Chapter 5: Autoregressive Integrated Moving Average(ARIMA) Models

Rong Fan

South Dakota State University

rong.fan@sdstate.edu

5.1 Introduction

Classical regression is often insufficient for explaining all of the interesting dynamics of a time series. We have seen a **structured error** terms for many regression models.

Exponential smoothing is established based on an assumption that the random noise is generated independently, which is often violated. That is successive observations may show serial dependence.

Indeed, the introduction of correlation leads to proposing the autoregressive (AR) and autoregressive moving average (ARMA) models. Adding nonstationary models to the mix leads to the autoregressive integrated moving average (ARIMA) model and seasonal ARIMA model, which were popularized in the great work by Box and Jenkins in 1970.

5.2 Linear Model for Stationary Time Series I

A linear filter is a linear operation from one time series x_t (which can be thought of as the input) to another time series y_t (which can be thought of as the output):

$$y_t = L(t) = \sum_{i=-\infty}^{\infty} \Psi_i x_{t-i} \tag{1}$$

where $t = ... - 2, -1, 0, 1, 2...$
The linear filter is a "process" which converts the input x_t into an output y_t, and that conversation is not instantaneous but involves all (present, past and future) values of the input in the form of a summation with different "weights", Ψ_i, on each x_i.

5.2 Linear Model for Stationary Time Series II

The linear filter (1) has the following properties:

1. **Time-invariant**: The coefficients Ψ_i do not depend on time.
2. **Physically realizable**: If $\Psi_i = 0$ for $i < 0$, it is physically realizable. That is the output y_t is a linear function of the current and past values of the input:

$$y_t = \Psi_{i=0} x_t + \Psi_1 x_{t-1} + \cdots .$$

3. **Stable**: If $\sum_{i=-\infty}^{\infty} |\Psi_i| < \infty$, it is stable.

Show that a moving average of span N, and the first difference are linear filters.

For example,

- When $\Psi_0 = \Psi_1 = \Psi_2 = 1/3$ and others are 0, we have $y_t = \frac{1}{3}(x_t + x_{t-1} + x_{t-2})$.
- When $\Psi_0 = -\Psi_1 = 1$ and others are 0, we have $y_t = x_t - x_{t-1}$.

5.2.1 Stationarity

Let us recall the (weak) stationarity. A time series is (weak) stationary if

1. The expected value does not depend on time.
2. The autocovariance for any lag k depends only on k. That is

$$\gamma(k) = Cov(y_t, y_{t+k}) = Cov(y_{t_1}, y_{t_1+k})$$

5.2.2 Stationary Time Series I

Let us recall the Backshift operator $\mathbf{B}$,

$$\mathbf{B}y_t = y_{t-1}, \quad \text{and} \quad \mathbf{B}^i y_t = y_{t-i}$$

thus a linear filter can be written as

$$y_t = \sum_{i=0}^{\infty} \Psi_i x_{t-i} = \sum_{i=0}^{\infty} \Psi_i \mathbf{B}^i x_t = \left(\sum_{i=0}^{\infty} \Psi_i \mathbf{B}^i \right) x_t$$

Then we call $\Psi(\mathbf{B}) = \sum_{i=0}^{\infty} \Psi_i \mathbf{B}^i$ a linear filter operator.

5.2.2 Stationary Time Series II

For a time-invariant, stable linear filter and a stationary input time series $\{x_t\}$ with $\mu_x = E(x_t)$, and $\gamma_x(k) = cov(x_t, x_{t+k})$, the output time series y_t is also a stationary time series with

$$E(y_t) = \mu_y = \sum_{i=-\infty}^{\infty} \Psi_i \mu_x.$$

and

$$cov(y_t, y_{t+k}) = \gamma_y(k) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \Psi_i \Psi_j \gamma_x(i - j + k)$$

5.2.2 Stationary Time Series III

It is easy to show that the following stable linear process y_t with white noise time series, ϵ_t, as the input time series x_t, is also stationary:

$$
\begin{aligned}
y_t &= \mu + \Psi_0 \epsilon_t + \Psi_1 \epsilon_{t-1} + \Psi_2 \epsilon_{t-2} + \cdots \\
&= \mu + \sum_{i=0}^{\infty} \Psi_i \epsilon_{t-i} \\
&= \mu + \sum_{i=0}^{\infty} \Psi_i \mathbf{B}^i \epsilon_t \\
&= \mu + \Big(\sum_{i=0}^{\infty} \Psi_i \mathbf{B}^i\Big) \epsilon_t \\
&= \mu + \Psi(\mathbf{B}) \epsilon_t
\end{aligned}
\tag{2}
$$

with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \sigma^2$, $cov(\epsilon_t, \epsilon_{t+k}) = 0$ if $k \neq 0$, i.e.

$$
\gamma_\epsilon(k) = \begin{cases} \sigma^2, & \text{if } k = 0 \\ 0, & \text{otherwise.} \end{cases} \tag{3}
$$

5.2.2 Stationary Time Series IV

The equation (2) is called **infinite moving average** and serves as a general class of models for any stationary time series.

Wold theorem (1938) states that any nondeterministic weakly stationary time series y_t can be represented as in equation (2), where Ψ_i satisfy $\sum_{i=0}^{\infty} \Psi_i^2 < \infty$

An intuitive explanation of this is that a stationary time series can be viewed as the weighted sum of the present and past random noises, plus a certain level (μ).

It can also be noticed that such $\{y_t\}$ has autocovariances as

$$\gamma_y(k) = Cov(y_t, y_{t+k}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \Psi_i \Psi_j \gamma_\epsilon(i+k-j) \tag{4}$$

$$\overset{j=i+k}{=} \sigma^2 \sum_{i=0}^{\infty} \Psi_i \Psi_{i+k} \tag{5}$$

In particular,

$$\gamma(0) = Var(y_t) = \sigma^2 \sum_{i=0}^{\infty} \Psi_i^2$$

5.3 Finite Order Moving Average(MA) Processes I

For a given constant ϕ, consider a time series $\{y_t\}$

$$y_t = \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} \tag{6}$$

The process $\{y_t\}$ is stationary if $|\phi| < 1$. (Why?)

Rewrite the (6), we have

$$
\begin{aligned}
y_t = \mu + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} &= \mu + \epsilon_t + \sum_{i=1}^{\infty} \phi^i \epsilon_{t-i} && (7)\\
&\overset{i=j+1}{=} \mu + \epsilon_t + \sum_{j=0}^{\infty} \phi^{j+1} \epsilon_{t-(j+1)} && (8)\\
&= \mu + \epsilon_t + \phi \sum_{j=0}^{\infty} \phi^j \epsilon_{(t-1)-j} && (9)\\
&= \mu + \epsilon_t + \phi\Big(\sum_{j=0}^{\infty} \phi^j \epsilon_{(t-1)-j} + \mu\Big) - \phi\mu && (10)\\
&= \phi y_{t-1} + (\mu - \phi\mu) + \epsilon_t && (11)\\
&= \phi y_{t-1} + \delta + \epsilon_t. && (12)
\end{aligned}
$$

5.3 Finite Order Moving Average(MA) Processes III

We call such a series

$$y_t = \phi y_{t-1} + \delta + \epsilon_t \tag{13}$$

where $\delta = \mu(1 - \phi)$, a **first-order autoregressive process**.

What is the process $\{y_t\}$, if $\phi = 1$?

5.3 Finite Order Moving Average(MA) Processes IV

Recall in Wold's Theorem, the coefficients $\{\psi_i\}$ in an infinite set of unknown parameters.

For practical purposes, we need models that

- Finite order moving average (MA) models
- Finite order autoregressive (AR) models
- A mixture of finite order autoregressive and moving average models (ARMA)

A series $\{y_t\}$ is called a moving average of order q, denoted as **MA(q)**, if

$$
\begin{aligned}
y_t =& \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \\
=& \mu + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \\
=& \mu + \left(1 - \sum_{i=1}^{q} \theta_i \mathbf{B}^i\right) \epsilon_t \\
=& \mu + \mathbf{\Theta}(B) \epsilon_t
\end{aligned}
\tag{14}
$$

where we define $\mathbf{\Theta}(B) = 1 - \sum_{i=1}^{q} \theta_i \mathbf{B}^i$

Notice a MA(q) is obtained when we let $\psi_0 = 1$, $\psi_i = -\theta_i$ for $i = 1, \cdots, q$, and $\psi_i = 0$ for $i > q$.

5.3 Finite Order Moving Average(MA) Processes VI

Since equation (14) is a special case of equation (6), a finite order moving average process (MA(q) is always stationary regardless of values of the weights. It has the following properties:

1. Stationarity
2. The mean of $\{y_t\}$ is μ, or $E(y_t) = \mu$.
3. The variance function is

$$Var(y_t) = \sigma^2(1 + \theta_1^2 + ... + \theta_q^2) \tag{15}$$

4. The autocovariance function is

$$Cov(y_t, y_{t+k}) = \begin{cases} \sigma^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q), & k = 1, \cdots, q \\ 0, & k > q \end{cases} \tag{16}$$

(In (5), we have set $j = i + k$. let $j = q$ in (16), we have $i = q - k$.)

5.3 Finite Order Moving Average(MA) Processes VII

In particular, the 3rd property implies the autocorrelation function (ACF) of MA(q) is

$$\rho(k) = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \begin{cases} 1, & k = 0 \\ \frac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1+\theta_1^2+\ldots+\theta_q^2}, & k = 1, 2, \ldots q \\ 0, & k > q \end{cases} \quad (17)$$

This shows a distinctive feature of ACF in identifying a MA(q) process, as it "vanishes" after lag q.
Remember we talked about the significance limits used by R are $\pm 2/\sqrt{N}$ is very stringent. The reason is that it is based on the assumption that the series is not autocorrelated, i.e. $\rho(k) = 0$.
A more accurate method by Bartlett uses a wider limits. (See textbook page 334 for a complete expression) When q not very big, e.g. q=1, q=2, these two limits are not very far away.

