STATS 207 (Time Series Analysis) HW2 R Code

Chih-Hsuan 'Carolyn' Kao (chkao831 at stanford dot edu)

Oct 19, 2020 (Fall 2020)

All references (exercises, examples, equations, etc.) are to Shumway & Stoffer

Q1. Exercise 2.1

In this case, time t is in quarters (1960.00, 1960.25, …) so one unit of time is a year.

(a)

Fit the regression model $x_t = \underbrace{\beta t}_{\text{trend}} + \underbrace{\alpha_1 Q_1(t) + \alpha_2 Q_2(t) + \alpha_3 Q_3(t) + \alpha_4 Q_4(t)}_{\text{seasonal}} + \underbrace{w_t}_{\text{noise}}$ where $Q_i(t) = 1$ if time t corresponds to quarter i = 1, 2, 3, 4, and zero otherwise. The $Q_i(t)$'s are called indicator variables. We will assume for now that w_t is a Gaussian white noise sequence. We will assume for now that w_t is a Gaussian white noise sequence.

```
library(astsa)
```

```
## Warning: package 'astsa' was built under R version 3.6.2
```

```
trend = time(jj) - 1970 # helps to 'center' time
Q = factor(cycle(jj)) # make (Q)uarter factors
reg = lm(log(jj) ~ 0 + trend + Q, na.action=NULL) # no intercept
#model.matrix(reg)
```

```
summary(reg)
```

```
##
## Call:
## lm(formula = log(jj) ~ 0 + trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##       Estimate Std. Error t value Pr(>|t|)
## trend 0.167172   0.002259   74.00   <2e-16 ***
## Q1    1.052793   0.027359   38.48   <2e-16 ***
## Q2    1.080916   0.027365   39.50   <2e-16 ***
## Q3    1.151024   0.027383   42.03   <2e-16 ***
## Q4    0.882266   0.027412   32.19   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9935, Adjusted R-squared:  0.9931
## F-statistic:  2407 on 5 and 79 DF,  p-value: < 2.2e-16
```

(b) If the model is correct, what is the estimated average annual increase in the logged earnings per share?

The estimated average annual increase in the logged earnings per share is

```
coefficients(summary(reg))[2][1] + coefficients(summary(reg))[3][1] + coefficients(summary(reg))[4][1] + co
efficients(summary(reg))[5][1]
```

```
## [1] 4.167
```

(c) If the model is correct, does the average logged earnings rate increase or decrease from the third quarter to the fourth quarter? And, by what percentage

does it increase or decrease?

```
coefficients(summary(reg))[5][1] - coefficients(summary(reg))[4][1]
```

```
## [1] -0.2687577
```

This is the average logged earnings rate decrease from the third quarter to the fourth quarter. In percentage, by

```
num <- (coefficients(summary(reg))[5][1] - coefficients(summary(reg))[4][1])/coefficients(summary(reg))[4][1
] * 100
print(paste0(num, ' %'))
```

```
## [1] "-23.3494385958917 %"
```

(d) What happens if you include an intercept term in the model in (a)? Explain why there was a problem.

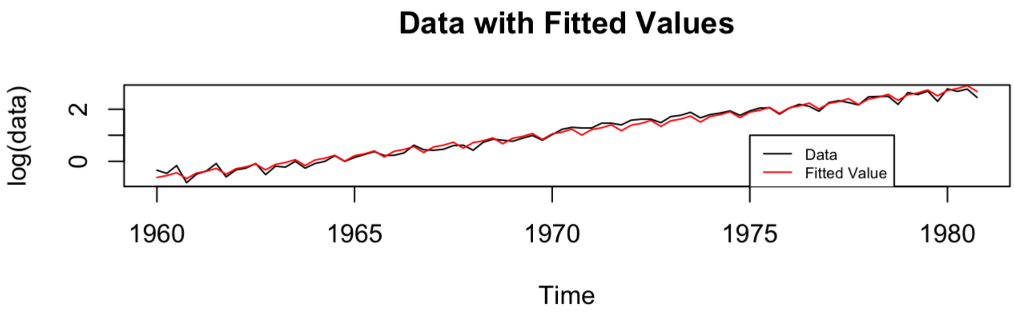
```
new_reg = lm(log(jj) ~ trend + Q, na.action=NULL)
summary(new_reg)
```

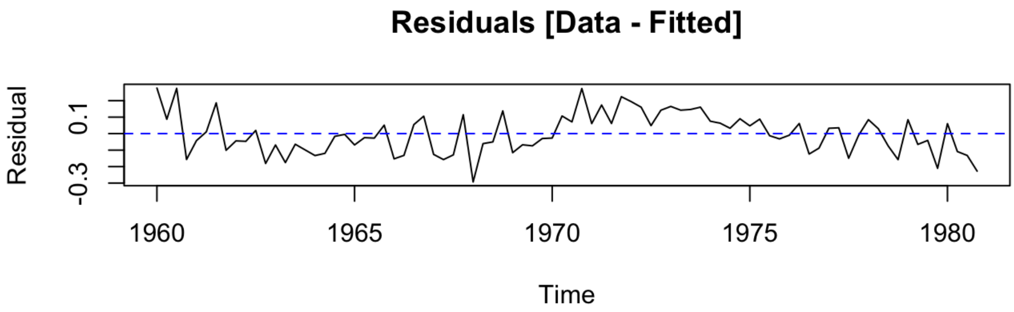
```
##
## Call:
## lm(formula = log(jj) ~ trend + Q, na.action = NULL)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29318 -0.09062 -0.01180  0.08460  0.27644
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.052793   0.027359  38.480  < 2e-16 ***
## trend        0.167172   0.002259  73.999  < 2e-16 ***
## Q2           0.028123   0.038696   0.727   0.4695
## Q3           0.098231   0.038708   2.538   0.0131 *
## Q4          -0.170527   0.038729  -4.403 3.31e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1254 on 79 degrees of freedom
## Multiple R-squared:  0.9859, Adjusted R-squared:  0.9852
## F-statistic:  1379 on 4 and 79 DF,  p-value: < 2.2e-16
```

