-1}) \qquad \textbf{10.8}$$

γ_2

$$\gamma_2 = COV(Y_t, Y_{t-2}) \qquad \textbf{10.9}$$

$$\vdots$$

γ_k

$$\gamma_k = COV(Y_t, Y_{t-k}) \qquad \textbf{10.10}$$

$$\vdots$$

In a stationary process, the autocovariances are only dependent on the lag. This is property (3) of a stationary process.

Because we are only concerned with the first and second-order moments of the process, this is also called ***second-order*** or ***weak stationarity***. If it is applied to all higher moments then it is called ***strong stationarity***, or just ***stationarity***. If, however, we assume the Y_t's are normally distributed then weak stationarity and stationarity are equivalent.

Autocorrelation in Time Series Data

Recall from Chapter 2 the relation between the covariance and correlation of two random variables.

COVARIANCE

$$\sigma_{XY} = COV(X,Y)$$

CORRELATION

$$\rho_{XY} = \frac{COV(X,Y)}{\sigma_X \sigma_Y}$$

With regard to ***autocovariance*** and ***autocorrelation at lag 1***, for example,

AUTOCOVARIANCE at lag 1

$$\gamma_1 = COV(Y_t, Y_{t-1}) \qquad \textbf{10.11}$$

AUTOCORRELATION at lag 1

$$\rho_1 = \frac{COV(Y_t, Y_{t-1})}{\sigma_{Y_t} \sigma_{Y_{t-1}}} \qquad \textbf{10.12}$$

Because $\sigma_{Y_t} = \sqrt{\gamma_0}$ and $\sigma_{Y_{t-1}} = \sqrt{\gamma_0}$, we have a denominator in equation 5.12 of

$\sigma_{Y_t}\sigma_{Y_{t-1}} = \sqrt{\gamma_0}\sqrt{\gamma_0} = \gamma_0$, the variance of the series Y_t

AUTOCORRELATION at lag 1

$$\rho_1 = \frac{COV(Y_t, Y_{t-1})}{\gamma_0} \quad \mathbf{10.13}$$

$$\rho_1 = \frac{\gamma_1}{\gamma_0} \quad \mathbf{10.14}$$

The autocorrelation at other lags are similary defined.

AUTOCORRELATION at lag 2

$$\rho_2 = \frac{COV(Y_t, Y_{t-2})}{\gamma_0} \quad \mathbf{10.15}$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} \quad \mathbf{10.16}$$

AUTOCORRELATION at lag 3

$$\rho_3 = \frac{COV(Y_t, Y_{t-3})}{\gamma_0} \quad \mathbf{10.17}$$

$$\rho_3 = \frac{\gamma_3}{\gamma_0} \quad \mathbf{10.18}$$

AUTOCORRELATION at lag k

$$\rho_k = \frac{COV(Y_t, Y_{t-k})}{\gamma_0} \quad \mathbf{10.19}$$

$$\rho_k = \frac{\gamma_k}{\gamma_0} \quad \mathbf{10.20}$$

Notice, also, then that

AUTOCORRELATION at lag 0

$$\boldsymbol{\rho_0 = \frac{COV(Y_t,Y_t)}{\gamma_0}} \qquad \mathbf{10.21}$$

$$\boldsymbol{\rho_0 = \frac{\gamma_0}{\gamma_0} = 1} \qquad \mathbf{10.22}$$

$\rho_0 = 1$ is makes sense because there is perfect correlation between a time series Y_t and itself.

The set of autocorrelations

$1, \rho_1, \rho_2, \rho_3, \ldots, \rho_k, \ldots$

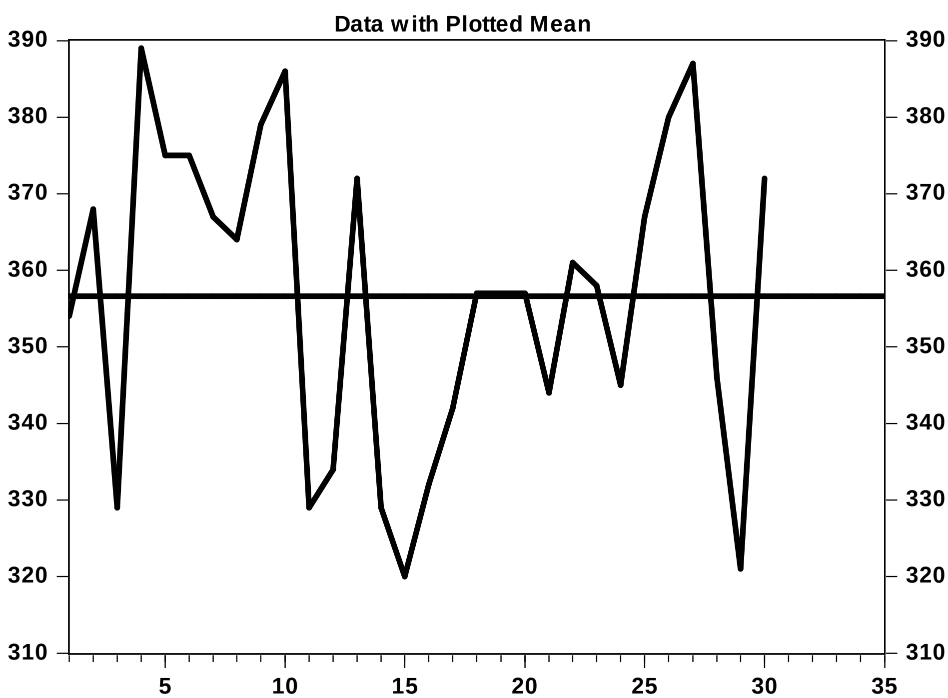
are independent of the magnitude of the numbers of the time series. They are only dependent on the lag *k*.