5.3.1 MA(1) Process I

When $q = 1$, we have the MA(1) process defined by

$$y_t = \mu + \epsilon_t - \theta\epsilon_{t-1}$$

$$\begin{aligned}
E(y_t) &= \mu + 0 + 0 = \mu \\
V(y_t) &= E(y_t^2) - [E(y_t)]^2 \\
&= E(\mu^2 + \epsilon_t^2 + \theta^2\epsilon_{t-1}^2 + \mu\epsilon_t - \mu\theta\epsilon_{t-1} - \theta\epsilon_t\epsilon_{t-1}) - \mu^2 \\
&= \mu^2 + E(\epsilon_t^2) + \theta^2 E(\epsilon_{t-1}^2) - \mu^2 \\
&= \sigma^2 + \sigma^2\theta^2 \\
&= \sigma^2(1 + \theta^2)
\end{aligned}$$

5.3.1 MA(1) Process II

The auto covariance function is

$$\gamma_y(k) = \begin{cases} \sigma^2(1+\theta^2), & k = 0 \\ -\sigma^2\theta, & k = 1 \\ 0, & k > 1 \end{cases} \tag{18}$$

which implies the ACF is

$$\rho_y(k) = \begin{cases} 1, & k = 0 \\ -\frac{\theta}{1+\theta^2}, & k = 1 \\ 0, & k > 1 \end{cases} \tag{19}$$

It cuts off after lag 1.
For example, consider MA(1) process with $\mu = 40$ and $\theta = -0.8$,

$$y_t = 40 + \epsilon_t + 0.8\epsilon_{t-1}$$

5.3.1 MA(1) Process III

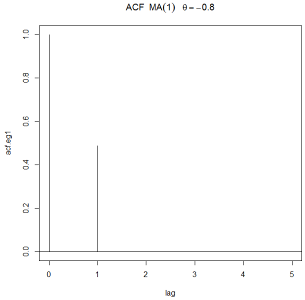
Notice the ACF depends only on the value of θ:

$$\rho(1) = -\frac{-0.8}{1 + 0.8^2} = 0.488. \tag{20}$$

5.3.1 MA(1) Process IV

The theoretical ACF is as follows and is obtained in R by

```
> nlag=5
> acf.eg1 = ARMAacf(ma=.8, lag.max=nlag);
> acf.eg1
        0         1         2         3         4         5
1.0000000 0.4878049 0.0000000 0.0000000 0.0000000 0.0000000
>
> plot(0:nlag, acf.eg1, type="h", xlab="lag", main=(expression(ACF~~MA(1)~~~theta==-0.8)))
> abline(h=0)
```



A realization of this model $y_t = 40 + \epsilon_t + 0.8\epsilon_{t-1}$ with its sample ACF is shown below

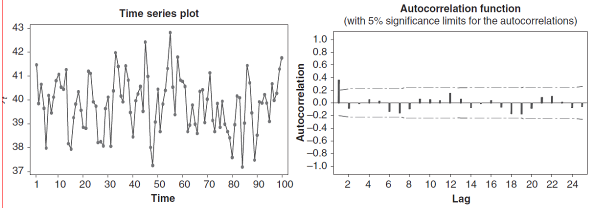


FIGURE 5.2 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$.

5.3.1 MA(1) Process VI

Notice the different between the theoretical one and the simulated one. A realization of the model $y_t = 40 + \epsilon_t - 0.8\epsilon_{t-1}$ with its sample ACF is shown below

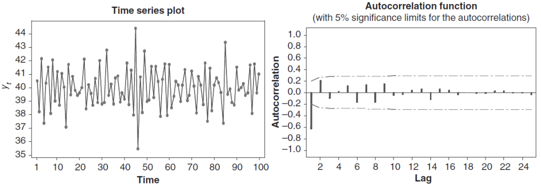


FIGURE 5.3 A realization of the MA(1) process, $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$.

Notice the different behavior of the two models at lag$= 1$. Also notice that they have the same variance.

5.3.2 MA(2) Process I

When $q = 2$, we have the MA(2) equation

$$y_t = \mu + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2}$$

Similarly, the ACF are given as

$$\rho(k) = \begin{cases} 1, & k = 0 \\ \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 1 \\ \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, & k = 2 \\ 0, & k > 2 \end{cases}$$

It cuts off at lag 2.
For example, consider MA(2) process with $\mu = 40$, $\theta_1 = -0.7$ and $\theta_2 = 0.28$,

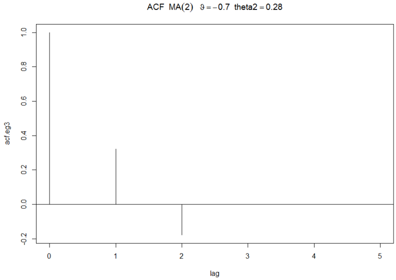
$$y_t = 40 + \epsilon_t + 0.7\epsilon_{t-1} - 0.28\epsilon_{t-2}$$

5.3.2 MA(2) Process II

The ACF can be computed accordingly: $\rho(1) = \frac{0.7-0.7*0.28}{1+0.7^2+0.28^2} = 0.321$ and $\rho(2) = -0.179$.

The theoretical ACF is as follows and is obtained in R by

```
> acf.eg3 = ARMAacf(ma=c(0.7, -0.28), lag.max=nlag);
> acf.eg3
        0         1          2         3         4         5
1.0000000 0.3213466 -0.1785259 0.0000000 0.0000000 0.0000000
> plot(0:nlag, acf.eg3, type="h", xlab="lag", main=(expression(ACF~~MA(2)~~~theta1==-0.7~~theta2==0.28)))
> abline(h=0)
```



A realization of this model with its sample ACF is shown below

5.3.2 MA(2) Process III

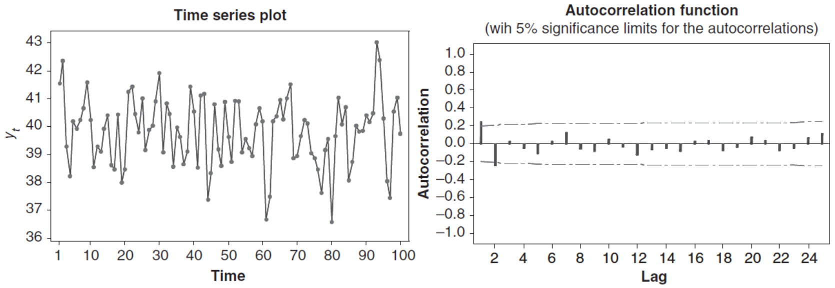


FIGURE 5.4 A realization of the MA(2) process, $y_t = 40 + \varepsilon_t + 0.7\varepsilon_{t-1} - 0.28\varepsilon_{t-2}$.

Notice the sample ACF cuts off after lag 2.

5.4 Finite Order Autoregressive(AR) Process I

A process $\{y_t\}$ is autoregressive of order p, denoted as AR(p), if it can be written as

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

where $\{\epsilon_t\}$ is a white noise process. δ is a constant (but δ is not the mean, or expecation), and $\phi_1 ... \phi_p$ are called autoregressive(AR) coefficients.

If we call $\Phi(\mathbf{B}) := 1 - \sum_{i=1}^{p} \phi_i \mathbf{B}^i$ an AR(p) operator, then we can rewrite

$$\Phi(\mathbf{B}) y_t = \delta + \epsilon_t$$

5.4.1 AR(1) Process I

We now take our attention to an simple example–AR(1):

$$y_t = \phi y_{t-1} + \delta + \epsilon_t$$

A successive substitution yields

$$\begin{aligned}
y_t =& \phi(\phi y_{t-2} + \delta + \epsilon_{t-1}) + \delta + \epsilon_t \\
=& \phi^2 y_{t-2} + (\phi + 1)\delta + \phi\epsilon_{t-1} + \epsilon_t \\
=& \phi^3 y_{t-3} + (\phi^2 + \phi + 1)\delta + \phi^2\epsilon_{t-2} + \phi\epsilon_{t-1} + \epsilon_t \\
=& \ldots\ldots \\
=& (\ldots + \phi^2 + \phi + 1)\delta + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}
\end{aligned}$$

5.4.1 AR(1) Process II

If $(... + \phi^2 + \phi + 1) = \frac{1}{1-\phi}$ is finite $\Leftrightarrow |\phi| < 1$. Then we have

$$y_t = \frac{1}{1-\phi}\delta + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}. \tag{21}$$

This is a infinite MA process, with weights ϕ^i that decay exponentially.

Thus an AR(1) process is stationary if $|\phi| < 1$. Then in this case, the mean of the process is

$$E(y_t) = \frac{\delta}{1-\phi}.$$

$$
\begin{aligned}
\gamma(k) =& cov(y_t, y_{t+k}) = E[(y_t - E(y_t))(y_{t+k} - E(y_{t+k}))] \\
=& E\left(\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}\right)\left(\sum_{i=0}^{\infty} \phi^i \epsilon_{t+k-i}\right)\right) \\
=& E(\epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \cdots + \phi^i\epsilon_{t-i} + \dots) \\
& (\epsilon_{t+k} + \phi\epsilon_{t+k-1} + \phi^2\epsilon_{t+k-2} + \cdots + \phi^{k-1}\epsilon_{t+1} \\
& + \phi^k\epsilon_t + \phi^{k+1}\epsilon_{t-1} + \phi^{k+2}\epsilon_{t-2} + \cdots + \phi^{k+i}\epsilon_{t-i} + \cdots) \\
=& E(\phi^k\epsilon_t^2 + \phi\phi^{k+1}\epsilon_{t-1}^2 + \phi^2\phi^{k+2}\epsilon_{t-2}^2 + \cdots + \phi^i\phi^{k+i}\epsilon_{t-i}^2 + \cdots) \\
=& (\sum_{i=0}^{\infty} \phi^i\phi^{i+k})\sigma^2 = \sigma^2\phi^k(\sum_{i=0}^{\infty}\phi^{2i}) = \frac{\sigma^2\phi^k}{1-\phi^2}.
\end{aligned} \tag{22}
$$

5.4.1 AR(1) Process IV

let $k = 0$, we have

$$\gamma(0) = var(y_t) = \frac{\sigma^2}{1 - \phi^2}.$$

The ACF for a stationary AR(1) process is calculated as

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = \phi^k, k = 0, 1, 2, \cdots$$

which shows the ACF of an AR(1) process has an exponential decay form.

