

# STAT33600 Homework 4

Sarah Adilijiang

## Section 3.2 Difference Equations

### 3.7 ACF of AR(2) series

(a)

$$x_t + 1.6x_{t-1} + 0.64x_{t-2} = w_t$$

$$E(x_t x_{t-h}) + 1.6E(x_{t-1} x_{t-h}) + 0.64E(x_{t-2} x_{t-h}) = E(w_t x_{t-h}) \quad (h > 0)$$

$$\rho(h) + 1.6\rho(h-1) + 0.64\rho(h-2) = 0 \quad (h = 1, 2, \dots) \quad (*)$$

And the roots for polynomial  $\phi(z) = 1 + 1.6z + 0.64z^2$  are the real and equal root  $z_0 = -1.25$ . Thus for the sequence  $\rho(h)$  which satisfies the function (\*), we have that:

$$\rho(h) = z_0^{-h}(c_1 + c_2 h) = (-1.25)^{-h}(c_1 + c_2 h) = (-0.8)^h(c_1 + c_2 h)$$

The initial conditions are:  $\rho(0) = 1$ ,  $\rho(1) = -1.6/(1 + 0.64) = -40/41$ . Use this to solve for  $c_1, c_2$ :  $c_1 = 1$ ,  $-0.8(c_1 + c_2) = -40/41$ , we get that:  $c_1 = 1$ ,  $c_2 = 9/41 \approx 0.22$ .

Therefore:

$$\rho(h) = (-0.8)^h(1 + 0.22h) \quad (h = 0, 1, 2, \dots)$$

Now we check the results using ARMAacf in R:

```
# results from my computation
h = (0:10)
rho = (-0.8)^h * (1+9/41*h)

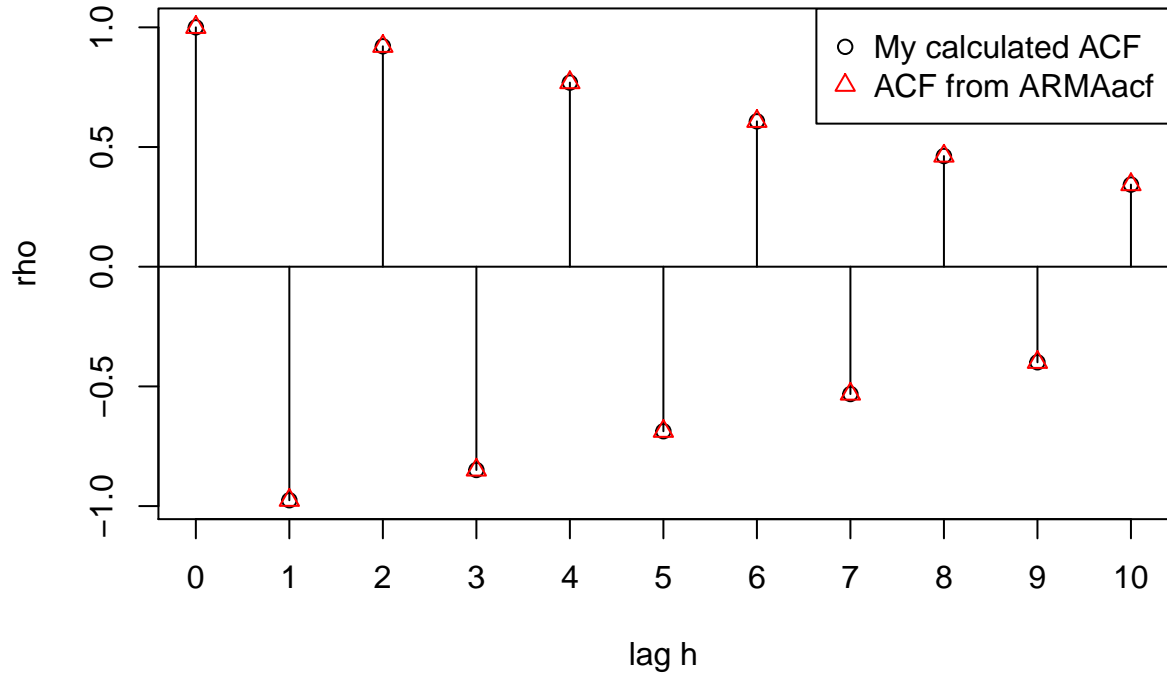
# results from ARMAacf function
ACF = ARMAacf(ar=c(-1.6,-0.64), ma=0, lag.max=10)

# compare results
rbind(rho, ACF)
```

```
##      0      1      2      3      4      5      6
## rho 1 -0.9756098 0.9209756 -0.8491707 0.7692488 -0.6873288 0.6074068
## ACF 1 -0.9756098 0.9209756 -0.8491707 0.7692488 -0.6873288 0.6074068
##      7      8      9     10
## rho -0.5319605 0.4623964 -0.3993796 0.3430736
## ACF -0.5319605 0.4623964 -0.3993796 0.3430736
```

The results from my computation are exactly the same as those from the ARMAacf function.

```
# plot calculated ACF and theoretical ACF
plot((0:10), rho, type="h", xlab="lag h", xaxt="n", col=1)
points((0:10), rho, pch=1, col=1)
points((0:10), ACF, pch=2, col=2)
legend("topright", legend=c("My calculated ACF", "ACF from ARMAacf"), pch=c(1,2), col=c(1,2))
axis(at=(0:10), side=1)
abline(h=0)
```



(b)

$$x_t - 0.40x_{t-1} - 0.45x_{t-2} = w_t$$

$$E(x_t x_{t-h}) - 0.40E(x_{t-1} x_{t-h}) - 0.45E(x_{t-2} x_{t-h}) = E(w_t x_{t-h}) \quad (h > 0)$$

$$\rho(h) - 0.40\rho(h-1) - 0.45\rho(h-2) = 0 \quad (h = 1, 2, \dots) \quad (*)$$

And the roots for polynomial  $\phi(z) = 1 - 0.40z - 0.45z^2$  are real and distinct roots  $z_1 = -2$ ,  $z_2 = 10/9$ . Thus for the sequence  $\rho(h)$  which satisfies the function (\*), we have that:

$$\rho(h) = c_1 z_1^{-h} + c_2 z_2^{-h} = c_1 (-2)^{-h} + c_2 (10/9)^{-h} = c_1 (-0.5)^h + c_2 (0.9)^h$$

The initial conditions are:  $\rho(0) = 1$ ,  $\rho(1) = 0.4/(1 - 0.45) = 8/11$ . Use this to solve for  $c_1, c_2$ :  $c_1 + c_2 = 1$ ,  $-0.5c_1 + 0.9c_2 = 8/11$ , we get that:  $c_1 = 19/154 \approx 0.12$ ,  $c_2 = 135/154 \approx 0.88$ .

