Problem1A

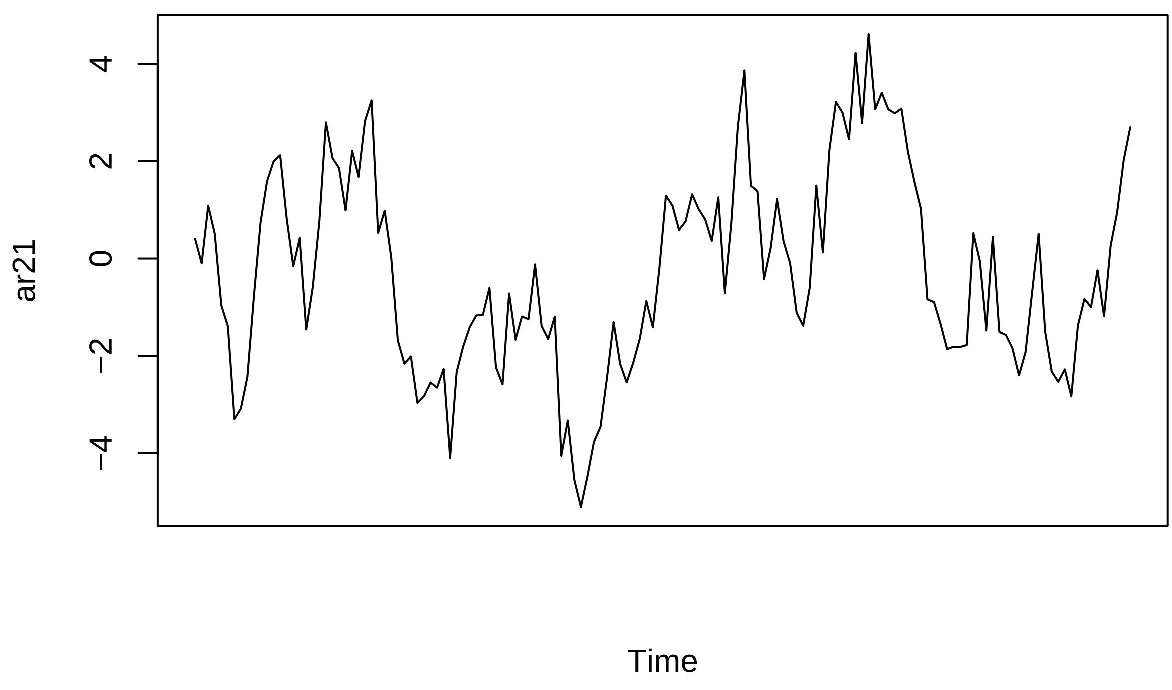
Fjolle Gjonbalaj

2 February 2021

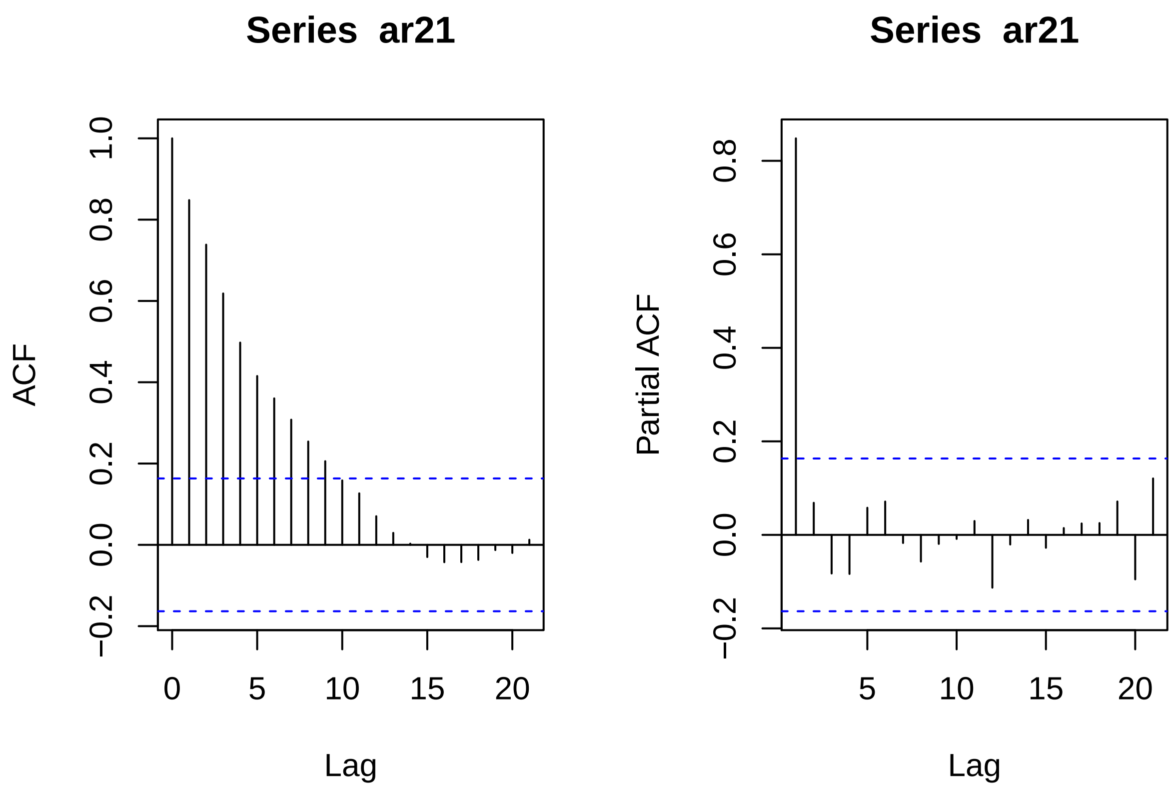
QUESTION 1 I.

Part a

```
ar21 <- arima.sim(list(order=c(2,0,0), ar = c(0.7, 0.2)), n=144)
plot.ts(ar21, axes=F); box(); axis(2)
```



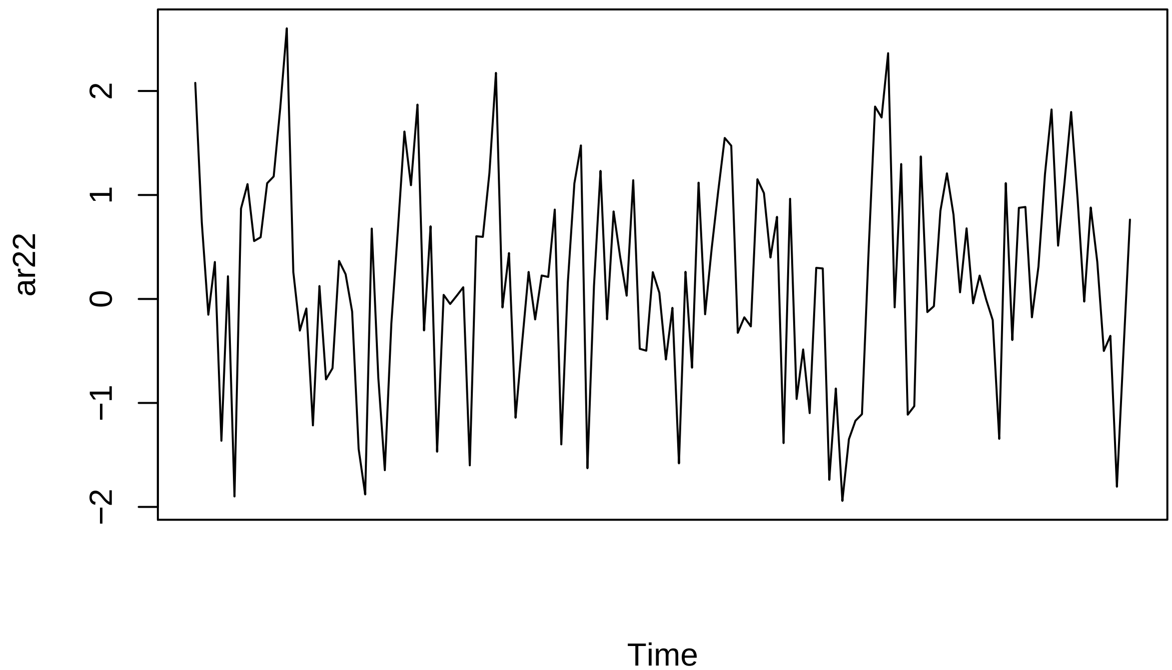

```
par(mfrow=c(1,2))
ar21.acf<-acf(ar21, plot=TRUE)
ar21.pacf<-pacf(ar21, plot=TRUE)
```