For example, consider the AR(1) model with $\delta = 8$ and $\phi = 0.8$,

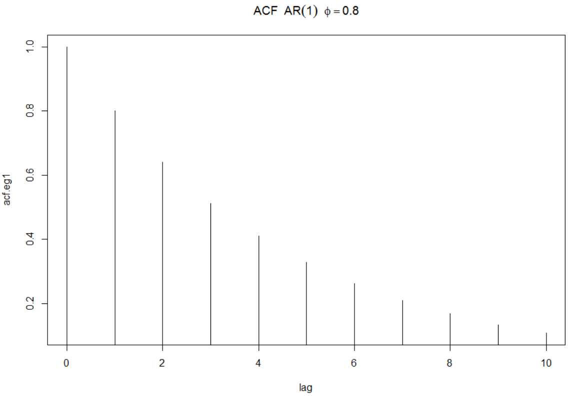
$$y_t = 8 + 0.8y_{t-1} + \epsilon_t$$

The theoretical plot of ACF is

```
> nlag=10
> acf.eg1 = ARMAacf(ar=0.8, lag.max=nlag)
> acf.eg1
        0         1         2         3         4         5         6         7         8         9
1.0000000 0.8000000 0.6400000 0.5120000 0.4096000 0.3276800 0.2621440 0.2097152 0.1677722 0.1342177
       10
0.1073742
```

Notice here unlike the MA case, the parameter **ar** in R code is just equal to ϕ.

5.4.1 AR(1) Process VI

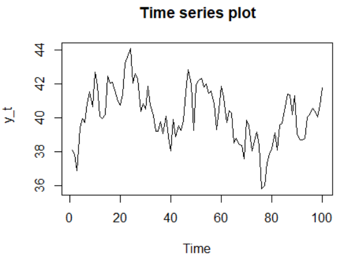


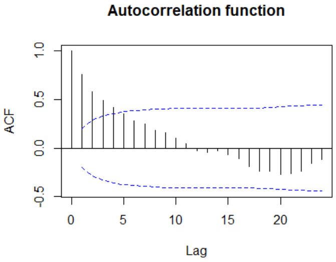
5.4.1 AR(1) Process VII

A realization of the process is shown below with its sample ACF. Recall the mean is $\mu = \frac{\delta}{1-\phi} = \frac{8}{1-0.8} = 40$, which is the reason there is a $+40$ in the end. Dropping $+40$ will change the mean, but will not affect the ACF.Since $\phi > 0$, we can observe some upward and downward trends with considerable lengths, although the overall trend is flat.

```
#realization of AR(1) process: y_t=8+0.8y_{t-1}+\epsilon_t
myseed=1
set.seed(myseed)
eg1.ar=arima.sim(list(order=c(1,0,0),ar=0.8),n=100)+40 #Generate a AR(1) process with mean=8/(1-0.8)=40
par(mfrow=c(1,2))
ts.plot(eg1.ar,ylab=expression(~~y_t), main="Time series plot")
acf(eg1.ar,lag=24,ci.type="ma", main="Autocorrelation function")
```

.



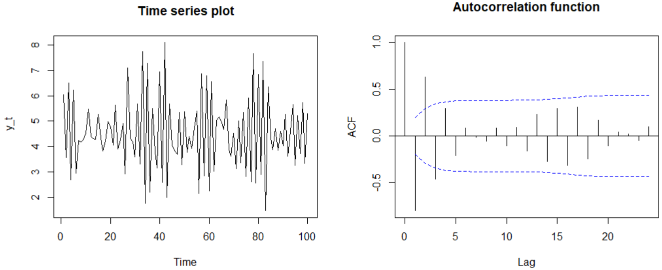


5.4.1 AR(1) Process VIII

For the AR(1) process $y_t = 8 - 0.8y_{t-1} + \epsilon_t$, we repeat as follows:

```
#realization of AR(1) process: y_t=8-0.8y_{t-1}+\epsilon_t
myseed=1
set.seed(myseed)
eg2.ar=arima.sim(list(order=c(1,0,0),ar=-0.8),n=100)+4.44 #Generate a AR(1) process with mean=8/(1+0.8)=4.4444
par(mfrow=c(1,2))
ts.plot(eg2.ar,ylab=expression(~~y_t), main="Time series plot")
acf(eg2.ar,lag=24,ci.type="ma", main="Autocorrelation function")
```

.



Question: where does the 4.444 come from? Notice the difference between ACFs of the two processes.

5.4.2 AR(2) Process I

An AR(2) process has the form

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t \tag{23}$$

As AR(1), we will first try to rewrite it in an infinite MA form. To do so, we recall the backshift operator **B**,

$$(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2) y_t = \delta + \epsilon_t, \tag{24}$$

or

$$\Phi(\mathbf{B}) y_t = \delta + \epsilon_t, \tag{25}$$

where $\Phi(\mathbf{B}) = 1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2$.

5.4.2 AR(2) Process II

IF there is another operator, say $\Psi(\mathbf{B})$ that is the inverse of $\Phi(\mathbf{B})$, i.e. $\Psi(\mathbf{B}) = \Phi^{-1}(\mathbf{B})$, or $\Psi(\mathbf{B})\Phi(\mathbf{B}) = 1$, the problem of (25) solved as

$$y_t = \Phi(\mathbf{B})^{-1}\delta + \Phi(\mathbf{B})^{-1}\epsilon_t = \Psi(\mathbf{B})\delta + \Psi(\mathbf{B})\epsilon_t, \tag{26}$$

where $\Psi(\mathbf{B})\delta$ gives the mean μ, and $\Psi(\mathbf{B})\epsilon_t$ gives the infinite MA.

Now the question becomes that how to find such an inverse operator, for AR(2), and in general for AR(p).

5.4.2 AR(2) Process III

The formulation is that by denoting $\Psi(B) = \psi_0 + \psi_1 \mathbf{B} + \psi_2 \mathbf{B}^2 + \ldots$, we want

$$(1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2)(\psi_0 + \psi_1 \mathbf{B} + \psi_2 \mathbf{B}^2 + \ldots) = 1. \tag{27}$$

Both sides are polynomials of $\mathbf{B}$, therefore by equalizing their corresponding coefficient, we obtain:

$$\psi_0 = 1 \tag{28}$$

$$\psi_1 - \phi_1 \psi_0 = 0 \tag{29}$$

$$\psi_j - \phi_1 \psi_{j-1} - \phi_2 \psi_{j-2} = 0 \text{ for all } j = 2, 3, \cdots \tag{30}$$

Therefore, inductively, one can get the values for all ψ_js.

5.4.2 AR(2) Process IV

Rewrite (30), we have

$$\psi_j - \phi_1 \mathbf{B}\psi_j - \phi_2 \mathbf{B}^2 \psi_j = 0 \tag{31}$$

$$\Rightarrow (1 - \phi_1 \mathbf{B} - \phi_2 \mathbf{B}^2)\psi_j = 0 \tag{32}$$

Note that the ψ_j in equation (31) satisfy the second-order linear difference equation and that they can be expressed as the solution to this equation in terms of the two roots m_1 and m_2 of the associated polynomial

$$m^2 - \phi_1 m - \phi_2 = 0 \tag{33}$$

5.4 AR(2) Process I

Remark

To get the expression for ψ_j, one needs to use a theory in Difference Equation which says ψ_j can be expressed in terms of the solutions, m_1 and m_2 to the quadratic equation

$$m^2 - \phi_1 m - \phi_2 = 0$$

Stationarity

- If the roots m_1 and m_2 are real and $|m_1| < 1$, and $|m_2| < 1$, then the AR(2) process is stationary.
- If the roots m_1 and m_2 are complex conjugates of the form $a + bi$, the condition for stationary is that $\sqrt{a^2 + b^2} < 1$.

5.4 AR(2) Process II

Equivalently, the AR(2) model is stationary if ϕ_1 and ϕ_2 satisfy the following 3 conditions (Proof can be found in the D2L):

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ |\phi_2| < 1. \end{cases}$$

5.4 AR(2) Process III

For example,

$$y_t = 0.4y_{t-1} + 0.5y_{t-2} + \epsilon_t$$

or

$$(1 - 0.4\mathbf{B} - 0.5\mathbf{B}^2)y_t = \epsilon_t$$

is such an AR(2) process.
The corresponding polynomial is

$$m^2 - 0.4m - 0.5 = 0$$

The R code polyroot(c(-0.5,-0.4,1)) gives the two roots:

$$-0.5348469 + 0i \quad 0.9348469 - 0i$$

Both roots are less than 1 in absolute value, therefore the AR(2) model $y_t = 0.4y_{t-1} + 0.5y_{t-2} + \epsilon_t$ is stationary.

5.4 AR(2) Process IV

Equivalently:

$$\begin{cases} \phi_1 + \phi_2 < 1 \Leftrightarrow 0.4 + 0.5 < 1 \\ \phi_2 - \phi_1 < 1 \Leftrightarrow 0.5 - 0.4 < 1 \\ |\phi_2| < 1 \Leftrightarrow |0.5| < 1. \end{cases}$$

5.4 AR(2) Process V

Assuming the AR(2) is stationary, then it has the properties:

1. The mean $E(y_t) = \mu = \frac{\delta}{1-\phi_1-\phi_2}$
2. ACF value $\rho(1) = \frac{\phi_1}{1-\phi_2}$
3. ACF values for $k \geq 2$ satisfy

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2),$$

which is called the Yule-Walker equations for $\rho(k)$.

Proof:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon(t) \tag{34}$$

$$E(y_t) = \delta + \phi_1 E(y_{t-1}) + \phi_2 E(y_{t-2}) + E(\epsilon(t)) \tag{35}$$

$$\mu = \delta + \phi_1 \mu + \phi_2 \mu \tag{36}$$

$$(36) \Rightarrow \mu = \frac{\delta}{1 - \phi_1 - \phi_2} \tag{37}$$

$$\begin{aligned}
\gamma(k) &= cov(y_t, y_{t-k}) \\
&= cov(\delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t, y_{t-k}) \\
&= \phi_1 cov(y_{t-1}, y_{t-k}) + \phi_2 cov(y_{t-2}, y_{t-k}) + cov(\epsilon_t, y_{t-k}) \qquad (38) \\
&= \phi_1 \gamma(k-1) + \phi_2 \gamma(k-2) + \begin{cases} \sigma^2, & \text{if} \quad k = 0 \\ 0, & \text{if} \quad k > 0 \end{cases}
\end{aligned}$$

5.4 AR(2) Process VII

Thus

$$\gamma(0) = \phi_1\gamma(-1) + \phi_2\gamma(-2) + \sigma^2 \tag{39}$$
$$= \phi_1\gamma(1) + \phi_2\gamma(2) + \sigma^2. \tag{40}$$

$$\gamma(k) = \phi_1\gamma(k-1) + \phi_2\gamma(k-2), k = 1, 2, \cdots \tag{41}$$

The equation (41) is called the **Yule-Walker** equations for $\gamma(k)$.

$$\rho(k) = \phi_1\rho(k-1) + \phi_2\rho(k-2), k = 1, 2, \cdots \tag{42}$$

The equation (42) is called the **Yule-Walker** equations for $\rho(k)$.