Therefore:

$$\rho(h) = 0.12(-0.5)^h + 0.88(0.9)^h \quad (h = 0, 1, 2, \dots)$$

Now we check the results using ARMAacf in R:

```
# results from my computation
h = (0:10)
rho = 19/154 * (-0.5)^h + 135/154 * (0.9)^h

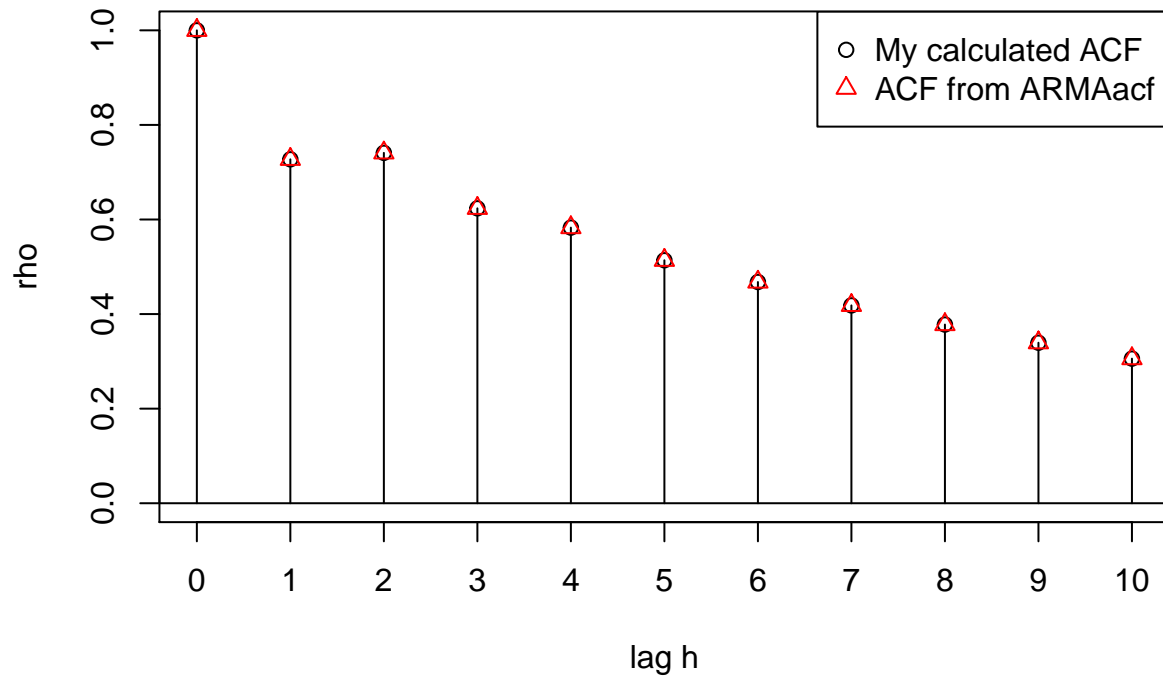
# results from ARMAacf function
ACF = ARMAacf(ar=c(0.4,0.45), ma=0, lag.max=10)
```

```
# compare results
rbind(rho, ACF)
```

```
##      0      1      2      3      4      5      6
## rho 1 0.7272727 0.7409091 0.6236364 0.5828636 0.5137818 0.4678014
## ACF 1 0.7272727 0.7409091 0.6236364 0.5828636 0.5137818 0.4678014
##
##      7      8      9     10
## rho 0.4183224 0.3778396 0.3393809 0.3057802
## ACF 0.4183224 0.3778396 0.3393809 0.3057802
```

The results from my computation are exactly the same as those from the ARMAacf function.

```
# plot calculated ACF and theoretical ACF
plot((0:10), rho, type="h", xlab="lag h", xaxt="n", col=1, ylim=c(0,1))
points((0:10), rho, pch=1, col=1)
points((0:10), ACF, pch=2, col=2)
legend("topright", legend=c("My calculated ACF", "ACF from ARMAacf"), pch=c(1,2), col=c(1,2))
axis(at=(0:10), side=1)
abline(h=0)
```



(c)

$$\begin{aligned}
 x_t - 1.2x_{t-1} + 0.85x_{t-2} &= w_t \\
 E(x_t x_{t-h}) - 1.2E(x_{t-1} x_{t-h}) + 0.85E(x_{t-2} x_{t-h}) &= E(w_t x_{t-h}) \quad (h > 0) \\
 \rho(h) - 1.2\rho(h-1) + 0.85\rho(h-2) &= 0 \quad (h = 1, 2, \dots) \quad (*)
 \end{aligned}$$

And the roots for polynomial  $\phi(z) = 1 - 1.2z + 0.85z^2$  are a complex conjugate pair  $z_1, z_2 = 12/17 \pm 14/17 i \approx 0.71 \pm 0.82i$ . Thus for the sequence  $\rho(h)$  which satisfies the function (\*), we have that:

$$\rho(h) = a|z_1|^{-h} \cos(h\theta + b) = a |12/17 + 14/17i|^{-h} \cos(h\theta + b) = a \left(\sqrt{\frac{17}{20}}\right)^h \cos(h\theta + b) \approx a (0.92)^h \cos(h\theta + b)$$

where

$$\theta = \arg(z_1) = \arg(12/17 + 14/17i) \approx 0.86$$

The initial conditions are:  $\rho(0) = 1$ ,  $\rho(1) = 1.2/(1 + 0.85) = 24/37$ , use this to solve for  $a, b$ :  $a \cos(b) = 1$ ,  $0.92a \cos(\theta + b) = 24/37$ , we get that:  $a \approx 1$ ,  $b = 0$  or  $a \approx -1$ ,  $b = \pi$ .

Using equation  $\cos(\alpha + \beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ , the computations for above results are shown below:

```
z = complex(real=12/17, imaginary=14/17)
theta = Arg(z)
a = sqrt( ( (24/37 * Mod(z) - cos(theta))/sin(theta) )^2 + 1 ); a
```

```
## [1] 1.002412
```

Therefore:

$$\rho(h) = (0.92)^h \cos(0.86h) \quad (h = 0, 1, 2, \dots)$$

or:

$$\rho(h) = -(0.92)^h \cos(0.86h + \pi) \quad (h = 0, 1, 2, \dots)$$

And these two equations are actually the same since  $\cos(0.86h + \pi) = -\cos(0.86h)$ .

