STAT 153 Homework 4

Donggyun Kim

12/18/2018

Theoretical exercise:

1. Consider an invertible MA(1) model $X_t = Z_t + \theta Z_{t-1}$ for some i.i.d. white noise process $\{Z_t\}$ with variance σ^2.

(a) Derive the explicit form of the minimum mean-square error one-step prediction

$$\tilde{X}_{n+1} = \mathbb{E}[X_{n+1}|X_n, X_{n-1}, X_{n-2}, ...]$$

for X_{n+1} based on the complete infinite past $X_n, X_{n-1}, X_{n-2}, \ldots$.

Since it is invertible,
$Z_{n+1} = \sum_{j=0}^{\infty} \theta^j X_{n+1-j} = X_{n+1} + \sum_{j=1}^{\infty} \theta^j X_{n+1-j}$
$\implies X_{n+1} = Z_{n+1} - \theta \sum_{j=0}^{\infty} \theta^j X_{n-j}$
$\tilde{X}_{n+1} = \mathbb{E}[Z_{n+1} - \theta \sum_{j=0}^{\infty} \theta^j X_{n-j}|X_n, X_{n-1}, X_{n-2}, ...] = -\theta \sum_{j=0}^{\infty} \theta^j X_{n-j}$
since Z_{n+1} is independent of $Z_n, Z_{n-1}, \ldots$, and mean zero white noise.

(b) Derive the mean squared error $\mathbb{E}[(\tilde{X}_{n+1} - X_{n+1})^2]$.

$\tilde{X}_{n+1} - X_{n+1} = -\theta \sum_{j=0}^{\infty} \theta^j X_{n-j} - X_{n+1} = -Z_{n+1}$
$\mathbb{E}[(\tilde{X}_{n+1} - X_{n+1})^2] = Var(Z_{n+1}) = \sigma^2$

(c) Now consider the truncated estimate $\tilde{X}_{n+1}^n$, which equals $\tilde{X}_{n+1}$ but with unobserved data being set to zero, that is, $0 = X_0 =$ \$X_{-1} = $\ldots$\$. Show that

$$E[(X_{n+1} - \tilde{X}_{n+1}^n)^2] = \sigma^2(1 + \theta^{2+2n})$$

$$\begin{aligned}
\tilde{X}_{n+1}^n &= \mathbb{E}[X_{n+1}|X_n, X_{n-1}, X_{n-2}, ..., X_1] = \mathbb{E}[-\theta \sum_{j=0}^{\infty} \theta^j X_{n-j}|X_n, X_{n-1}, \ldots, X_1] \\
&= -\theta \sum_{j=0}^{n-1} \theta^j X_{n-j} + \mathbb{E}[-\theta^{n+1} \sum_{j=0}^{\infty} \theta^j X_{-j}] = -\theta \sum_{j=0}^{n-1} \theta^j X_{n-j} - \theta^{n+1}\mathbb{E}[Z_0] \\
&= -\theta \sum_{j=0}^{n-1} \theta^j X_{n-j}
\end{aligned}$$

$$\begin{aligned}
X_{n+1} - \tilde{X}_{n+1}^n &= \sum_{j=0}^{n} \theta^j X_{n+1-j} = \sum_{j=0}^{\infty} \theta^j X_{n+1-j} - \sum_{j=n+1}^{\infty} \theta^j X_{n+1-j} \\
&= Z_{n+1} - \theta^{n+1} \sum_{j=0}^{\infty} \theta^j X_{-j} = Z_{n+1} - \theta^{n+1} Z_0
\end{aligned}$$

$\mathbb{E}[(Z_{n+1} - \theta^{n+1} Z_0)^2] = \sigma^2 + \theta^{2(n+1)}\sigma^2 = \sigma^2(1 + \theta^{2+2n})$

(d) Comment on how well the truncated estimate $\tilde{X}^n_{n+1}$ works compared to $\tilde{X}_{n+1}$.

For invertible MA(1) process, $|\theta| < 1$ and $\sum_{j\geq 0} |\theta^j| < \infty$. Thus, $\lim_{n\to\infty} \theta^{2+2n} = 0$. For sufficiently large sample size, truncated estimator $\tilde{X}^n_{n+1}$ works as well as eistimator $\tilde{X}_{n+1}$.

2. Consider an invertible MA(q) model $X_t = \theta(B)Z_t$ for some white noise $\{Z_t\}$ with variance σ^2.

(a) Show that for any m > q the best linear predictor of X_{n+m} based on $X_1, \cdots, X_n$ is always zero.

Invertible MA(q) model $\implies$ Causal AR(∞) model.
The best linear predictor of causal AR(∞) model based on $X_1, \cdots, X_n$ is, for large n, approximately equal to the best linear predictor of causal AR(n) model.
The coefficients of the best linear predictor are determined by variance matirx of X and covariance matirx between Y($=X_{n+m}$) and X.
For $m > q$ and $1 \leq i \leq n$, $Cov(X_{n+m}, X_i) = 0$.
Therefore, the best linear predictor of X_{n+m} is always zero.

(b) Now assume that the white noise {Z_t} is also i.i.d.. Show that for any m > q the best predictor (minimum mean-square error forecast) of X_{n+m} based on the full history $X_n, X_{n-1}, X_{n-2}, ...$ is also zero.