5.4 AR(2) Process VIII

1. k=1

$$\rho(1) = \phi_1\rho(0) + \phi_2\rho(-1) \tag{43}$$
$$= \phi_1(1) + \phi_2\rho(1) \tag{44}$$
$$\Rightarrow \rho(1) = \frac{\phi_1}{1-\phi_2} \tag{45}$$

2. k=2,

$$\rho(2) = \phi_1\rho(1) + \phi_2$$

3. k=3

$$\rho(3) = \phi_1\rho(2) + \phi_2\rho(1)$$

$$\vdots$$

Remark

$$m^2 - \phi_1 m - \phi_2 = 0 \tag{46}$$

which has the same coefficients as the Yule-Walker equation, has two roots, say m_1 and m_2. m_1 and m_2 can determine the behavior of ACF of a AR(2).

5.4 ACF of AR(2) Process I

There are THREE cases.

1. m_1 and m_2 are distinct, real roots, then

$$\rho(k) = c_1 m_1^k + c_2 m_2^k, k = 0, 1, 2, \cdots$$

where c_1 and c_2 are particular constants and can, for example, be obtained from $\rho(0)$ and $\rho(1)$. The ACF is a mixture of two exponential decay terms.

For example,

$$y_t = 0.4y_{t-1} + 0.5y_{t-2} + \epsilon_t$$

or

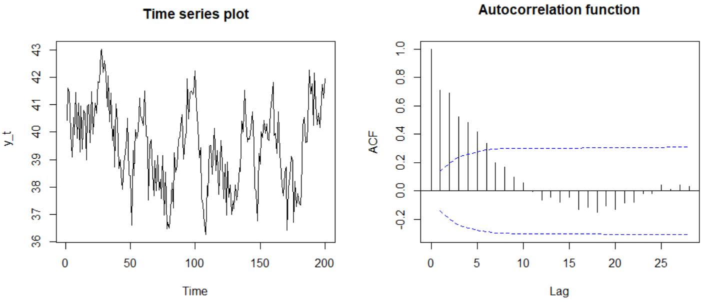
$$(1 - 0.4\mathbf{B} - 0.5\mathbf{B}^2)y_t = \epsilon_t$$

is such an AR(2) process.

5.4 ACF of AR(2) Process II

```
#realization of AR(2) process: y_t=4+0.4 y_{t-1}+0.5 y_{t-2}+\epsilon_t
#Fig 5.7
myseed=10
set.seed(myseed)
#Generate a AR(2) process with mean=delta/(1-phi_1-\phi_2)=4/(1-0.4-0.5)=40
eg3.ar=arima.sim(list(ar = c(0.4,0.5)), n = 200)+4/(1-0.4-0.5)
par(mfrow=c(1,2))
ts.plot(eg3.ar,ylab=expression(~~y_t), main="Time series plot")
acf(eg3.ar,lag=28,ci.type="ma", main="Autocorrelation function")
```

The ACF plot exhibits a mixture of two exponential decay terms.



5.4 ACF of AR(2) Process III

2. m_1 and m_2 are complex conjugates, $a \pm bi$, then

$$\rho(k) = R^k[c_1 \cos(\lambda k) + c_2 \sin(\lambda k)]$$

where $R = |m_i| = \sqrt{a^2 + b^2}$, where a,b are the real and imaginary part of m_i. The λ satisfies $\cos(\lambda) = \frac{a}{R}$ and $\sin(\lambda) = \frac{b}{R}$. The ACF has the form of a damped sinusoid, with damping factor and frequency λ, that is the period $\frac{2\pi}{\lambda}$

For example,

$$y_t = 0.8y_{t-1} - 0.5y_{t-2} + \epsilon_t$$

is such an AR(2) process. The corresponding polynomial is

$$m^2 - 0.8m + 0.5 = 0$$

The R code polyroot(c(0.5,-0.8,1)) gives the two roots:

$$0.4 + 0.5830952i \quad 0.4 - 0.5830952i$$

$\sqrt{0.4^2 + 0.5830952^2} = 0.7071068$, therefore the AR(2) model $y_t = 0.8y_{t-1} - 0.5y_{t-2} + \epsilon_t$ is stationary.

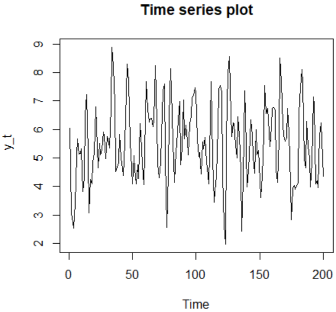
5.4 ACF of AR(2) Process V

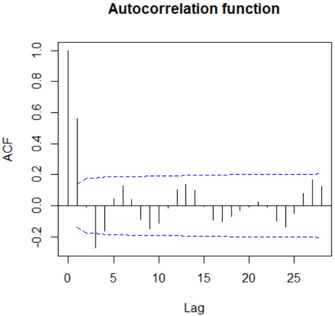
Equivalently:

$$\begin{cases} \phi_1 + \phi_2 < 1 \Leftrightarrow 0.8 - 0.5 < 1 \\ \phi_2 - \phi_1 < 1 \Leftrightarrow -0.5 - 0.8 < 1 \\ |\phi_2| < 1 \Leftrightarrow |-0.5| < 1. \end{cases}$$

5.4 ACF of AR(2) Process VI

```
#realization of AR(2) process: y_t=4+0.8 y_{t-1}-0.5 y_{t-2}+\epsilon_t
#Fig 5.8
myseed=10
set.seed(myseed)
#Generate a AR(2) process with mean=delta/(1-phi_1-\phi_2)=4/(1-0.8+0.5)=5.714
eg4.ar=arima.sim(list(ar = c(0.8,-0.5)), n = 200)+4/(1-0.8+0.5)
par(mfrow=c(1,2))
ts.plot(eg4.ar,ylab=expression(~~y_t), main="Time series plot")
acf(eg4.ar,lag=28,ci.type="ma", main="Autocorrelation function")
```





5.4 ACF of AR(2) Process VII

3. m_1 and m_2 are two identical roots, then

$$\rho(k) = (c_1 + c_2 k) m_1^k$$

The ACF will exhibit an exponential decay pattern.In practice, the ACF of case 3 has little difference to the ACF of case 1.

5.4 ACF of AR(2) Process VIII

For example,

$$y_t = 0.8y_{t-1} - 0.16y_{t-2} + \epsilon_t$$

is such an AR(2) model.The corresponding polynomial is

$$m^2 - 0.8m + 0.16 = 0$$

The R code polyroot(c(0.16,-0.8,1)) gives the same two roots:

$$0.4 + 0i \quad 0.4 + 0i$$

Since $|0.4| < 1$, the AR(2) model $y_t = 0.8y_{t-1} - 0.16y_{t-2} + \epsilon_t$ is stationary.

5.4 ACF of AR(2) Process IX

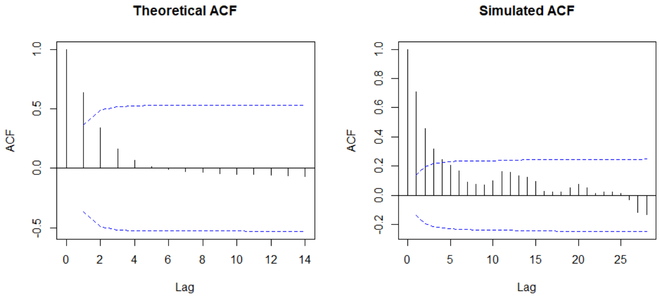
Equivalently:

$$\begin{cases} \phi_1 + \phi_2 < 1 \Leftrightarrow 0.8 + 0.16 < 1 \\ \phi_2 - \phi_1 < 1 \Leftrightarrow -0.16 - 0.8 < 1 \\ |\phi_2| < 1 \Leftrightarrow |-0.16| < 1. \end{cases}$$

5.4 ACF of AR(2) Process X

```
##Generate a AR(2) process y_t=4+0.8 y_{t-1}-0.16 y_{t-2}+\epsilon_t
#with mean=delta/(1-phi_1-\phi_2)=4/(1-0.8+0.16)=11.1111
#Theoretical and simulated ACF
par(mfrow=c(1,2))
#Theoretical ACF
acf.theo=ARMAacf(ar=c(0.8,-0.16), lag.max = nlag)
acf(acf.theo,type = "correlation",ci.type="ma",main="Theoretical ACF")

#simulated ACF
myseed=10
nlag=28
eg5.ar.sim=arima.sim(list(ar = c(0.8,-0.16)), n = 200)+4/(1-0.8+0.16)
acf(eg5.ar.sim,lag=nlag,ci.type="ma", main="Simulated ACF")
```



5.4.3 Finite Order Autoregressive(AR) Process I

Now let us come back to the general AR(p) process. Following a similar but more complicated derivation, the AR(p) process $\{y_t\}$ is stationary if the roots of

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - ... - \phi_p = 0 \tag{47}$$

are less than one in absolute value. Then the mean of the process

$$E(y_t) = \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - ... - \phi_p}. \tag{48}$$

The Yule-Walker equation for ACF is

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + ... + \phi_p \rho(k-p) = \sum_{i=1}^{p} \phi_i \rho(k-i). \tag{49}$$

for $k = 1, 2, 3, \cdots$.

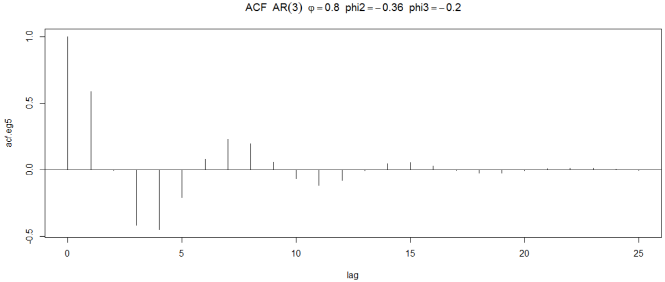
5.4.3 Finite Order Autoregressive(AR) Process II

Notice here we need to use the fact that $\rho(-k) = \rho(k)$ for first p steps. The expression for $\rho(k)$ is determined by the roots m_1, $m_2 ... m_p$ of equation (47).

$$\rho(k) = c_1 m_1^k + c_2 m_2^k + ... + c_p m_p^k$$

Similarly, the ACF will have behaviors like exponential decay or damped sinusoid or a mixture.