Now we check the results using ARMAacf in R:

```
# results from my computation
h = (0:10)
rho = (0.92)^h * cos(0.86*h)

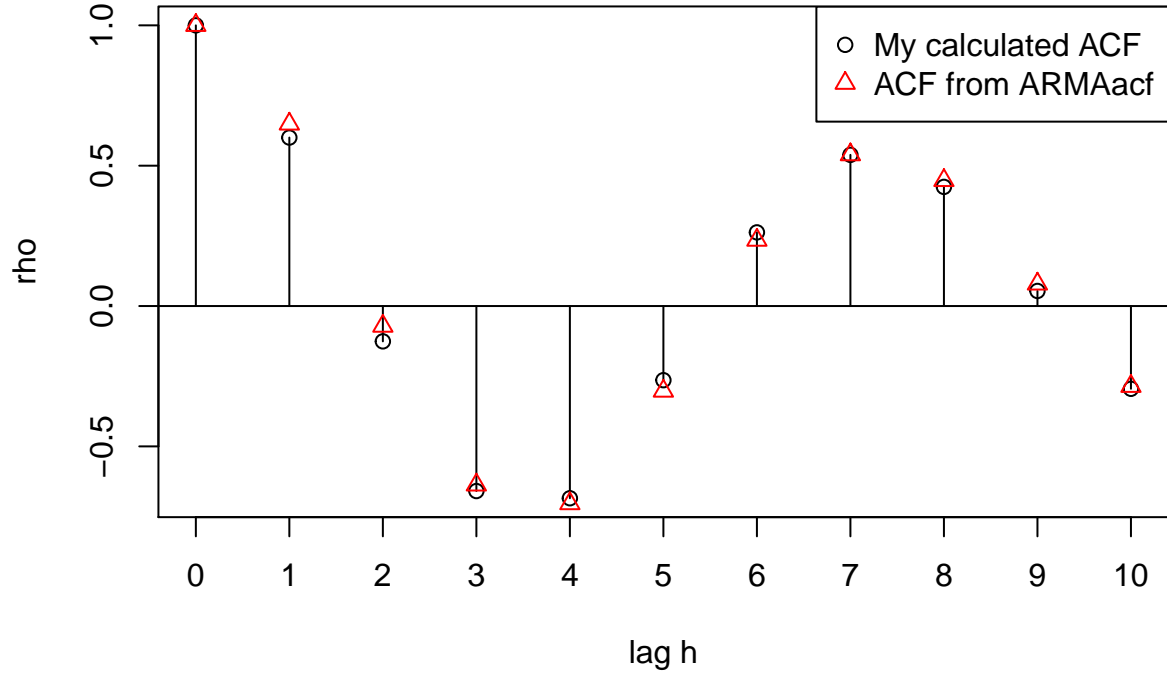
# results from ARMAacf function
ACF = ARMAacf(ar=c(1.2,-0.85), ma=0, lag.max=10)

# compare results
rbind(rho, ACF)
```

```
##      0      1      2      3      4      5      6
## rho 1 0.6002425 -0.12581795 -0.6590878 -0.6847326 -0.2641593 0.2624384
## ACF 1 0.6486486 -0.07162162 -0.6372973 -0.7038784 -0.3029514 0.2347550
##      7      8      9     10
## rho 0.5386378 0.4244987 0.05370127 -0.2948282
## ACF 0.5392146 0.4475158 0.07868654 -0.2859646
```

The results from my computation are roughly similar with those from the ARMAacf function.

```
# plot calculated ACF and theoretical ACF
plot((0:10), rho, type="h", xlab="lag h", xaxt="n", col=1)
points((0:10), rho, pch=1, col=1)
points((0:10), ACF, pch=2, col=2)
legend("topright", legend=c("My calculated ACF", "ACF from ARMAacf"), pch=c(1,2), col=c(1,2))
axis(at=(0:10), side=1)
abline(h=0)
```



## Section 3.3 Autocorrelation and Partial Autocorrelation

### 3.8

#### (a) Verification of ACF function in ARMA(1,1)

The ARMA(1,1) process:

$$x_t = \phi x_{t-1} + \theta w_{t-1} + w_t \quad (|\phi| < 1)$$

The autocorrelation function for  $h \geq 2$  is:

$$\begin{aligned} \gamma(h) &= \text{Cov}(x_{t+h}, x_t) = \text{Cov}(\phi x_{t+h-1} + \theta w_{t+h-1} + w_{t+h}, x_t) \\ &= \phi \gamma(h-1) + \theta \text{Cov}(w_{t+h-1}, \sum_{j=0}^{\infty} \psi_j w_{t-j}) + \text{Cov}(w_{t+h}, \sum_{j=0}^{\infty} \psi_j w_{t-j}) \\ &= \phi \gamma(h-1) + \theta \psi_{1-h} \sigma_w^2 + \psi_{-h} \sigma_w^2 \end{aligned}$$

So the autocorrelation function satisfies:

$$\gamma(h) - \phi \gamma(h-1) = 0 \quad (h = 2, 3, \dots)$$

And its general solution is:

$$\gamma(h) = \phi \gamma(h-1) = \phi^2 \gamma(h-2) = \dots = \phi^{h-1} \gamma(1) = c \phi^h \quad (h = 1, 2, \dots), \quad \text{where } c = \frac{\gamma(1)}{\phi} \quad (*)$$

Since:

$$x_t = \phi x_{t-1} + \theta w_{t-1} + w_t = \sum_{j=0}^{\infty} \psi_j w_{t-j}$$

so we have:  $\psi_0 = 1, \psi_1 = \theta + \phi$

Thus the initial conditons are:

$$\gamma(1) = \phi\gamma(0) + \theta\psi_0\sigma_w^2 + \psi_{-1}\sigma_w^2 = \phi\gamma(0) + \theta\sigma_w^2 \quad (1)$$

$$\gamma(0) = \phi\gamma(-1) + \theta\psi_1\sigma_w^2 + \psi_0\sigma_w^2 = \phi\gamma(1) + [\theta(\theta + \phi) + 1]\sigma_w^2 \quad (2)$$

Substitutue (1) into (1), we get:

$$\gamma(0) = \phi(\phi\gamma(0) + \theta\sigma_w^2) + [\theta(\theta + \phi) + 1]\sigma_w^2 = \phi^2\gamma(0) + (\theta^2 + 2\phi\theta + 1)\sigma_w^2$$

$$\Rightarrow \gamma(0) = \sigma_w^2 \frac{\theta^2 + 2\phi\theta + 1}{1 - \phi^2}$$

$$\Rightarrow \gamma(1) = \phi\sigma_w^2 \frac{\theta^2 + 2\phi\theta + 1}{1 - \phi^2} + \theta\sigma_w^2 = \sigma_w^2 \frac{(\theta + \phi)(1 + \theta\phi)}{1 - \phi^2}$$

Take the value of  $\gamma(1)$  into the function (\*), we can get that:

$$\gamma(h) = c \phi^h = \phi^{h-1}\gamma(1) = \phi^{h-1}\sigma_w^2 \frac{(\theta + \phi)(1 + \theta\phi)}{1 - \phi^2} \quad (h = 1, 2, \dots)$$

Therefore, the ACF function is:

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \begin{cases} 1 & \text{if } h = 0 \\ \frac{(\theta + \phi)(1 + \theta\phi)}{\theta^2 + 2\phi\theta + 1} \phi^{h-1} & \text{if } h \geq 1 \end{cases}$$

## (b) Compare ARMA(1,1) with ARMA(1,0) and ARMA(0,1)

(1) ARMA(1,0) = AR(1)

For AR(p), the results in textbook (3.50) shows that the general ACF of an AR(P) is:

$$\rho(h) = z_1^{-h}P_1(h) + z_2^{-h}P_2(h) + \dots + z_r^{-h}P_r(h) \quad (h \geq p)$$

where  $z_1, \dots, z_r$  are the roots of  $\phi(z) = 1 - \phi_1z - \phi_2z^2 - \dots - \phi_pz^p$ , each with multiplicity  $m_1, \dots, m_r$ , respectively, where  $m_1 + \dots + m_r = p$ . And  $P_j(h)$  is a polynomial in  $h$  of degree  $m_j - 1$ .

So for AR(1) series:

$$x_t = \phi x_{t-1} + w_t \quad (\phi_1 = \phi, |\phi| < 1)$$

The ACF of an AR(1) is:

$$\rho(h) = z_1^{-h}P_1(h) = (1/\phi)^{-h} \times c = c \phi^h \quad (h \geq 1)$$

where  $z_1 = 1/\phi$  is the root of  $\phi(z) = 1 - \phi z$ , each with multiplicity  $m_1 = 1$ . And  $P_1(h) = ch^0 = c$  is a polynomial in  $h$  of degree  $m_1 - 1 = 0$ .

