Chapter 18: Time Series

18.1 *Stationary Data Series*

In this chapter we consider a series of observation taken from a single entity over time much as we assumed in Section 17.5. The entity generating the data might be a particular company, Web site, household, market, geographic region or anything else that maintains a fixed identity over time. Our observations look like y_1 , y_2, ···, y_n with a joint density $\Pr(y_1 , y_2, \cdots, y_n)$. When data are collected over time, there is a very important concept that is called *stationarity* and in fact the concept shows up in other places in this book, notably Equation (15.1). For our purposes, we define the stationarity of a time series as

$$\Pr(y_t , y_{t+1}, \cdots, y_{t+k}) = \Pr(y_{t+m} , y_{t+m+1}, \cdots, y_{t+m+k}), \tag{18.1}$$

for all t, j and k. Given that, it must be the case also that for $m = \pm1, \pm2, \cdots$

$$\Pr(y_t) = \Pr(y_{t+m})$$

which then further implies that

$$E(y_t) = E(y_{t+m})$$

and

$$V(y_t) = V(y_{t+m}).$$

Presumably under stationarity it is the case as well that

$$\Pr(y_t , y_{t+1}) = \Pr(y_{t+m} , y_{t+m+1}) \tag{18.2}$$

which would then make obvious the notion that

$$\mathrm{Cov}(y_t , y_{t+1}) = \mathrm{Cov}(y_{t+m} , y_{t+m+1}) = \gamma_1.$$

In general, since

$$\Pr(y_t , y_{t+j}) = \Pr(y_{t+m} , y_{t+m+j}) \tag{18.3}$$

the following is implied

$$\mathrm{Cov}(y_t , y_{t+j}) = \mathrm{Cov}(y_{t+m} , y_{t+m+j}) = \gamma_j.$$

The parameter γ_j is known as the *autocovariance* at lag j. Putting all of these results together, we can say that

$$E\begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{bmatrix} = E(\mathbf{y}) = \begin{bmatrix} \mu \\ \mu \\ \cdots \\ \mu \end{bmatrix}$$

and

$$V(\mathbf{y}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix}$$

Like all covariance matrices, V(**y**) is symmetric. If $E(y_t)$ does not depend on t, which it should not with a stationary series, then we would ordinarily expect to find the series in the neighborhood of μ. History tends to repeat itself, probabilistically. By the definition of covariance [Equation (4.7)]:

$$\gamma_j = E[(y_t - \mu)(y_{t+j} - \mu)].$$

If $\gamma_j > 0$ we would expect that a higher than usual observation would be followed by another higher than usual observation. We can standardize the covariances by defining the autocorrelation,

$$\rho_j = \frac{\gamma_j}{\sqrt{\gamma_0}\sqrt{\gamma_0}} = \frac{\gamma_j}{\gamma_0}.$$

As usual, $\rho_0 = 1$. The structure of the autocorrelations will greatly help us in understating the behavior of the series, **y**.

18.2 *A Linear Model for Time Series*

The time series models that we will be covering are called *discrete linear stochastic processes* and are of the form

$$y_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots . \tag{18.4}$$

In effect, an observation within the series is conceptualized as being the result of a possibly linear combination of random inputs. The e_t values are assumed identically distributed with

$$E(e_t) = 0 \text{ and}$$

$$V(e_t) = \sigma_e^2.$$

Further, we will assume that

$$\mathrm{Cov}(e_t, e_{t+j}) = 0 \tag{18.5}$$

for all $j \neq 0$. These e_t values are independent inputs and are often called *white noise*. We also assume that

$$\sum_{i=0}^{\infty} \psi_i = c \text{ and that}$$

$$\psi_0 = 1.$$

Given the preceding long list of notation and assumptions, what is the expectation and variance of our data? As was pointed out before, it is still the case the $E(y_t) = \mu$ since we can combine Equation (18.4) and the assumption that $E(e_t) = 0$. As for the variance of $V(y_t)$,

$$V(y_t) = E(y_t - \mu)^2$$

$$= E(\mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots - \mu)^2 \tag{18.6}$$

where the two μ's will just cancel. Squaring the remaining terms, we can collect them into two sets:

$$V(y_t) = E(e_t^2 + \psi_1^2 e_{t-1}^2 + \psi_2^2 e_{t-2}^e + \cdots) + E(\text{all cross terms}).$$