With the intercept included, the Q1 is explicitly taken away. For the remaining quarters, the intercept effect is included. There was a problem because this hinders our ability to interpret each quarter performance.

(e) Graph the data, x, and superimpose the fitted values, say $\hat{x}$. Examine the residuals, $x - \hat{x}$, and state your conclusions.

```
par(mfrow=c(2,1))
plot(log(jj), main="Data with Fitted Values", ylab = 'log(data)')
lines(fitted(reg), col="red") # fitted
legend(1975,1,
       c("Data","Fitted Value"),
       lty=c(1,1),
       col = c('black', 'red'),
       cex = 0.6)

res <- log(jj)-fitted(reg)
plot(res, main="Residuals [Data - Fitted]", ylab = 'Residual')
abline(h=0, lty=2, col="blue")
```





Does it appear that the model fits the data well (do the residuals look white)?

From the first plot above, the fitted value fits the data well. From the residual plot, the residuals look white in terms of the pattern, as well as the decaying nature of ACF plot.

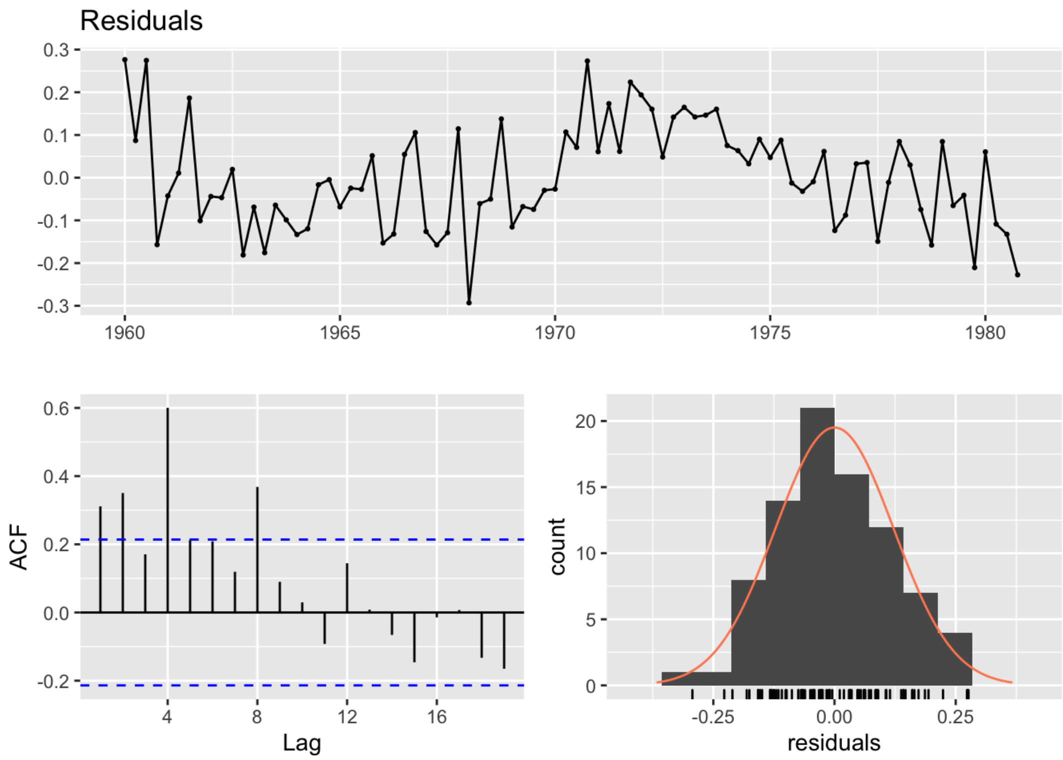
```
forecast::checkresiduals(res)
```

```
## Registered S3 method overwritten by 'xts':
##   method     from
##   as.zoo.xts zoo
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
## Registered S3 methods overwritten by 'forecast':
##   method             from
##   fitted.fracdiff    fracdiff
##   residuals.fracdiff fracdiff
```