Use the initial condition  $\rho(0) = c = 1$ , we get:  $c = 1$ .

Therefore, the ACF is:

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \phi^h = \phi \phi^{h-1} & \text{if } h \geq 1 \end{cases}$$

(2) ARMA(0,1) = MA(1)

For MA(q), the results in textbook (3.43) shows that the ACF of an MA(q) is:

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\sum_{j=0}^{q-h} \theta_j \theta_{j+h}}{1+\theta_1^2+\dots+\theta_q^2} & \text{if } 1 \leq h \leq q \\ 0 & \text{if } h > q \end{cases}$$

So for MA(1) series:

$$x_t = \theta w_{t-1} + w_t \quad (\theta_1 = \theta, \theta_0 = 1)$$

The ACF of an MA(1) is:

$$\rho(h) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{if } h = 1 \\ 0 & \text{if } h > 1 \end{cases}$$

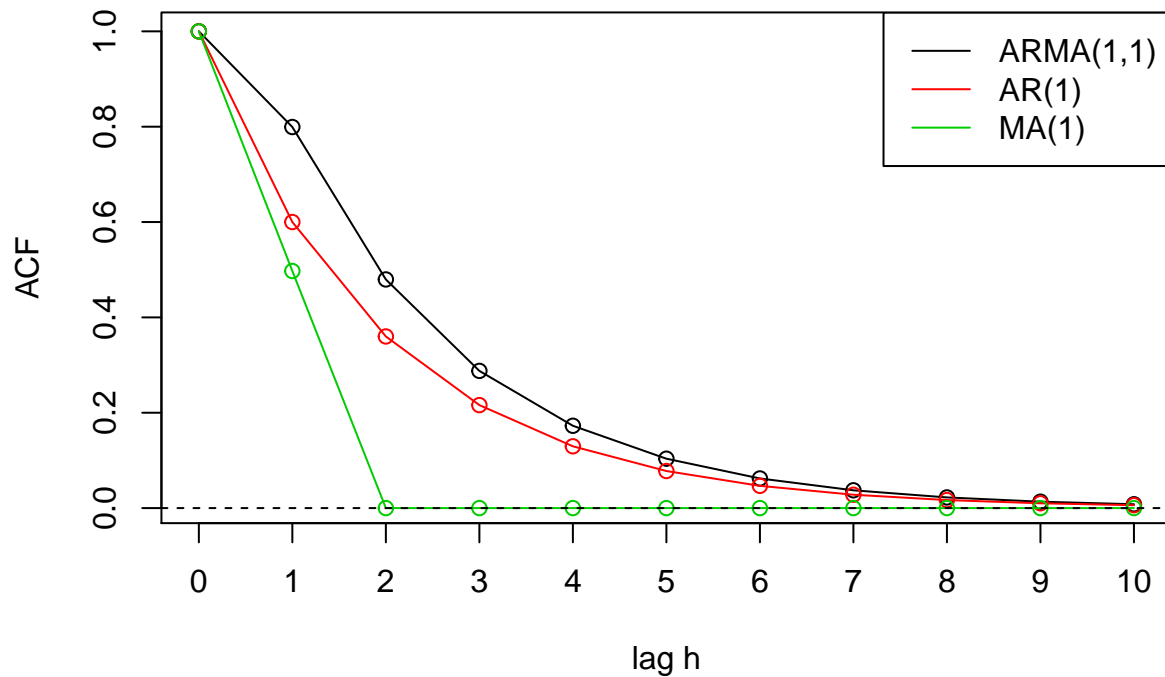
(3) Comparing ARMA(1,1) with AR(1) and MA(1)

The ACF of an MA(1) only has first two non-zero values and equals to 0 for all  $h > 1$ . Therefore, it is very easy to distinguish MA(1) series using only the sample ACF function plot.

However, the ACF of ARMA(1,1) and AR(1) are both in the form of  $\rho(h) = c\phi^{h-1}$  for  $h \geq 1$ , whose difference is only within the constant part  $c$ . And for both of them,  $\rho \rightarrow 0$  exponentially fast as  $h \rightarrow \infty$ . As a result, their sample ACF plot will look very similar to each other. Therefore, we will be not able to tell the difference between ARMA(1,1) and AR(1) series using only the sample ACF function plot.

### (c) Plot ACFs

```
# plot ACFs
ACF_ARMA = ARMAacf(ar=0.6, ma=0.9, lag.max=10)
ACF_AR = ARMAacf(ar=0.6, ma=0, lag.max=10)
ACF_MA = ARMAacf(ar=0, ma=0.9, lag.max=10)
plot((0:10), ACF_ARMA, col=1, xlab="lag h", ylab="ACF", xaxt="n")
axis(at=(0:10), side=1)
lines((0:10), ACF_ARMA, col=1)
points((0:10), ACF_AR, col=2)
lines((0:10), ACF_AR, col=2)
points((0:10), ACF_MA, col=3)
lines((0:10), ACF_MA, col=3)
abline(h=0, lty=2)
legend("topright", legend=c("ARMA(1,1)", "AR(1)", "MA(1)"), lty=1, col=c(1,2,3))
```



Comments:

From the ACF plots, we can see that the ACF plot of MA(1) is easy to be distinguished from the other two plots, and it is also easy to identify  $p = 1$  here because the cutting off of  $\rho(h)$  after  $q$  lags is the signature of the MA( $q$ ) series.

However, as I discussed above, it is unlikely that we will be able to tell the difference between an ARMA(1,1) and an AR(1) based solely on an ACF estimated from a sample, because their general pattern of  $\rho(h)$  verses  $h$  are very much similar. Therefore, we need PACF to further distinguish them.

### 3.9

```
ar_ma_parameters = list(arma=c(0.6,0.9), ar=c(0.6,0), ma=c(0,0.9))
model_name = c("ARMA(1,1)","AR(1)","MA(1)")
n = c(100,1000)

for (i in 1:3) {
  AR = ar_ma_parameters[[i]][1]
  MA = ar_ma_parameters[[i]][2]

  par(mfrow=c(1,2))
  for (j in 1:2) {
    set.seed(123)
    sim = arima.sim(model=list(ar=AR, ma=MA), n=n[j])
    ACF_theoretical = ARMAacf(ar=AR, ma=MA, lag.max=10, pacf=FALSE)[-1]
    ACF_sample = acf(sim, lag.max=10, type="correlation", plot=FALSE)$acf[-1]
    PACF_sample = acf(sim, lag.max=10, type="partial", plot=FALSE)$acf # or just use function pacf
```