We can quickly dispense of all the cross terms from Equation (18.6) because, by assumption [Equation (18.5)] the e_t are independent. Worrying just about the first part of the above equation, and noting that the expectation of a sum is equal to the sum of the expectation[Equation (4.4)], we can then say that

$$V(y_t) = \sigma_e^2 \sum_{i=0}^{\infty} \psi_i^2. \tag{18.7}$$

Are you game for figuring out the covariance at lag j of two data points from the series? Here goes. We note that the covariance between y_t and y_{t-j} is $E[(y_t - \mu)(y_{t-j} - \mu)]$. Once again, all values of μ will cancel leaving us with

$$\gamma_j = E[(e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots)(e_{t-j} + \psi_1 e_{t-j-1} + \psi_2 e_{t-j-2} + \cdots)]$$

$$= E[(\psi_j e_{t-j}^2) + (\psi_{j+1}\psi_1 e_{t-j-1}^2) + (\psi_{j+2}\psi_2 e_{t-j-2}^2) + \cdots] + E(\text{all cross terms}).$$

In this case, E(all cross terms) refers to any term involving $E(e_t, e_{t-m})$ for $m \neq 0$ and once again, with independent e_t all such covariances vanish. That leaves us with the very manageable Equation (18.8)

$$\begin{aligned} \gamma_j &= \sigma_e^2(\psi_j + \psi_{j+1}\psi_1 + \psi_{j+1}\psi_2 + \cdots) \\ &= \sigma_e^2 \sum_{i}^{\infty} \psi_i \psi_{i+j} \end{aligned} \tag{18.8}$$

Neither the variance in Equation (18.7) nor the covariances in Equation (18.8) can exist unless the infinite sum in those two equations is equal to a finite value. That an infinite series can be finite is seen in the reasoning that runs between Equation (15.17) and (15.17). We will return to this concept momentarily, but first we will assume that $\psi_i = \phi^i$, with $|\phi| < 1$. Then

$$y_t = \mu + e_t + \phi e_{t-1} \ \phi^2 e_{t-2} + \cdots .$$

It can be shown that

$$\sum_{i=0}^{\infty} \psi_i = \sum_{i=0}^{\infty} \phi^i = \frac{1}{1-\phi}.$$

That this is so can be seen by defining s = $\sum_{i=0}^{\infty} \phi^i = 1 + \phi + \phi^2 + \phi^3 + \cdots$, and then multiplying by ϕ so that ϕs - s = 1. Solving for s leads to the result, s = $1/1-\phi$. Combining this result with Equation (18.7), y_t then has a variance of

$$\gamma_0 = \frac{\sigma_e^2}{1-\phi^2}$$

and from Equation (18.8), autocovariances of

$$\gamma_j = \frac{\sigma_e^2 \phi^j}{1-\phi^2}.$$

Needless to say, this will only work for with $|\phi| < 1$, as otherwise, the variance will blow up. If ϕ = 1 our model becomes

$$\begin{aligned} y_t &= \mu + e_t + e_{t-1} + e_{t-2} + \cdots \\ &= \mu + e_{t-1} + e_{t-2} + \cdots + e_t \\ &= y_{t-1} + e_t \end{aligned}$$

and so forth, as we could now substitute for y_{t-1} above. Obviously, the variance of a series with ϕ = 1 blows up.

18.3 *Moving Average Processes*

A moving average model is characterized by a finite number of non-zero values ψ_i with $\psi_i = 0$ for i > q. The model will then look like the following,

$$y_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots + \psi_q e_{t-q}.$$

The tradition in this area calls for us to modify the notation somewhat and utilize $\theta_i = -\psi_i$ which then modifies the look of the model slightly to

$$y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}.$$

Such as model is often called a *Moving Average (q) process*, or MA(q) for short. As an example, consider the MA(1):

$$y_t = \mu + e_t - \theta_1 e_{t-1}$$

which can also be written with the *Backshift operator*, symbolized with the letter B and presented also in Equation (17.23):

$$y_t = \mu + (1 - \theta_1 B)e_t,$$

i. e.