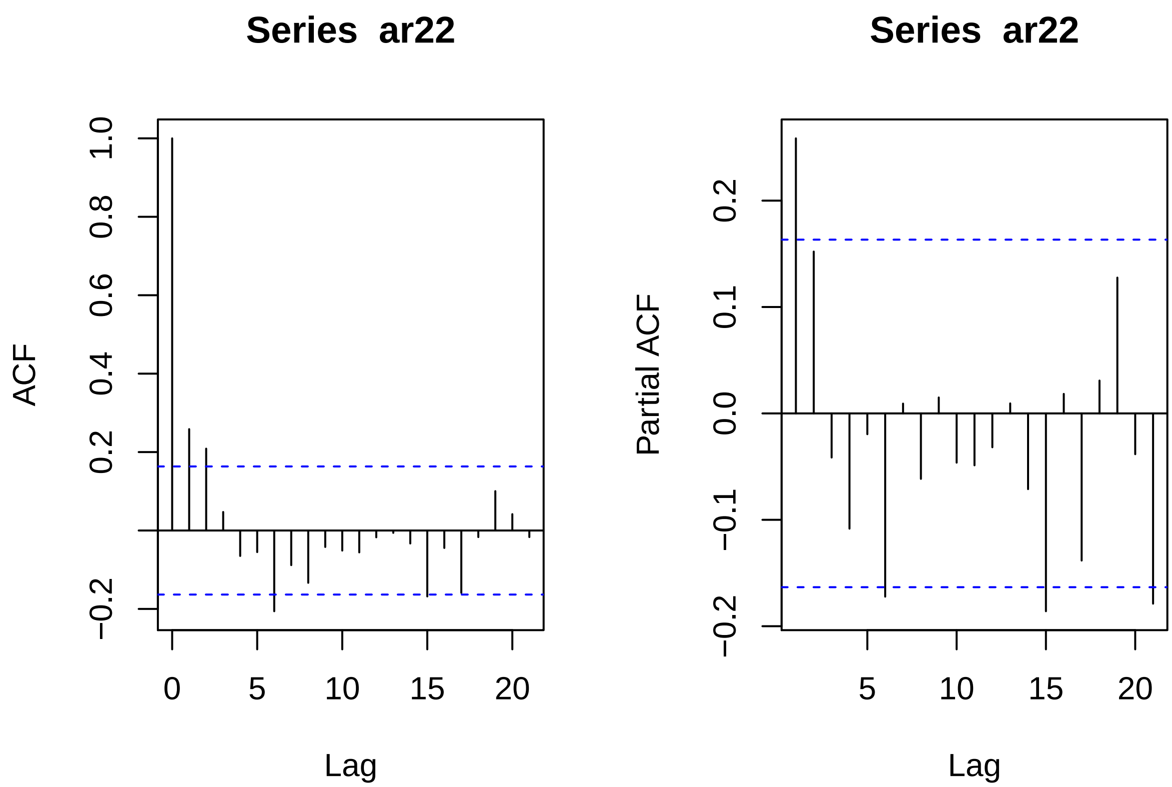
In the acf function of this AR(2) model we observe a geometric decay In the pacf function of this AR(2) model, on the other hand, we notice that the first and possibly the second lag are also significant. This indicates that we need to use at least one lag in our model. Whether we need to include a second lag or not needs to be determined by a more formal time series test.

Part b

```
ar22 <- arima.sim(list(order=c(2,0,0), ar = c(0.2, 0.2)), n=144)
plot.ts(ar22, axes=F); box(); axis(2)
```



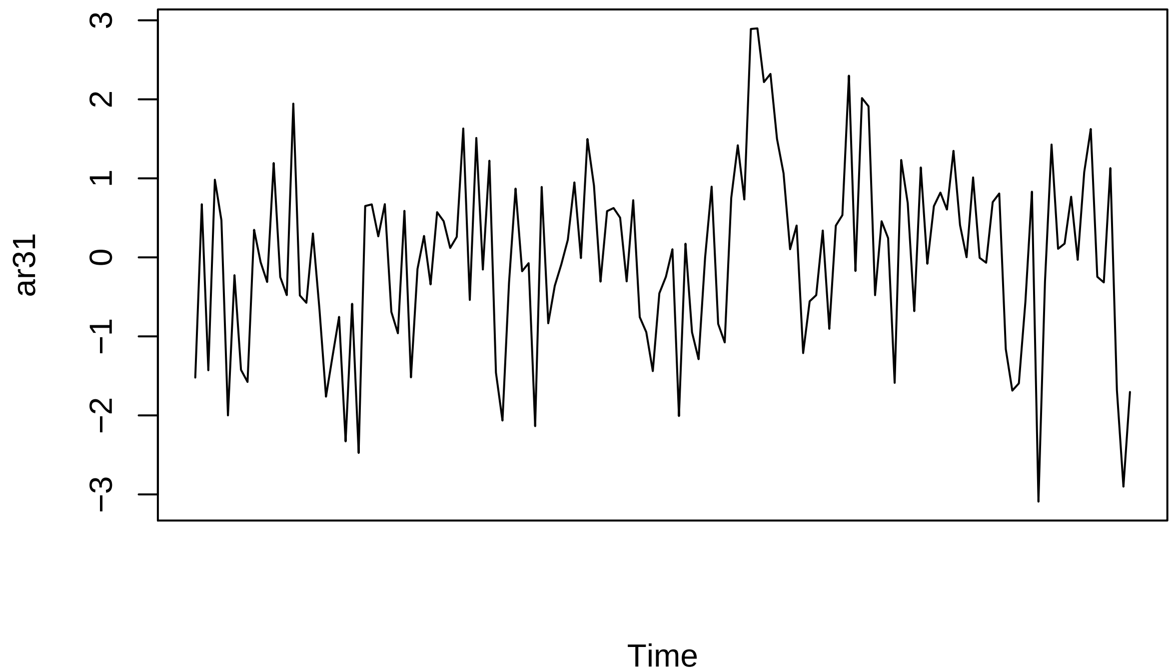

```
par(mfrow=c(1,2))
ar22.acf<-acf(ar22, plot=TRUE)
ar22.pacf<-pacf(ar22, plot=TRUE)
```