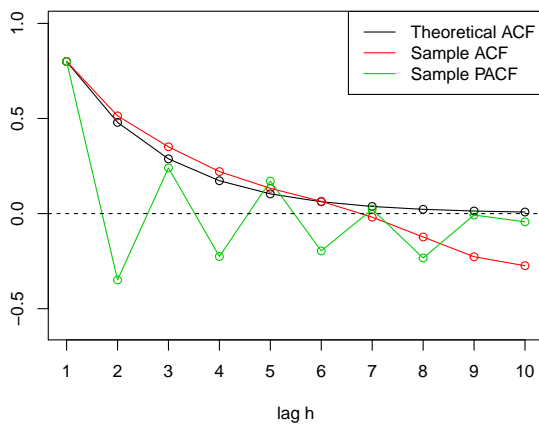


```

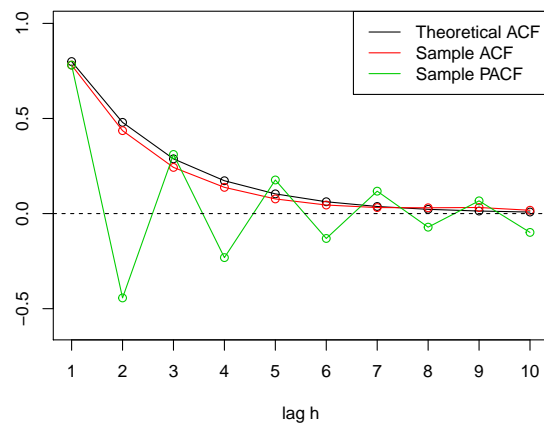
plot((1:10), ACF_theoretical, col=1, xlab="lag h", ylab="", ylim = c(-0.6,1),
     main = paste(model_name[i], "model, n =", n[j]), xaxt="n")
axis(at=(1:10), side=1)
lines((1:10), ACF_theoretical, col=1)
points((1:10), ACF_sample, col=2)
lines((1:10), ACF_sample, col=2)
points((1:10), PACF_sample, col=3)
lines((1:10), PACF_sample, col=3)
abline(h=0, lty=2)
legend("topright", legend=c("Theoretical ACF", "Sample ACF", "Sample PACF"), lty=1, col=c(1,2,3))
}
}

```

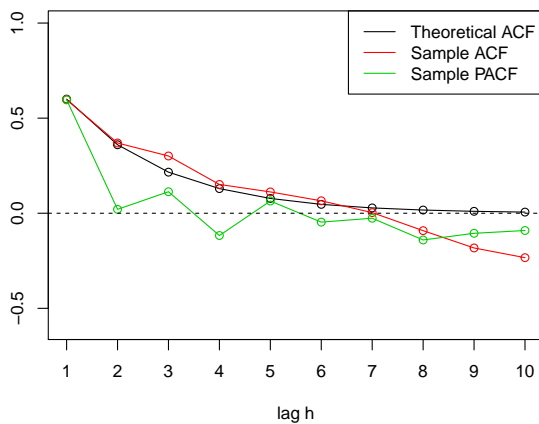
ARMA(1,1) model, n = 100



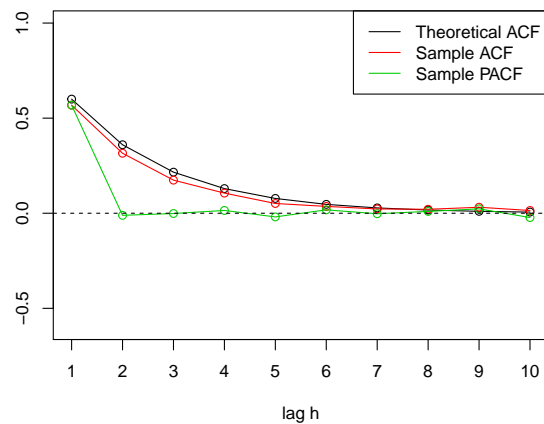
ARMA(1,1) model, n = 1000

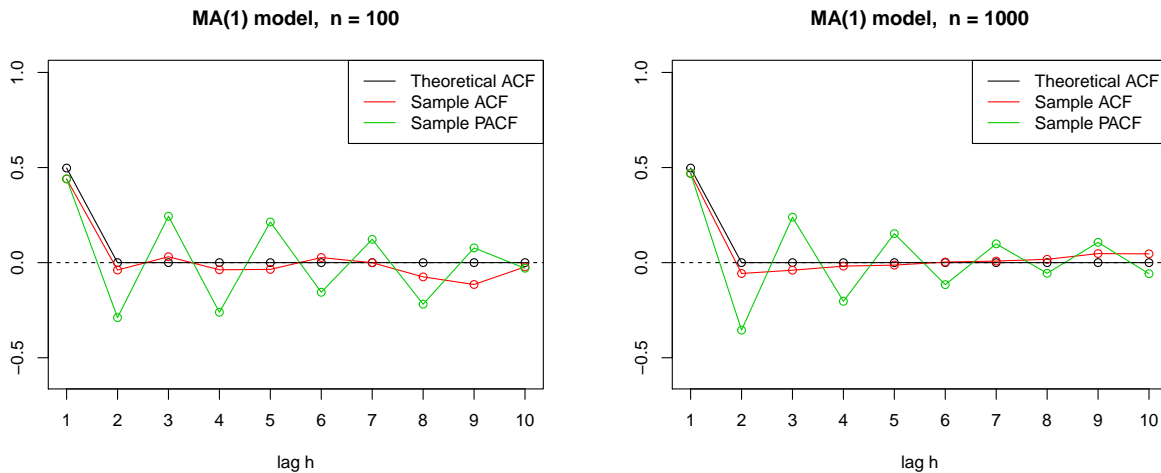


AR(1) model, n = 100



AR(1) model, n = 1000





Comments:

Note: when generating 100 samples, the sample ACFs matches the theoretical ACFs only up to lag = 4 to 6, and after that they are very different from each other. Therefore, to compare them up to lag = 10, I also generated a sample of size  $n = 1000$ , which gives a better comparison for lag up to 10. Actually, the window size for estimating ACF is at the level of  $h = n^{\frac{1}{3}}$ , so it's more reasonable to use  $n = h^3 = 10^3 = 1000$ .

Below I will mainly discuss the situations for sample size  $n = 1000$ .

#### (1) ARMA(1,1)

The sample ACF matches the theoretical ACF quite well. The sample ACF and sample PACF both tail off as lag  $h$  increases to 10.

#### (2) AR(1)

The sample ACF matches the theoretical ACF quite well. The sample ACF tails off as lag  $h$  increases to 10, while the sample PACF cuts off after lag  $p = 1$ .

#### (3) MA(1)

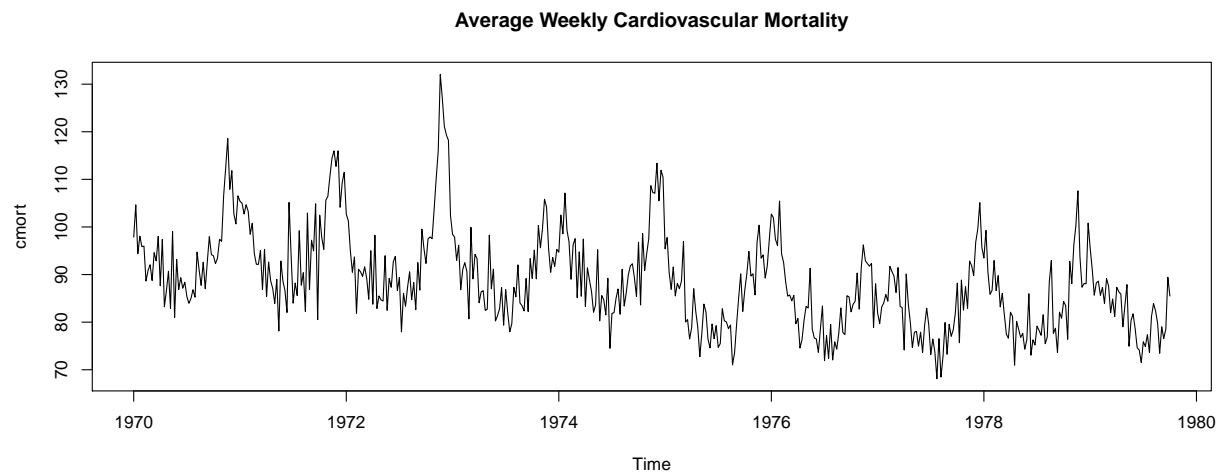
The sample ACF matches the theoretical ACF roughly well. The sample ACF cuts off after lag  $q = 1$ , while the sample PACF tails off as lag  $h$  increases to 10.