$X_n, X_{n-1}, X_{n-2}, ...$ can be represented in terms of $Z_n, Z_{n-1}, Zn-2, ...$ and
$X_{n+m} = Z_{n+m} + \theta_1 Zn + m - 1 + \cdots + \theta^q Z_{n+m-q}$.
For $m > q$, $n + m - q \geq n + 1$. Thus,

$$\begin{aligned}
\mathbb{E}[X_{n+m}|X_n, X_{n-1}, \dots] &= \mathbb{E}[Z_{n+m} + \theta_1 Zn + m - 1 + \cdots + \theta^q Z_{n+m-q}|Z_n, Z_{n-1}, \dots] \\
&= \mathbb{E}[Z_{n+m} + \theta_1 Zn + m - 1 + \cdots + \theta^q Z_{n+m-q}] \\
&= 0
\end{aligned}$$

3. Consider a causal, zero mean AR(1) model $X_t - \phi X_{t-1} = Z_t$ for some white noise $\{Z_t\}$ with variance σ^2.

(a) Derive the general form of the best linear predictor $\tilde{X}_{n+m}$ in terms of $X_1, \dots, X_n$.

$$\begin{aligned}
\tilde{X}_{n+m} &= \phi \tilde{X}_{n+m-1} \\
&= \phi^2 \tilde{X}_{n+m-2} \\
&= \cdots \\
&= \phi^m X_n
\end{aligned}$$

(b) Show that

$$\mathbb{E}[(X_{n+m} - \tilde{X}_{n+m})^2] = \sigma^2 \frac{1-\phi^{2m}}{1-\phi^2}$$

$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$
$X_{n+m} - \tilde{X}_{n+m} = \sum_{j=0}^{\infty} \phi^j Z_{n+m-j} - \phi^m \sum_{j=0}^{\infty} \phi^j Z_{n-j} = \sum_{j=0}^{m-1} \phi^j Z_{n+m-j}$

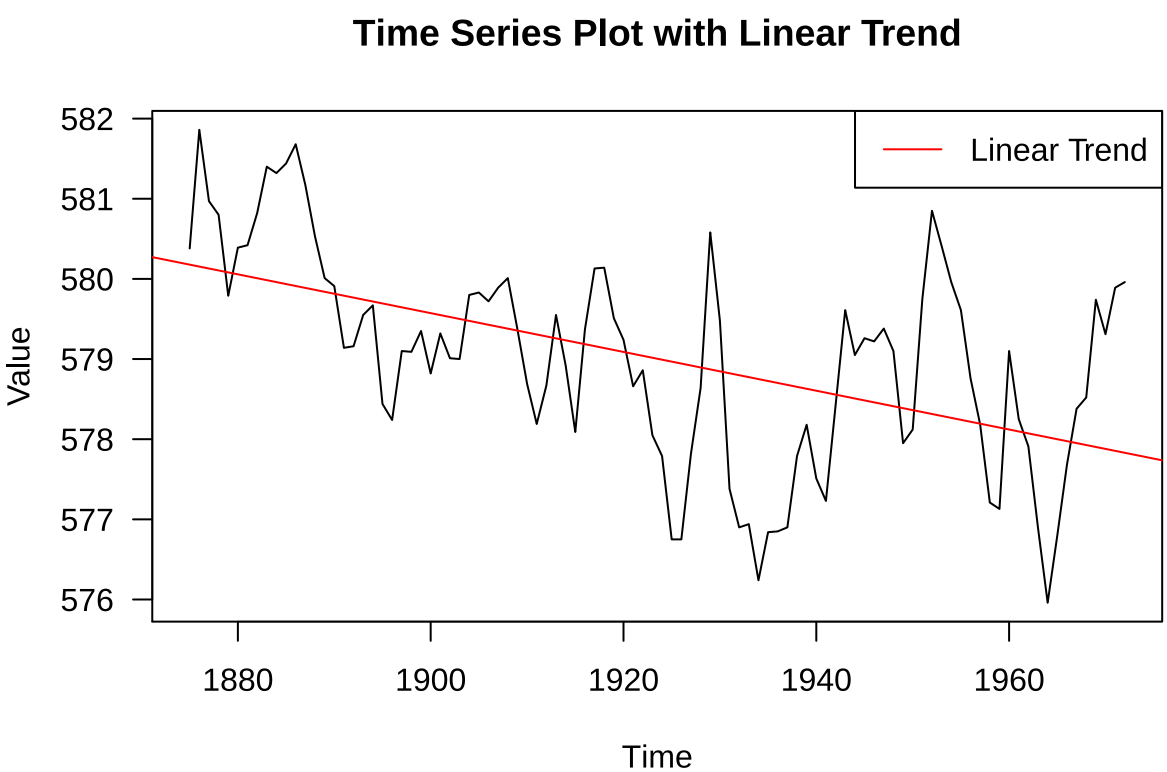
$$
\begin{aligned}
\mathbb{E}[(X_{n+m} - \tilde{X}_{n+m})^2] &= Var(X_{n+m} - \tilde{X}_{n+m}) + (\mathbb{E}[X_{n+m} - \tilde{X}_{n+m}])^2 \\
&= Cov(\sum_{j=0}^{m-1} \theta^j Z_{n+m-j}, \sum_{k=0}^{m-1} \theta^k Z_n + m - k) \\
&= \sigma^2(1 + \phi^2 + \phi^4 + \cdots + \phi^{2(m-1)}) \\
&= \sigma^2 \frac{1 - \phi^{2m}}{1 - \phi^2}
\end{aligned}
$$

Computer exercise:

1. Consider the LakeHuron dataset in R.

(a) Fit a linear trend function to the data and obtain residuals.

```
dat <- LakeHuron
t <- time(dat)
lm_trend <- lm(dat ~ t)
# plot
plot(dat, main = "Time Series Plot with Linear Trend", ylab = "Value", las = 1)
abline(lm_trend, col = "red")
legend("topright", legend = "Linear Trend", col = "red", lty = 1)
```




```
# residuals
r <- lm_trend$residuals
```