$$Be_t = e_{t-1}, \tag{18.9}$$

$$B \cdot (Be_t) = B^2 e_t = e_{t-1} \text{ and} \tag{18.10}$$

$$B^0 e_t = e_t. \tag{18.11}$$

We will have much cause to use the backshift operator in this chapter. For now, it will be interesting to look at the autocovariances of the MA(1) model. These will be

$$\gamma_1 = E[(e_t - \theta_1 e_{t-1})(e_{t-1} - \theta_1 e_{t-2})]$$

$$= \sigma_e^2(-\theta_1).$$

OK, that's a nice result. What about the autocovariance at lag 2?

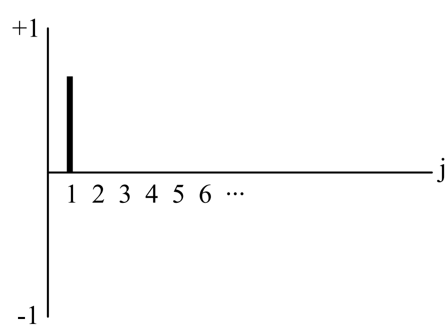
$$\gamma_2 = E[(e_t - \theta_1 e_{t-1})(e_{t-2} - \theta_1 e_{t-3})]$$

$$= 0.$$

Since none of the errors overlap with the same subscript, everything vanishes as the errors are assumed independent. Thus we note that for the MA(1),

$$\gamma_j = \begin{cases} -\sigma_e^2\theta_1 & \text{for } j = 1 \\ 0 & \text{for } j > 1 \end{cases}.$$

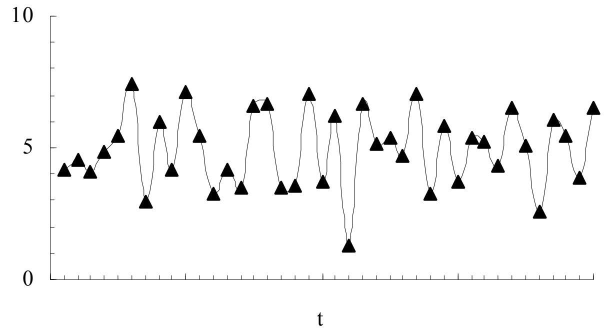
We can plot the autocorrelation function, which plots the value of the autocorrelations at various lags, j. In the case of the MA(1), the theoretical pattern is unmistakable:



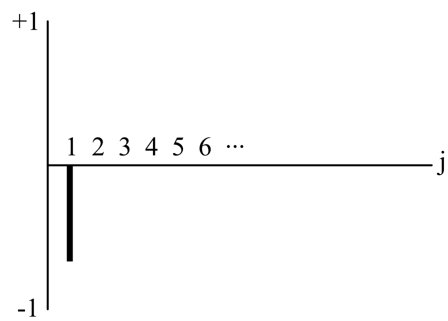
As we will see later in the chapter, the *correlogram*, as a diagram such as the one above is called, is an important mechanism to identify the underlying structure of a time series. For the sake of curiosity, it will be nice to look at a simulated MA(1) process with $\theta_1 = -.9$ and $\mu = 5$. The model would be

$$y_t = e_t + .9e_{t-1}$$

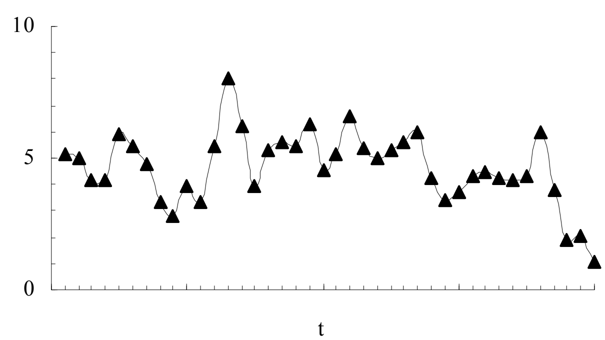
and $\sigma_e^2 = 1$, $\gamma_0 = \sigma_e^2(1+\theta_1^2) = 1.81$, $\gamma_1 = -\sigma_e^2(\theta_1) = .9$, $\rho_1 = \gamma_1/\gamma_0 = .5$ and $\rho_j = 0$ for all j > 1. An example of this MA(1) process, produced using a random number generator is shown below:



If $\theta_1 = +.9$ so that $\rho_1 = -.5$ the correlogram would appear as



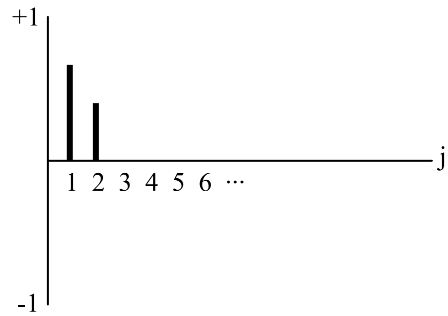
with the spike heading off in the negative, rather than the positive direction. The plot of the time series would by more jagged, since a positive value of y_t would tend to be associated with a negative value of y_{t-1}.



For an arbitrary value of q, an MA(q) process will have autocovariances

$$\gamma_j = \begin{cases} -\sigma_e^2(-\theta_j + \theta_1\theta_{j+1} + \cdots + \theta_{q-j}\theta_q & \text{for } j = 1,\ 2,\ \cdots,\ q \\ 0 & \text{for } j > q. \end{cases}$$

For example the MA(2) process will have a correlogram that has two *spikes*:



18.4 *Autoregressive Processes*

Recall that any discrete linear stochastic process can be expressed as

$$y_t = \mu + e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \cdots$$

as was Equation (18.4). Needless to say this implies that we can express the errors as

$$e_t = y_t - \mu - \psi_1 e_{t-1} - \psi_2 e_{t-2} - \cdots .$$

Our assumption of stationarity requires that the same basic model that holds for e_t must hold true for e_{t-1} which would then be

$$e_{t-1} = y_{t-1} - \mu - \psi_1 e_{t-2} - \psi_2 e_{t-3} - \cdots .$$

If we substitute the model for e_{t-1} into the model for y_t we get

$$y_t = \mu + e_t + \psi_1[y_{t-1} - \mu - \psi_1 e_{t-2} - \cdots] + \psi_2 e_{t-2} + \cdots$$

$$= \mu(1 - \psi_1) + e_t + \psi_1 y_{t-1} + (\psi_2 - \psi_1^2)e_{t-2} + \cdots$$

You can keep doing this - now we substitute an expression for e_{t-2} and so forth until all the e_t terms are banished and all that remains are y_t values, with various coefficients. Arbitrarily naming these coefficients with the letter π, we get something that looks like

$$y_t = \pi_1 y_{t-1} + \pi_2 y_{t-2} + \cdots + \delta + e_t. \tag{18.12}$$

Our discrete linear stochastic process can be expressed as a possibly infinite series of past random disturbances [i. e. Equation (18.4)]. If the series is finite, we call it an MA process. Any discrete linear stochastic process can also be expressed as a possibly infinite series of its own past values disturbances [i. e. Equation (18.12)]. If the series is finite, we will call it an *autoregressive* *process*, also known as an *AR* process. This is illustrated below, where we have modified Equation (18.12) by assuming that $\pi_i = 0$ for i > p:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \delta + e_t.$$

To the paragraph above, I would add that a finite AR is equivalent to an infinite MA and a finite MA is equivalent to an infinite AR. Below we will prove the first of these two assertions. But before we do that, it should be noted that all of this gives the data analyst a lot of flexibility in creating a parsimonious model.

The AR(1) model looks like

$$y_t = \phi_1 y_{t-1} + \delta + e_t \tag{18.13}$$

$$(1 - \phi_1 B) y_t = \delta + e_t \tag{18.14}$$

If we take Equation (18.13) and substitute the equivalent expression for y_{t-1}, we have

$$y_t = \phi_1 [\phi_1 y_{t-2} + \delta + e_{t-1}] + \delta + e_t$$

and then again

$$y_t = \phi_1 [\phi_1 ([\phi_1 y_{t-3} + \delta + e_{t-2}) + \delta + e_{t-1}] + \delta + e_t$$

and so on until we see that we end up with

$$y_t = \frac{\delta}{1-\phi_1} + e_t + \phi_1 e_{t-1} + \phi_1^2 e_{t-2} + \phi_1^3 e_{t-3} + \cdots$$

which is an infinite MA process. As claimed, an AR(1) leads to an infinite MA.