In conclusion, in all three plots, the sample ACF matches the theoretical ACF well, and the sample ACF and sample PACF do behave similarly as the general results given in Table 3.1.

## Section 3.4 Forecasting

### 3.10

```
library(astsa)
plot(cmort, main="Average Weekly Cardiovascular Mortality")
```



(a)

```
# fit an AR(2) model to x_t using linear regression by OLS
( regr = ar.ols(cmort, order=2, demean=FALSE, intercept=TRUE) )

##
## Call:
## ar.ols(x = cmort, order.max = 2, demean = FALSE, intercept = TRUE)
##
## Coefficients:
##      1      2
## 0.4286 0.4418
##
## Intercept: 11.45 (2.394)
##
## Order selected 2  sigma^2 estimated as  32.32
regr$asy.se.coef # standard errors of the estimates

## $x.mean
## [1] 2.393673
##
## $ar
## [1] 0.03979433 0.03976163
```

Results:

The fitted model using linear regression by OLS is:

$$x_t = \hat{\phi}_0 + \hat{\phi}_1 x_{t-1} + \hat{\phi}_2 x_{t-2} + w_t = 11.45 + 0.4286 x_{t-1} + 0.4418 x_{t-2} + w_t$$

where  $\hat{\sigma}_w^2 = 32.32$

(b)

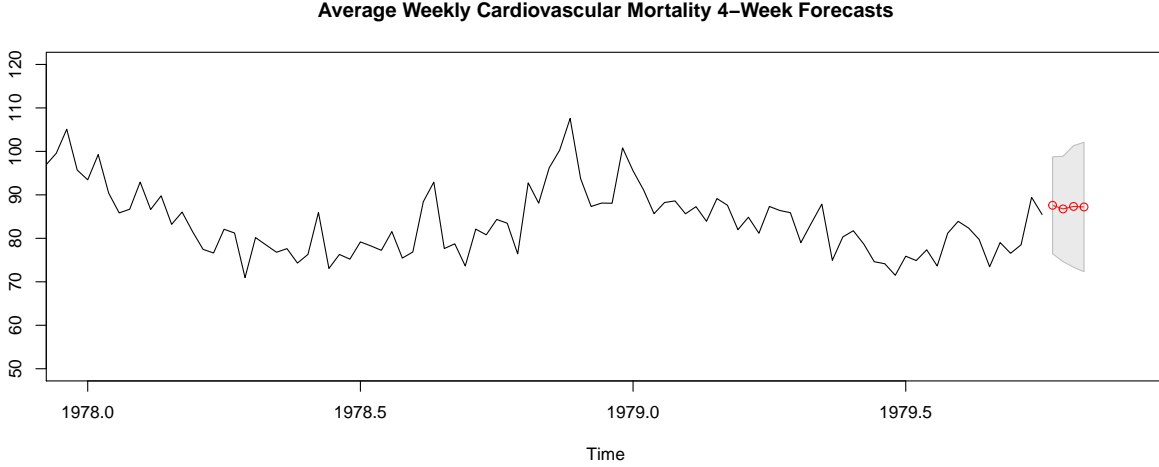
```
# forecasts
forecasts = predict(regr, n.ahead=4)
forecasts
```

```
## $pred
## Time Series:
## Start = c(1979, 41)
## End = c(1979, 44)
## Frequency = 52
## [1] 87.59986 86.76349 87.33714 87.21350
##
## $se
## Time Series:
## Start = c(1979, 41)
## End = c(1979, 44)
## Frequency = 52
## [1] 5.684848 6.184973 7.134227 7.593357
```

```
# 95% Confidence Interval
pred = forecasts$pred
se = forecasts$se
z = qnorm(0.975, lower.tail=TRUE)
CI = cbind(pred-z*se, pred+z*se)
colnames(CI) = c("lower", "upper")
CI
```

```
## Time Series:
## Start = c(1979, 41)
## End = c(1979, 44)
## Frequency = 52
##          lower      upper
## 1979.769 76.45777  98.74196
## 1979.788 74.64117  98.88581
## 1979.808 73.35431 101.31997
## 1979.827 72.33079 102.09621
```

```
# plot
ts.plot(cmort, pred, col=1:2, xlim=c(1978,1979.9), ylim=c(50,120),
        main="Average Weekly Cardiovascular Mortality 4-Week Forecasts")
xx = c(time(pred+z*se), rev(time(pred+z*se)))
yy = c(pred-z*se, rev(pred+z*se))
polygon(xx, yy, border=8, col=gray(0.6, alpha=0.2))
lines(pred, type="p", col=2)
```



Results:

Assume the fitted model in (a) is the true model, the above figure shows the result of forecasting the average weekly cardiovascular mortality series over a 4-week horizon, i.e. forecast  $x_{n+m}^n$  for  $m = 1, 2, 3, 4$ .

The predicted values for the next 4 weeks are: 87.59986, 86.76349, 87.33714, 87.21350, respectively. And their corresponding 95% prediction intervals are: (76.45777, 98.74196), (74.64117, 98.88581), (73.35431, 101.31997), (72.33079, 102.09621), respectively.

### 3.12

Note that the matrix  $\mathbf{\Gamma}_n = \{\gamma(k-j)\}_{j,k=1}^n$  is an  $n \times n$  positive semi-definite (nonnegative definite) matrix.

Now we want to show that if  $\gamma(0) > 0$  and  $\gamma(h) \rightarrow 0$  as  $h \rightarrow 0$ , then the matrix  $\mathbf{\Gamma}_n$  is positive definite.

Let's suppose there exists a matrix  $\mathbf{\Gamma}_n$  that is not positive definite, then this  $\mathbf{\Gamma}_n$  is a singular matrix. Since  $\det(\mathbf{\Gamma}_n) = \det(\{\gamma(0)\}) = \gamma(0) > 0$ , so  $\mathbf{\Gamma}_1$  is nonsingular. Thus the singular matrix  $\mathbf{\Gamma}_n$  must have  $n \geq 1$ .