(b) Fit an AR(1) model to the residuals using the R function `arima()`.

```
my_arima <- arima(r, order = c(1, 0, 0))
my_arima
```

```
##
## Call:
```

```
## arima(x = r, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.7829     0.0797
## s.e.  0.0634     0.3178
##
## sigma^2 estimated as 0.4972:  log likelihood = -105.29,  aic = 216.58
```

$X_t - 0.0797 = 0.7829(X_{t-1} - 0.0797) + Z_t$
$X_t = (1 - 0.7829) \times 0.0797 + 0.7829X_{t-1} + Z_t$

(c) Assume that your fitted model coincides with the true generating model of the data. Obtain predictions for the residuals for the future m = 30 time points.

$\tilde{X}_t = (1 - \phi) \times \alpha + \phi(\tilde{X}_{t-1})$

```
phi <- my_arima$coef[[1]]
intercept <- (1 - phi) * my_arima$coef[[2]]
r_future <- rep(0, 30)
r_future[1] <- intercept + phi * r[[98]]
for (i in 2:30){
  r_future[i] <- intercept + phi * r_future[i-1]
}
r_future
```

```
##  [1] 1.68457482 1.33612055 1.06332549 0.84976190 0.68256896 0.55167830
##  [7] 0.44920768 0.36898631 0.30618325 0.25701650 0.21852523 0.18839150
## [13] 0.16480064 0.14633203 0.13187347 0.12055426 0.11169278 0.10475537
## [19] 0.09932427 0.09507241 0.09174375 0.08913784 0.08709774 0.08550061
## [25] 0.08425025 0.08327139 0.08250506 0.08190513 0.08143545 0.08106776
```

(d) Compare your predictions with those obtained by the `predict()` function in R.

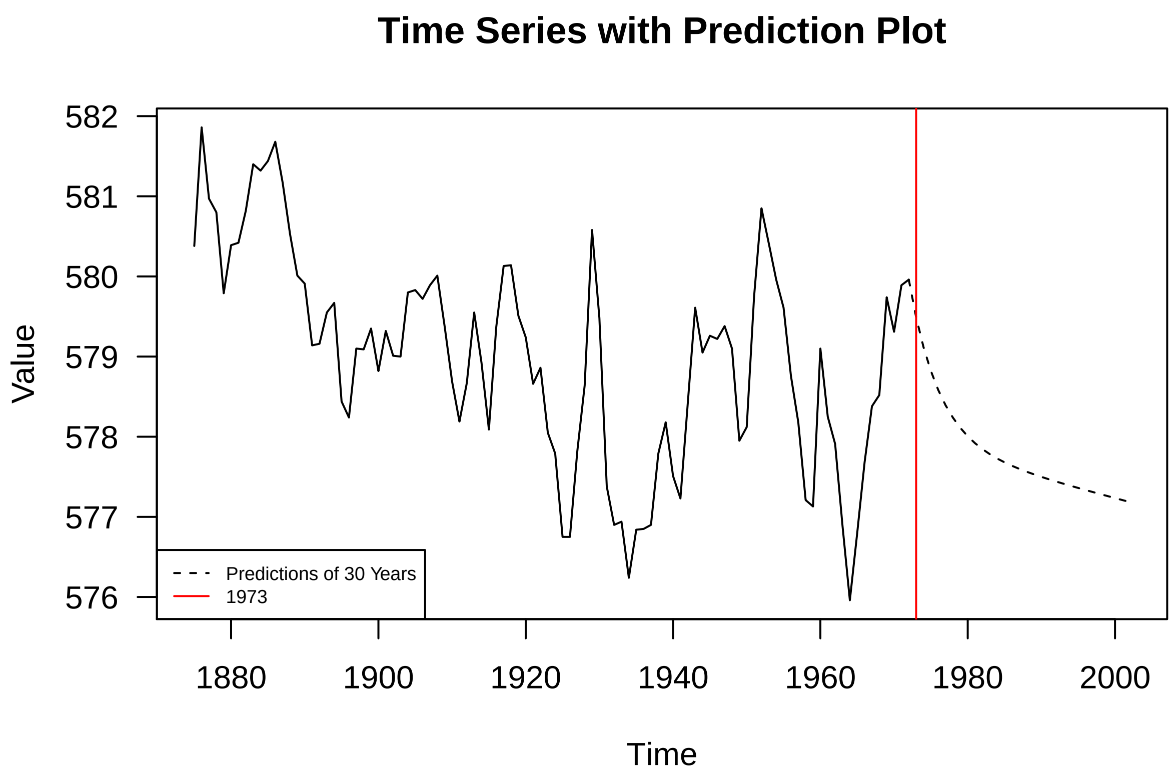
```
predict(my_arima, n.ahead = 30)$pred
```

```
## Time Series:
## Start = 99
## End = 128
## Frequency = 1
##  [1] 1.68457482 1.33612055 1.06332549 0.84976190 0.68256896 0.55167830
##  [7] 0.44920768 0.36898631 0.30618325 0.25701650 0.21852523 0.18839150
## [13] 0.16480064 0.14633203 0.13187347 0.12055426 0.11169278 0.10475537
## [19] 0.09932427 0.09507241 0.09174375 0.08913784 0.08709774 0.08550061
## [25] 0.08425025 0.08327139 0.08250506 0.08190513 0.08143545 0.08106776
```

They are the same.

