Ch. 3: Moving Average Model

Contents

February 22, 2017

1

Residual Analysis

[ToC]

Residual Analysis

[ToC]

After time series model is fit, we want to check the residual for

1. uncorrelated?
2. Heteroscedasticity (non-constant conditional variance)
3. Normality - (not important in ARMA)

given sequence Y_t, we want to see if it is an uncorrelated sequence or not.

1. Plot ACF/PACF
2. Ljung-Box test for randomness
3. McLeod-Li test

1.1 Ljung-Box test

[ToC]

With sample ACF $\hat{\rho}(h)$,

$$Q = n(n+2)\sum_{h=1}^{H} \frac{\hat{\rho}(h)^2}{n-h}$$

Under the null hypothesis of uncorrelation, $Q \sim \chi^2(H)$.

Use as one-sided upper tailed test.

Parameter H must be chosen by hand. Use couple of different values.

Example: L-B test

```
D  <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/gtemp.txt")
D1 <- ts(D, start=c(1880), freq=1)
D2 <- diff(D1)

Plot(D2)

Fit1 <- ar(D2)

A <- Fit1$residuals

Box.test(A,   lag = 15, type = "Ljung-Box")
Box.test(A,   lag = 20, type = "Ljung-Box")
Box.test(A,   lag = 25, type = "Ljung-Box")
```

McLeod-Li test

[ToC]

Instead of testing Y_t, test Y_t^2 for uncorrelation.

Use same test statistic as Ljung-Box, replace Y_t with Y_t^2.

With sample ACF of Y_t^2, $\hat{\rho}(h)$,

$$Q = n(n+2) \sum_{h=1}^{H} \frac{\hat{\rho}(h)^2}{n-h}$$

1.2 Example: McLeod-Li test

```
A <- Fit1$residuals

Box.test(A^2, lag = 15, type = "Ljung-Box")
Box.test(A^2, lag = 20, type = "Ljung-Box")
```

1.3 Test for Normality

[ToC]

Given sequence Y_t, we want to test if it came from Normal distribution. (μ and σ unspecified).

1. q-q normal plot
2. Jarque-Bera test

qq plot

Jarque-Bera Test for Normality

Normal distribution has skewness 0, and kurtosis of 3. for any μ and σ^2.

$$\text{Skewness} = \frac{E[(Y_t - \mu)^3]}{\text{Var}(Y_t)^{3/2}} \qquad \text{Kurtosis} = \frac{E[(Y_t - \mu)^4]}{\text{Var}(Y_t)^2}$$

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right)$$

where S is sample skewness, and K is sample kurtosis.

Asymptotically JB has $\chi^2(2)$ distribution under the normality.

Example: Testing Normality

```
qqnorm(Fit1$resid)

jarque.bera.test(Fit1$resid)


#--- This line copy and paste Basic Functions on class web page
source("http://gozips.uakron.edu/~nmimoto/689/TS_R-90.txt")


#- Use the function as below

ehat <- Fit1$resid

Randomness.tests(ehat)
```

1.4 Summary 1

[ToC]

Residuals from any time-sereis analysis must be checked for:

1. Randomness (for uncorrelation) - (e.g. Ljung-Box test)
2. Heteroscedasticity (Randomness of squared series) - (e.g. McLeod-Li test)
3. Normality - (e.g. qqplot and Jarque-Bera test)
4. Below R code will load the all-in-one function (you need package `tseries` installed already)

```
ehat <- Fit1$resid

source("http://gozips.uakron.edu/~nmimoto/477/TS_R-90.txt")

Randomness.tests(ehat)
```

2

Moving Average Model

[ToC]

2.1 MA(1)

[ToC]

- First order moving average process

$$X_t = e_t - \theta_1 e_{t-1}, \qquad\qquad e_t \sim WN(0, \sigma^2)$$

  and θ_1 is real valued constant.

- Using the backward operator, we can write this as

$$X_t = \underbrace{(1 - \theta_1 B)}_{\Theta(B)} e_t.$$

- Watch the sign in front of θ_1! This is Cryer's convension.