For example, the AR(3) process $y_t = 0.8y_{t-1} - 0.36y_{t-2} - 0.2y_{t-3} + \epsilon_t$ has ACF plot as



5.4 Partial Autocorrelation Function, PACF I

For an Autoregressive process AR(p), the ACF is not very informative to determine the order p.

Partial Autocorrelation Function (PACF) is a quantity defined with the motivation to identify the order p of an AR process.

We assume that $\{y_t\}$ is a stationary process with given ACF values.

5.4 Partial Autocorrelation Function, PACF II

We define PACF as follows: for any value k, consider the $k \times k$ system of Yule-Walker equations

$$\rho(j) = \phi_{1k}\rho(j-1) + \phi_{2k}\rho(j-2) + ... + \phi_{kk}\rho(j-k) \quad j = 1, 2...k$$

The partial autocorrelation coefficient at lag k is ϕ_{kk}, that is the last solution to the $k \times k$ system of equations. The the PACF is defined to be the collection of these ϕ_{kk}, i.e. $\{\phi_{kk}\}_{k=1}^{\infty}$.

5.4 Partial Autocorrelation Function, PACF III

In matrix notation,

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(k-2) \\ \rho(2) & \rho(1) & 1 & \dots & \rho(k-3) \\ \dots & \dots & \dots & \dots & \dots \\ \rho(k-1) & \rho(k-2) & \rho(k-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{1k} \\ \phi_{2k} \\ \vdots \\ \phi_{kk} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(k) \end{bmatrix}$$

or simply

$$\mathbf{R}_{k\times k}\phi_{k\times 1} = \rho_{k\times 1}$$

so the parital autocorrelation coefficient at lag k is the last entry of $\phi_{k\times 1} = \mathbf{R}^{-1}\rho_{k\times 1}$.

5.4 Partial Autocorrelation Function, PACF IV

For k=1, then $\phi_{11} = \rho(1)$

For k=2, then

$$\begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_{12} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} \rho(1) \\ \rho(2) \end{bmatrix}$$

implies $\phi_{22} = \frac{\rho(2)-\rho(1)^2}{1-\rho^2(1)}$

and so on...
The PACF cuts off at lag p for an AR(p) process. That is,

$$\phi_{kk} = 0, \quad \text{for } k > p$$

5.4 Partial Autocorrelation Function, PACF V

Remark

This intuitively makes sense, as the Yule-Walker equation for AR(p)

$$\rho(k) = \phi_1 \rho(k-1) + \phi_2 \rho(k-2) + ... + \phi_p \rho(k-p)$$

has p terms on right side. For some $k > p$, say $k = 3$, and $p = 2$, we would have

$$\rho(3) = \phi_1 \rho(2) + \phi_2 \rho(1) + 0$$

meaning the last entry, $\phi_{33} = 0$. And in general, $\phi_{kk} = 0$, for $k > p$.

This suggest that the PACF can be used in identifying the order of an AR process similar to how ACF is used for an MA process.

The sample PACF is obtained by using sample ACF with the procedure described above.

$$\hat{\phi}_{11} = \hat{\rho}(1)$$

$$\hat{\phi}_{22} = \frac{\hat{\rho}(2) - \hat{\rho}(1)^2}{1 - \hat{\rho}(1)^2}$$

$$\cdots\cdots$$

It can be shown that for N observations from an AR(p) process, the sample PACF $\hat{\phi_{kk}}$ satisfies that

$$E(\hat{\phi_{kk}}) = 0, \quad \text{and} \quad Var(\hat{\phi_{kk}}) = \frac{1}{N} \quad \text{approximatedly}$$

so 95% limits for $Var(\hat{\phi_{kk}})$ are given by $\pm 1.96/\sqrt{N}$. (In practice, $\pm 2/\sqrt{N}$ is commonly used.)

An Examples on PACF for AR(2) I

Recall the AR(2) process we have talked about

$$y_t = y_{t-1} - 0.24y_{y-2} + 1 + \epsilon_t$$

which has ACF values $\rho(0) = 1$, $\rho(1) = 0.806$, $\rho(2) = 0.566$, $\rho(3) = 0.373$, from which we can compute the PACF as follows:

1. For k=1, $\phi_{11} = \rho(1) = 0.806$.
2. For k=2, then

$$\begin{bmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_{12} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0.806 \\ 0.806 & 1 \end{bmatrix} \begin{bmatrix} \phi_{12} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} 0.806 \\ 0.566 \end{bmatrix}$$

   from which we solve for $\phi_{22} = -0.239$

An Examples on PACF for AR(2) II

3. For k=3, then

$$\begin{bmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{bmatrix} \begin{bmatrix} \phi_{13} \\ \phi_{23} \\ \phi_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0.806 & 0.566 \\ 0.806 & 1 & 0.806 \\ 0.566 & 0.806 & 1 \end{bmatrix} \begin{bmatrix} \phi_{13} \\ \phi_{23} \\ \phi_{33} \end{bmatrix} = \begin{bmatrix} 0.806 \\ 0.566 \\ 0.373 \end{bmatrix}$$

from which we solve for

$$\begin{bmatrix} \phi_{13} \\ \phi_{23} \\ \phi_{33} \end{bmatrix} = \begin{bmatrix} 1 \\ -0.24 \\ 0 \end{bmatrix}$$

It means that the PACF at lag 3 is 0 for this AR(2) process. One can compute a few more PACF values and confirm that they are all zero after lag 3.

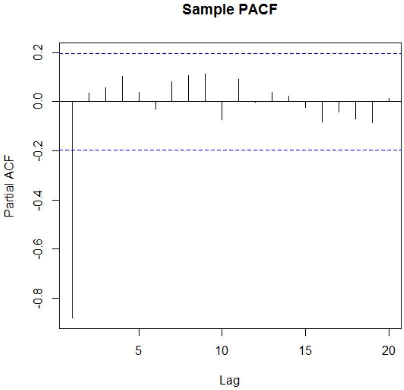
An Examples on PACF for AR(2) III

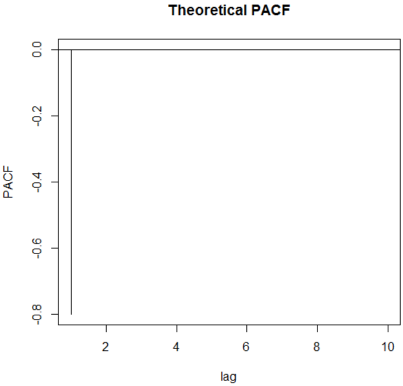
Exercise: Suppose an AR(2) process is given below. Repeat the procedures as above.

$y_t = 0.8y_{t-1} - 0.15y_{y-2} - 2 + \epsilon_t$

5.4 Plots of PACF, and Sample PACF for a AR(1) process

AR(1) process: $y_t = -0.8y_{t-1} + \epsilon_t$

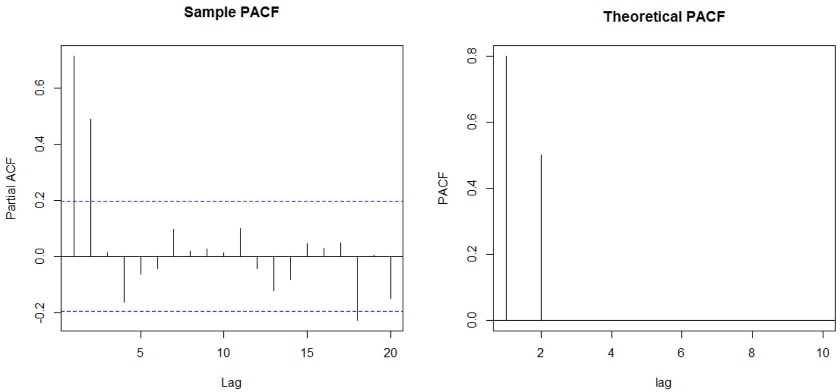




The PACF cuts off after the first lag.

5.4 Plots of PACF, and Sample PACF for a AR(2) process

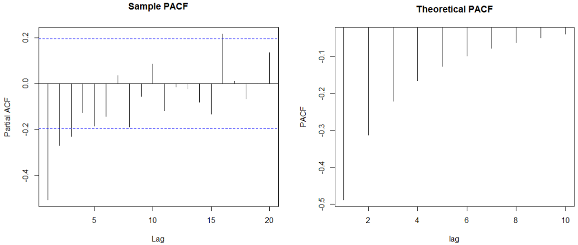
AR(2) process: $y_t = 0.4y_{t-1} + 0.5y_{t-2} + \epsilon_t$



The PACF cuts off after lag 2.

MA(1) process: $y_t = \epsilon_t - 0.8\epsilon_{t-1}$

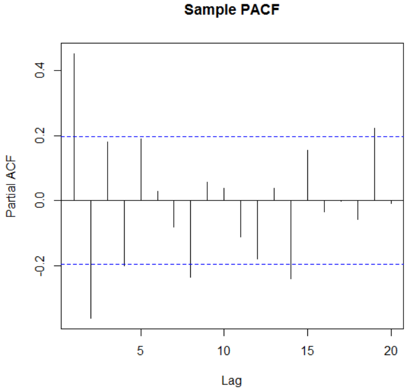
```
> eg1 = arima.sim(list(order=c(0,0,1), ma=-0.8), n=100)
> pacf(eg1, main="Sample PACF")
> eg1.pacf=ARMAacf(ma=-0.8, lag.max=10, pacf=T)
> eg1.pacf
 [1] -0.48780488 -0.31225605 -0.22147781 -0.16519352 -0.12666946 -0.09871330
 [7] -0.07768408 -0.06150597 -0.04888195 -0.03894205
> plot(eg1.pacf, type="h",xlab="lag",ylab="PACF",main="Theoretical PACF")
```

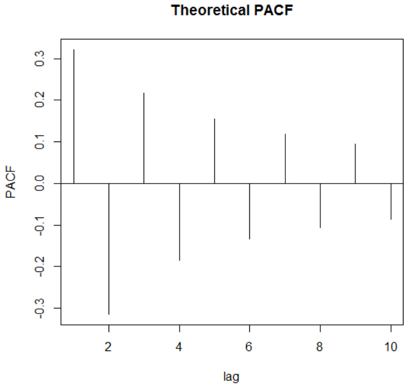


They exhibit an exponential decay pattern.