(e) Obtain predictions for the original data for the future m = 30 time period.

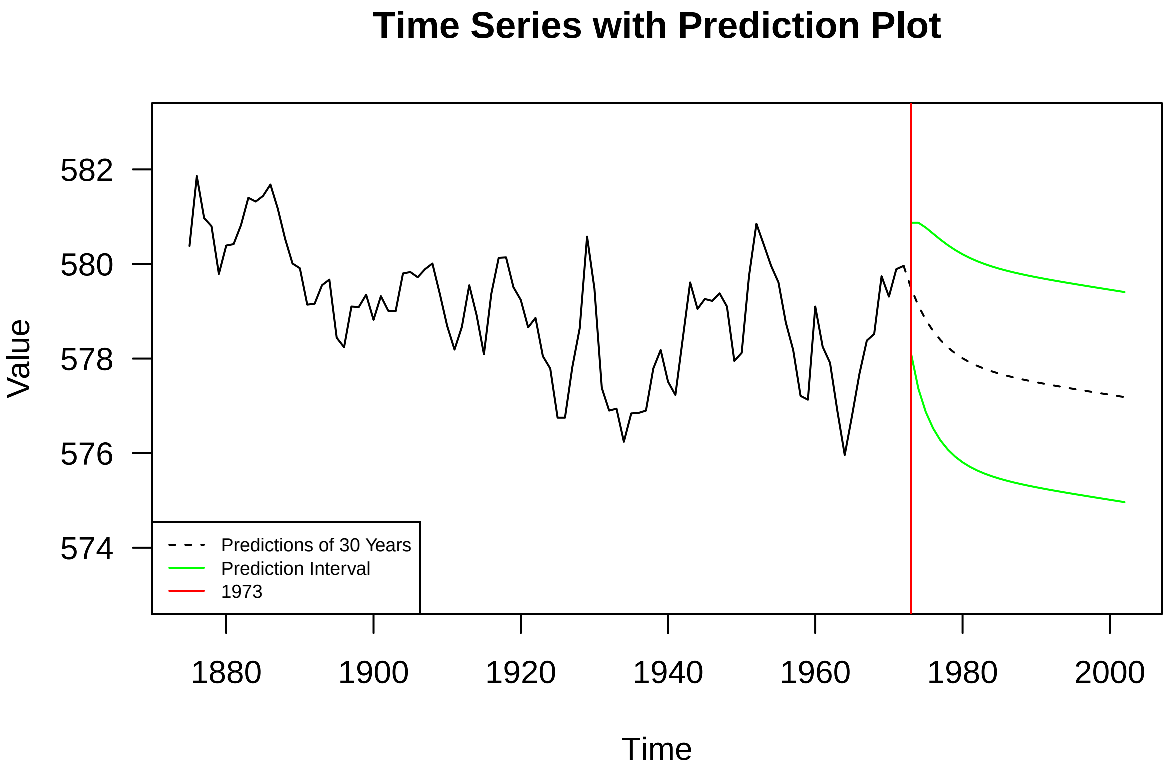
```
trend_future <- predict(lm_trend, newdata = data.frame(t = 1973:2002))
predictions <- trend_future + r_future
plot(dat, xlim = c(1875, 2002), ylab = "Value", las = 1,
     main = "Time Series with Prediction Plot")
lines(x = 1972:2002, y = c(dat[98], predictions), lty = 2)
abline(v = 1973, col = "red")
legend("bottomleft", legend = c("Predictions of 30 Years", "1973"),
       lty = 2:1, col = c("black", "red"), cex = 0.6)
```



(f) Under Gaussian noise assumption, obtain prediction intervals by using your results from the theoretical exercises.