Calculating $\hat{\rho}_k$

As with other parameters we estimate $\hat{\rho}_k$ through sample data. We shall use the time series from Table 6.3

Table 10.4

Period	***Actual***	***Period***	***Actual***
1	354	16	332
2	368	17	342
3	329	18	357
4	389	19	357
5	375	20	357
6	375	21	344
7	367	22	361
8	364	23	358
9	379	24	345
10	386	25	367
11	329	26	380
12	334	27	387
13	372	28	346
14	329	29	321
15	320	30	372



Because we are calculating the autocorrelations from sample data this is formally called the ***sample autocorrelation function,*** the ***SACF.***

First, we calculate the sample mean, $\bar{Y}$, the estimate of μ.

$$\bar{Y} = \frac{\sum Y_t}{T}$$

$$\bar{Y} = \frac{354 + 368 + 329 + \ldots + 321 + 372}{30}$$

$$\bar{Y} = 356.6$$

Next, we calculate the sample variance, $\hat{\gamma}_0$, the estimate of the population variance, γ_0.

$$\hat{\gamma}_0 = \frac{\sum(Y_t - \bar{Y})^2}{T-1}$$

$$\hat{\gamma}_0 = \frac{\sum(Y_t - 356.6)^2}{29}$$

$$\hat{\gamma}_0 = \frac{(354-356.6)^2 + (368-356.6)^2 + (329-356.6)^2 + \ldots + (372-356.6)^2}{29}$$

$$\hat{\gamma}_0 = 416.98$$

Sample Autocorrelation at lag 1

We now calculate the sample autocovariance at lag 1. By definition,

Autocovariance at lag 1

$$\hat{\gamma}_1 = COV(Y_t, Y_{t-1}) \qquad \textbf{10.23}$$

Equivalently,

$$\hat{\gamma}_1 = \frac{\sum_{2}^{30}(Y_t - \bar{Y})(Y_{t-1} - \bar{Y})}{n-1}$$

In this example,

$$\hat{\gamma}_1 = \frac{\sum_{2}^{30}(Y_t - 356.6)(Y_{t-1} - 356.6)}{28}$$

Expanded, this means,

$$\hat{\gamma}_1 = \frac{(Y_2 - 356.6)(Y_1 - 356.6) + (Y_3 - 356.6)(Y_2 - 356.6) + \ldots + (Y_{30} - 356.6)(Y_{29} - 356.6)}{28}$$

Note that because there are now 29 pairs of numbers $n = 29$, so $n - 1 = 28$.

$$\hat{\gamma}_1 = 81.31$$

Then, because $\hat{\gamma}_0 = 416.98$ and $\hat{\gamma}_1 = 81.31$, by definition

$$\hat{\rho}_1 = \frac{\hat{\gamma}_1}{\gamma_0}$$

Autocorrelation at lag 1

$$\hat{\rho}_1 = \frac{81.31}{416.98} = .195$$

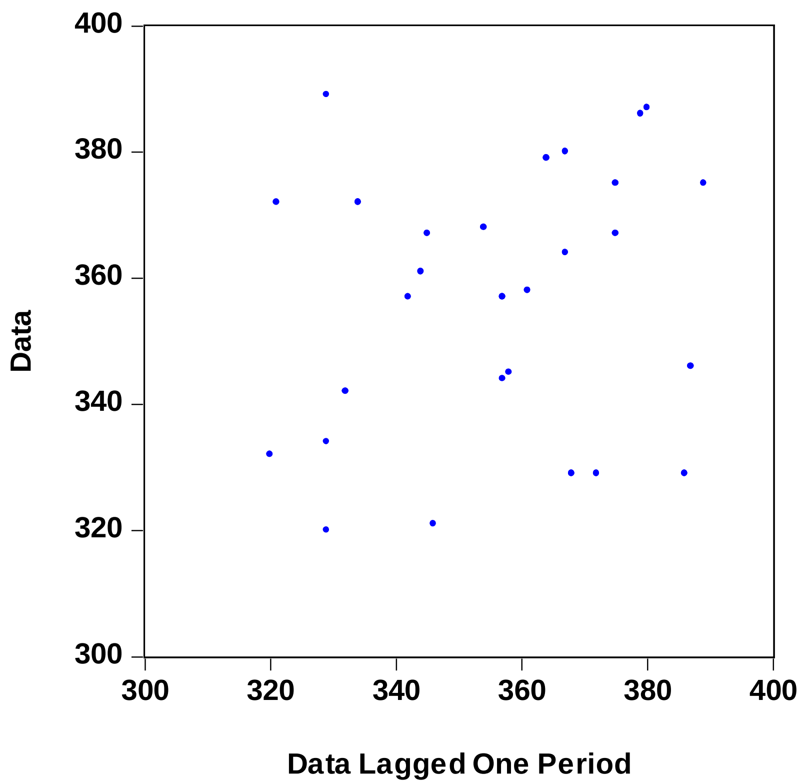
It is often helpful to think of the autocovariance at lag 1 in the following manner.

We consider the original series, Y_t, and then the original series, lagged, or shifted, by one period.

Table 10.5

Period	*Original Series* Y_t	*Original Series, Lagged by 1 period* Y_{t-1}
1	*354*	*
2	*368*	*354*
3	*329*	*368*
4	*389*	*329*
5	*375*	*389*
6	*375*	*375*
⋮	⋮	*375*
		⋮

We have created a set of bivariate data, (Y_t,Y_{t-1}), and so we can graph that data.



The autocovariance and the autocorrelation at lag 1 are the covariance and correlation of that set of bivariate data.