Suppose  $\mathbf{\Gamma}_{m+1}$  is the first singular matrix in the sequence  $\mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_n$ , so  $\det(\mathbf{\Gamma}_{m+1}) = 0$  and the matrices  $\mathbf{\Gamma}_1, \dots, \mathbf{\Gamma}_m$  are all nonsingular and positive definite matrices with non-zero determinants. Thus we can get that  $x_{m+1}$  is a linear combination of  $x_1, x_2, \dots, x_m$ . Let's set:

$$x_{m+1} = \sum_{i=1}^m c_i x_i = \mathbf{c}^T \mathbf{x}_m, \quad \text{where } \mathbf{x}_m = (x_1, \dots, x_m)^T, \quad \mathbf{c} = (c_1, \dots, c_m)^T$$

Then because  $x_t$  is a stationary series, we also have that:

$$x_{m+h+1} = \sum_{i=1+h}^{m+h} c_i x_i = \mathbf{c}^T \mathbf{x}_{m+h}, \quad \text{where } \mathbf{x}_{m+h} = (x_{1+h}, \dots, x_{m+h})^T, \quad \mathbf{c} = (c_1, \dots, c_m)^T \quad \forall h \geq 1$$

Because  $\mathbf{x}_h = (x_{1+h}, \dots, x_m, x_{m+1}, \dots, x_{m+h})^T$ , where each element of all the last number  $h$  sequences  $(x_{m+1}, \dots, x_{m+h})^T$  is a linear combination of  $x_1, x_2, \dots, x_m$ . As a result, all the  $x_n$  are a linear combination of  $x_1, x_2, \dots, x_m$  for all  $n \geq m+1$ . We can write it as:

$$x_n = \sum_{i=1}^m a_{ni} x_i = \mathbf{a}_n^T \mathbf{x}_m, \quad \text{where } \mathbf{x}_m = (x_1, \dots, x_m)^T, \quad \mathbf{a}_n = (a_{n1}, \dots, a_{nm})^T \quad \forall n \geq m+1$$

Assume the eigenvalues of the matrix  $\mathbf{\Gamma}_m$  are  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ . Because  $\mathbf{\Gamma}_m$  is nonsingular and positive definite, so all of its eigenvalues are positive, i.e.  $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ , and it can be eigendecomposed (factorized) as:  $\mathbf{\Gamma}_m = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$ , where  $\mathbf{Q}\mathbf{Q}^{-1} = \mathbf{I}$  and  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ .

Therefore, we can derive that:

$$\begin{aligned}\gamma(0) &= \text{Var}(x_n) = \text{Var}(\mathbf{a}_n^T \mathbf{x}_m) = \mathbf{a}_n^T \mathbf{\Gamma}_m \mathbf{a}_n = \mathbf{a}_n^T \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1} \mathbf{a}_n \\ &\geq \lambda_1 \mathbf{a}_n^T \mathbf{Q}\mathbf{Q}^{-1} \mathbf{a}_n = \lambda_1 \mathbf{a}_n^T \mathbf{a}_n = \lambda_1 \sum_{i=1}^m a_{ni}^2 \\ &\Rightarrow \|\mathbf{a}_n\|^2 = \sum_{i=1}^m a_{ni}^2 \leq \frac{\gamma(0)}{\lambda_1}\end{aligned}$$

On the other hand, because  $\gamma(h) \rightarrow 0$  as  $h \rightarrow 0$ , then we can also derive that:

$$\begin{aligned}0 < \gamma(0) &= \text{Cov}(x_n, x_n) = \text{Cov}\left(\sum_{i=1}^m a_{ni} x_i, x_n\right) = \sum_{i=1}^m a_{ni} \text{Cov}(x_i, x_n) = \sum_{i=1}^m a_{ni} \gamma(n-i) \\ &\leq \sum_{i=1}^m |a_{ni}| |\gamma(n-i)| \leq \|\mathbf{a}_n\| \left( \sum_{i=1}^m \gamma(n-i)^2 \right)^{\frac{1}{2}} \leq \left( \frac{\gamma(0)}{\lambda_1} \right)^{\frac{1}{2}} \left( \sum_{i=1}^m \gamma(n-i)^2 \right)^{\frac{1}{2}} \rightarrow 0 \quad (\text{as } n \rightarrow \infty)\end{aligned}$$

But  $\gamma(0) > 0$  is a constant, so there is a conflict between  $\gamma(0) > 0$  and  $\gamma(h) \rightarrow 0$  as  $h \rightarrow 0$  in this case. As a result, there do not exist any  $\mathbf{\Gamma}_n$  that is not positive definite if  $\gamma(0) > 0$  and  $\gamma(h) \rightarrow 0$  as  $h \rightarrow 0$ . Therefore, all the matrix  $\mathbf{\Gamma}_n$  are positive definite if  $\gamma(0) > 0$  and  $\gamma(h) \rightarrow 0$  as  $h \rightarrow 0$ .

### 3.13

(i)

The prediction equations in (3.63) is:

$$\mathbf{\Gamma}_n \phi_n = \gamma_n$$

where  $\mathbf{\Gamma}_n = \{\gamma(k-j)\}_{j,k=1}^n$ ,  $\phi_n = (\phi_{n1}, \dots, \phi_{nn})^T$ ,  $\gamma_n = (\gamma(1), \dots, \gamma(n))^T$ .

Set  $n = h$ , and then divide both sides with  $\gamma(0)$ , we get:

$$\mathbf{R}_h \phi_h = \rho_h$$

where  $\mathbf{R}_h = \{\rho(k-j)\}_{j,k=1}^h$ ,  $\phi_h = (\phi_{h1}, \dots, \phi_{hh})^T$ ,  $\rho_h = (\rho(1), \dots, \rho(h))^T$ .

Note that, in this problem, the  $\phi_h = (\phi_{h1}, \dots, \phi_{hh})^T = (\alpha_{h1}, \dots, \alpha_{hh})^T = \alpha_h$ . To make it consistent with the form with (3.63), here I use  $\phi_h$  instead of  $\alpha_h$ .

Then we can write the above equation in the following form:

$$\begin{aligned}&\Rightarrow \begin{pmatrix} \mathbf{R}_{h-1} & \tilde{\rho}_{h-1} \\ \tilde{\rho}_{h-1}^T & \rho(0) \end{pmatrix} \begin{pmatrix} \phi_{h-1} \\ \phi_{hh} \end{pmatrix} = \begin{pmatrix} \rho_{h-1} \\ \rho(h) \end{pmatrix} \\ &\Rightarrow \mathbf{R}_{h-1} \phi_{h-1} + \tilde{\rho}_{h-1} \phi_{hh} = \rho_{h-1} \quad (1)\end{aligned}$$

$$\tilde{\rho}_{\mathbf{h}-1}^T \phi_{\mathbf{h}-1} + \phi_{hh} = \rho(h) \quad (2)$$

From (1), we get:

$$\phi_{\mathbf{h}-1} = \mathbf{R}_{\mathbf{h}-1}^{-1} (\rho_{\mathbf{h}-1} - \tilde{\rho}_{\mathbf{h}-1} \phi_{hh})$$

Then from (2), we get:

$$\phi_{hh} = \rho(h) - \tilde{\rho}_{\mathbf{h}-1}^T \phi_{\mathbf{h}-1} = \rho(h) - \tilde{\rho}_{\mathbf{h}-1}^T \mathbf{R}_{\mathbf{h}-1}^{-1} (\rho_{\mathbf{h}-1} - \tilde{\rho}_{\mathbf{h}-1} \phi_{hh})$$

$$\Rightarrow \phi_{hh} = \frac{\rho(h) - \tilde{\rho}_{\mathbf{h}-1}^T \mathbf{R}_{\mathbf{h}-1}^{-1} \rho_{\mathbf{h}-1}}{1 - \tilde{\rho}_{\mathbf{h}-1}^T \mathbf{R}_{\mathbf{h}-1}^{-1} \tilde{\rho}_{\mathbf{h}-1}} \quad (*)$$