5.4 Plots of PACF, and Sample PACF for a MA(2) process

MA(2) process: $y_t = \epsilon_t + 0.7\epsilon_{t-1} - 0.28\epsilon_{t-2}$





They also exhibit an exponential decay pattern in absolute value.

5.4 Invertibility of MA(q) I

Consider a simple MA(1) model $y_t = \epsilon_t - \theta_1 \epsilon_{t-1}$, which can be rewritten as

$$
\begin{aligned}
\epsilon_t =& y_t + \theta_1 \epsilon_{t-1} \\
=& y_t + \theta_1 [y_{t-1} + \theta_1 \epsilon_{t-2}] \\
=& y_t + \theta_1 y_{t-1} + \theta_1^2 \epsilon_{t-2} \\
=& \ldots \\
=& \sum_{i=0}^{\infty} \theta_1^i y_{t-i}
\end{aligned}
$$

which is convergent for $|\theta_1| < 1$. Then it is called an invertible moving average process. In general, a MA(q) process is invertible if it can be written as a convergent infinite AR form

$$
y_t = \delta + \sum_{i=1}^{\infty} \pi_i y_{t-i} + \epsilon_t
$$

5.4 Invertibility of MA(q) II

We define invertibility for model identification. That is, given MA(q) model with its ACF values up to lag q, there are several sets of coefficients θ_1, $\theta_2 ... \theta_q$ which give several MA(q) model equations. There is only one solution that will satisfy the invertibility, thus by convention, we always choose the invertible MA(q) model for uniqueness.

5.4 Invertibility of MA(q) III

For example, consider the MA(1) process with $\rho(1) = 0.4$. The ACF for an MA(1) process is defined as

$$\rho_y(k) = \begin{cases} 1, & k = 0 \\ -\frac{\theta}{1+\theta^2}, & k = 1 \\ 0, & k > 1 \end{cases} \tag{50}$$

Therefore,

$$\rho(1) = \frac{-\theta_1}{1 + \theta_1^2} = 0.4. \tag{51}$$

Solving for θ_1, we have $\theta_1 = -0.5$ and $\theta_1 = -2$. Only $|\theta_1| = 0.5 < 1$ gives an invertible MA(1) process. The unique MA(1) process is

$$y_t = \mu + \epsilon_t - (-0.5)\epsilon_{t-1} = \mu + \epsilon_t + 0.5\epsilon_{t-1}. \tag{52}$$

5.4 Final Remark

Condition for invertibility of MA(q)

All roots m_1, $m_2 ... m_q$ of

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - ... - \theta_q = 0$$

are less than 1 in absolute value.

The ACF and PACF do have very distinct and indicative properties of MA and AR models respectively. Therefore, in model identification, we use both sample ACF and sample PACF simultaneously.

5.5 Autoregressive-Moving Average (ARMA) process

If we "combined" an AR(p) component and an MA(q) component, we get a autoregressive moving average (ARMA(p,q)) model, which looks like

$$y_t - \sum_{i=1}^{p} \phi_i y_{t-i} = \delta + \epsilon_t - \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{53}$$

which ϵ_t is a white noise process.
The equation (53) has an equivalent operator form

$$\Phi(\mathbf{B}) y_t = \delta + \Theta(\mathbf{B}) \epsilon_t \tag{54}$$

where $\Phi(\mathbf{B}) = 1 - \sum_{i=1}^{p} \phi_i \mathbf{B}^i$ and $\Theta(\mathbf{B}) = 1 - \sum_{i=1}^{q} \theta_i \mathbf{B}^i$ are AR and MA operators respectively.

5.5.1 Stationarity of ARMA process

The stationarity of an ARMA process depends on the AR component.

Condition for Stationarity

All roots m_1, $m_2 ... m_p$ of

$$m^p - \phi_1 m^{p-1} - \phi_2 m^{p-2} - ... - \phi_p = 0$$

are less than 1 in absolute value.

In this case, the ARMA process has an infinite MA form

$$y_t = \Phi(\mathbf{B})^{-1}\delta + \Phi(\mathbf{B})^{-1}\Theta(\mathbf{B})\epsilon_t = \mu + \Psi(\mathbf{B})\epsilon_t \tag{55}$$

Notice here $\Psi(\mathbf{B})$ is an infinite MA operator that satisfies $\Theta(\mathbf{B}) = \Phi(\mathbf{B})\Psi(\mathbf{B})$.

5.5.2 Invertibility of ARMA process

The invertibility of an ARMA process depends on the MA component.

Condition for Invertibility

All roots m_1, $m_2 ... m_p$ of

$$m^q - \theta_1 m^{q-1} - \theta_2 m^{q-2} - ... - \theta_q = 0$$

are less than 1 in absolute value.

In this case, the ARMA process has an infinite AR form

$$\Phi(\mathbf{B}) y_t = \delta + \Theta(\mathbf{B}) \epsilon_t \tag{56}$$

$$\Theta(\mathbf{B})^{-1} \Phi(\mathbf{B}) y_t = \Theta(\mathbf{B})^{-1} \delta + \epsilon_t \tag{57}$$

$$\Pi(\mathbf{B}) y_t = a + \epsilon_t \tag{58}$$

where $\Theta(\mathbf{B})^{-1} \Phi(\mathbf{B}) = \Pi(\mathbf{B})$ and $\Theta(\mathbf{B})^{-1} \delta = a$.

5.5.3 Example – ARMA(1,1) process I

When $p = q = 1$, we have the ARMA(1,1) equation

$$y_t - \phi_1 y_{t-1} = \delta + \epsilon_t - \theta_1 \epsilon_{t-1} \tag{59}$$

It is invertible if $|\theta_1| < 1$.

It is stationary if $|\phi_1| < 1$.

In this case, the mean is

$$\mu = \frac{\delta}{1 - \phi_1}.$$

5.5.3 Example – ARMA(1,1) process II

For example, $\delta = 1$, $\phi_1 = 0.8$ and $\theta_1 = -0.6$ yield

$$y_t - 0.8y_{t-1} = 1 + \epsilon_t + 0.6\epsilon_{t-1} \tag{60}$$

Equivalently,

$$y_t = 1 + 0.8y_{t-1} + \epsilon_t + 0.6\epsilon_{t-1} \tag{61}$$

It is stationary and therefore has an infinite MA form.

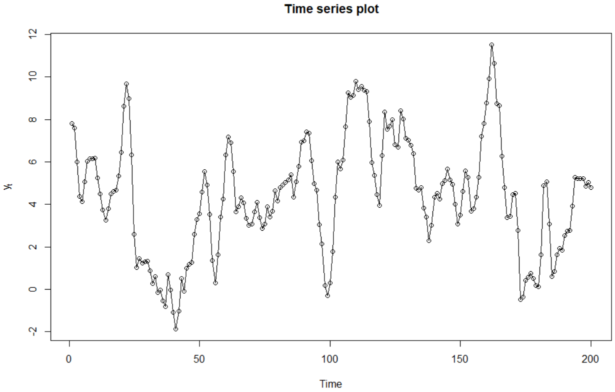
It is invertible and therefore has an infinite AR form.

It has mean

$$\frac{\delta}{1-\phi} = \frac{1}{1-0.8} = 5.$$

5.5.3 Example – ARMA(1,1) process III

Model: $y_t = 1 + 0.8y_{t-1} + \epsilon_t + 0.6\epsilon_{t-1}$



5.5.3 Example – ARMA(1,1) process IV

To make this example more general, we work on the general stationary and invertible ARMA(1,1).

$$(1 - \phi B)y_t = \delta + (1 - \theta B)\epsilon_t$$

We will just provide the results without proving it. The ACF values satisfy

$$\rho(1) = \frac{(1 - \phi\theta)(\phi - \theta)}{1 - 2\phi\theta + \theta^2}$$

$$\rho(k) = \phi\rho(k-1) \quad k = 2, 3, 4, \cdots$$

5.5.3 Example – ARMA(1,1) process V

Thus the ACF values of $y_t - 0.8y_{t-1} = 1 + \epsilon_t + 0.6\epsilon_{t-1}$ are solved as

$$\rho(1) = 0.893, \rho(2) = 0.714, \rho(3) = 0.572, \cdots$$

From the ACF values, one can solve its PACF values are

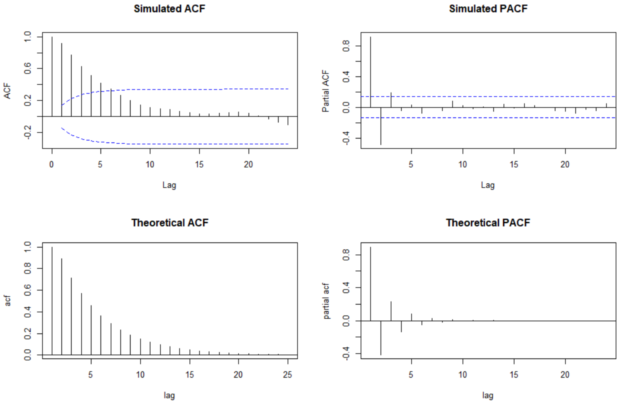
$$\phi_{11} = 0.893, \phi_{22} = -0.411, \phi_{33} = 0.227, \cdots$$

.