Sample Autocovariance and Autocorrelation at lag 2

We now calculate the sample autocovariance at lag 2. By definition,

Autocovariance at lag 2

$\hat{\gamma}_2 = COV(Y_t, Y_{t-2})$	**10.24**

Equivalently,

$$\hat{\gamma}_2 = \frac{\sum(Y_t - \bar{Y})(Y_{t-2} - \bar{Y})}{n-1}$$

In this example,

$$\hat{\gamma}_2 = \frac{\sum(Y_t - 356.6)(Y_{t-2} - 356.6)}{27}$$

Expanded, this means,

$$\hat{\gamma}_2 = \frac{(Y_3 - 356.6)(Y_1 - 356.6) + \ldots + (Y_{30} - 356.6)(Y_{28} - 356.6)}{27}$$

Note that $n = 28$, so $n - 1 = 27$, because there are now 28 pairs of numbers

$$\hat{\gamma}_2 = -22.52$$

Then, because $\hat{\gamma}_0 = 416.98$ and $\hat{\gamma}_2 = 22.5$, by definition

$$\hat{\rho}_2 = \frac{\hat{\gamma}_2}{\gamma_0}$$

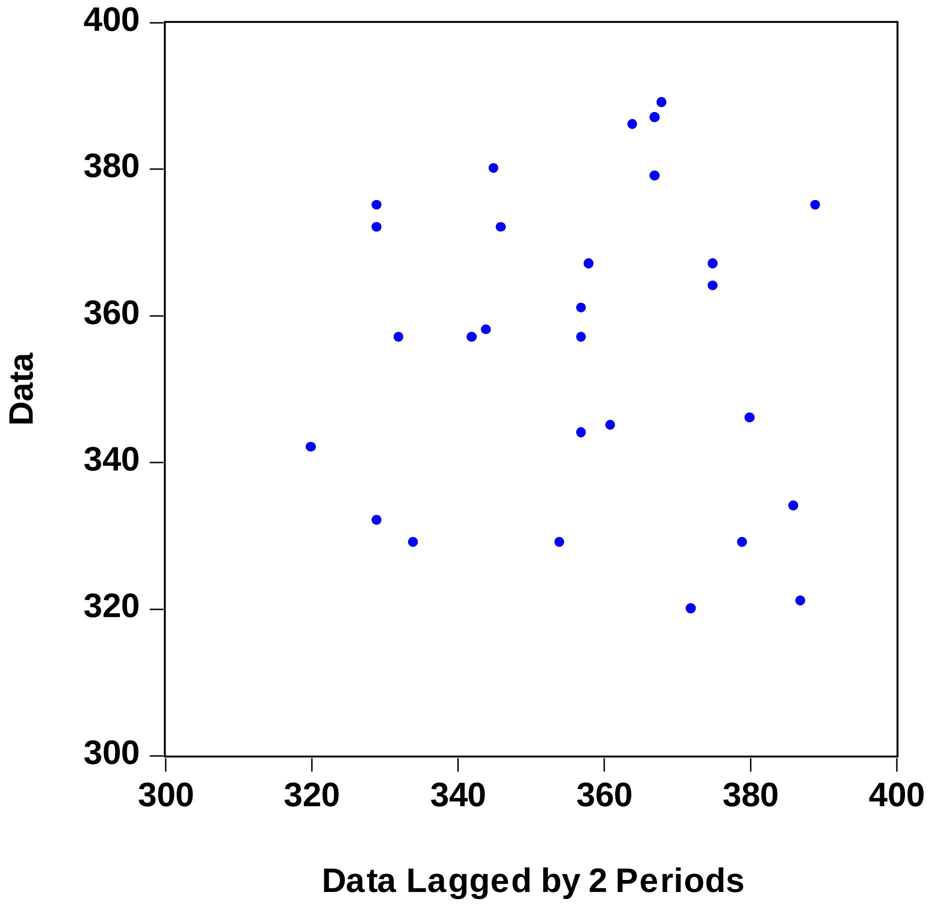
Autocorrelation at lag 2

$$\hat{\rho}_2 = \frac{22.52}{416.98} = .054$$

As before we can also think of the original series having been lagged by 2 periods, now creating a set of bivariate data.

Table 10.6

Period	*Original Series*	*Original Series, Lagged by 2 periods*
1	354	*
2	368	*
3	329	354
4	389	368
5	375	329
6	375	389
.	.	375
.	.	375
.	.	.