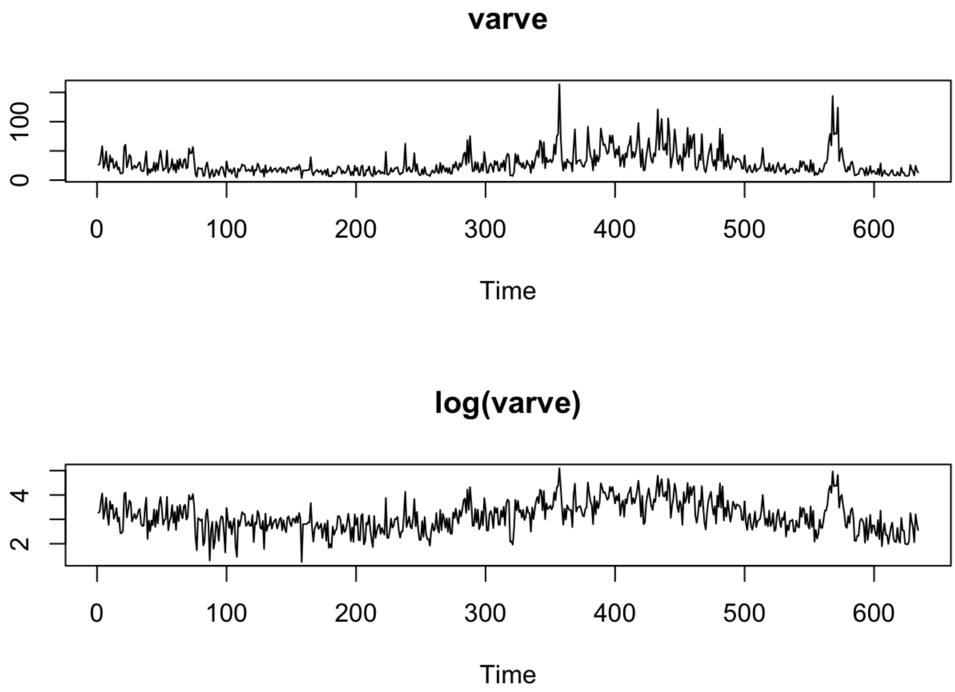
```
## Warning in modeldf.default(object): Could not find appropriate degrees of
## freedom for this model.
```



Q2. Exercise 2.8

The glacial varve record plotted below exhibits some nonstationarity that can be improved by transforming to logarithms and some additional nonstationarity that can be corrected by differencing the logarithms.

```
library("astsa")
par(mfrow = c(2,1))
x<-varve
plot(x, main="varve", ylab="")
y<- log(varve)
plot(y, main="log(varve)", ylab="" )
```



(a) Argue that the glacial varves series, say x_t, exhibits heteroscedasticity by computing the sample variance over the first half and the second half of the data.

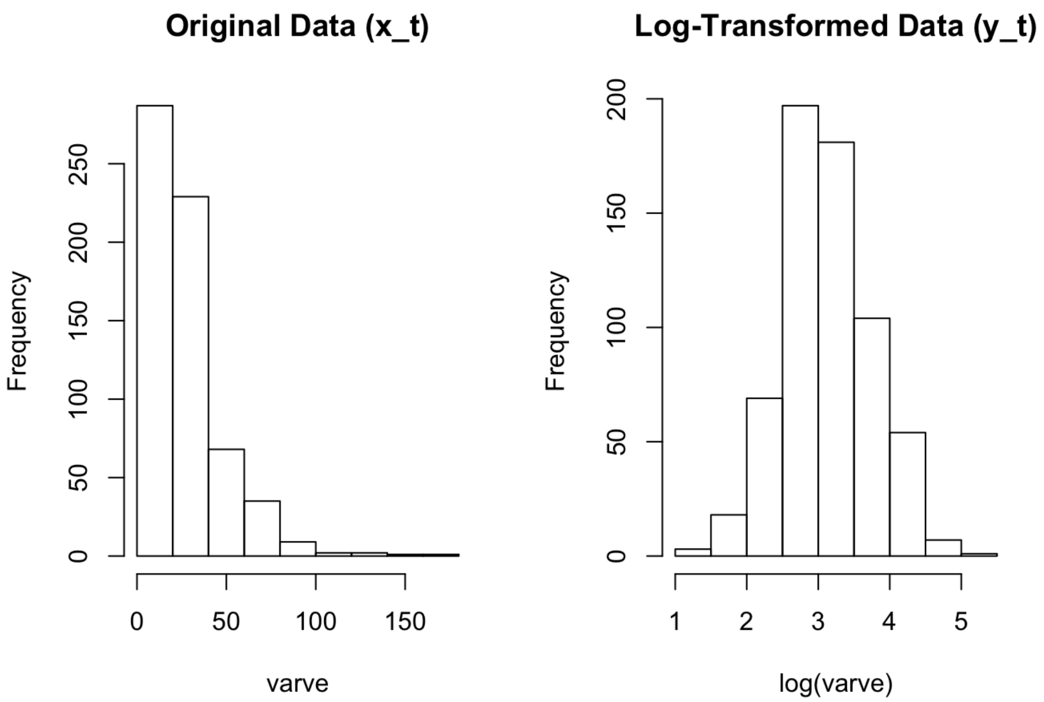
```r
n = length(varve)
varve1 = varve[1:n/2]
varve2 = varve[(n/2+1):n]

content <- matrix(c(var(varve1), var(varve2)),ncol=2,byrow=TRUE)
rownames(content)<-c("Sample Variance")
colnames(content)<-c("1st Half", "2nd Half")
tab <- as.table(content)
tab
```

```
##                 1st Half 2nd Half
## Sample Variance 132.5010 594.4904
```

From here, it is seen that Var(2nd_Half) >> Var(1st_Half) of the data, indicating the heteroscedasticity nature.

Argue that the transformation $y_t = log x_t$ stabilizes the variance over the series. Plot the histograms of x_t and y_t to see whether the approximation to normality is improved by transforming the data.