Mean of MA(1)

$$
\begin{aligned}
E(Y_t) &= E(e_t - \theta_1 e_{t-1}) \\
&= E(e_t) - \theta_1 E(e_{t-1}) \\
&= 0
\end{aligned}
$$

ACVF at lag 0

$$\begin{aligned} \mathrm{Var}(Y_t) &= \mathrm{Var}(e_t - \theta_1 e_{t-1}) \\ &= \mathrm{Var}(e_t) + \theta_1^2 \mathrm{Var}(e_{t-1}) \\ &= \sigma^2 + \theta_1^2 \sigma^2 \\ &= (1 + \theta_1^2)\sigma^2 \end{aligned}$$

ACVF at lag 1

$$
\begin{aligned}
\gamma(1) &= \text{Cov}\ (X_t, X_{t+1}) \\
&= \text{Cov}\ (e_t - \theta_1 e_{t-1},\quad e_{t+1} - \theta_1 e_t) \\
&= \text{Cov}\ (e_t,\, e_{t+1}) - \ \text{Cov}\ (e_t,\, \theta_1 e_t) \\
&\quad - \ \text{Cov}\ (\theta_1 e_{t-1},\, e_{t+1}) + \ \text{Cov}\ (\theta_1 e_{t-1},\, \theta_1 e_t) \\
&= -\theta_1\ \text{Cov}\ (e_t,\, e_t) \\
&= -\theta_1 \sigma^2
\end{aligned}
$$

ACVF at lag 2

$$\begin{aligned}\gamma(2) &= \text{Cov}\,(Y_t, Y_{t+2}) \\ &= \text{Cov}\,(e_t - \theta_1 e_{t-1},\quad e_{t+2} - \theta_1 e_{t+1}) \\ &= 0\end{aligned}$$

ACVF and ACF of MA(1)

- ACVF

$$\gamma(h) = \begin{cases} \sigma^2(1+\theta_1^2) & \text{if } h = 0 \\ -\sigma^2\theta_1 & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1 \end{cases}$$

- ACF

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ -\theta_1/(1+\theta_1^2) & \text{if } h = \pm 1 \\ 0 & \text{if } |h| > 1 \end{cases}$$

MA(q)

- Moving Average process of order q is

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q},$$

where $e_t \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ is real valued constant.

- Using the backward operator, we write this as

$$Y_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)e_t.$$

We shorten the notation further, and write

$$Y_t = \Theta(B)Z_t.$$

where $\Theta(z) = (1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q)$.

Mean and Var of MA(q)

- Mean

$$E(X_t) \quad = \quad E(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}) \quad = \quad 0.$$

- Variance

$$\begin{aligned} \mathrm{Var}(X_t) &= \mathrm{Var}(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}) \\ &= \sigma^2 + \theta_1 \, \sigma^2 + \theta_2 \, \sigma^2 + \cdots + \theta_q \, \sigma^2 \\ &= \Big(1 + \theta_1 + \theta_2 + \cdots + \theta_q \Big) \, \sigma^2 \end{aligned}$$

ACVF and ACF of MA(q)

$$\begin{aligned}
\gamma(h) &= \text{Cov}\ (X_t, X_{t+h}) \\
&= \text{Cov}\ (e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}, \\
&\qquad\qquad e_{t+h} - \theta_1 e_{t-1+h} - \theta_2 e_{t-2+h} - \cdots - \theta_q e_{t-q+h})
\end{aligned}$$

Theoretical ACVF:

$$\gamma(h) = \begin{cases}
\sigma^2(1 + \theta_1^2 \cdots + \theta_q^2) & \text{if } h = 0 \\
\sigma^2(-\theta_1 + \theta_2\theta_1 + \theta_3\theta_2 + \cdots + \theta_{q-1}\theta_q) & \text{if } h = 1 \\
\sigma^2(-\theta_2 + \theta_3\theta_1 + \cdots + \theta_{q-2}\theta_q) & \text{if } h = 2 \\
\vdots & \\
\sigma^2(-\theta_q) & \text{if } h = q \\
0 & \text{if } |h| > q
\end{cases}$$

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \gamma(h)/\gamma(0) & \text{if } -q \leq h \leq q \\ 0 & \text{if } |h| > q \end{cases}$$

Theorem:

Every stationary q-corr TS with mean 0 can be represented as MA(q) process.