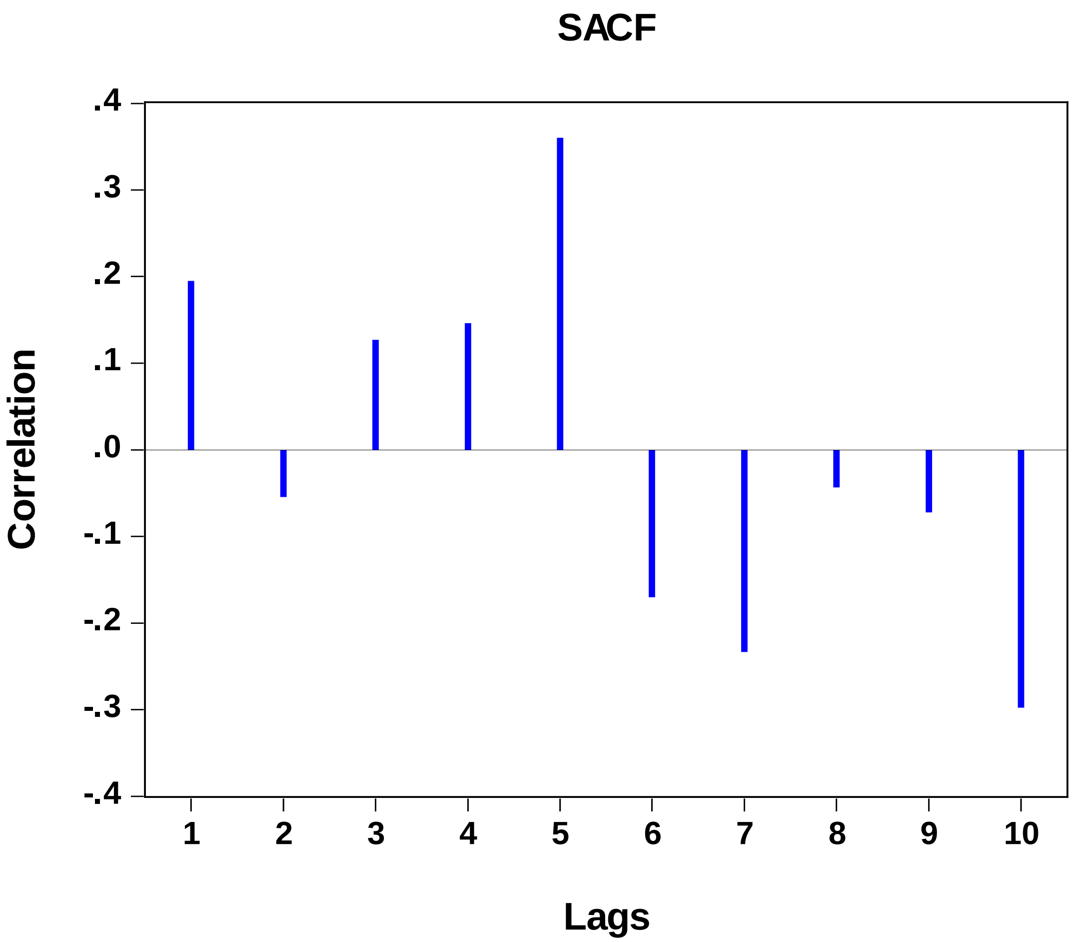
We calculate up to 10 lags, the sample autocorrelation function,

Table 10.7

SACF

Lags	1	2	3	4	5	6	7	8	9	10
	.195	-.054	.127	.146	.036	-.170	-.233	-.043	-.072	-.297

We can also graph the *SACF*, usually called a *correlogram.*



Interpreting the *SACF*

The *SACF* provides the autocorrelations for each lag. We interpret this as the "memory" of the series of past periods. For example, because $\hat{\rho}_1 = .195$, this tells us that at 1 period back, the series has about 20% memory. In other words, the serial correlation is about 20% between 1 period.

Because $\hat{\rho}_2 = -0.054$, this tells us that at 2 periods back, the series has about -5% memory. In other words, the serial correlation is about 5% between 2 periods.

If the series has a very short "memory" then $\hat{\rho}_1, \hat{\rho}_2, \ldots$ will decrease rapidly to zero. If the time series has a long memory, the autocorrelations will decrease at a slower rate.

The Variance of $\hat{\rho}_k$

$\hat{\rho}_k$ will vary depending on the sample data from the time series.

Variance of $\hat{\rho}_k$

$$\boldsymbol{VAR(\hat{\rho}_k) = \frac{1}{n}} \qquad \textbf{10.25}$$

Hence, the standard deviation is

$\boldsymbol{s_{\hat{\rho}_k}}$

$$\boldsymbol{s_{\hat{\rho}_k} = \sqrt{VAR(\hat{\rho}_k)} = \frac{1}{\sqrt{n}}} \qquad \textbf{10.26}$$

An approximate 95% confidence interval around 0 is then

$$s_{\hat{\rho}_k} = \frac{\pm 2}{\sqrt{n}}$$

In the example we have been considering, $n = 30$, so a 95% confidence interval is

$$s_{\hat{\rho}_k} = \frac{\pm 2}{\sqrt{30}} = \pm .365$$

Recopying the *ACF* of *Actual*, and this time placing confidence limits on the correlogram, we see that none of the autocorrelations at any lags are significantly different from 0.

Introduction to Box-Jenkins Models

Having analyzed the autocorrelation structure of time series data in the previous 4 sections, we would like to identify now the appropriate time series model given the autocorrelation structure of the data. It is the class of Box-Jenkins models that have been developed to fully use the information of the autocorrelation structure. The remainder of this chapter is thus an introduction to this class of models call Box-Jenkins (or *ARIMA)* models. In Chapter 6 we show how to use the sample autocorrelation function (the *SACF)* to help determine which Box-Jenkins model is the most appropriate.

The Backshift Operator

Because much of our notation deals with variables like

$Y_{t-1}, Y_{t-2}, \ldots$

$e_{t-1}, e_{t-2}, \ldots$

we need a notational device, the ***Backshift Operator***, denoted ***B***, to shift an observation one period back.

For example, if we have Y_{17}, then Y_{16} , the previous observation, one period back, is written in backshift notation.

$$\boldsymbol{B}Y_{17} = Y_{16}$$

B applied to Y_{17} shifted back to Y_{16}.