The simulated and theoretical ACF and PACF of the model is shown below:

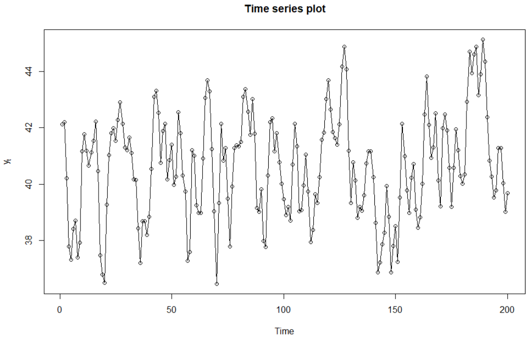
5.5.3 Example – ARMA(1,1) process VI

$y_t = 1 + 0.8y_{t-1} + \epsilon_t + 0.6\epsilon_{t-1}$ with mean $\mu = \frac{\delta}{1-\phi} = \frac{1}{1-0.8} = 5$

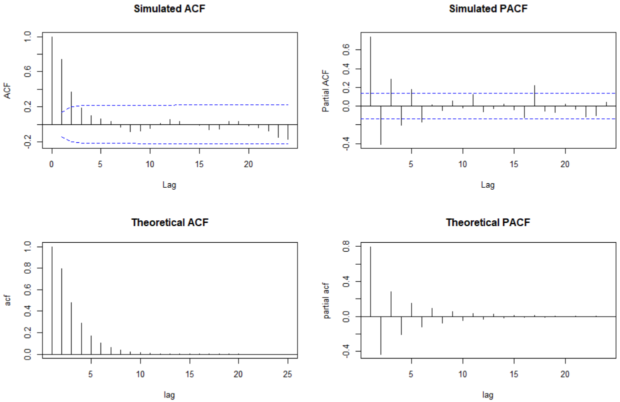


5.5.3 Example – ARMA(1,1) process VII

ARMA(1,1) model: $y_t = 16 + 0.6y_{t-1} + \epsilon_t + 0.8\epsilon_{t-1}$ has mean $\frac{\delta}{1-\phi} = \frac{16}{1-0.6} = 40$.

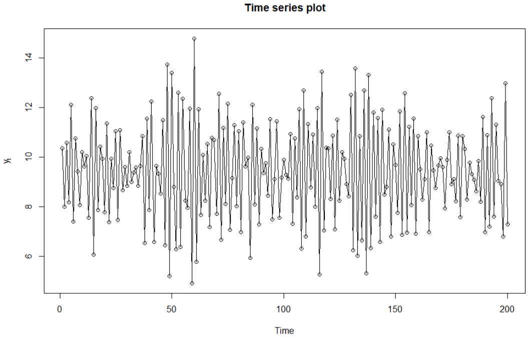


$$y_t = 16 + 0.6y_{t-1} + \epsilon_t + 0.8\epsilon_{t-1}$$

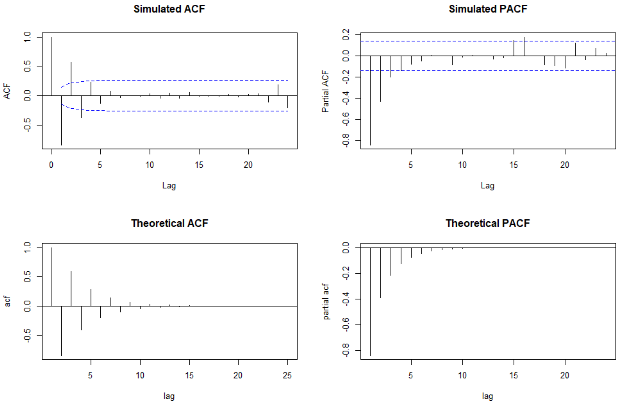


5.5.3 Example – ARMA(1,1) process IX

ARMA(1,1) model: $y_t = 16 - 0.7y_{t-1} + \epsilon_t - 0.6\epsilon_{t-1}$ has mean $\frac{\delta}{1-\phi} = \frac{16}{1-(-0.7)} = 9.411$.



$$y_t = 16 - 0.7y_{t-1} + \epsilon_t - 0.6\epsilon_{t-1}$$



Define a function **acf.pacf.arma()** for generating theoretical and simulated ACF and PACF.

```
### Define a function for theoretical and simulated ACF and PACF.
```{r}
library(tinytex)
# Define and ARMA(1,1) function
acf.pacf.arma <- function(delta, phi, theta, nlag = 24) {
  sim <- arima.sim(list(ar = phi, ma = theta), n = 200) + delta / (1 - phi)
  # Time series plot of y(t)
  par(mfrow = c(1, 1))
  ts.plot(sim, ylab = "yt", main = "Time series plot", type = "o")
  par(mfrow = c(2, 2))
  # simulated ACF and PACF
  acf(sim, lag = nlag, ci.type = "ma", type = "correlation", main = "Simulated ACF")
  acf(sim, lag = nlag, type = "partial", main = "Simulated PACF")
  # Theoretical ACF and PCAF
  thm.acf <- ARMAacf(ar = phi, ma = theta, lag.max = nlag, pacf = F)
  plot(thm.acf, type = "h", xlab = "lag", ylab = "ACF", main = "Theroretical ACF")
  abline(h = 0)
  thm.pacf <- ARMAacf(ar = phi, ma = theta, lag.max = nlag, pacf = T)
  plot(thm.pacf, type = "h", xlab = "lag", ylab = "Partial ACF", main = "Theoretical PACF")
  abline(h = 0)
}
```
```

```
#Define an ARMA(1,1) function
acf.pacf.arma=function(delta,phi,theta,nlag=24){
sim = arima.sim(list(ar = phi,ma=theta), n = 200)+ delta(1-phi)
#Time series plot of y(t)
par(mfrow=c(1,1))
ts.plot(sim,ylab='y_t', main='Time series plot',type='o')
par(mfrow=c(2,2))
#simulated ACF and PACF
acf(sim,lag=nlag, ci.type='ma', type='correlation', main='Simulated
ACF')
acf(sim,lag=nlag, type='partial', main='Simulated PACF')
#Theoretical ACF and PACF
thm.acf=ARMAacf(ar=phi, ma=theta, lag.max=nlag, pacf=F)
plot(thm.acf, type='h',xlab='lag',ylab = 'ACF', main='Theoretical ACF')
abline(h=0)
thm.pacf=ARMAacf(ar=phi, ma=theta, lag.max=nlag, pacf=T)
```

5.5.3 R function and exercises III

```
plot(thm.pacf,type='h', xlab='lag', ylab='Partial ACF',
main="Theoretical PACF")
abline(h=0)
}
```

Example 1:
$y_t = 16 + 0.6y_{t-1} + \epsilon_t + 0.8\epsilon_{t-1}$
acf.pacf.arma(delta=16,phi=0.6,theta=0.8)

Example 2:
$y_t = 16 - 0.7y_{t-1} + \epsilon_t - 0.6\epsilon_{t-1}$
acf.pacf.arma(delta=16,phi=-0.7,theta=-0.6)

Exercise:
State why the ARMA(1,1) process

$$y_t + 0.5y_{t-1} = 2 + \epsilon_t - 0.4\epsilon_{t-1}$$

is stationary and invertible. Then find its mean and ACF values for lags 1, 2, 3 by the method introduced in the example. Then check your answers using R function(s).

Theoretical ACF and PACF

	ACF	$PACF$
$MA(q)$	cuts of after lag q	exp. decay and/or damped sinusoid
$AR(p)$	exp. decay and/or damped sinusoid	cuts of after lag p
$ARMA(p,q)$	exp. decay and/or damped sinusoid	exp. decay and/or damped sinusoid

Sample ACF and PACF

	Sample ACF	Sample $PACF$
$MA(q)$	$E(\hat{\rho}) \approx 0$ for $k > q$ $Var(\hat{\rho}) \approx \frac{1}{T}$ for $k > q$	exp. decay and/or damped sinusoid
$AR(p)$	exp. decay and/or damped sinusoid	$E(\hat{\rho}) \approx 0$ for $k > p$ $Var(\hat{\rho}) \approx \frac{1}{T}$ for $k > p$
$ARMA(p,q)$	exp. decay and/or damped sinusoid	exp. decay and/or damped sinusoid

5.6 Nonstationary processes I

A series with a trend (linear, quadratic, etc...) is not stationary. But it is possible that the differences between the series and trend is stationary. For example, the first differences of a linear trend process is stationary.

Denote

$$w_t = y_t - y_{t-1} = (1 - \mathbf{B})y_t \tag{62}$$

then w_t is stationary.

Or even first difference is still nonstationary, but the second differences

$$w_t^{(2)} = w_t - w_{t-1} = (1 - \mathbf{B})^2 y_t \tag{63}$$

form a stationary series.

5.6 Nonstationary processes II

In general, we may have that the *dth differences*

$$w_t^{(d)} = (1 - \mathbf{B})^d y_t \tag{64}$$

form a stationary series.

In practice, usually d=1, and occasionally d=2.

We say that the differencing operator reduces the nonstationary series to stationarity.
Since $w_t^{(d)}$ is stationary, say it is an ARMA(p,q) process, then it is expressed as

$$\Phi(\mathbf{B}) w_t^{(d)} = \delta + \Theta(\mathbf{B}) \epsilon_t \tag{65}$$

5.6 Nonstationary processes III

which implies for $\{y_t\}$,

$$\Phi(\mathbf{B})(1-\mathbf{B})^d y_t = \delta + \Theta(\mathbf{B})\epsilon_t. \tag{66}$$

$\{y_t\}$ *is called an Autoregressive Integrated Moving Average (ARIMA) process of orders p, d, and q, or ARIMA(p,d,q),where*

- *p is the number of autoregressive terms,*
- *d is the number of nonseasonal differences needed for stationarity,*
- *q is the number of lagged forecast errors in the prediction equation.*

Notice here $\{y_t\}$ *is a nonstationary process. The order of differencing d is understood as the smallest integer that produces a stationary series.*

For different values of p, d, and q:

- *ARIMA(1,0,0) = first-order autoregressive model*
- *ARIMA(0,1,0) without constant = random walk without growth*
- *ARIMA(0,1,0) with constant = random walk with growth*
- *ARIMA(1,1,0) = differenced first-order autoregressive model*
- *ARIMA(0,1,1) without constant = simple exponential smoothing*
- *ARIMA(0,1,1) with constant = simple exponential smoothing with growth*
- *ARIMA(0,2,1) or (0,2,2) without constant = linear exponential smoothing*
- *ARIMA(1,1,2) with constant = damped-trend linear exponential smoothing Spreadsheet implementation*

5.6.1 ARIMA(0,1,0) I

The process

$$y_t = y_{t-1} + \delta + \epsilon_t \tag{67}$$

is called a random walk process. This is a ARIMA(0,1,0) process. When $\delta = 0$, it can be viewed as that for each time period an object moves for a distance which is a random noise ϵ_t. When $\delta \neq 0$, it can be viewed as moving for a fixed distance δ then plus a random noise.