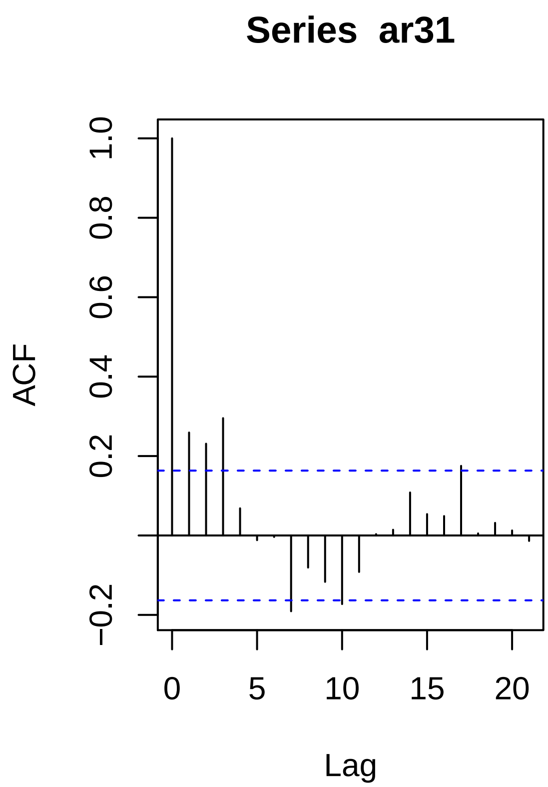
Different from part a, in the acf function of this AR(2) model we observe that the acf function does not show a geometric decay. In the pacf function of this AR(2) model we notice that the there does not seem to be much significance on the lags. The result of these differences are most likely due to the lag coefficients being very low

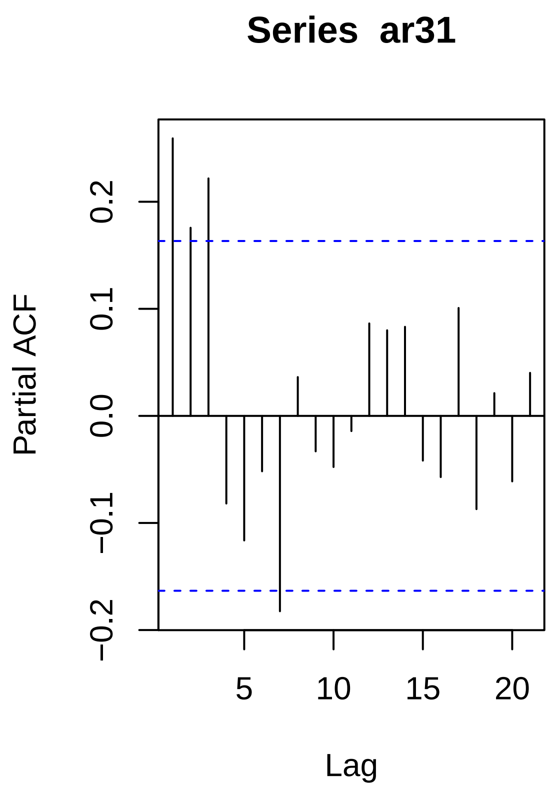
Part c

```
ar31 <- arima.sim(list(order=c(3,0,0), ar = c(0.2, 0.2, 0.2)), n=144)
plot.ts(ar31, axes=F); box(); axis(2)
```




```
par(mfrow=c(1,2))
ar31.acf<-acf(ar31, plot=TRUE)
ar31.pacf<-pacf(ar31, plot=TRUE)
```

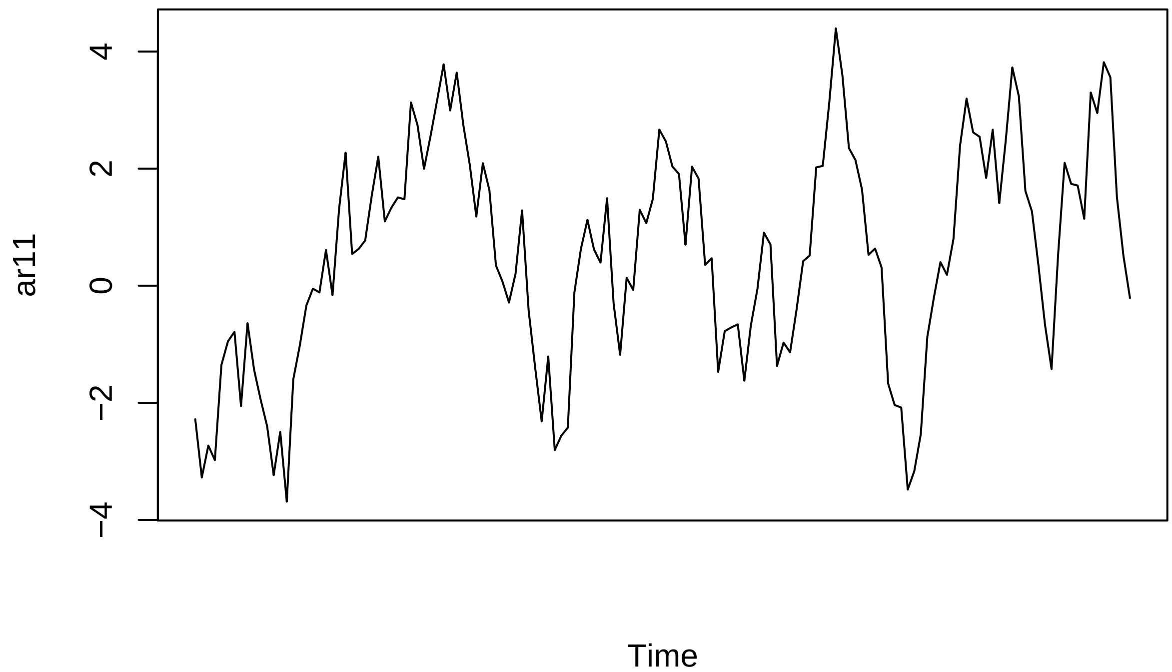




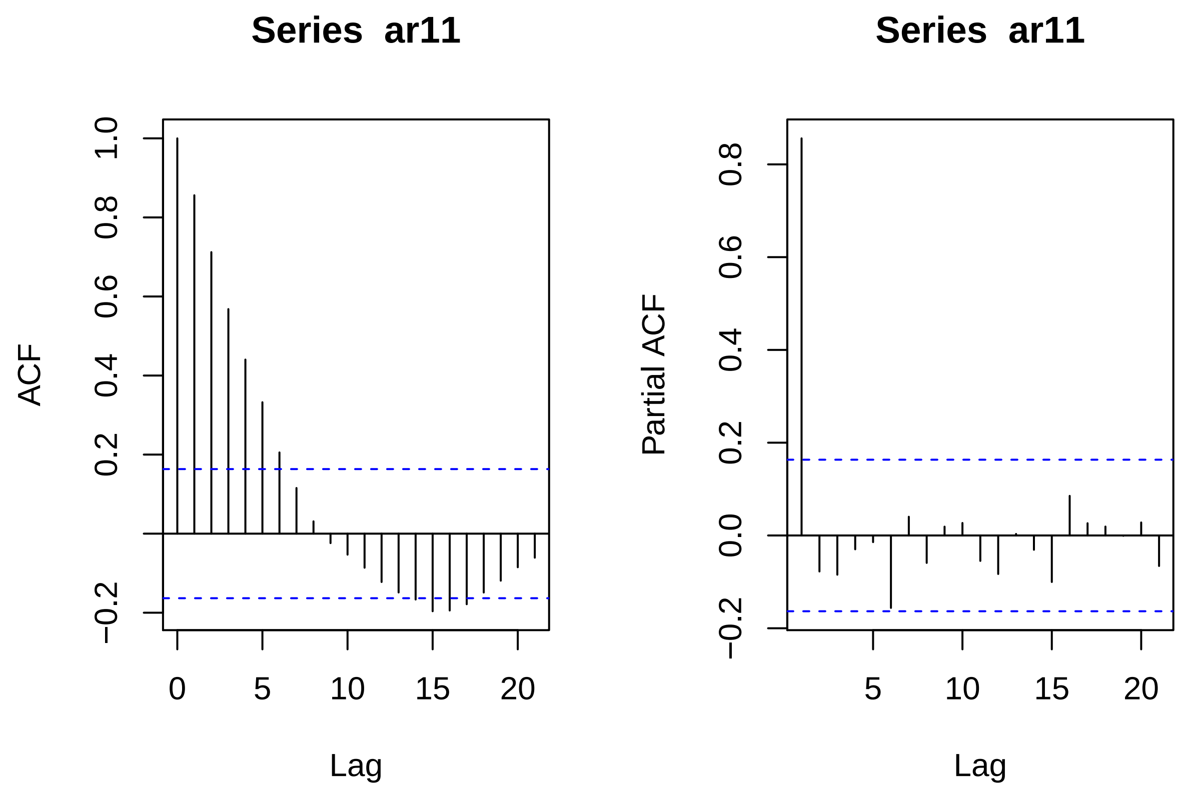
Again here, due to the coefficients being very low, the ACF function fails to show a geometric decay. However, the pacf function seems to indicate that only the first two lags of the AR(3) model are significant.