In general,

The Backshift Operator, B

$\boldsymbol{BY_t = Y_{t-1}}$	**10.27**

The Backshift Operator, B

$\boldsymbol{Be_t = e_{t-1}}$	**10.28**

Notice, then, that

$$\boldsymbol{B}(\boldsymbol{B}Y_t) = Y_{t-2}$$

B applied twice shifts Y_t back two periods.

$$\begin{aligned} \boldsymbol{B}(\boldsymbol{B}Y_t) = \boldsymbol{B}^2(Y_t) &= Y_{t-2} \\ \boldsymbol{B}^3(Y_t) &= Y_{t-3} \\ &\vdots \\ \boldsymbol{B}^k(Y_t) &= Y_{t-k}. \end{aligned}$$

For example, the three-period moving average one-step ahead forecast from Chapter 4:

$$\hat{Y}_t(1) = \frac{Y_t + Y_{t-1} + Y_{t-2}}{3}$$

$$\hat{Y}_t(1) = \tfrac{1}{3}Y_t + \tfrac{1}{3}Y_{t-1} + \tfrac{1}{3}Y_{t-2}$$

Using backshift notation

$$\hat{Y}_t(1) = \tfrac{1}{3}Y_t + \tfrac{1}{3}BY_t + \tfrac{1}{3}B^2Y_t$$

$$\hat{Y}_t(1) = \left(\tfrac{1}{3} + \tfrac{1}{3}B + \tfrac{1}{3}B^2\right)Y_t$$

Autoregressive (*AR*) Time Series Models

Abbreviated *AR*, an $AR(1)$ model is of the form

$$Y_t = \beta_0 + \beta_1 Y_{t-1}$$

In time series methods we adjust the notation slightly to

$$Y_t = \delta + \phi_1 Y_{t-1}$$

An *AR* process:

***AR*(*1*) Process**

$$\boldsymbol{Y_t = \delta + \phi_1 Y_{t-1} + \epsilon_t} \qquad \textbf{10.29}$$

(δ = "delta", ϕ_1 = "phi-one")

In backshift notation,

$$Y_t = \delta + \phi_1 B Y_t + \epsilon_t$$

Or,

$$Y_t - \phi_1 B Y_t = \delta + \epsilon_t$$

AR(1) $\quad (1 - \phi_1 B) Y_t = \delta + \epsilon_t$

***AR*(*2*) Process**

$$\boldsymbol{Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t} \qquad \textbf{10.30}$$

In backshift notation,

$$Y_t = \delta + \phi_1 B Y_t + \phi_2 B^2 Y_t + \epsilon_t$$

Or,

$$Y_t - \phi_1 B Y_t - \phi_2 B^2 Y_t = \delta + \epsilon_t$$

AR(2) $\quad (1 - \phi_1 B - \phi_2 B^2) Y_t = \delta + \epsilon_t$

***AR*(*3*) Process**

$$\boldsymbol{Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t} \qquad \textbf{10.31}$$

In general, an *AR(p)* model is

***AR*(*p*) Model**

$$\boldsymbol{Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \ldots + \phi_p Y_{t-p} + \epsilon_t} \qquad \textbf{10.32}$$

Notice that *AR* models are generalizations of the moving average models discussed earlier in Chapter 4. For example, the three period moving average model we write as

$$Y_t = \tfrac{1}{3} Y_{t-1} + \tfrac{1}{3} Y_{t-2} + \tfrac{1}{3} Y_{t-3} + \epsilon_t$$

This is an *AR(3)* model with $\delta = 0$, $\phi_1 = \frac{1}{3}$, $\phi_2 = \frac{1}{3}$, and $\phi_3 = \frac{1}{3}$.

The weighted three period moving average model,

$$Y_t = .6 Y_{t-1} + .3 Y_{t-2} + .1 Y_{t-3} + \epsilon_t$$

is an *AR(3)* model with $\delta = 0$, $\phi_1 = .6$, $\phi_2 = .3$, and $\phi_3 = .1$.

Moving Average (*MA*) Models

Unfortunately, the name "moving average model" is misleading and confusing in this case. The following moving average models – the kinds used in the Box-Jenkins methodology – bear no relation to those discussed above. Moving average models are more similar to exponential smoothing.

A Box-Jenkins moving average model is a model in which the new forecast is based on the error of forecast of previous forecast or forecasts (similar to exponential smoothing).

For example, a moving average model at lag 1, *MA(1)*, is

$$Y_t = \mu + \alpha\epsilon_{t-1} + \epsilon_t$$

We adjust the notation slightly, we write $-\theta_1$ instead of α. The reason for the minus sign will be explained shortly.

MA*(*1*) *Process

$$\boldsymbol{Y_t = \mu - \theta_1\epsilon_{t-1} + \epsilon_t} \qquad \textbf{10.33}$$

In backshift notation,

$$Y_t = \mu - \theta_1 B\epsilon_t + \epsilon_t$$

Or,

MA(1) $$Y_t = \mu + (1 - \theta_1 B)\epsilon_t$$

An *MA(2)* process

MA*(*2*) *Process

$$\boldsymbol{Y_t = \mu - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} + \epsilon_t} \qquad \textbf{10.34}$$

An *MA(2)* process shows that an observation made at time *t*, Y_t is based on the mean of the process, μ, the error of forecast one period back, the error two periods back, and the random noise term.

In backshift notation, $$Y_t = \mu - \theta_1 B\epsilon_t - \theta_2 B^2\epsilon_t + \epsilon_t$$

MA(2) Or, $$Y_t = \mu + (1 - \theta_1 B - \theta_2 B^2)\epsilon_t$$

It may be apparent now why we use $-\theta$ as the *MA* coefficients. Let us compare an *MA*(*2*) equation with an *AR*(*2*) equation.

MA(2) $$Y_t = \mu + (1 - \theta_1 B - \theta_2 B^2)\epsilon_t$$

AR(2) $$(1 - \phi_1 B - \phi_2 B^2)Y_t = \delta + \epsilon_t$$

Each equation is now written with a polynomial expression using *B*. The *MA* equations have the expression using *B* on the right, and the *AR* equations have the expression on the left. By using $-\theta$ we have uniformity in the expressions.

MA*(*3*) *Process

$$\boldsymbol{Y_t = \mu - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \theta_3\epsilon_{t-3} + \epsilon_t} \qquad \textbf{10.35}$$

And so on.