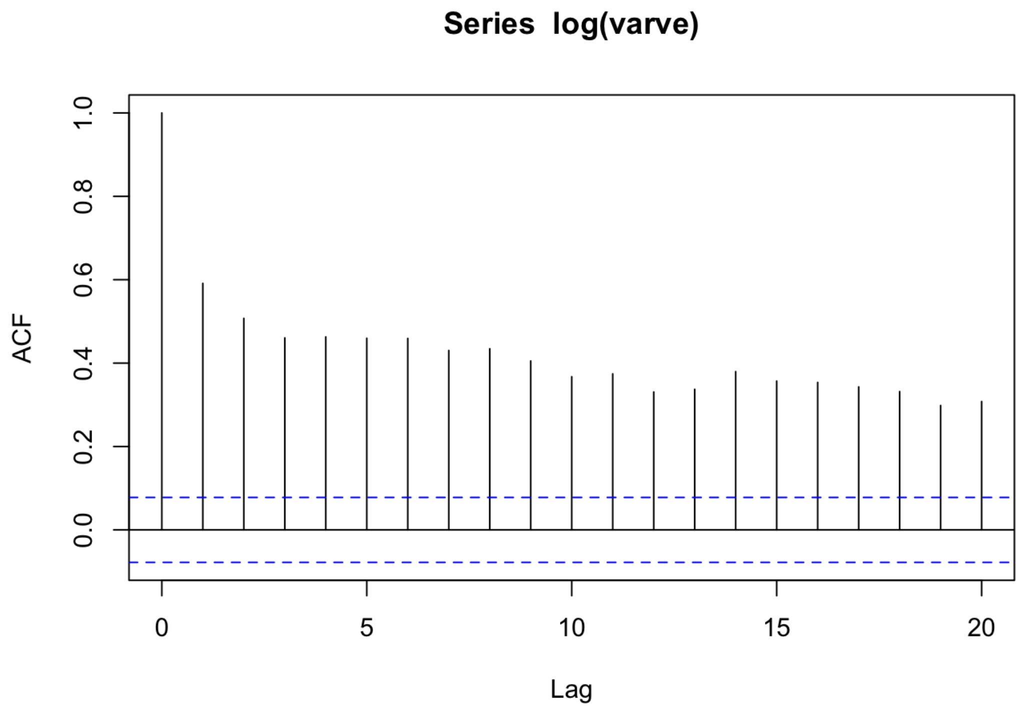
```r
par(mfrow = c(1,2))
hist(varve, main="Original Data (x_t)")
hist(log(varve), main="Log-Transformed Data (y_t)")
```



From here, the logarithmic transformation of the original data statibilizes the variance, as shown in the right. Such a transformation made the raw data relatively more normal.

(b) Examine the sample ACF of y_t and comment.

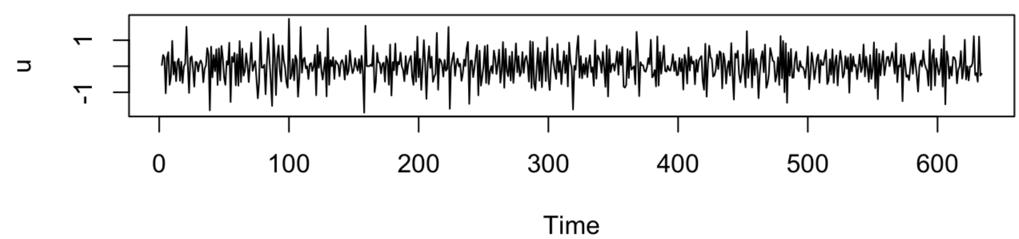
```r
acf(log(varve), lag.max = 20)
```

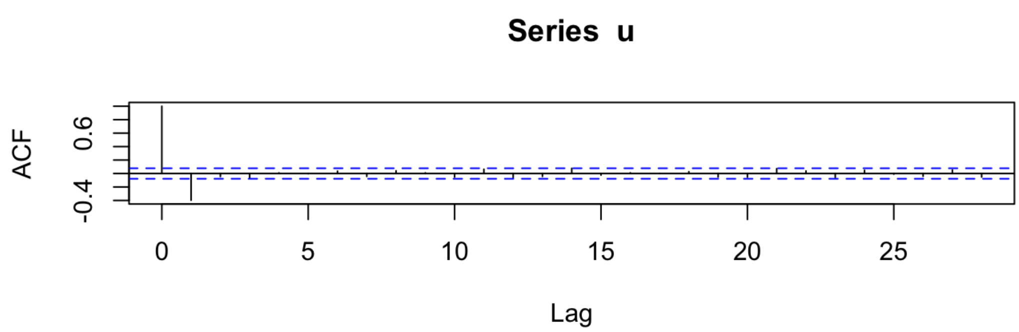


From the plot above, it is seen that the ACF doesn't decay fast enough, illustrating significant data dependency in neighbor lags.

(c) Compute the difference $u_t = y_t - y_{t-1}$, examine its time plot and sample ACF, and argue that differencing the logged varve data produces a reasonably stationary series. Can you think of a practical interpretation for u_t?

```
par(mfrow = c(2,1))
u = diff(log(varve), 1)
plot(u)
acf(u)
```





Footnote (*):

$$log(1+p) = p - \frac{p^2}{2} + \frac{p^3}{3} - ... for -1 < p \leq 1.$$

If p is near zero, the higher-order terms in the expansion are negligible.