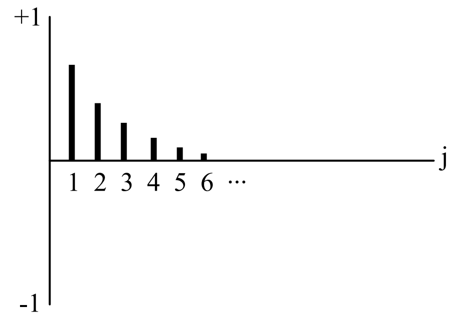
What are the moments of an AR(1) process? We have

$$E(y_t) = \frac{\delta}{1-\phi_1},$$

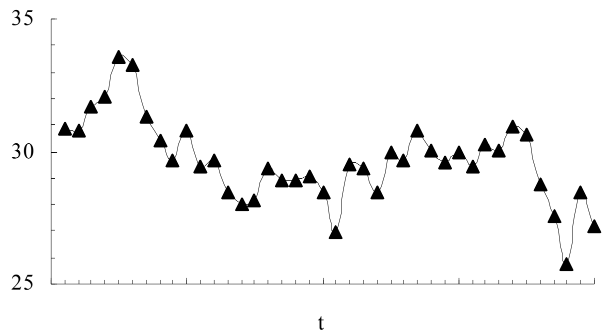
$$\gamma_j = \phi_1^j \frac{\sigma_e^2}{1-\phi_1^2} \quad \text{and}$$

$$\rho_j = \phi_1^j.$$

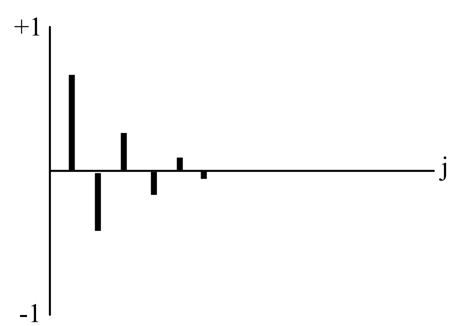
For the AR(1), the autocorrelations decline exponentially. An idealized correlogram is shown below:



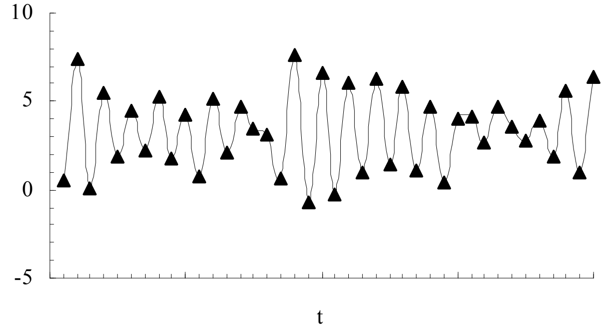
The autocorrelations damp out slowly. Next we show a random realization of the AR(1) model $y_t = .8y_{t-1} + 6 + e_t$:



Another example is identical to the first, but the sign on ϕ_2 is reversed. The correlogram appears below



and then we see a random realization of the series:



18.5 *Details of the Algebra of the Backshift Operator*

One of the most beautiful aspects of time series analysis is the use of backshift notation. Say we have an AR(1) with parameter ϕ_1. We can express the model as

$$(1 - \phi_1 B)y_t = e_t + \delta.$$

Putting the model in reduced form we have

$$y_t = (1-\phi_1 B)^{-1}\delta + (1-\phi_1 B)^{-1} e_t.$$

But what does it mean to invert a function with "B" in it? It produces an infinite series. To see that, start with the basic fact that

$$(1-\phi_1 B)^{-1} = \frac{1}{1-\phi_1 B}.$$

So far so good. However, the series

$$s = \sum_{i=0}^{\infty} \phi_1^i B^i = 1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots$$

and the series

$$\phi_1 B \cdot s = \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots$$

differ by 1. Thus

$$s - \phi_1 B \cdot s = 1$$

and therefore

$$\begin{aligned} s = (1-\phi_1 B)^{-1} &= \frac{1}{1-\phi_1 B} \\ &= \sum_{i=0}^{\infty} \phi_1^i B^i = 1 + \phi_1 B + \phi_1^2 B^2 + \phi_1^3 B^3 + \cdots. \end{aligned} \tag{18.15}$$

Stationarity, and the need to avoid infinities in the infinite sum, require that

$$|\phi_1| < 1. \tag{18.16}$$

This is equivalent to saying that the root of that $1 - \phi_1 B = 0$ must lie outside the unit circle.