The generalized moving average model at lag *q*, *MA(q)*

MA(q) Process

$$\boldsymbol{Y_t = \mu - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \theta_3\epsilon_{t-3} - \ldots - \epsilon_{t-q} + \epsilon_t} \quad \textbf{10.36}$$

Autoregressive Moving Average (*ARMA)* Processes

A time series Y_t may be a combination of autoregressive (*AR*) and moving average (*MA*) terms. These are also called ***mixed models*** or ***ARMA models***.

For example, a combination of an *AR(1)* and an *MA(1)*, denoted as an *ARMA(1, 1).*

ARMA(1,1)

$$\boldsymbol{Y_t = \delta + \phi_1 Y_{t-1} - \theta_1\epsilon_{t-1} + \epsilon_t} \quad \textbf{10.37}$$

(δ is a combination of the previous δ and the previous μ)

In backshift notation

ARMA(1,1)

$$Y_t = \delta + \phi_1 Y_{t-1} - \theta_1\epsilon_{t-1} + \epsilon_t$$

$$Y_t = \delta + \phi_1 BY_t - \theta_1 B\epsilon_t + \epsilon_t$$

$$Y_t - \phi_1 BY_t = \delta - \theta_1 B\epsilon_t + \epsilon_t$$

$$(1 - \phi_1 B)Y_t = \delta + (1 - \theta_1 B)\epsilon_t$$

Another example is a combination of an *AR(2)* and an *MA(1)*, denoted as an *ARMA(2, 1).*

ARMA(2,1)

$$\boldsymbol{Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} - \theta_1\epsilon_{t-1} + \epsilon_t} \quad \textbf{10.38}$$

In backshift notation

ARMA(2,1)

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} - \theta_1\epsilon_{t-1} + \epsilon_t$$

$$Y_t = \delta + \phi_1 BY_t + \phi_2 B^2 Y_t - \theta_1 B\epsilon_t + \epsilon_t$$

$$Y_t - \phi_1 BY_t - \phi_2 B^2 Y_t = \delta - \theta_1 B\epsilon_t + \epsilon_t$$

$$(1 - \phi_1 B - \phi_2 B^2)Y_t = \delta + (1 - \theta_1 B)\epsilon_t$$

And so on.

ARMA(1, 2)

$$Y_t = \delta + \phi_1 Y_{t-1} - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} + \epsilon_t$$

$$Y_t = \delta + \phi_1 BY_t - \theta_1 B\epsilon_t - \theta_2 B^2\epsilon_t + \epsilon_t$$

$$(1 - \phi_1 B)Y_t = \delta + (1 - \theta_1 B - \theta_2 B^2)\,\epsilon_t$$

ARMA(2, 2)

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} + \epsilon_t$$

$$Y_t = \delta + \phi_1 BY_t + \phi_2 B^2 Y_t - \theta_1 B\epsilon_t - \theta_2 B^2\epsilon_t + \epsilon_t$$

$$(1 - \phi_1 B - \phi_2 B^2)Y_t = \delta + (1 - \theta_1 B - \theta_2 B^2)\,\epsilon_t$$

The general notation for *ARMA(p, q)* is written

ARMA(p,q)

$$\boxed{Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_{t-p} Y_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} + \epsilon_t \qquad \textbf{10.39}}$$

Stationary Time Series and Achieving Stationarity

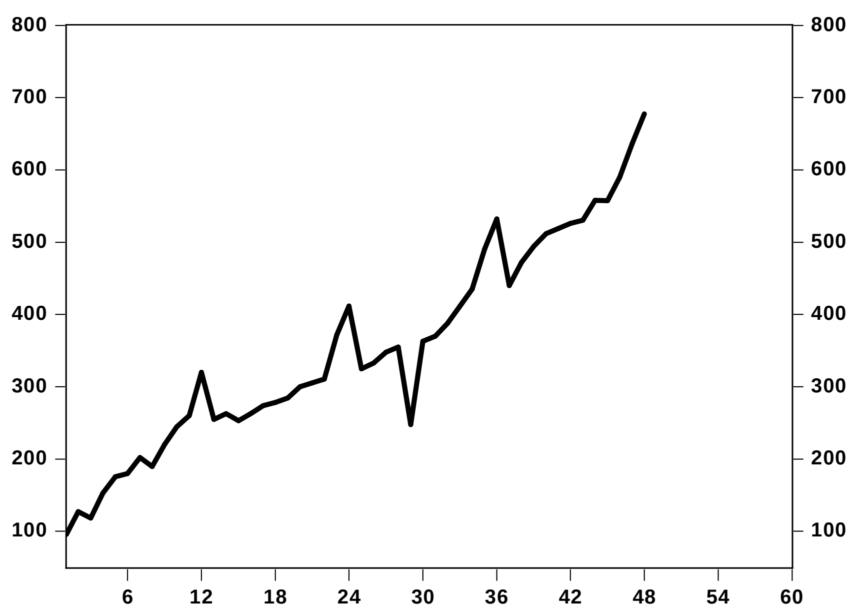
All of the *ARMA* models described above assume stationary data. Stationary data, roughly speaking, is "horizontal" data; data that has no trend up or down, no changing variance.

However, most business and economic time series are non-stationary in that the data series usually has some trend. To achieve stationarity, in most cases, we must at least remove the trend. For example, the time series of Sales Volume has major trend, so is not stationary.

Table 10.8

Four Years of Time Series Data

Period	Year 1	Period	Year 2	Period	Year 3	Period	Year 4
1	95.5	13	255.0	25	325.0	37	440.0
2	127.0	14	263.0	26	333.0	38	472.0
3	118.0	15	253.0	27	348.0	39	494.5
4	153.0	16	263.0	28	355.0	40	512.0
5	175.5	17	274.0	29	248.0	41	519.0
6	180.0	18	278.0	30	363.0	42	526.0
7	202.0	19	284.5	31	370.0	43	530.5
8	190.0	20	300.0	32	388.0	44	558.0
9	220.5	21	305.5	33	412.0	45	557.5
10	245.0	22	310.5	34	435.0	46	590.0
11	260.0	23	372.0	35	490.5	47	636.4
12	320.0	24	412.0	36	532.5	48	677.4



Removing Trend through Differencing

The usual method of removing trend from a time series is through ***differencing***. We reprint the first few observations of Sales Volume data to illustrate differencing.