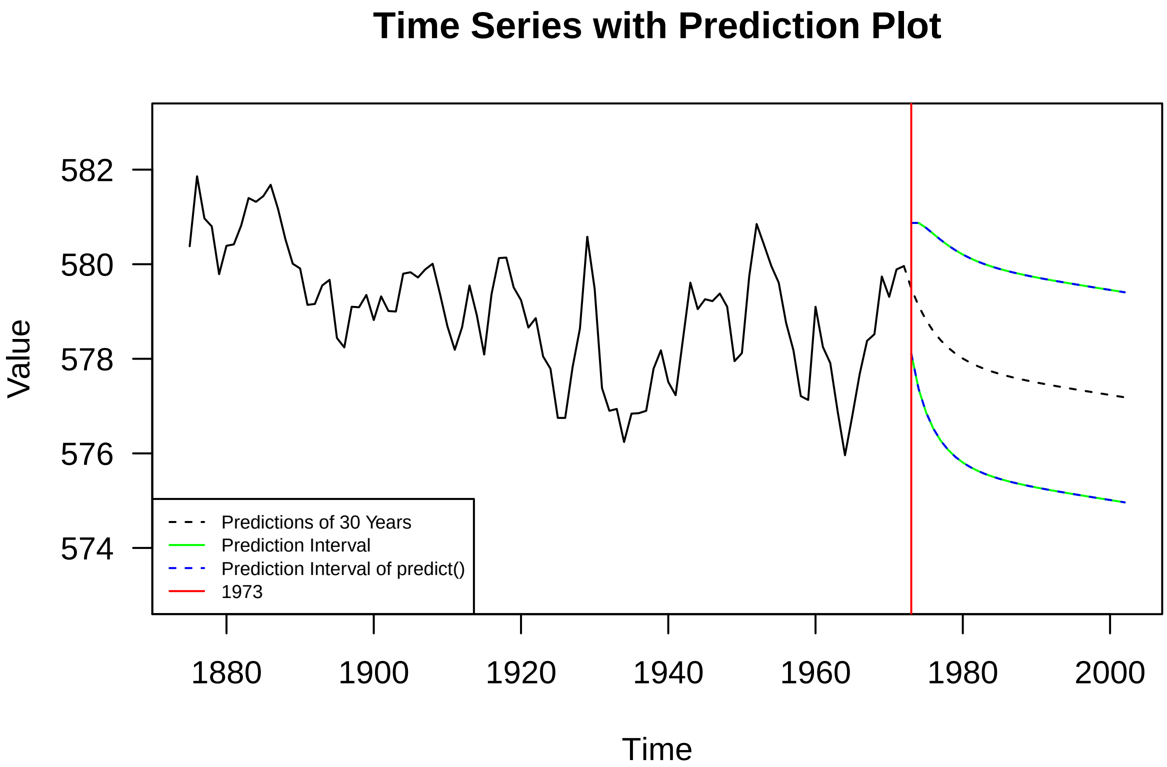
$$\mathbb{E}[(X_{n+m} - \tilde{X}_{n+m})^2] = \sigma^2 \frac{1 - \phi^{2m}}{1 - \phi^2}$$

```
sigma2 <- my_arima$sigma2
upper <- predictions + 1.96 * sqrt(sigma2 * ((1 - phi^(2*(1:30))) / (1 - phi^2)))
lower <- predictions - 1.96 * sqrt(sigma2 * ((1 - phi^(2*(1:30))) / (1 - phi^2)))

plot(dat, xlim = c(1875, 2002), ylim = c(573, 583), ylab = "Value", las = 1,
     main = "Time Series with Prediction Plot")
lines(x = 1972:2002, y = c(dat[98], predictions), lty = 2)
lines(x = 1973:2002, y = upper, col = "green")
lines(x = 1973:2002, y = lower, col = "green")
abline(v = 1973, col = "red")
legend("bottomleft",
       legend = c("Predictions of 30 Years", "Prediction Interval", "1973"),
       lty = c(2, 1, 1), col = c("black", "green", "red"), cex = 0.6)
```



(g) Compare your prediction intervals with those obtained from the `predict()` function.

```
se_pf <- predict(my_arima, n.ahead = 30)$se
upper_pf <- predictions + 1.96 * se_pf
lower_pf <- predictions - 1.96 * se_pf

plot(dat, xlim = c(1875, 2002), ylim = c(573, 583), ylab = "Value", las = 1,
     main = "Time Series with Prediction Plot")
lines(x = 1972:2002, y = c(dat[98], predictions), lty = 2)
lines(x = 1973:2002, y = upper, col = "green")
lines(x = 1973:2002, y = lower, col = "green")
lines(x = 1973:2002, y = upper_pf, col = "blue", lty = 2)
lines(x = 1973:2002, y = lower_pf, col = "blue", lty = 2)
abline(v = 1973, col = "red")
legend("bottomleft",
       legend = c("Predictions of 30 Years", "Prediction Interval",
                  "Prediction Interval of predict()", "1973"),
       lty = c(2, 1, 2, 1), col = c("black", "green", "blue", "red"), cex = 0.6)
```

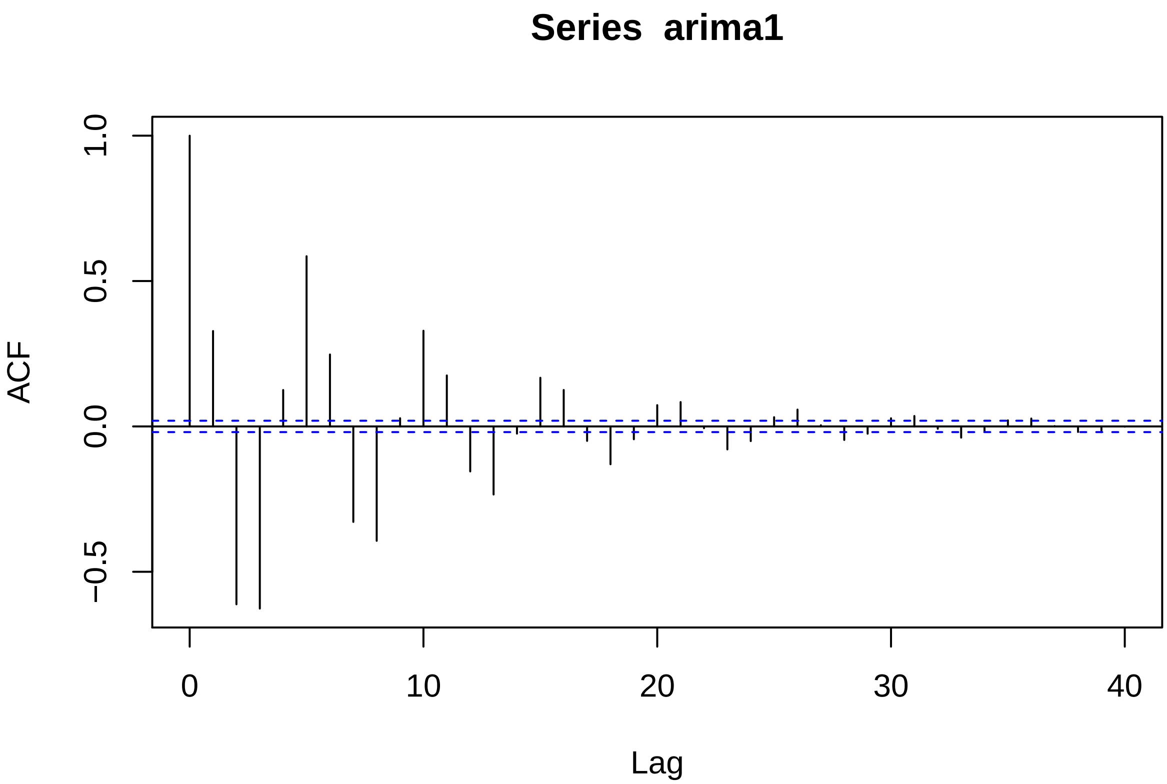


Prediction Intervals are the same.