Part d

```
ar11 <- arima.sim(list(order=c(1,0,0), ar = c(0.9)), n=144)
plot.ts(ar11, axes=F); box(); axis(2)
```



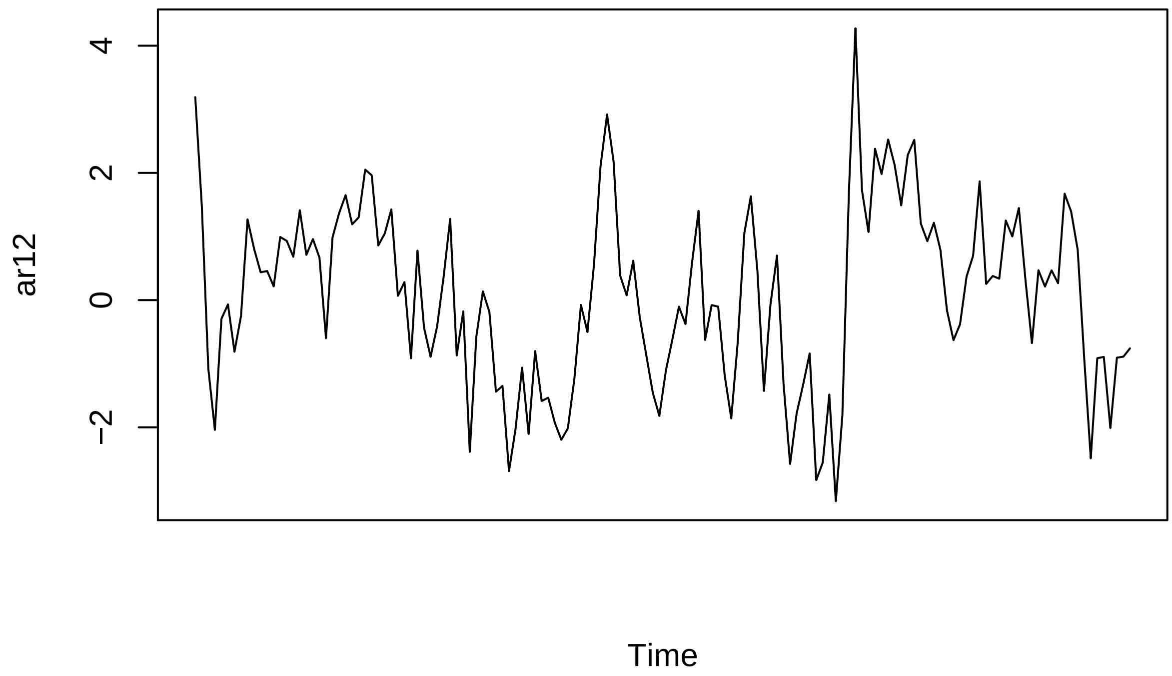

```r
par(mfrow=c(1,2))
ar11.acf<-acf(ar11, plot=TRUE)
ar11.pacf<-pacf(ar11, plot=TRUE)
```



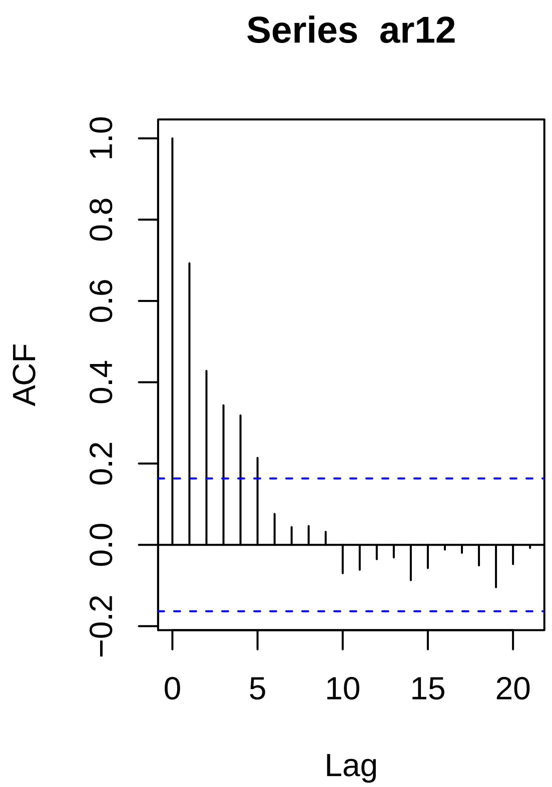
Here the acf function does show a geometric decay, as we would also expect. The pacf indicates significance for the first lag only.

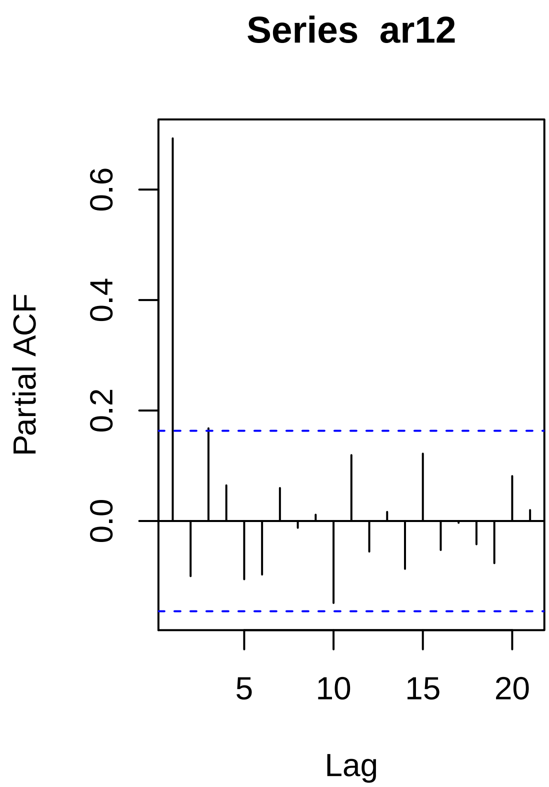
Part e

```
ar12 <- arima.sim(list(order=c(1,0,0), ar = c(0.7)), n=144)
plot.ts(ar12, axes=F); box(); axis(2)
```




```r
par(mfrow=c(1,2))
ar12.acf<-acf(ar12, plot=TRUE)
ar12.pacf<-pacf(ar12, plot=TRUE)
```