Example: MA(3)

$$X_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_3 e_{t-3}$$

$$\gamma(h) = \begin{cases} \sigma^2(1 + \theta_1^2 + \theta_2^2 + \theta_3^2) & \text{if } h = 0 \\ \sigma^2(-\theta_1 + \theta_2\theta_1 + \theta_3\theta_2) & \text{if } h = 1 \\ \sigma^2(-\theta_2 + \theta_3\theta_1) & \text{if } h = 2 \\ \sigma^2(-\theta_3) & \text{if } h = 3 \\ 0 & \text{if } |h| > \end{cases}$$

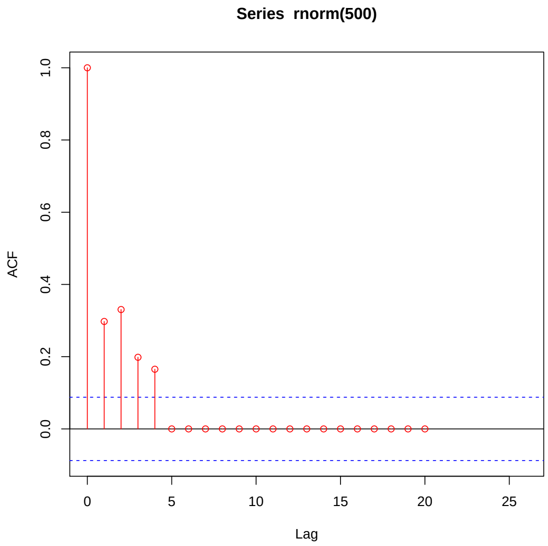
$$\rho(h) = \gamma(h)/\gamma(0)$$

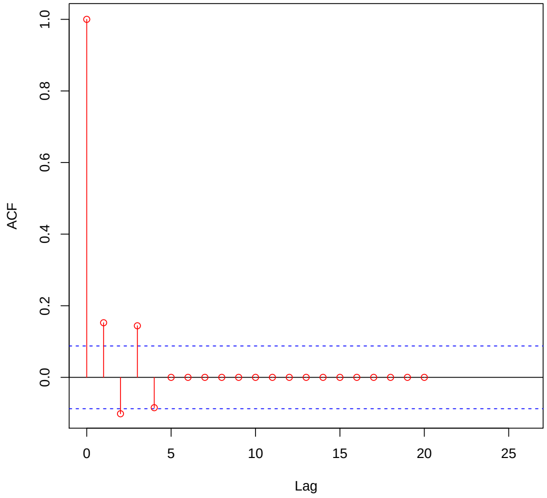
Example: Model ACF of MA(4) - Watch the Sign!

To Summary 2

```
#   [Cryer]'s sign convention is (1-phi B) Y_t = (1-theta) e_t.
#   [Brockwell]'s             is (1-phi B) Y_t = (1+theta) e_t.
#
#   For AR(1) (1-.5 B) Y_t = e_t,          ACF should be all positive.
#             (1+.5 B) Y_t = e_t,          ACF should be alternating in sign.
#
#   For MA(1)          Y_t = (1-.5 B) e_t  ACF at lag 1 should be negative
#                      Y_t = (1+.5 B) e_t  ACF at lag 1 should be positive

#--- Theoretical ACF and PACF of MA ---
#                sign of MA parameter is like [Brockwell]:  Y_t = (1+theta B) e_t

    T1 = c(.2,  .3, .2,  .2)
    T2 = c(.2, -.3, .2, -.2)
    T3 = c(-.2, -.3, -.2, -.2)

    Theta = T1

    MArho1 <- ARMAacf(ma = Theta, lag.max=20, pacf=FALSE)
    MApacf1<- ARMAacf(ma = Theta, lag.max=20, pacf=TRUE)

    layout(matrix(1:2, 1,2))
    plot(0:20, MArho1, type="h", col="red");   abline(h=0)
    plot(1:20, MApacf1, type="h", col="red");   abline(h=0)
```

```
#--- Theoretical ACVF of MA
#
#   sign of MA parameter is like [Cryer]:  Y_t = (1 - theta B) e_t
#

  library(ltsa)


  MAgam2 <- tacvfARMA(theta= c(.2, .3, .2, .2), maxLag=20, sigma2=1)  #- Theoretical ACVF
  MArho2 <- MAgam2/MAgam2[1]                                          #- Theoretical ACF

  plot(0:20, MArho2, type="h", col="red");   abline(h=0)
```