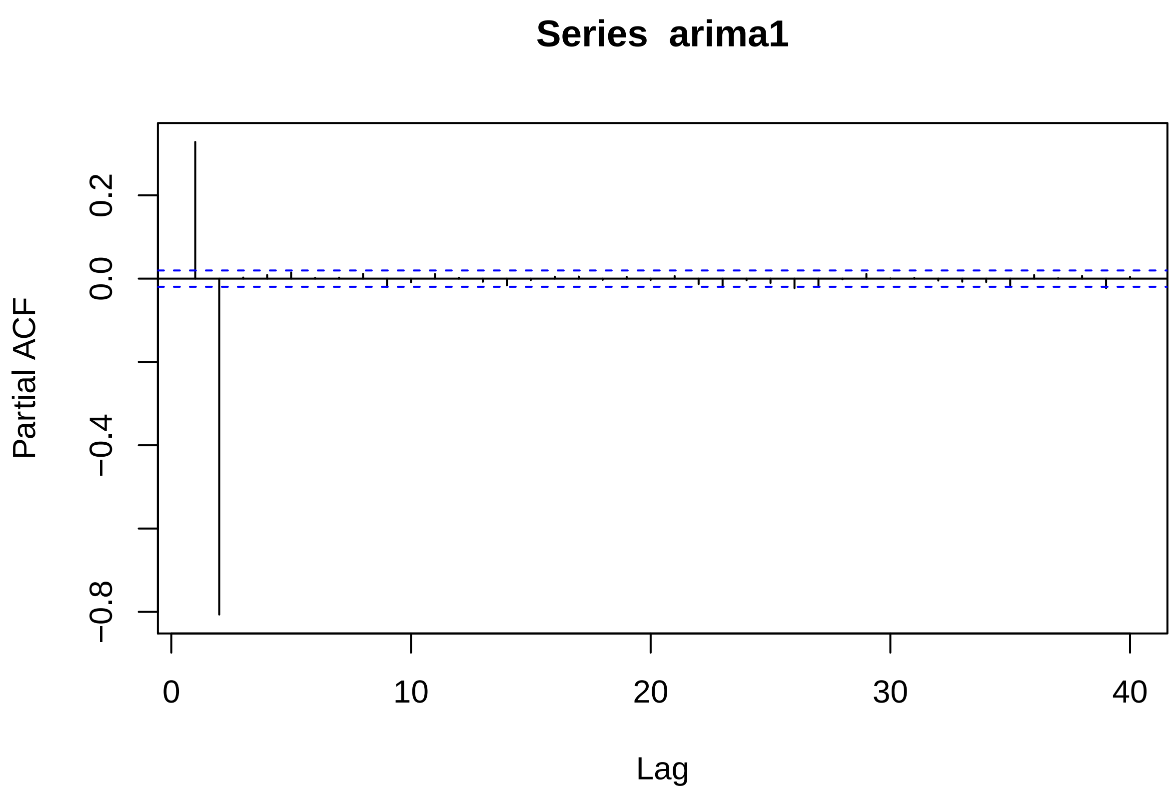
2. Plot and describe the important characteristics of sample ACF and sample PACF for each of the following ARMA models. Use 10,000 observations in each case.

(a) $X_t = \frac{3}{5}X_{t-1} - \frac{4}{5}X_{t-2} + Z_t$

```
arima1 <- arima.sim(list(ar = c(3/5, -4/5)), n = 10000)
acf(arima1)
```