(ii)

The definition of PACF at lag  $h$  is defined as the cross-correlation between  $\epsilon_t$  and  $\delta_{t-h}$ :

$$\phi_{hh} = \frac{E(\epsilon_t \delta_{t-h})}{\sqrt{E(\epsilon_t^2) E(\delta_{t-h}^2)}}$$

From the equations in (3.64), (3.65) and (3.66), we can have that:

$$x_t^{h-1} = \phi_{h1} x_{t-1} + \dots + \phi_{h-1, h-1} x_{t-(h-1)} = \phi_{\mathbf{h}-1}^T \mathbf{x} = (\mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \gamma_{\mathbf{h}-1})^T \mathbf{x} = \gamma_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x}$$

where  $\mathbf{x} = (x_{t-1}, \dots, x_{t-(h-1)})^T$  and matrix  $\mathbf{\Gamma}_{\mathbf{h}-1}$  is symmetric.

$$\Rightarrow \epsilon_t = x_t - x_t^{h-1} = x_t - \gamma_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x}$$

Similarly, for the regression of  $x_{t-h}$  on  $\mathbf{x} = (x_{t-1}, \dots, x_{t-(h-1)})^T$ , we also get that:

$$\Rightarrow \delta_{t-h} = x_{t-h} - x_{t-h}^{h-1} = x_{t-h} - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x}$$

Therefore,

$$E(\epsilon_t^2) = Var(\epsilon_t) = E(x_t - x_t^{h-1})^2 = E(x_t - \gamma_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x})^2 = \gamma(0) - \gamma_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \gamma_{\mathbf{h}-1}$$

$$E(\delta_{t-h}^2) = Var(\delta_{t-h}) = E(x_{t-h} - x_{t-h}^{h-1})^2 = E(x_{t-h} - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x})^2 = \gamma(0) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \tilde{\gamma}_{\mathbf{h}-1}$$

$$E(\epsilon_t \delta_{t-h}) = Cov(\epsilon_t, \delta_{t-h}) = E[(x_t - x_t^{h-1})(x_{t-h} - x_{t-h}^{h-1})]$$

$$= E[(x_t - \gamma_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x})(x_{t-h} - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \mathbf{x})] = \gamma(h) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \gamma_{\mathbf{h}-1}$$

Because  $\mathbf{\Gamma}_{\mathbf{h}-1}$  is symmetric, we also have that:

$$E(\epsilon_t^2) = \gamma(0) - \gamma_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \gamma_{\mathbf{h}-1} = \gamma(0) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \tilde{\gamma}_{\mathbf{h}-1} = E(\delta_{t-h}^2)$$

$$\begin{aligned}\Rightarrow \phi_{hh} &= \frac{E(\epsilon_t \delta_{t-h})}{\sqrt{E(\epsilon_t^2)E(\delta_{t-h}^2)}} = \frac{\gamma(h) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \gamma_{\mathbf{h}-1}}{\gamma(0) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \tilde{\gamma}_{\mathbf{h}-1}} \\ &= \frac{(\gamma(h) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \gamma_{\mathbf{h}-1})/\gamma(0)}{(\gamma(0) - \tilde{\gamma}_{\mathbf{h}-1}^T \mathbf{\Gamma}_{\mathbf{h}-1}^{-1} \tilde{\gamma}_{\mathbf{h}-1})/\gamma(0)} = \frac{\rho(h) - \tilde{\rho}_{\mathbf{h}-1}^T \mathbf{R}_{\mathbf{h}-1}^{-1} \rho_{\mathbf{h}-1}}{1 - \tilde{\rho}_{\mathbf{h}-1}^T \mathbf{R}_{\mathbf{h}-1}^{-1} \tilde{\rho}_{\mathbf{h}-1}}\end{aligned}$$

which is in the same format of the equation (\*), so the result is proved.

### 3.14

(a)

Due to the Law of Total Expectation, i.e.  $E(X) = E(E(X|Y))$ , we can write the MSE as:

$$MSE = E[(y - g(x))^2] = E[E[(y - g(x))^2|x]]$$

So to minimize MSE, we need to minimize  $E[(y - g(x))^2|x]$ .

Take the derivative with respect to  $g(x)$ :

$$\frac{\partial}{\partial g(x)} E[(y - g(x))^2|x] = -2 E[(y - g(x))|x] = -2 (E(y|x) - g(x)) = 0$$

We get:

$$g(x) = E(y|x)$$

Thus the MSE is minimized by the choice  $g(x) = E(y|x)$ .

(b)

For the model:

$$y = x^2 + z, \quad \text{where } x \sim \text{iid}N(0, 1), \quad z \sim \text{iid}N(0, 1), \quad x \text{ independent of } z$$

The minimizer of MSE is:

$$g(x) = E(y|x) = E(x^2 + z|x) = E(x^2|x) + E(z|x) = x^2 + E(z) = x^2$$

Thus the minimized MSE is:

$$MSE = E[(y - g(x))^2] = E[(y - x^2)^2] = E(z^2) = \text{Var}(z) + (E(z))^2 = 1$$

(c)

Since:

$$E(g(x)) = E(a + bx) = a + bE(x) = a$$

And due to the Law of Total Expectation, i.e.  $E(X) = E(E(X|Y))$ , we have:

$$E(g(x)) = E(E(y|x)) = E(y) = E(x^2 + z) = E(x^2) = \text{Var}(x) = 1$$

So we get:  $a = 1$



Then, take the derivative of MSE with respect to  $b$ , we have:

$$\begin{aligned}\frac{\partial}{\partial b} MSE &= \frac{\partial}{\partial b} E[(y - g(x))^2] = \frac{\partial}{\partial b} E[(y - a - bx)^2] \\ &= -2 E[(y - 1 - bx)x] = -2 E(xy - x - bx^2) = 0 \\ \Rightarrow b &= \frac{E(xy) - E(x)}{E(x^2)} = \frac{E(xy)}{E(x^2)} = E(xy) = E((x^2 + z)x) = E(x^3 + xz) = E(x^3) + E(xz) = 0\end{aligned}$$

Therefore, the MSE is:

$$\begin{aligned}MSE &= E[(y - g(x))^2] = E[(y - 1)^2] = E(y^2 - 2y + 1) = E(y^2) - 2E(y) + 1 = E(y^2) - 1 \\ &= E((x^2 + z)^2) - 1 = E(x^4) + 2E(x^2)E(z) + E(z^2) - 1 = E(x^4) + E(z^2) - 1 = 3 + 1 - 1 = 3\end{aligned}$$

Interpretation:

In this example, the MSE of the best linear predictor,  $g(x) = a + bx$ , is three times that of the optimal minimum mean square error predictor,  $g(x) = E(y|x)$  (the conditional expectation). Because here the true  $y = x^2 + z$  is not a Gaussian, thus the the best linear predictor will not close to the minimum mean square error predictor. Note that if the process is Gaussian, then the best linear predictors and the minimum mean square error predictors will be the same.