To interpret u_t, it is the yearly increase of the varve thicknesses. Statistically, by the Footnote (*),

$$u_t = \nabla y_t = y_t - y_{t-1} = log(x_t) - log(x_{t-1}) \approx \frac{x_t - x_{t-1}}{x_{t-1}}$$

which can also be practically interpreted as the marginal change in the varve thicknesses.

Q4. Exercise 3.9

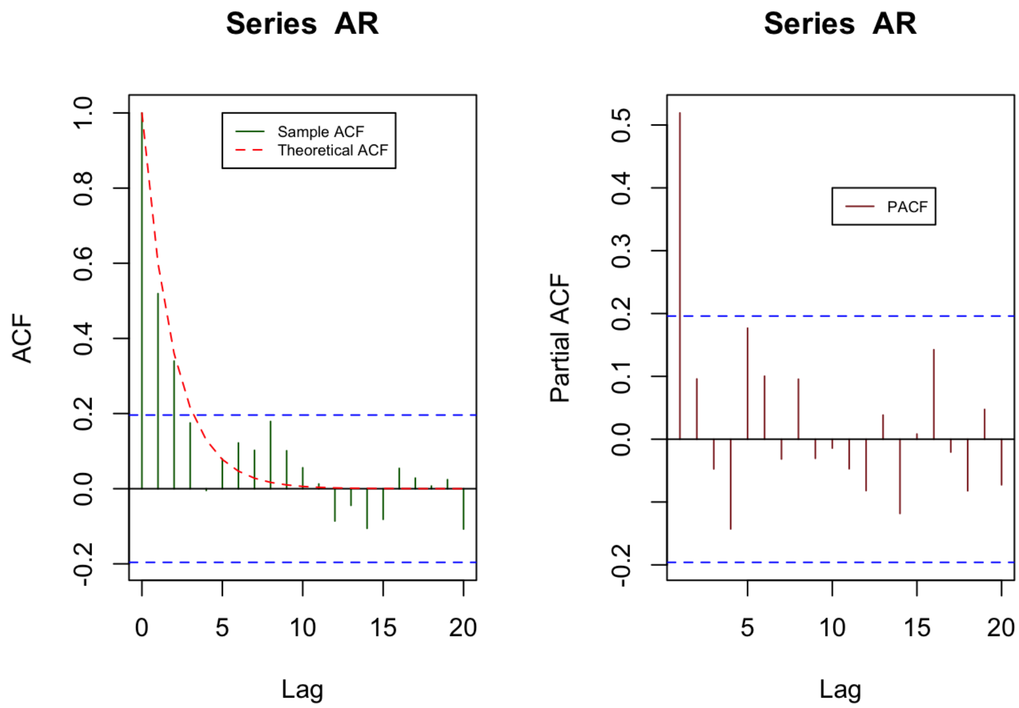
Generate n = 100 observations from each of the three models ARMA(1,1), ARMA(1,0) and ARMA(0,1).

```
AR <- arima.sim(list(order=c(1,0,0), ar = 0.6), n = 100)
MA <- arima.sim(list(order=c(0,0,1), ma = 0.9), n = 100)
ARMA <- arima.sim(list(order=c(1,0,1), ar = 0.6, ma = 0.9), n = 100)
```

ARMA(1,0)

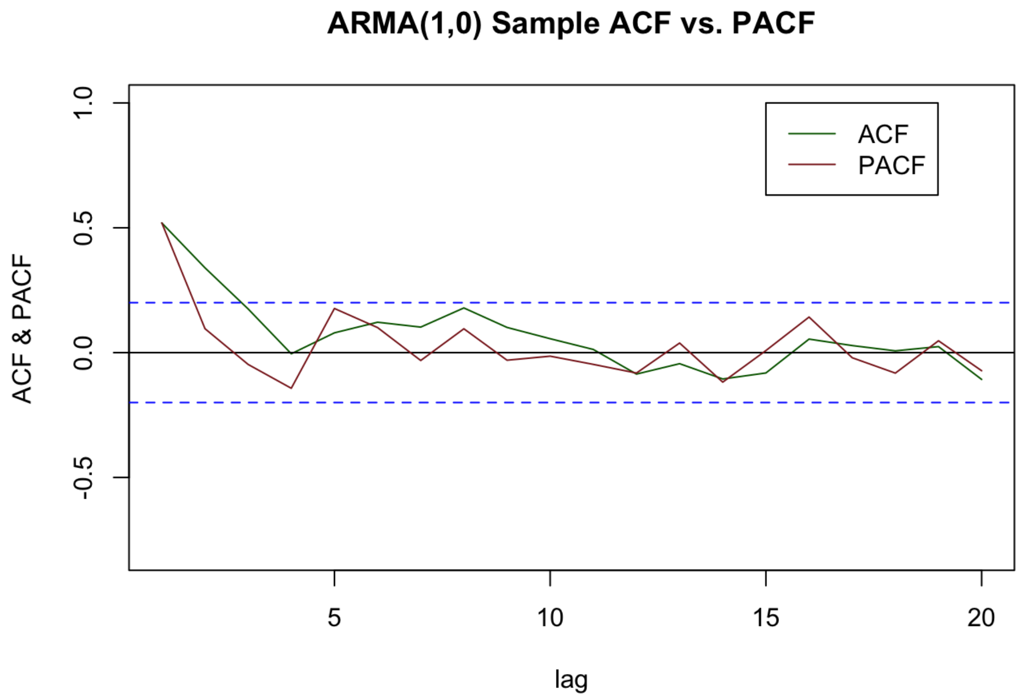
Compute the sample ACF for each model and compare it to the theoretical values.

```
par(mfcol= c(1,2))

ARMA10_theoACF <- ARMAacf(ar = 0.6, lag.max = 20)
ARMA10_sampleACF <- acf(AR, plot = FALSE)

#[LEFT] black: sample ACF; red: theoretical ACF
plot(ARMA10_sampleACF, col = 'dark green')
lines(ARMA10_sampleACF$lag,
      ARMA10_theoACF,
      type = "l",
      lty = 2,
      col = 'red')
legend(5,1,
       c("Sample ACF","Theoretical ACF"),
       lty=c(1,2),
       col = c('dark green', 'red'),
       cex = 0.6)

#[RIGHT] PACF
pacf(AR, col = 'brown4')
legend(10,0.4,
       c("PACF"),
       lty=c(1),
       col = c('brown4'),
       cex = 0.6)
```