Series rnorm(500)
ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Lag


ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
Lag

Simulation: MA(q)

To Summary 2

```
#--- Basic Simulation with MA(q) ---

  T1 = c(.5)
  T2 = c(.2, -.3, .2, -.2)

  Theta = T1

  x  <- arima.sim(n = 250, list(ma = c(0.5) )    )

  MArho1 <- ARMAacf(ma = Theta, lag.max=20, pacf=FALSE)
  MApacf1<- ARMAacf(ma = Theta, lag.max=20, pacf=TRUE)

  layout(matrix(c(1,1,2,3), 2,2, byrow=TRUE))
  plot(x, type="o");  abline(h=0)

  acf(x);    lines(0:20, MArho1,  type="p", col="red")

  pacf(x);   lines(1:20, MApacf1, type="p", col="red")
```

Characteristic of MA(q)

- ACF cuts off at q
- PACF tails off

2.2 Summary 2

[ToC]

---

1. MA(q) process is defiend as

$$\begin{aligned} X_t &= (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)e_t. \\ &= \Theta(B)e_t \end{aligned}$$

   where $\Theta(z)$ is the characteristic polynomial of MA(q).

2. MA(q) has ACF that cuts-off after lag q, and PACF that decays down.

3. MA(q) has mean of 0. Theoretical ACVF:

$$\gamma(h) = \begin{cases} \sigma^2(1 + \theta_1^2 \cdots + \theta_q^2) & \text{if } h = 0 \\ \sigma^2(-\theta_1 + \theta_2\theta_1 + \theta_3\theta_2 + \cdots + \theta_{q-1}\theta_q) & \text{if } h = 1 \\ \sigma^2(-\theta_2 + \theta_3\theta_1 + \cdots + \theta_{q-2}\theta_q) & \text{if } h = 2 \\ \vdots & \\ \sigma^2(-\theta_q) & \text{if } h = q \\ 0 & \text{if } |h| > q \end{cases}$$

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \gamma(h)/\gamma(0) & \text{if } -q \le h \le q \\ 0 & \text{if } |h| > q \end{cases}$$

4. You can find code for Theoretical ACF and ACVF of MA(q) on page 26.

5. You can find code for Simulating MA(q) process and looking at Saompe ACF and PACF on page 28.

Invertibility of MA(p)

[ToC]

3.1 Invertible Representation of MA(q)

[ToC]

- Suppose we have MA(1) model **(watch the sign!)**

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

  This is already a causal representation.

- We can rewrite the equation and get

$$X_t - \theta_1 e_{t-1} \quad = \quad e_t.$$

- That means we can do the same for X_{t-1}, and write

$$X_{t-1} - \theta_1 e_{t-2} \quad = \quad e_{t-1}$$

  We can substitute this into e_{t-1} above.

- Substituting this expression of e_{t-1} into the first equation,

$$
\begin{aligned}
X_t - \theta_1 e_{t-1} &= e_t. \\
X_t - \theta_1\big(X_{t-1} - \theta_1 e_{t-2}\big) &= e_t \\
X_t - \theta_1 X_{t-1} - \theta_1^2 e_{t-2} &= e_t
\end{aligned}
$$

- Repeat this n times, we get

$$X_t - \theta_1 X_{t-1} - \cdots - \theta_1^{n-1} X_{t-n+1} - \theta_1^n e_{t-n} \quad = \quad e_t$$

If $|\theta_1| < 1$, then we can write

$$e_t \quad = \quad \sum_{i=0}^{\infty} \theta_1^i \, X_{t-i}$$

This is called invertible representation.