Table 10.9

Period	$\boldsymbol{Y_t}$
1	*95.5*
2	*127.0*
3	*118.0*
4	*153.0*
5	*175.5*
$\vdots$	$\vdots$

The difference between Period 1 and Period 2, denoted Z_2, is

$$Z_2 = Y_2 - Y_1$$
$$Z_2 = 127.0 - 95.5 = 31.5$$

The difference between Period 2 and Period 3, denoted Z_3, is

$$Z_3 = Y_3 - Y_2$$
$$Z_3 = 118.0 - 127.0 = -.9$$

$$Z_4 = Y_4 - Y_3$$
$$Z_4 = 153.0 - 118.0 = 35.0$$
$$\vdots$$

In general terms,

First Order Differencing

$$\boldsymbol{Z_t = Y_t - Y_{t-1}} \qquad \textbf{10.40}$$

Using backshift notation:

$$Z_t = Y_t - BY_t$$

$$Z_t = (1 - B)Y_t$$

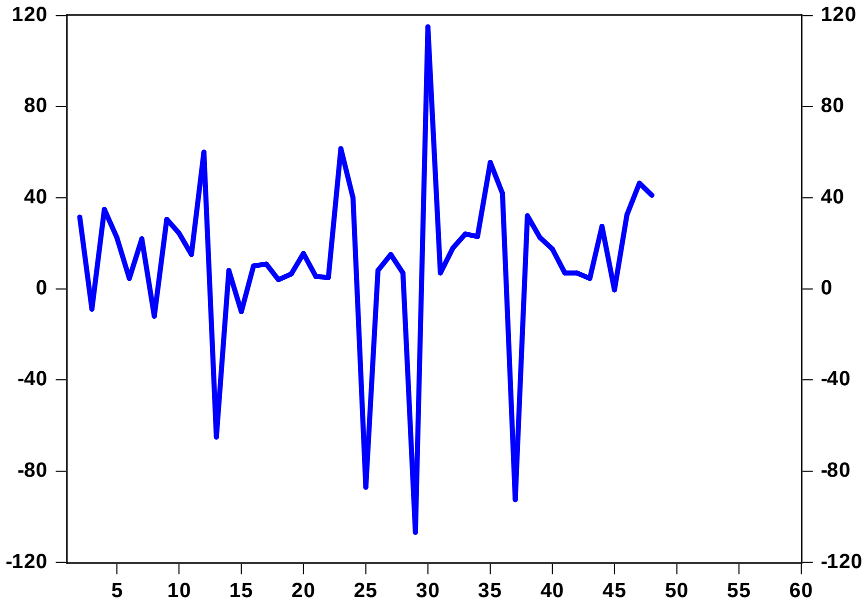
Thus Table 10.8 expands to

Table 10.10

Period	$\boldsymbol{Y_t}$	$\boldsymbol{Z_t}$
1	95.5	*
2	127.0	31.5
3	118.0	-9.0
4	153.0	35.0
5	175.5	22.5
$\vdots$	$\vdots$	$\vdots$

Figure 10.7 is a plot of the first difference of the data. Figure 10.7 has a much more constant mean, as in a stationary series.

Differencing of a Time Series



This type of differencing is called ***first order differencing***, or ***first difference at lag 1***. The ***second difference*** is the difference of the difference. (And what a difference!) Using the above example, the second difference of Y_t is the difference of Z_t. We denote the second difference by W_t, so

$$W_t = Z_t - Z_{t-1}$$

$$W_3 = Z_3 - Z_2 = -9 - (31.5) = -40.5$$

$$W_4 = Z_4 - Z_3 = 35 - (-9) = 44$$

Notice, that because

$$Z_t = Y_t - Y_{t-1},$$

$$Z_{t-1} = Y_{t-1} - Y_{t-2}$$

then,

$$W_t = Z_t - Z_{t-1}$$

$$W_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2})$$

$$W_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

Using backshift notation

$$W_t = Y_t - 2Y_{t-1} + Y_{t-2} = Y_t - 2BY_t + B^2Y_t$$

$$W_t = (1 - 2B + B^2)Y_t \qquad 10.41$$

$$W_t = (1 - B)(1 - B)Y_t \qquad 10.42$$

$$W_t = (1 - B)^2Y_t$$

Because the expression $(1 - B)$ has the effect of first differencing, $(1 - B)^2$ has the effect of second differencing. The third difference is $(1 - B)^3$, and so on.

Notice that there is an important distinction between

$(1 - B)^2 Y_t$

and $(1 - B^2) Y_t$

$(1 - B)^2 Y_t$ is the second difference of Y_t at lag 1.