Compute the sample PACF for each of the generated series and compare the sample ACFs and PACFs with the general results given in Table 3.1.

```
ARMA10_pacf <- pacf(AR, plot=F)
plot(ARMA10_sampleACF$acf[-1], #truncate 1 out to matched x y length
     type='l',
     lty=1,
     ylim=c(-.8,1),
     ylab='ACF & PACF',
     xlab='lag',
     main="ARMA(1,0) Sample ACF vs. PACF",
     col = 'dark green')
abline(h=0)
abline(h=c(.2,-.2),lty=2, col = 'blue')
lines(ARMA10_sampleACF$lag[-1],ARMA10_pacf$acf, col = 'brown4')
legend(15,1,
       c("ACF","PACF"),
       lty=c(1,1),
       col = c('dark green', 'brown4'))
```



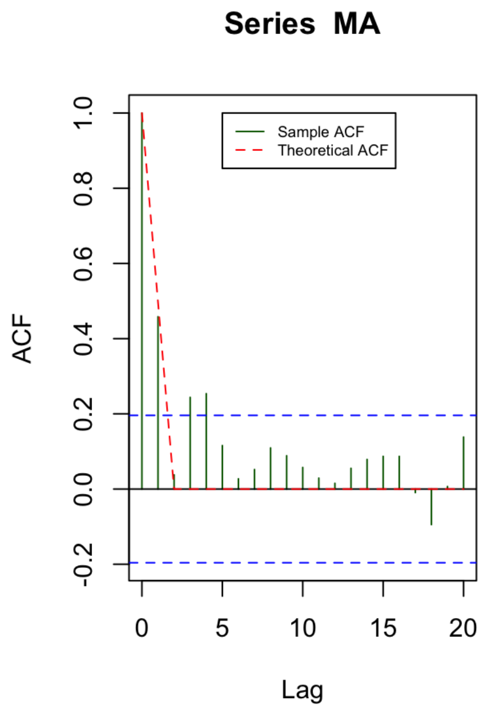

```
content <- matrix(c('Tails Off', 'Cuts off after lag p'),ncol=1,byrow=TRUE)
rownames(content)<-c("ACF","PACF")
colnames(content)<-c("AR(p)")
tab <- as.table(content)
knitr::kable(tab) #and then add some comments using previous graph and this table 3.1
```

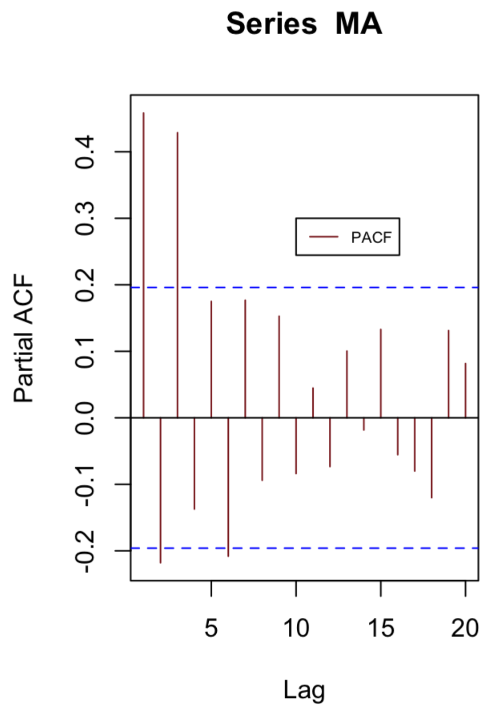
	AR(p)
ACF	Tails Off
PACF	Cuts off after lag p

ARMA(0,1)

Compute the sample ACF for each model and compare it to the theoretical values.