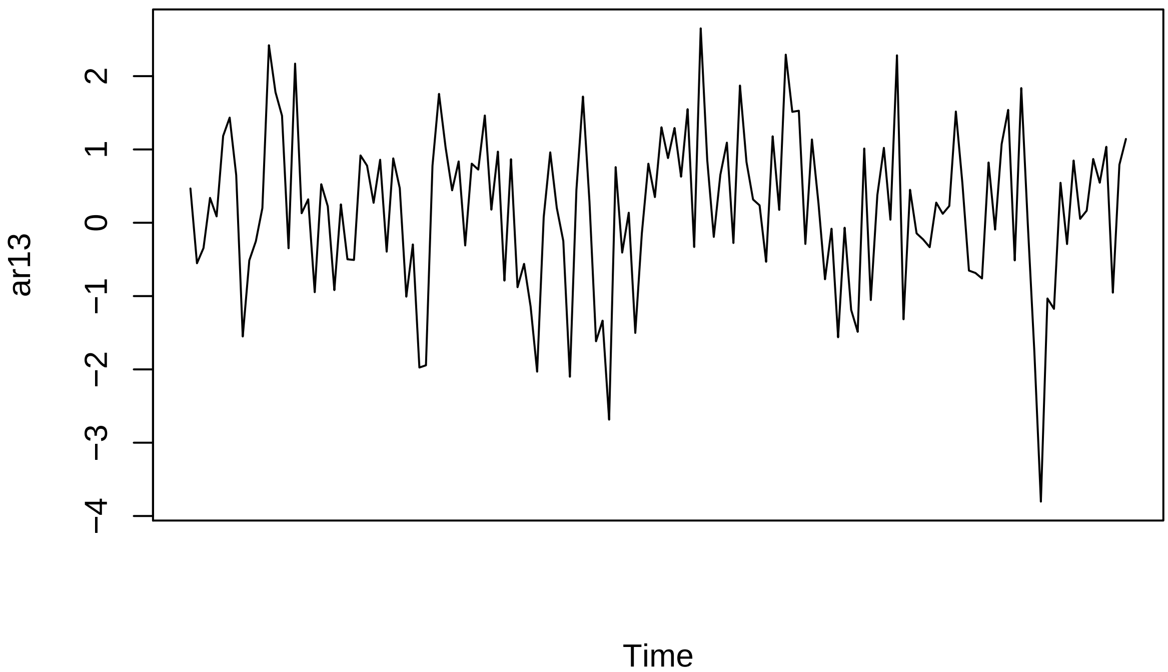




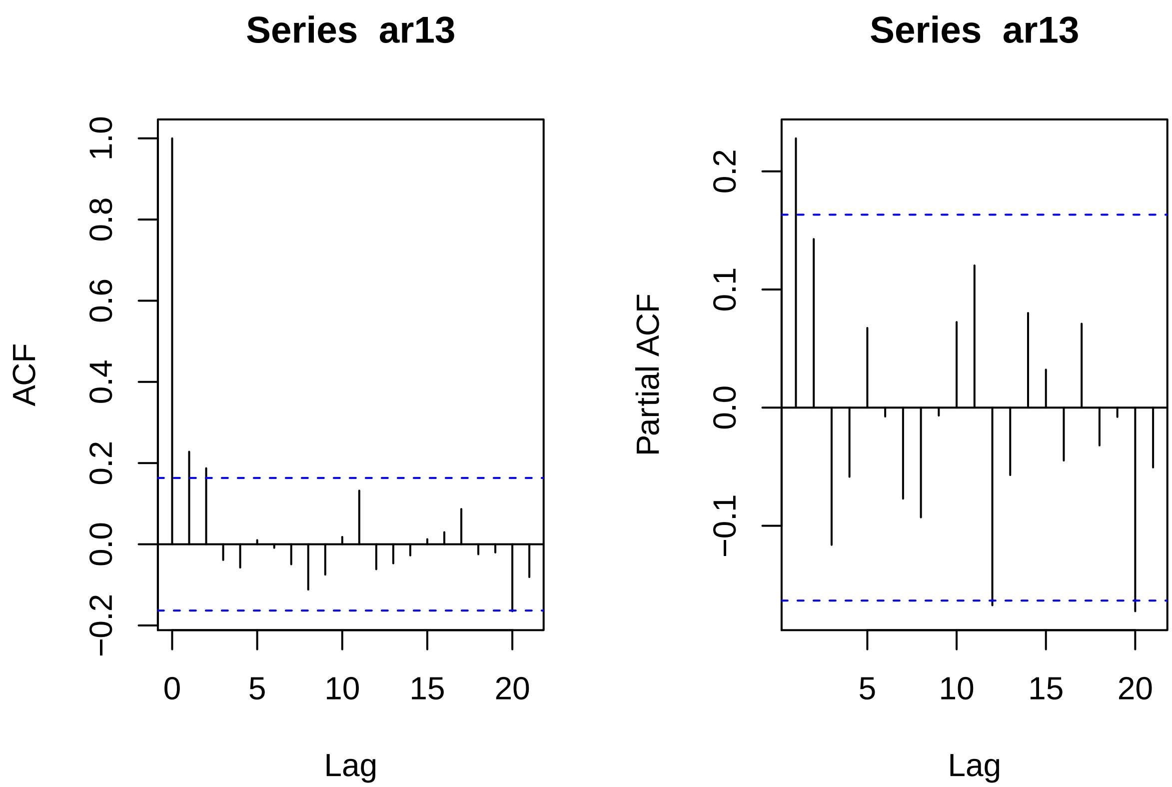
The acf function does show a geometric decay, as we would also expect. The pacf indicates significance for the first lag only. This is slightly less obvious than on the AR(1) model above due to the slightly lower coeeficient of 0.7 on the first lag.

Part f

```
ar13 <- arima.sim(list(order=c(1,0,0), ar = c(0.2)), n=144)
plot.ts(ar13, axes=F); box(); axis(2)
```



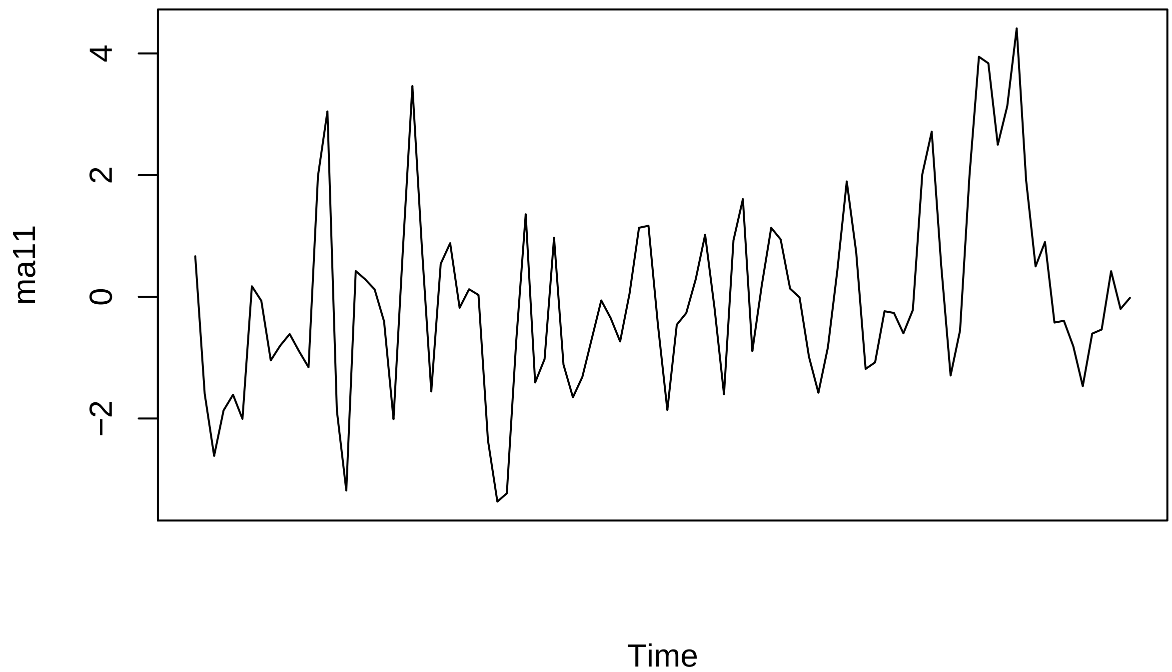

```r
par(mfrow=c(1,2))
ar13.acf<-acf(ar13, plot=TRUE)
ar13.pacf<-pacf(ar13, plot=TRUE)
```