Whereas $(1 - B^2) Y_t = Y_t - Y_{t-2}$

$Y_t - Y_{t-2}$ is the <u>first</u> difference of Y_t at <u>lag</u> <u>2.</u>

ARIMA models

If a time series requires differencing to achieve a constant mean then we integrate the differencing with the *ARMA* model. For example, suppose we have a time series with trend that requires first differencing to achieve stationarity. We construct the new differenced series Z_t, where

$$Z_t = Y_t - Y_{t-1}$$

$$Z_t = (1 - B) Y_t$$

We then determine that an *ARMA(*2,1) is an appropriate model for the Z_t series.

$$Z_t = \delta + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} - \theta_1 \epsilon_{t-1} + \epsilon_t \qquad 10.44$$

Hence,

$$Z_t - \phi_1 B Z_t - \phi_2 B^2 Z_t = \delta - \theta_1 B \epsilon_t + \epsilon_t$$

$$(1 - \phi_1 B - \phi_2 B^2) Z_t = \delta + (1 - \theta_1 B)\epsilon_t \qquad 10.45$$

But Z_t was a differenced series, so we substitute $(1 - B) Y_t$ in for Z_t in equation 5.17

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B) Y_t = \delta + (1 - \theta_1 B)\epsilon_t \qquad 10.46$$

We have integrated first differencing with an *ARMA(2,1)* model, we denote this by

ARIMA(2, 1, 1)

<u>AR</u>IMA(<u>2</u>,1,1)
AR(2)

AR<u>I</u>MA(2,<u>1</u>,1)
<u>I</u>ntegrated with <u>1st</u> difference

ARI<u>MA</u>(2,1,<u>1</u>)
MA(1)

In general, we write

ARIMA(p,d,q)	**10.47**

meaning an ***AR of order p, differencing of order d,*** and an ***MA of order q.***

So, for example, an *ARIMA(1,2,2)* means

AR(1), 2nd differencing, and *MA(2)*

In backshift notation,

ARIMA(1,2,2)

$$(1 - \phi_1 B)(1 - B)^2 Y_t = \delta + (1 - \theta_1 B - \theta_2 B^2)\epsilon_t$$

Expanded

$$(1 - \phi_1 B)(1 - 2B + B^2)Y_t = \delta + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

$$\Big(1 - (2 + \phi_1)B + (1 + 2\phi_1)B^2 - \phi_1 B^3\Big)Y_t = \delta + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

ARIMA(1, 2, 2)

$$Y_t = \delta + (2 + \phi_1)Y_{t-1} - (1 + 2\phi_1)Y_{t-2} + \phi_1 Y_{t-3} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} + \epsilon_t$$

As another example of an *ARIMA* model let us consider *ARIMA(0,1,1)* .

*1*st differencing and an *MA(1)*.

In backshift notation,

ARIMA(0, 1, 1)

$$(1 - B)Y_t = (1 - \theta_1)\epsilon_t$$

Expanded

$$Y_t = Y_{t-1} - \theta_1 \epsilon_{t-1} + \epsilon_t \qquad 10.48$$

Now a one-step ahead forecast for an *ARIMA*(*0,1,1*) model is

$$\hat{Y}_t(1) = Y_t - \theta_1 e_t \qquad 10.49$$

Recall that $e_t = Y_t - \hat{Y}_{t-1}(1)$

So by substitution equation 5.49 becomes

$$\hat{Y}_t(1) = Y_t - \theta_1\Big(Y_t - \hat{Y}_{t-1}(1)\Big) \qquad 10.50$$

Now if we let $\alpha = 1 - \theta_1$, then equation 5.50 becomes

$$\hat{Y}_t(1) = \alpha Y_t + (1 - \alpha)\hat{Y}_{t-1}(1) \qquad 10.51$$

Equation 5.51 is identical to equation 4. [See page of Chapter 4] when exponential smoothing was first defined. In other words, an *ARIMA(0,1,1)* is equivalent to *single exponential smoothing*. This is one example of the significant generality of *ARIMA* models.

ARIMA(p,d,q) models were first codified and published in the literature by two British statisticians, George E. P. Box, and Gwilym M. Jenkins. Specifically, their book ***Time Series Analysis: Forecasting and Control***, published in the late 1960's provided the first organized treatment of *ARIMA* models. Consequently, the term "*Box-Jenkins Models"* and "*ARIMA Models"* are synonymous. Box is currently a professor of statistics at the University of Wisconsin, and sadly, Jenkins died rather young in 1981.

Problems and Questions

10.1

Listed below is the constant mean time series from Chapter 6, Problem 6.1, of 24 observations. Plot the data and determine the mean and variance of the the data set.

Period	Observation	Period	Observation
1	95	13	118
2	100	14	86
3	87	15	86
4	123	16	112
5	90	17	85
6	96	18	101
7	75	19	135
8	78	20	120
9	106	21	76
10	104	22	115
11	89	23	90
12	83	24	92

Remember that we shall denote the variance of the series by γ_0.

10.2

Using the series of Problem 10.1, lag the series by one period, i.e., create Y_{t-1}, and print that alongside the original series, Y_t. Then determine the covariance and correlation between Y_t and Y_{t-1}. This is the autocovariance and the autocorrelation at lag 1, denoted γ_1.

10.3

Using the series of Problem 10.1, lag the series by two periods, creating Y_{t-2}, and print that alongside the original series, Y_t. Then determine the covariance and correlation between Y_t and Y_{t-2}. This is the autocovariance and the autocorrelation at lag 2, denoted γ_2.

10.4

Repeat the process of Problems 10.2 and 10.3 up to 8 lags, so that γ_1, γ_2,...,γ_8 , ρ_1, ρ_2,...,ρ_8 are all determined.

Then create the *SACF* or correlogram up to 8 lags.

10.5

With computer software determine the *SACF to LAG 8* of the data series.

10.6

Determine the variance and standard deviation of the autocorrelations. Are any of the "spikes" significant? Explain.

For each of the models listed below

10.7

Write the model using backshift notation.

10.8

Write the model in expanded notation.

a. *AR*(2)
b. *MA*(3)
c. *ARMA*(1,1)
d. *ARMA*(2,0)
e. *ARMA*(1,2)

f. *ARMA*(2,2)
g. *ARIMA*(1,1,0)
h. *ARIMA*(0,1,1)
i. *ARIMA*(1,0,2)
j. *ARIMA*(1,1,1)
h. *ARIMA*(2,1,2)
i. *ARIMA*(1,2,1)