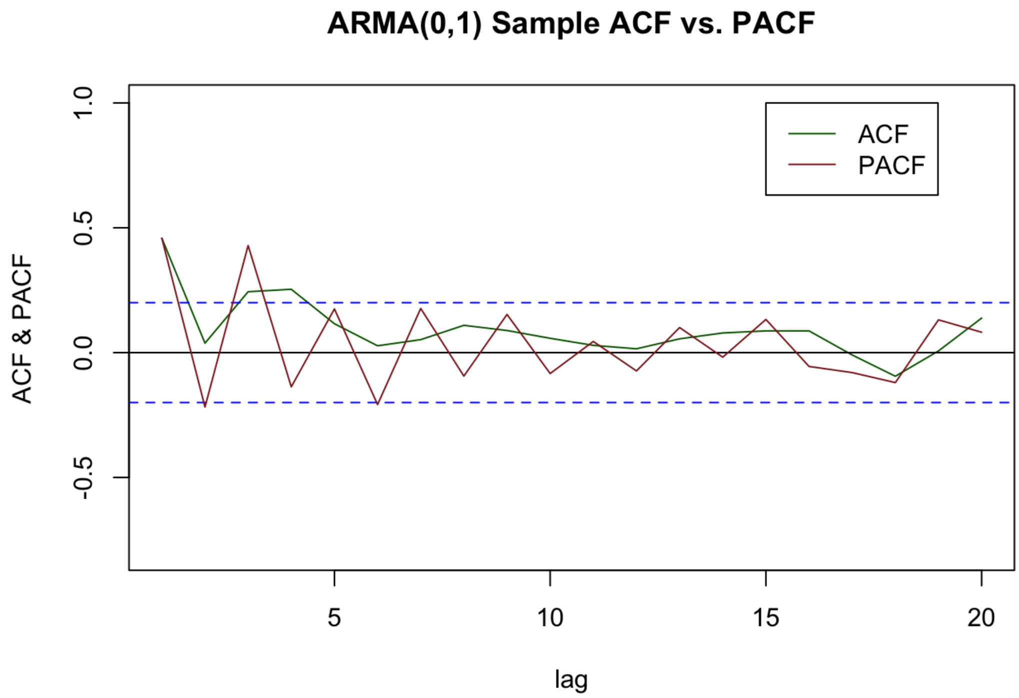
```
par(mfcol= c(1,2))

ARMA01_theoACF <- ARMAacf(ma = 0.9, lag.max = 20)
ARMA01_sampleACF <- acf(MA, plot = FALSE)

#[LEFT] black: sample ACF; red: theoretical ACF
plot(ARMA01_sampleACF, col = 'dark green')
lines(ARMA01_sampleACF$lag,
      ARMA01_theoACF,
      type = "l",
      lty = 2,
      col = 'red')
legend(5,1,
       c("Sample ACF","Theoretical ACF"),
       lty=c(1,2),
       col = c('dark green', 'red'),
       cex = 0.6)

#[RIGHT] PACF
pacf(MA, col = 'brown4')
legend(10,0.3,
       c("PACF"),
       lty=c(1),
       col = c('brown4'),
       cex = 0.6)
```





Compute the sample PACF for each of the generated series and compare the sample ACFs and PACFs with the general results given in Table 3.1.

```
ARMA01_pacf <- pacf(MA, plot=F)
plot(ARMA01_sampleACF$acf[-1], #truncate 1 out to matched x y length
     type='l',
     lty=1,
     ylim=c(-.8,1),
     ylab='ACF & PACF',
     xlab='lag',
     main="ARMA(0,1) Sample ACF vs. PACF",
     col = 'dark green')
abline(h=0)
abline(h=c(.2,-.2),lty=2, col = 'blue')
lines(ARMA01_sampleACF$lag[-1],ARMA01_pacf$acf, col = 'brown4')
legend(15,1,
       c("ACF","PACF"),
       lty=c(1,1),
       col = c('dark green', 'brown4'))
```



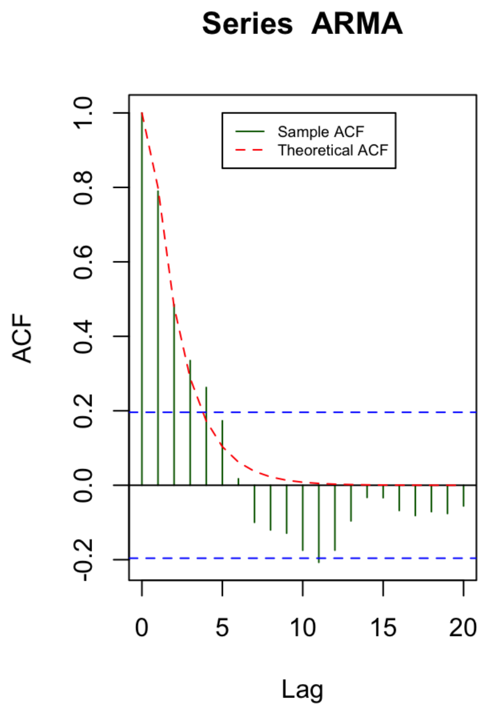

```
content <- matrix(c('Cuts off after lag q', 'Tails off'),ncol=1,byrow=TRUE)
rownames(content)<-c("ACF","PACF")
colnames(content)<-c("MA(q)")
tab <- as.table(content)
knitr::kable(tab) #and then add some comments using previous graph and this table 3.1
```