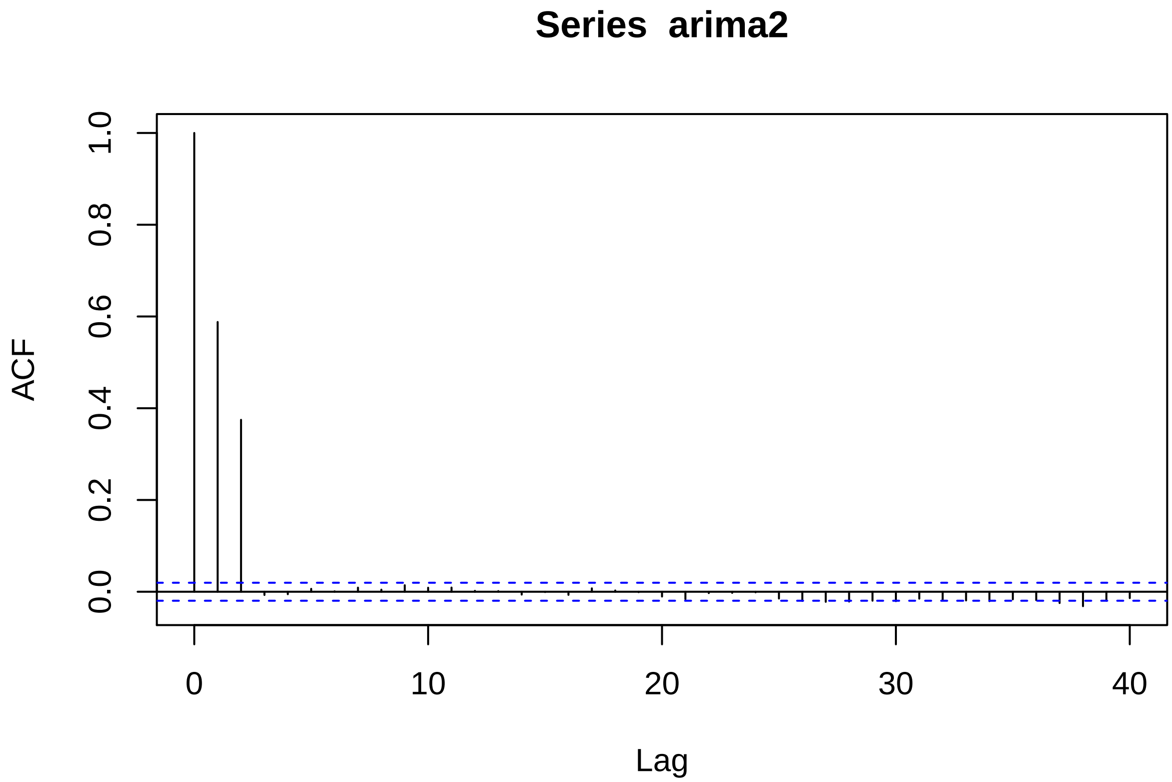

```r
pacf(arima1)
```



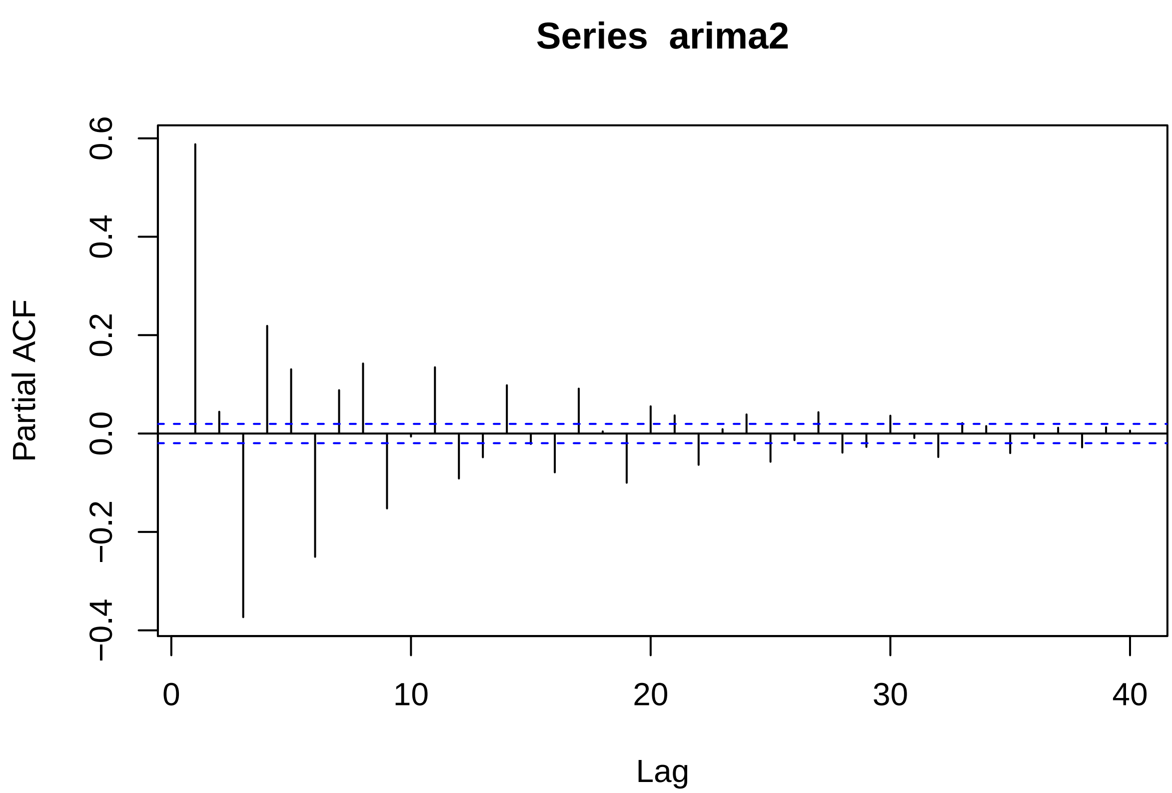
For AR(2) process, ACF tapers off and PACF cuts off after lag 2.

(b) $X_t = Z_t + 0.8Z_{t-1} + 1.1Z_{t-2}$

```r
arima2 <- arima.sim(list(ma = c(0.8, 1.1)), n = 10000)
acf(arima2)
```