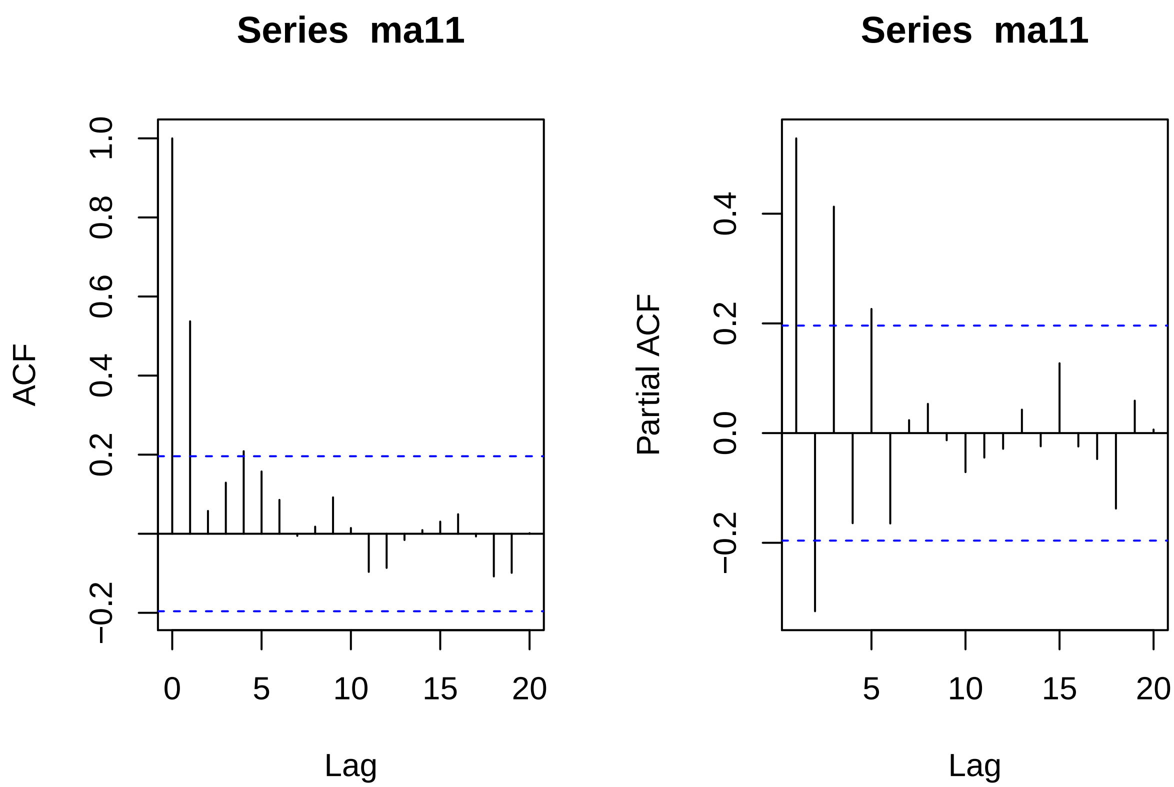
With a lag coefficient of 0.2, the results from the acf and pacf are less clear. A more formal time series test might be helpful here.

Part g

```
ma11 <- arima.sim(n = 100, list(ma=0.9), innov=rnorm(100))
plot.ts(ma11, axes=F); box(); axis(2)
```



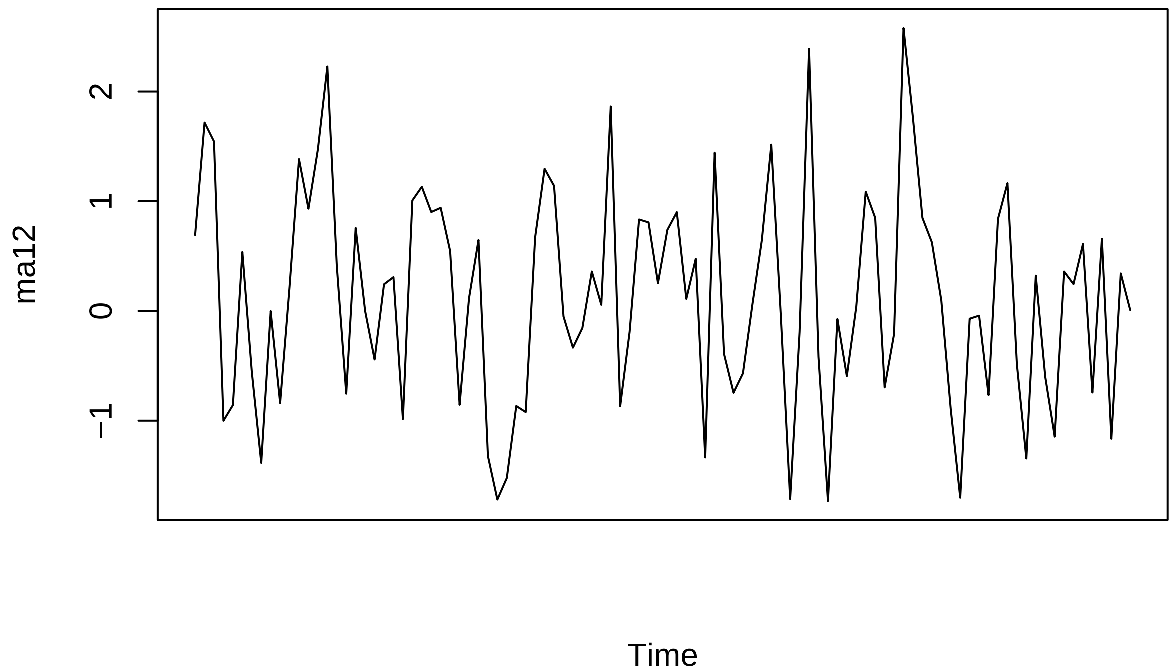

```r
par(mfrow=c(1,2))
ma11.acf<-acf(ma11, plot=TRUE)
ma11.pacf<-pacf(ma11, plot=TRUE)
```