For example, the process

$$y_t = 1 + y_{t-1} + \epsilon_t. \tag{68}$$

Plots of simulation, sample ACF, and PACF are shown on next page. (check out the code carefully!)

5.6.1 ARIMA(0,1,0) II

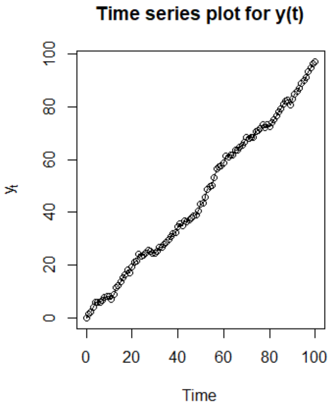
```
   #Generate white noise w(t)=1+ϵ_t
rw.diff=arima.sim(list(order=c(0,0,0)), n=100)+1
   #y_t=w_t + y_{t-1} = w_t + w_{t-1} + y_{t-2} = w_t + w_{t-1} + ··· + w_1 + y_0
rw.inv=diffinv(rw.diff)
   #Generate time series plot of y(t) and w(t)
par(mfrow=c(1,2))
rw.inv.plot=plot(rw.inv,xlab='Time', ylab=TeX('$
y_t$'), main="Time series plot for y(t)",type='o')
rw.plot=plot(rw.diff,xlab='Time', ylab=TeX('$
w_t$'), main="Time series plot for w(t)",type='o')
   #Generate ACF and PACF for y(t) and w(t)
par(mfrow=c(2,2))
acf(rw.inv,lag.max=26,ci.type='ma',xlab='Lag', ylab='ACF', main='ACF
for y(t)')
acf(rw.diff,lag.max=26,ci.type='ma',main='ACF for w(t)')
pacf(rw.inv,lag.max=26,xlab='Lag', ylab='PACF', main='PACF for y(t)')
```

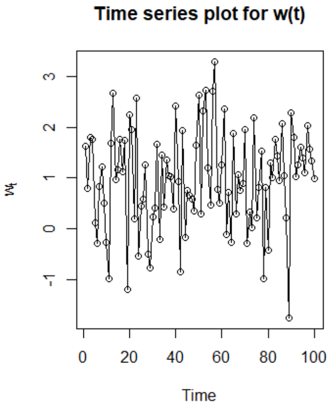
5.6.1 ARIMA(0,1,0) III

pacf(rw.diff,lag.max=26,xlab='Lag', ylab='PACF', main='PACF for w(t)')

5.6.1 ARIMA(0,1,0) IV

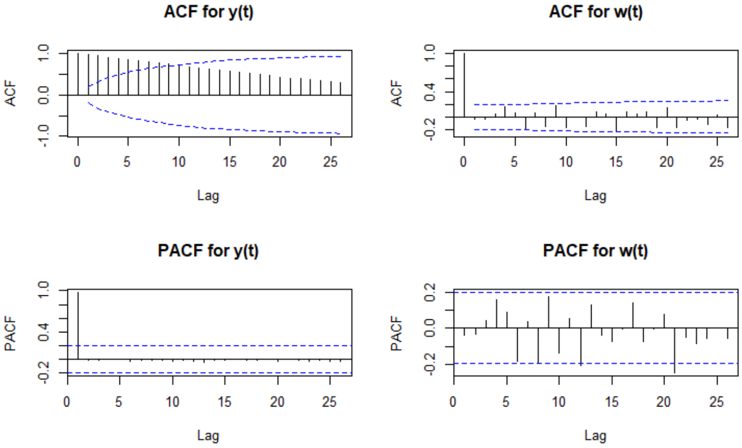
$$y_t = 1 + y_{t-1} + \epsilon_t.$$





5.6.1 ARIMA(0,1,0) V

$$y_t = 1 + y_{t-1} + \epsilon_t.$$



5.6 ARIMA(0,1,1) I

The ARIMA(0,1,1) (a.k.a IMA(1,1)) process is given by

$$y_t - y_{t-1} = \delta + \epsilon_t - \theta\epsilon_{t-1}. \tag{69}$$

The (69) can be written as,

$$(1 - \mathbf{B})y_t = \delta + (1 - \theta\mathbf{B})\epsilon_t. \tag{70}$$

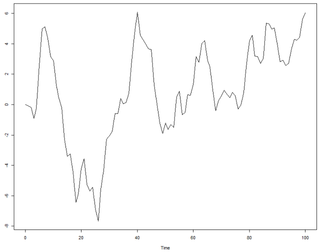
It has an equivalent infinite AR representation:

$$y_t = \alpha + (1 - \theta)(y_{t-1} + \theta y_{t-2} + \theta^2 y_{t-3} + \cdots) + \epsilon_t \tag{71}$$

This suggests that an ARIMA(0,1,1) can be written as an exponentially weighted moving average (EWMA) of all past values.

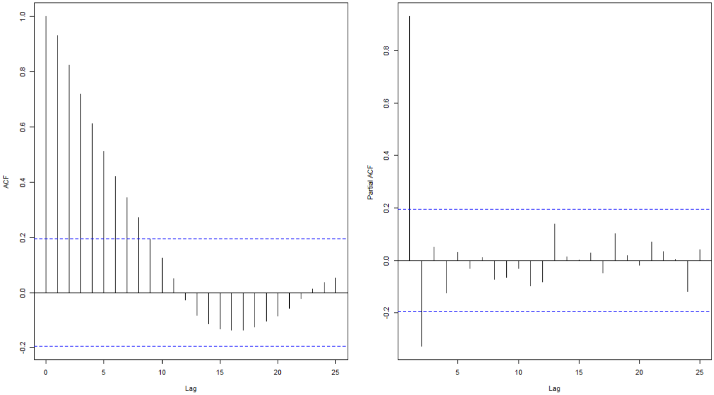
5.6 ARIMA(0,1,1) II

Let us suppose a time series data looks like the following:



The change in the mean (i.e. nonstationarity) is not as obvious as the previous example. It is tempting to ignore the change and look at sample ACF and PACF.

5.6 ARIMA(0,1,1) III

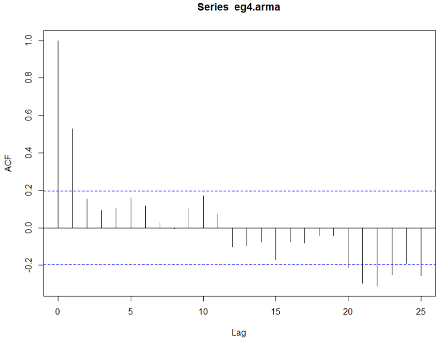


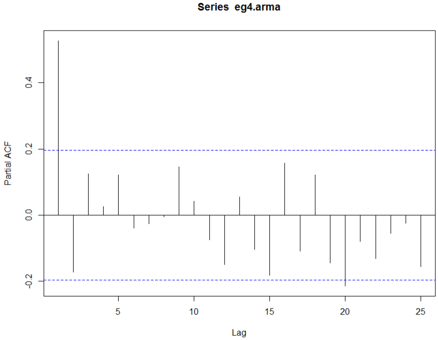
5.6 ARIMA(0,1,1) IV

Then fit an AR(2) model. We might even have a good fit. But we should fit the model (which we will talk about in 5.7) and check the roots of the associated polynomial to make sure that none of the roots are closed to 1. Here one root is found to be 0.86, we may suspect that it is too closed to 1 and therefore try to take a difference of the original series.

We take a difference then plot the sample ACF and PACF for the difference.

5.6 ARIMA(0,1,1) V





The difference forms a MA(1) process meaning the original series follow a ARIMA(0,1,1) model.

5.6 ARIMA(1,1,1) I

ARIMA(1,1,1) has the equation, with $|\phi| < 1$ and $|\theta| < 1$,

$$(1 - \phi\mathbf{B})(1 - \mathbf{B})y_t = \delta + (1 - \theta\mathbf{B})\epsilon_t \tag{72}$$

For example, $\delta = 0$ $\phi = 0.8$ and $\theta = 0.4$, then we have

$$(1 - 0.8\mathbf{B})(1 - \mathbf{B})y_t = (1 - 0.4\mathbf{B})\epsilon_t \tag{73}$$

implying

$$(1 - 1.8\mathbf{B} + 0.8\mathbf{B}^2)y_t = \delta + (1 - \theta\mathbf{B})\epsilon_t \tag{74}$$

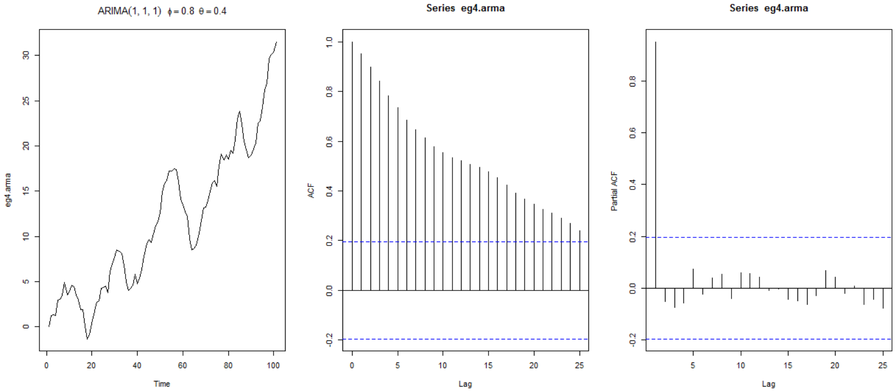
which has the "same" form as ARMA(2,1).

But notice the AR part corresponds a polynomial

$$m^2 - 1.8m + 0.8 = 0$$

which has a unit root, m=1.

A stationary ARMA model does not have this feature.



5.6 ARIMA(1,1,1) III

Then if we do fit an ARMA(2,1) model,

```
> eg.arima.mle=arima(eg4.arma, order=c(2,0,1))
> eg.arima.mle$coef
      ar1       ar2       ma1  intercept
0.1360263 0.8579723 0.9999792 15.0555655
```

and the resulting equation $m^2 - 0.1360263m - 0.8579723 = 0$ has a root 0.9968 which is extremely closed to 1. This is a sign of using ARIMA models. In additional, the time series plot also suggests a difference may be needed.

In general, an ARIMA(p,d,q) model has the same form as an ARMA(p+d,q) model, but its polynomial has d unit roots. Hence ARIMA models are also called "unit-root processes".