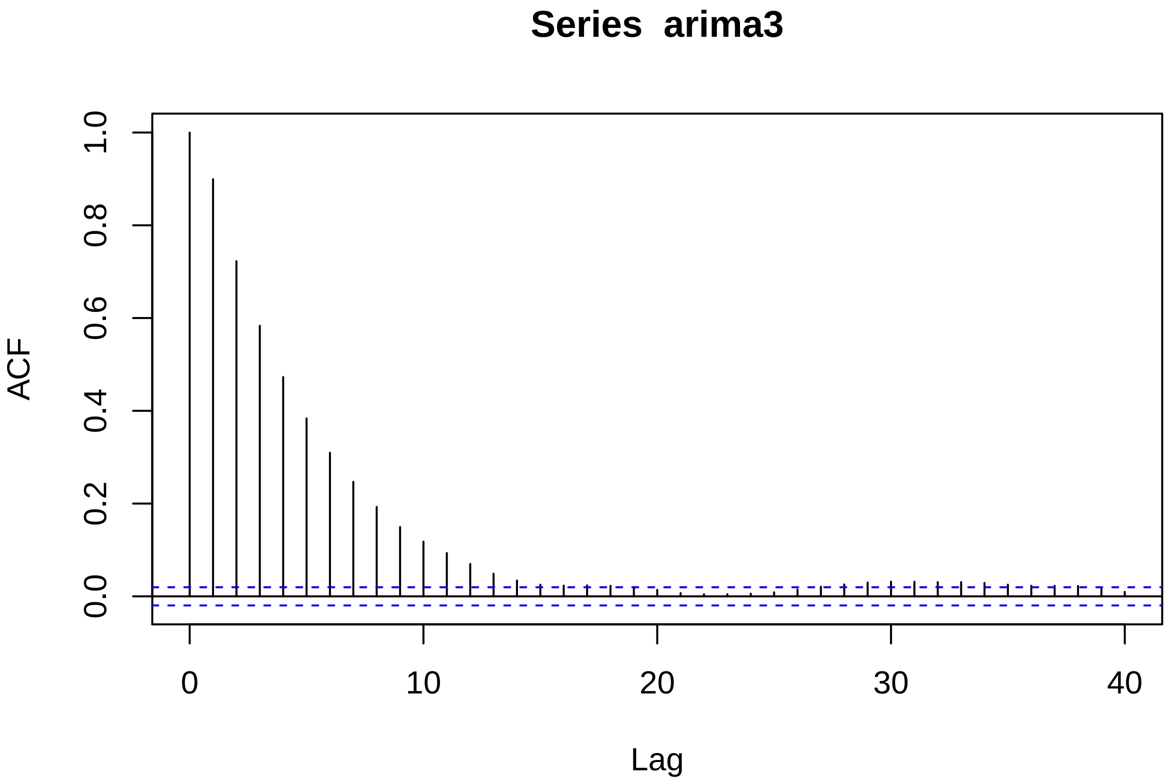

```r
pacf(arima2)
```



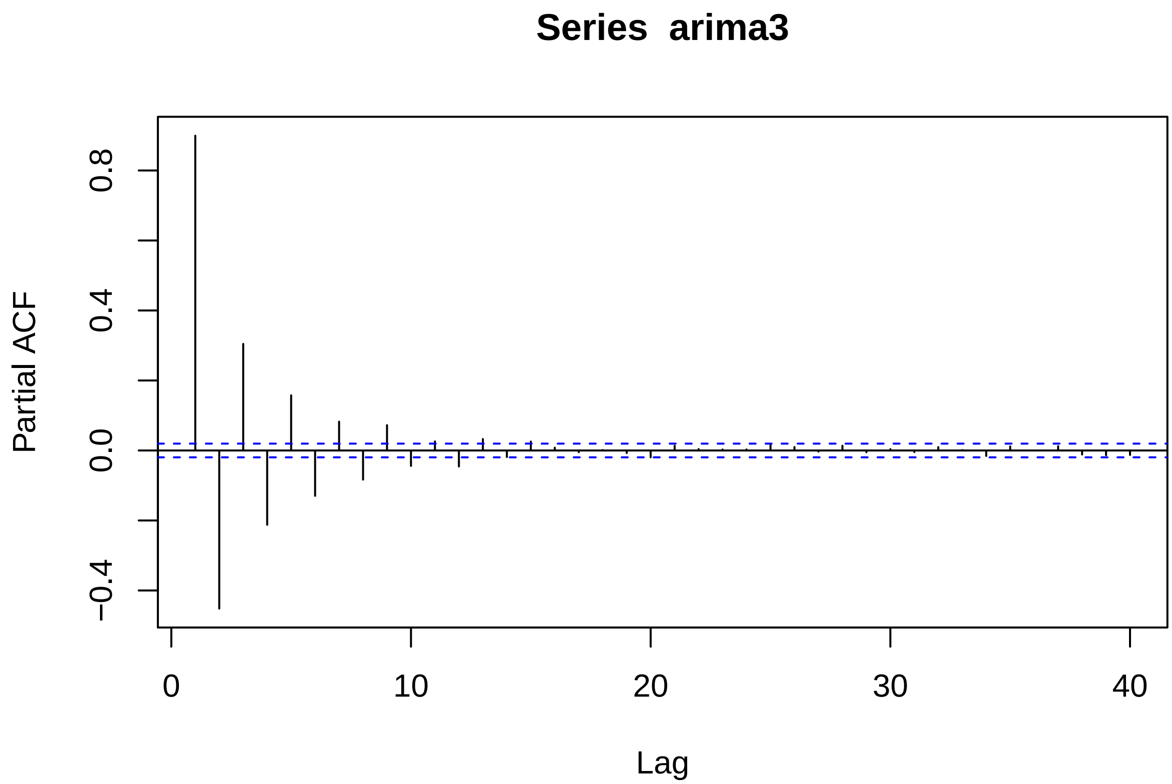
For MA(2) process, ACF cuts off after lag 2 and PACF tapers off.

(c) $X_t = \frac{4}{5}X_{t-1} + Z_t + \frac{4}{5}Z_{t-1}$

```
arima3 <- arima.sim(list(ar = 4/5, ma = 4/5), n = 10000)
acf(arima3)
```




```
pacf(arima3)
```



For ARMA(1,1), both ACF and PACF taper off.