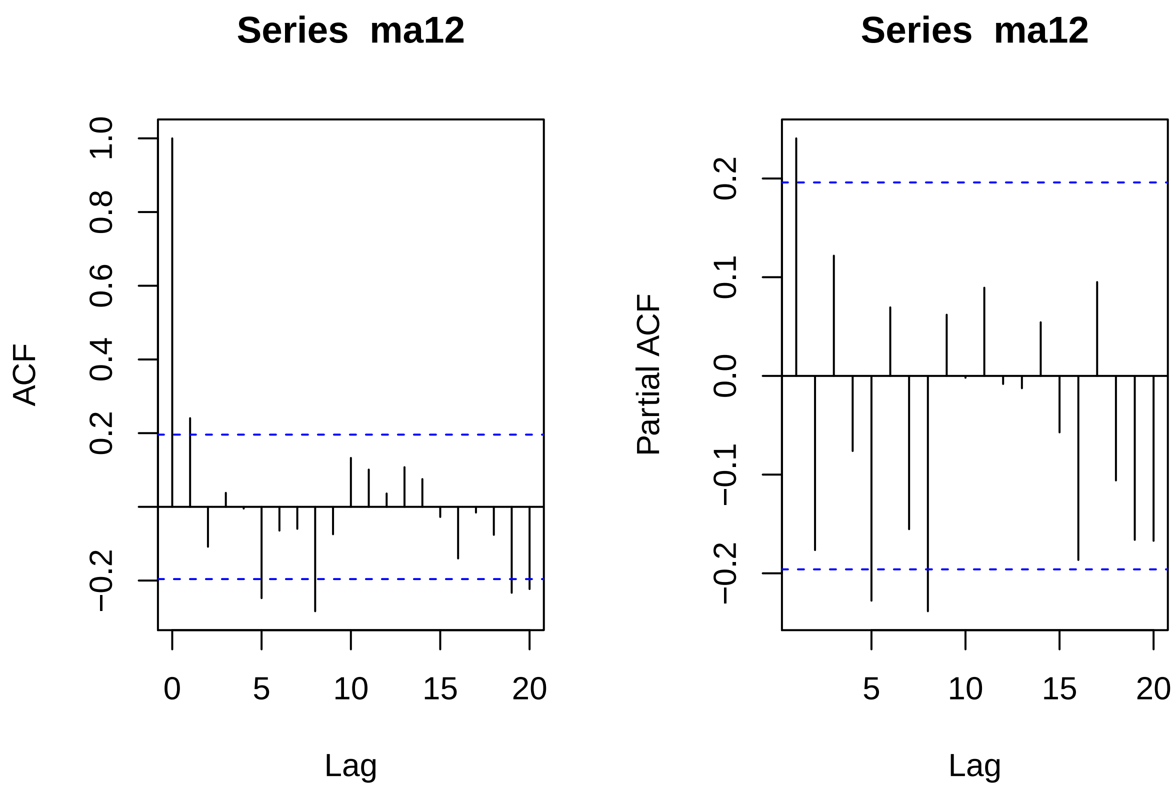
Here the acf function indicates significance in the first lag, and pacf indicates geometric decay.

Part h

```
ma12 <- arima.sim(n = 100, list(ma=0.2), innov=rnorm(100))
plot.ts(ma12, axes=F); box(); axis(2)
```




```r
par(mfrow=c(1,2))
ma12.acf<-acf(ma12, plot=TRUE)
ma12.pacf<-pacf(ma12, plot=TRUE)
```

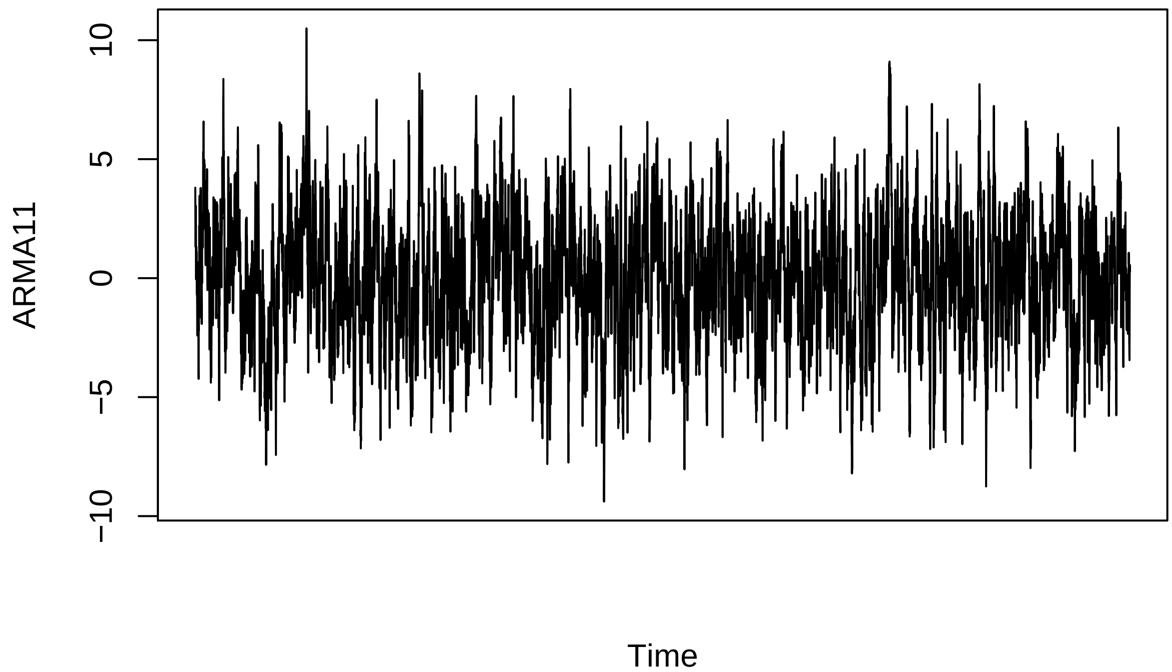


The results are not very clear. More formal test needed.

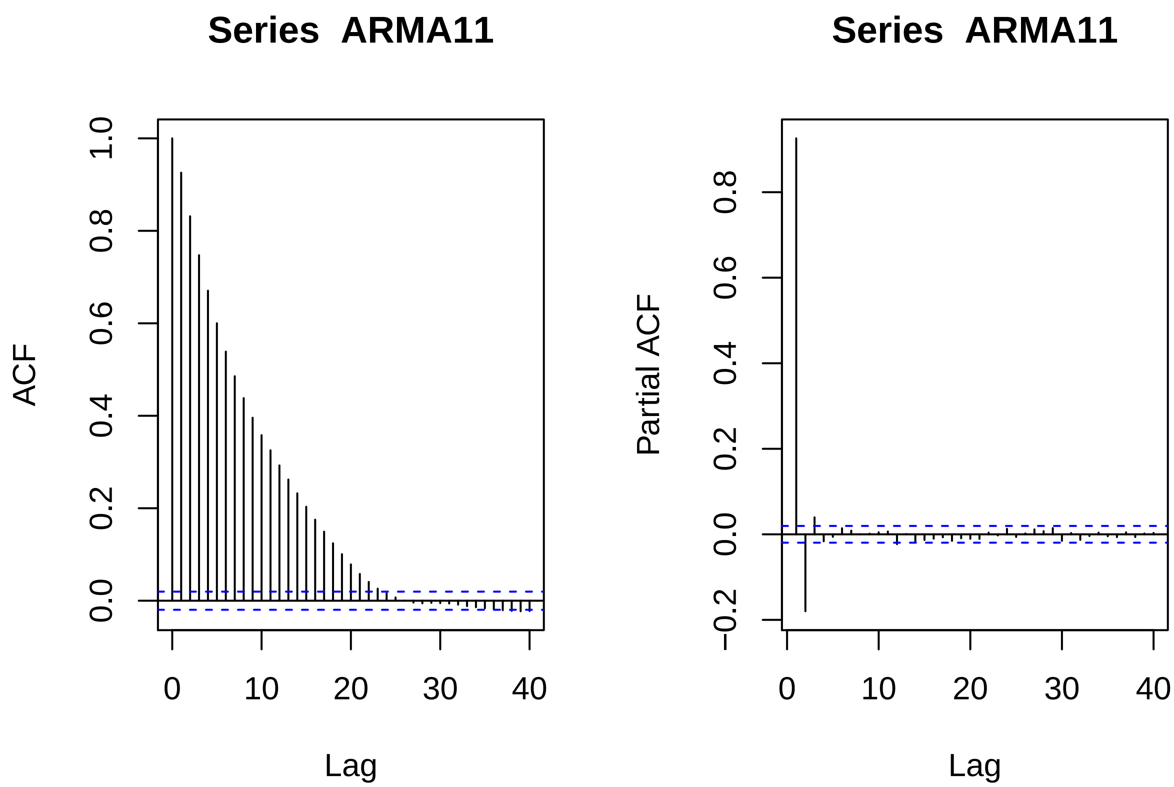
Part i

```
ARMA1=list(order=c(1,0,1), ar=0.9, ma=0.2)
ARMA11=arima.sim(n=10000, model=ARMA1)

plot.ts(ARMA11, axes=F); box(); axis(2)
```