18.6 *The AR(2) Process*

The AR(2) model is

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \delta + e_t$$

$$y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} = \delta + e_t$$

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \delta + e_t$$

which is stationary if the roots of

$$1 - \phi_1 B - \phi_2 B^2 = 0$$

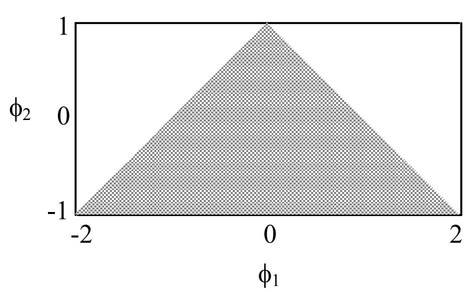
lie outside the unit circle, which is to say

$$\phi_1 + \phi_2 < 1, \tag{18.17}$$

$$\phi_1 - \phi_2 < 1 \text{ and} \tag{18.18}$$

$$|\phi_2| < 1. \tag{18.19}$$

Below, the graph shows the permissible region as a shaded triangle:



18.7 *The General AR(p) Process*

In general, an AR model of order p can be expressed as

$$(1-\phi_1 B-\phi_2 B^2-\phi_p B^p)y_t=\delta+e_t$$

$$\phi(B)y_t=\delta+e_t$$

$$y_t=\frac{\delta+e_t}{\phi(B)}.$$

Note that here we have introduced a new way of writing $1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$, namely to call it simply $\phi(B)$. The autocorrelations and the ϕ_i are related to each other via what are known as the *Yule-Walker Equations*:

$$\rho_1=\phi_1+\phi_2\rho_1+\cdots\phi_p\rho_{p-1}$$

$$\rho_2=\phi_1\rho_1+\phi_2+\cdots\phi_p\rho_{p-2}$$

$$\cdots=\cdots$$

$$\rho_p=\phi_1\rho_{p-1}+\phi_2\rho_{p-2}+\cdots\phi_p$$

which can be used to estimate $\hat{\phi}_j$ values.

18.8 *The ARMA(1,1) Mixed Process*

Consider the model

$$y_t=\delta+\phi_1 y_{t-1}+e_t-\theta_1 e_{t-1}$$

$$(1-\phi_1 B)y_t=\delta+(1-\theta_1 B)e_t.$$

Here we have both an autoregressive and a moving average component. The AR part results in an infinite MA model with

$$y_t=\frac{\delta}{1-\phi_1 B}+\frac{1-\theta_1 B}{1-\phi_1 B}e_t.$$

In compact notation we can say that $\psi(B) = \phi^{-1}(B) \cdot \theta(B)$. The MA part results in an infinite AR model with

$$\frac{1-\phi_1 B}{1-\theta_1 B}y_t-\frac{\delta}{1-\theta_1 B}=e_t.$$

Again we can compactify the notation noting that $\pi(B) = \phi(B) \cdot \theta^{-1}(B)$. Mixed models let you achieve parsimony as you can represent an infinite MA with a finite AR and vice versa. The situation that we have at hand can be graphed as follows:

$$e_t \longrightarrow \boxed{\psi(B)} \longrightarrow y_t$$

We conceptualize of our observed series of data as being driven by are series of random shocks, of random values or white noise inputs. These inputs are then passed through a filter with various properties and that eventually leads to an output, which consists of our data. Modeling the data requires that we come up with a parsimonious description, one with few model parameters, of the filter, i.e $\psi(B)$.

What stationarity is to the AR side, invertibility is to the MA side. Invertibility requires that the roots of

$$1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q = 0$$

lie outside of the unit circle.