Invertible Representation for MA(q)

- $$\begin{aligned} e_t &= \sum_{i=0}^{\infty} \pi_i X_{t-i} \\ &= \left(\pi_0 + \pi_1 B + \pi_2 B^2 + \pi_3 B^3 + \cdots\right) X_t \\ &= \Pi(B)\, X_t \end{aligned}$$

- So for our MA(q) model

$$\begin{aligned} X_t &= \Theta(B) e_t \\ \Pi(B)\, X_t &= e_t \end{aligned}$$

- Very similar to causal representation in AR(p), we have identity,

$$\Pi(z) = \frac{1}{\Theta(z)}$$

$$(\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots) = \frac{1}{(1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q)}$$

- We can calculate coefficients by moving all the polynomials to the left and matching

$$(1 - \theta_1 z - \theta_2 z^2 - \cdots - \theta_q z^q)(\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots) \;=\; 1$$

- For example, if we have MA(1),

$$(1 - \theta_1 z)(\pi_0 + \pi_1 z + \pi_2 z^2 + \pi_3 z^3 + \cdots) \;=\; 1$$

- To find out the coeffeicnets π_i, we can match the order of z.

$$\begin{aligned}
\pi_0 &= 1 && \text{(coefficient without } z \text{)} \\
-\theta_1 \pi_0 + \pi_1 &= 0 && \text{(coefficient of } z\text{)} \\
-\theta_1 \pi_1 + \pi_2 &= 0 && \text{(coefficient of } z^2\text{)} \\
-\theta_1 \pi_2 + \pi_3 &= 0 && \text{(coefficient of } z^3\text{)} \\
&\vdots
\end{aligned}$$

$$
\begin{aligned}
\pi_0 &= 1 \\
\pi_1 &= \theta_1 \pi_0 \\
\pi_2 &= \theta_1 \pi_1 \\
\pi_3 &= \theta_1 \pi_2 \\
&\vdots
\end{aligned}
$$

$$
\begin{aligned}
\pi_0 &= 1 \\
\pi_1 &= \theta_1 \\
\pi_2 &= \theta_1^2 \\
\pi_3 &= \theta_1^3 \\
&\vdots
\end{aligned}
$$

Invertible Representation of MA(1)

- Now we get invertible representation for MA(1),

$$e_t = \sum_{i=0}^{\infty} \pi_i X_{t-i} = \sum_{i=0}^{\infty} \theta_1^i X_{t-i}.$$

- To write the today's error (innovation) e_t, we need infinetely many past observation X_{t-i}.

- When can we do this?

3.2 Invertibility Condition

[ToC]

- If all roots of characteristic polynomial

$$\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q$$

  is outside of the unit circle, then MA(q) admits a invertible representation.

- Same as causal condition in AR(p).

- Watch the sign! We have same sign as $\Phi(z)$ in AR(p), because we started with Brockwell's convention; $X_t = e_t + \theta_1 e_{t-1}$.

- We assume all MA(q) we deal with are invertible.

Why Invertibility Condition is Important

- If we have MA(3) process,

$$X_t = e_t - \hat{\theta}_1 e_{t-1} - \hat{\theta}_2 e_{t-2} - \hat{\theta}_3 e_{t-3}$$

  with estimated parameter $\hat{\theta}_i$, we want to get residuals $\hat{e}_t$, and check their randomness.

- In AR(p), this was intuitive. e.g. if we had AR(2),

$$X_t - \hat{\phi}_1 X_{t-1} - \hat{\phi}_2 X_{t-2} = \hat{e}_t$$

- In MA(q), we must use invertible expression,

$$\hat{e}_t = \hat{\pi}_0 X_t + \hat{\pi}_1 X_{t-1} + \hat{\pi}_2 X_{t-2} + \cdots$$

Why Invertibility Condition is Important - 2

- The best 1-step linear predictor of AR(3) process was

$$\hat{X}(1) = \phi_1 X_n + \phi_2 X_{n-1} + \phi_3 X_{n-2}.$$

  This was intuitive since

$$X_{t+1} = \phi_1 X_n + \phi_2 X_{n-1} + \phi_3 X_{n-2} + e_{t+1}.$$

- Would the same happens for MA(q)?