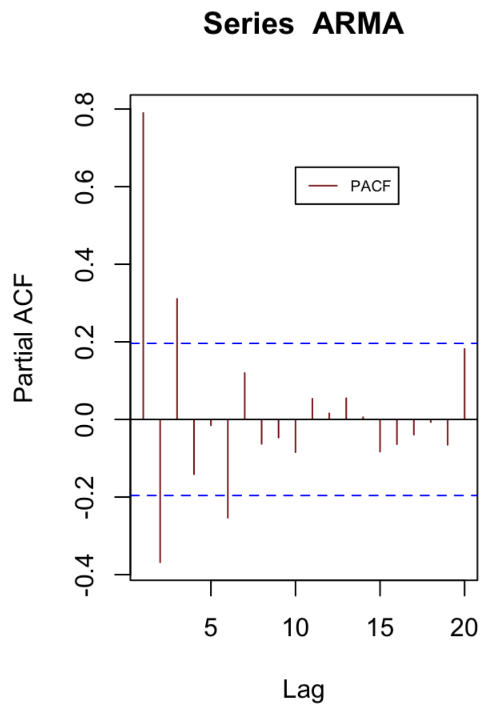
	MA(q)
ACF	Cuts off after lag q
PACF	Tails off

ARMA(1,1)

Compute the sample ACF for each model and compare it to the theoretical values.

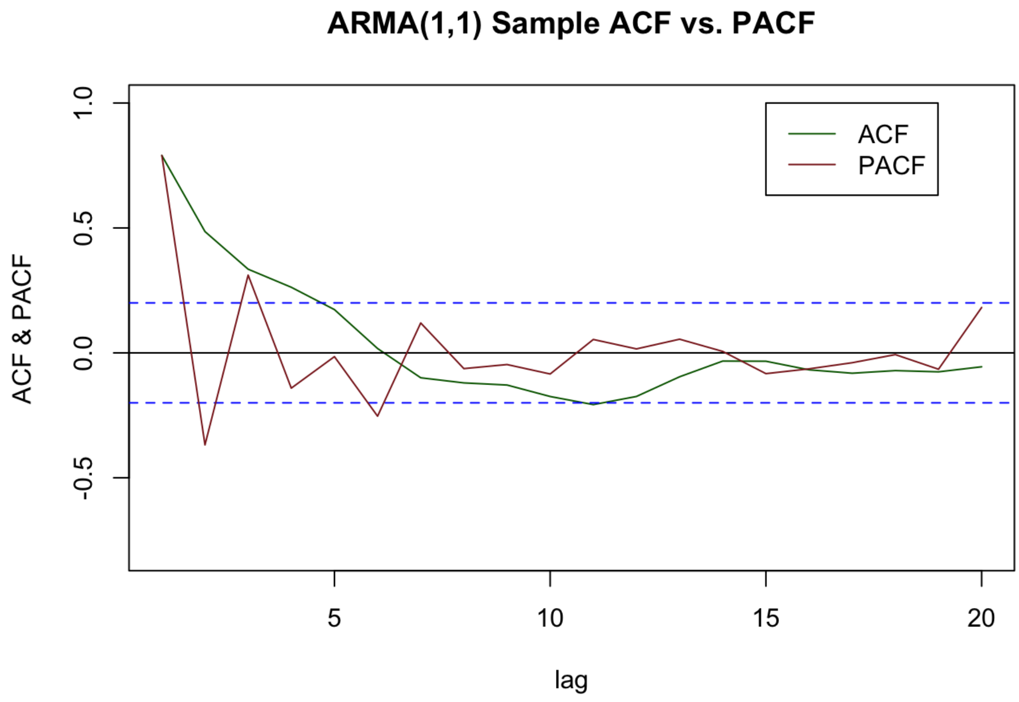
```
par(mfcol= c(1,2))

ARMA11_theoACF <- ARMAacf(ar = 0.6, ma = 0.9, lag.max = 20)
ARMA11_sampleACF <- acf(ARMA, plot = FALSE)

#[LEFT] black: sample ACF; red: theoretical ACF
plot(ARMA11_sampleACF, col = 'dark green')
lines(ARMA11_sampleACF$lag,
      ARMA11_theoACF,
      type = "l",
      lty = 2,
      col = 'red')
legend(5,1,
       c("Sample ACF","Theoretical ACF"),
       lty=c(1,2),
       col = c('dark green', 'red'),
       cex = 0.6)

#[RIGHT] PACF
pacf(ARMA, col = 'brown4')
legend(10,0.65,
       c("PACF"),
       lty=c(1),
       col = c('brown4'),
       cex = 0.6)
```





Compute the sample PACF for each of the generated series and compare the sample ACFs and PACFs with the general results given in Table 3.1.

```
ARMA11_pacf <- pacf(ARMA, plot=F)
plot(ARMA11_sampleACF$acf[-1], #truncate 1 out to matched x y length
     type='l',
     lty=1,
     ylim=c(-.8,1),
     ylab='ACF & PACF',
     xlab='lag',
     main="ARMA(1,1) Sample ACF vs. PACF",
     col = 'dark green')
abline(h=0)
abline(h=c(.2,-.2),lty=2, col = 'blue')
lines(ARMA11_sampleACF$lag[-1],ARMA11_pacf$acf, col = 'brown4')
legend(15,1,
       c("ACF","PACF"),
       lty=c(1,1),
       col = c('dark green', 'brown4'))
```




```
content <- matrix(c('Tails off', 'Tails off'),ncol=1,byrow=TRUE)
rownames(content)<-c("ACF","PACF")
colnames(content)<-c("ARMA(p,q)")
tab <- as.table(content)
knitr::kable(tab) #and then add some comments using previous graph and this table 3.1
```

	ARMA(p,q)
ACF	Tails off
PACF	Tails off

Processing math: 100%