18.9 *The ARIMA(1,1,1) Model*

A series may be relatively homogeneous, looking pretty much the same at all time periods, but it may end up being non-stationary simply because it shows no permanent affinity for a particular level or mean. Even though the original series of data may not be stationary, differences between successive observations may be stationary:

$$d_t = y_t - y_{t-1} = (1 - B)y_t.$$

Simply put, we can apply an ARMA model to the d_t. When we do so, this is called an ARIMA model with the middle I referring to the fact that it is integrated. If the first differences are not stationary, the second differences might be, i. e.

$$d'_t = d_t - d_{t-1} = (1 - B)(1 - B)y_t.$$

The ARIMA(1,1,1) process, with the middle number referring to the number of differences that are taken can be described as

$$d_t = \phi_1 d_{t-1} - \theta_1 e_{t-1} + e_t$$

$$y_t - y_{t-1} = \phi_1 (y_t - y_{t-1}) - \theta e_{t-1} + e_t$$

$$y_t = (1 + \phi_1) y_{t-1} - \phi_1 y_{t-2} - \theta_1 e_{t-1} + e_t.$$

Thus we see that the ARIMA(1,1,1) is an ARMA(2,1) where the first ARMA AR parameter is equal to $1 + \phi_1$ while the second ARMA(2,1) AR parameter is $-\phi_1$. These parameters violate the rules for stationarity in Equations (18.17), (18.18) and (18.19). Similarly, an ARIMA(0,1,1) process looks like

$$y_t = y_{t-1} + e_t - \theta_1 e_{t-1}$$

which violates the stationarity rule for an AR(1) [Equation (18.16)] right off the top since "ϕ_1" = 1!

Thus we see the importance of differencing the series first, if necessary, prior to fitting an ARMA model.

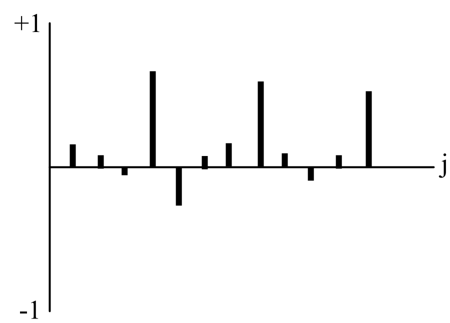
We can wrap up this section with another brief note about the backshift notation and the ARIMA(1,1,1) model. Such a model can be written quite elegantly as

$$(1 - \phi_1 B)(1 - B)y_t = \delta + (1 - \theta_1 B)e_t$$

In the model, the constant term δ implies that the average change will have the same sign as δ and the series will drift in the direction of the sign of δ.

18.10 *Seasonality*

Differencing, AR or MA parameters may be needed at various lags. For quarterly data, you may need to look at lags of 4, or for monthly data, lags of 12, which may occur whenever there are yearly patterns in data. For example, the following pattern seen in quarterly data:



may require that you difference the data at a lag of 4, i.e analyze $d_t = (1 - B^4)y^t$.

18.11 *Identifying ARIMA(p,d,q) Models*

In addition to the cues afforded in the autocorrelations, we can also look at what are known as the *partial autocorrelations*. For each lag j, you estimate ρ_j controlling for the first j - 1 values ρ_{j-1}, ρ_{j-2}, $\cdots$, ρ_1.

For a nonstationary process, the autocorrelations will be large at very long lags. On the other hand, over-differencing tends to produce an MA(1) with $\theta_1 = 1$.

For an AR process, the autocorrelations will decline exponentially. The partial autocorrelations will exhibit significant spikes at the first p lags.

For an MA process, the autocorrelations will exhibit significant spikes at the first q lags. The partial autocorrelations will exhibit exponential decline.

For a mixed process, the autocorrelations as well as the partial autocorrelations will decline exponentially.

It is generally a good idea to run the error of your model through the same diagnostic process to make sure that it is indeed acting like white noise. In effect, one adds a term to the ARIMA model, and then looks at the error to see if it is white noise yet. The process is repeated until the error is completely whitened.

References

Box, George E. P., Gwilym M. Jenkins (1976) *Time Series Analysis. Revised Edition*. Oakland, CA: Holden-Day.