- Take MA(1). We have

$$X_{t+1} = e_{t+1} + \theta_1 e_t.$$

Since e_{t+1} is uncorrelated new innovation (error) with mean 0, we would guess that the best predictor may look like

$$\hat{X}(1) = -\theta_1 e_t.$$

- e_t is yesterday's innovation. It has already been generated, but it's not observable. Then we can replace it with our invertible representation,

$$\hat{X}(1) = -\theta_1 \hat{e}_t \qquad \hat{e}_t = \sum_{i=1}^{n} \pi_i X_{t-i}.$$

- For MA(1), we know π_i are

$$\begin{aligned}\hat{X}(1) &= -\theta_1\big(\pi_0 X_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + \pi_{t-1} X_1\big) \\ &= -\theta_1\big(X_t + \theta_1 X_{t-1} + \theta_1^2 X_{t-2} + \cdots + \theta_1^{t-1} X_1\big)\end{aligned}$$

This actually turns out to be best linear predictor for X_{t+1}. (minimum MSE)

3.3 AR with non-causal ϕ_1.

1. **Nothing is lost:** We assume that our ARMA model is causal and invertible. However if they are not, you can always rewrite the ARMA into causal and invertible ARMA with new set of white noise.

2. Suppose our AR(1),

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

has $|\phi_1| > 1$. Then this is not causal AR. Suppose there's an stationary solution to this equation. ACVF formula is still valid, and

$$\gamma(0) = \frac{\sigma^2}{1-\phi_1^2}, \qquad \gamma(h) = \phi_1^{|h|}\gamma(0)$$

3. Then, you can rewite this AR(1) as

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

where ε_t is another WN$(0, \sigma_\varepsilon^2)$.

"New" WN sequence is

$$\varepsilon_t \quad = \quad Y_t - \frac{1}{\phi_1} Y_{t-1}.$$

Let's see ACVF of ε_t to see if it's really WN.

$$\begin{aligned}
\gamma(1) = E(\epsilon_t \epsilon_{t-1}) \quad &= \quad E\Big[\Big(Y_t - \frac{1}{\phi_1} Y_{t-1}\Big)\Big(Y_{t-1} - \frac{1}{\phi_1} Y_{t-2}\Big)\Big] \\
&= \quad E(Y_t Y_{t-1}) - \frac{1}{\phi_1} E(Y_t Y_{t-2}) - \frac{1}{\phi_1} E(Y_{t-1}^2) + \frac{1}{\phi_1^2} E(Y_{t-1} Y_{t-2}) \\
&= \quad \gamma(1) - \frac{1}{\phi_1}\gamma(2) - \frac{1}{\phi_1}\gamma(0) + \frac{1}{\phi_1^2}\gamma(1) \\
&= \quad \phi_1 \gamma(0) - \frac{1}{\phi_1}\phi_1^2 \gamma(0) - \frac{1}{\phi_1}\gamma(0) + \frac{1}{\phi_1^2}\phi_1\gamma(0) \quad = \quad 0
\end{aligned}$$

$\gamma(h)$ $h > 1$ cancells out simillarly.

Variance of ε_t is

$$
\begin{aligned}
V(\varepsilon_t) &= E\left[\left(Y_t - \frac{1}{\phi_1}Y_{t-1}\right)^2\right] \\
&= E(Y_t^2) - 2\frac{1}{\phi_1}E(Y_tY_{t-1}) + \frac{1}{\phi_1^2}E(Y_{t-1}^2) \\
&= \gamma(0) - 2\frac{1}{\phi_1}\gamma(1) + \frac{1}{\phi_1^2}\gamma(0) \quad = \quad \gamma(0)\left(1 - 2 + \frac{1}{\phi_1^2}\right) \\
&= \frac{\sigma^2}{1-\phi_1^2}\left(\frac{1}{\phi_1^2} - 1\right) \quad = \quad \frac{\sigma^2}{1-\phi_1^2}\left(\frac{1-\phi_1^2}{\phi_1^2}\right) \quad = \quad \frac{\sigma^2}{\phi_1^2}
\end{aligned}
$$

So the stationary solution to

$$Y_t = \phi_1 Y_{t-1} + e_t \qquad \text{with } e_t \sim \mathrm{WN}(0, \sigma^2)$$

And

$$Y_t = \frac{1}{\phi_1} Y_{t-1} + \varepsilon_t \qquad \text{with } \varepsilon_t \sim \mathrm{WN}(0, \sigma^2/\phi_1^2)$$

are the same. This is why we don't consider non-causal ARMA process.

3.4 Summary 3

[ToC]

---

1. MA(q) process is defiend with equation

$$X_t = \Theta(B)e_t$$

you can check the roots of the characteristic polynomial

$$\Theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q.$$

If all root are outside of the (complex) unit circle, then MA(q) admits a invertible representation.

2. Invertible representation allows to write today's error using infinite sum of past observations.

$$e_t = \sum_{i=0}^{\infty} X_{t-i}$$

3. Bcause we can't directly observe e_t, invertibility is importatnt in calculating residuals and forecasts.

Fitting MA(q) model

[ToC]

Fitting MA model

- Order selection
    1. Select best value for q using AIC, AICc, BIC, Log Likelihood
- Estimation
    1. Maximum Likelihood Method
- Prediction

4.1 Order Selection

1. Plot of ACF and PACF

2. AIC and AICc

3. BIC

4. Log-Likelihood

4.2 AICc

$$\text{AICC} = -2\log(\text{Max Likehood}) + \frac{2(k+1)(k+2)}{n-k-2}$$

where $k = p + q + 1$ if $demean = T$, and $k = p + q$ if $demean = F$.

4.3 BIC

Bayesian Information Criteria

$$\text{BIC} = -2\log(\text{Max Likehood}) + k\log(n)$$

where $k = p + q + 1$ if $demean = T$, and $k = p + q$ if $demean = F$.

Simulation Study: Using AICc better?

How much better is AICC over AIC? In what scenario?

- Simulate AR(2)
- Pick AR(P) model based on AICC, AIC, BIC.
- Repeat 1000 times, see how many times each criteria picked right p.

```
#- install.packages("forecast")

library(forecast)

Result1 <- 0
Result2 <- 0
Result3 <- 0
for (i in 1:1000){

  #Y <- arima.sim(list(ar = c(.6, -.6), ma=c(.8) ), 100 ) + 10  #.75 vs .79
  #Y <- arima.sim(list(ar = c(.6, -.3), ma=c(.5) ), 100 ) + 10

  Y <- arima.sim(list(ma=c(.8, -.4) ), 100 ) + 10

  #- picks model based on AICC, AIC, and BIC
  Fit1 <- auto.arima(Y, max.order=6, max.d=0, max.D=0,  ic=c("aicc"))
  Fit2 <- auto.arima(Y, max.order=6, max.d=0, max.D=0,  ic=c("aic"))
  Fit3 <- auto.arima(Y, max.order=6, max.d=0, max.D=0,  ic=c("bic"))

  print(i) #- print order on screen (optional)

  Result1[i] <- (sum(Fit1$arma==c(0,2,0,0,1,0,0))==7)
  Result2[i] <- (sum(Fit1$arma==c(0,2,0,0,1,0,0))==7)
  Result3[i] <- (sum(Fit1$arma==c(0,2,0,0,1,0,0))==7)
}

c(sum(Result1),  sum(Result2),  sum(Result3))/1000
```

4.4 Example: Daily overshorts

57 consecutive daily overshorts from an underground gasoline tank at a filling station in Colorado.

```
library(forecast)

D1 <- read.csv("http://gozips.uakron.edu/~nmimoto/pages/datasets/oshorts.csv", header=F)
D  <- ts(D1, start=1, freq=1)

layout(matrix(c(1,1,2,3), 2,2, byrow=T))
plot(D, type="o")
acf(D)
pacf(D)

Fit1 <- auto.arima(D, d=0  )  #- search for best ARMA(p,q) using AICc (default)
Fit1

Fit2 <- auto.arima(D, d=0, ic="bic" )  #- search for best ARMA(p,q) using BIC
Fit2

Fit3 <- Arima(D, order=c(2,0,0))   #- Force to fit AR(2)

Fit4 <- Arima(D, order=c(2,0,0), include.mean=FALSE)   #- Force to fit AR(2) without intercept

source("http://gozips.uakron.edu/~nmimoto/477/TS_R-90.txt")

Randomness.tests(Fit1$resid)
```

```
#--- Predicting Temparature 10 step ahead

  Y <- D

  Fit1 <- Arima(Y, order=c(0,0,1)   )   #- Force to fit MA(1)
  Y.h  <- predict(Fit1, n.ahead=10)
  Yhat <- Y.h$pred
  Yhat.CIu <- Yhat+1.96*Y.h$se
  Yhat.CIl <- Yhat-1.96*Y.h$se

  ts.plot(cbind(Y, Yhat, Yhat.CIu, Yhat.CIl ), type="o", col=c("black","red","blue","blue"))
  abline(h=0)
  abline(h=mean(Y), col="blue")
```

```
#--- Rolling 1-step predicton of last 29 years

  Rolling.len = 10
  Window.size = 47

  p = 0
  d = 0
  q = 1

  Y <- D


  Yhat <- Yhat.CIu <- Yhat.CIl <- Y2<- 0 #- Initialize
  for (i in 1:Rolling.len) {
    window.bgn <- i
    window.end <- i+Window.size-1
    Fit1 <- Arima(Y[window.bgn:window.end], order=c(p,d,q) ) #- Force to fit AR(p)
    Y.h <- predict(Fit1, n.ahead=1)
    Yhat[i] <- Y.h$pred
    Yhat.CIu[i] <- Yhat[i]+1.96*Y.h$se
    Yhat.CIl[i] <- Yhat[i]-1.96*Y.h$se
  }
```

```
X  <- window(Y, start=time(Y)[1], end=time(Y)[Window.size])
Y2 <- window(Y, start=time(Y)[Window.size+1], end=time(Y)[Window.size+Rolling.len])

Yhat       <- ts(Yhat,     start=c(floor(time(Y)[Window.size+1]), cycle(Y)[Window.size+1]), freq=frequency(Y)    )
Yhat.CIu   <- ts(Yhat.CIu, start=c(floor(time(Y)[Window.size+1]), cycle(Y)[Window.size+1]), freq=frequency(Y)    )
Yhat.CIl   <- ts(Yhat.CIl, start=c(floor(time(Y)[Window.size+1]), cycle(Y)[Window.size+1]), freq=frequency(Y)    )

Pred.error <- Y2-Yhat

Pred.rMSE = sqrt( mean( (Pred.error)^2 ) ) #- prediction root Mean Squared Error
Pred.rMSE

mean(Pred.error)

layout(matrix(c(1,1,1,2,2,3), 2, 3, byrow=TRUE))
plot(Y, type="o", col="blue", main="Rolling 1-step prediction with window=100" ) #- Entire dataset
lines(X, type="o")
lines(Yhat, type="o", col="red")
lines(Yhat.CIu, type="l", col="red", lty=2)
lines(Yhat.CIl, type="l", col="red", lty=2)

plot(Pred.error, type="o", main="Prediction Error (Blue-Red)")
abline(h=c(-1.96, 1.96), col="blue", lty=2)

acf(Pred.error)
```

1. You can compare different AR(p), MA(q) models with AIC, AICC, BIC, Log Likelihood

2. Check the residuals for model adequacy.

```
#--- Predicting Temparature 10 step ahead

  D  <- arima.sim(list(ma=c(.6, .3) ), 100 ) + 10

  Y <- D

  Fit1 <- Arima(Y, order=c(0,0,2)   )   #- Force to fit MA(2)
  Fit1

  Y.h  <- predict(Fit1, n.ahead=10)
  Yhat <- Y.h$pred
  Yhat.CIu <- Yhat+1.96*Y.h$se
  Yhat.CIl <- Yhat-1.96*Y.h$se

  ts.plot(cbind(Y, Yhat, Yhat.CIu, Yhat.CIl ), type="o", col=c("black","red","blue","blue"))
  abline(h=0)
  abline(h=mean(Y), col="blue")
```