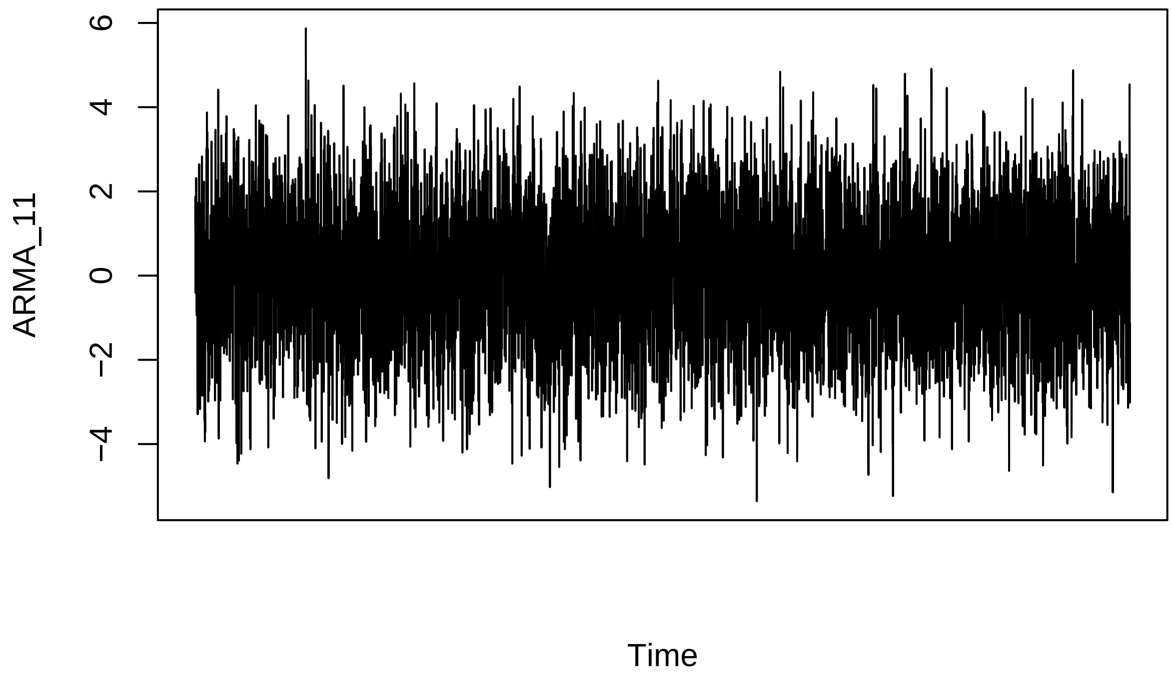

```
par(mfrow=c(1,2))
ARMA11.acf<-acf(ARMA11, plot=TRUE)
ARMA11.pacf<-pacf(ARMA11, plot=TRUE)
```



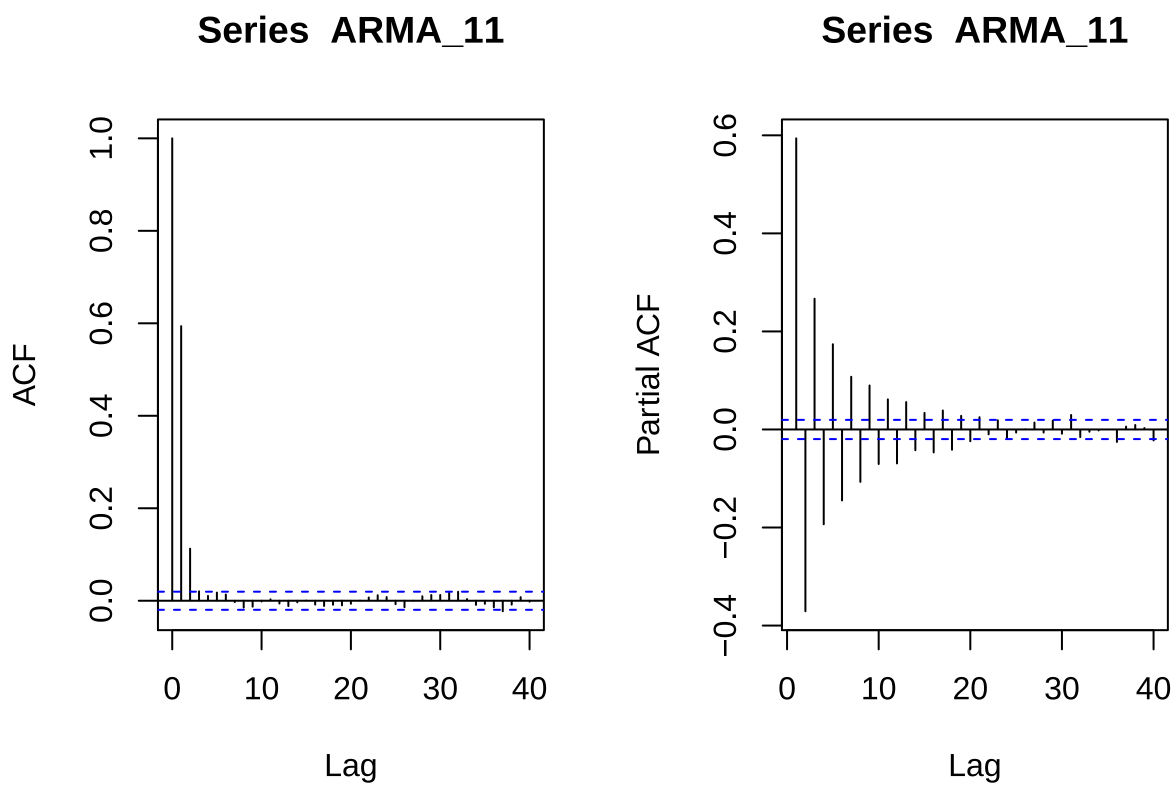
While the acf function does show geometric decay as we would also expect, the pacf function fails to do so. This might well be due to the low coefficient in the ma(1).

Part j

```
ARMA_1=list(order=c(1,0,1), ar=0.2, ma=0.9)
ARMA_11=arima.sim(n=10000, model=ARMA_1)
plot.ts(ARMA_11, axes=F); box(); axis(2)
```




```
par(mfrow=c(1,2))
ARMA_11.acf<-acf(ARMA_11, plot=TRUE)
ARMA_11.pacf<-pacf(ARMA_11, plot=TRUE)
```



Here the result is flipped from that in i. That is, the acf function fails to show geometric decay due to the low coefficient in ar(1), while the pacf function successfully shows the geometric decay we would expect to see.

QUESTION 1 II.

We generally know know the following:

1. An AR model is geometric in its ACF function, and significant till p lags on its PACF function.
2. The MA model, is significant till p lags on its ACF function, and geometric on its PACF function
3. The ARMA model is geometric in both its ACF and its PACF.

Hence, we would generally be able to eyeball the right model by just looking at the acfs and pacfs of our models, as long as the lag coefficients are large enough. With low lag coefficients the results are vague and eyeballing the model will be difficult.

PROBLEM 1B.

DATA SERIES INFORMATION:

use http://www.stata-press.com/data/r12/wpi1.dta

Time series data in the 'Example datasets' option of stata software> Time Series Reference Manual> arima

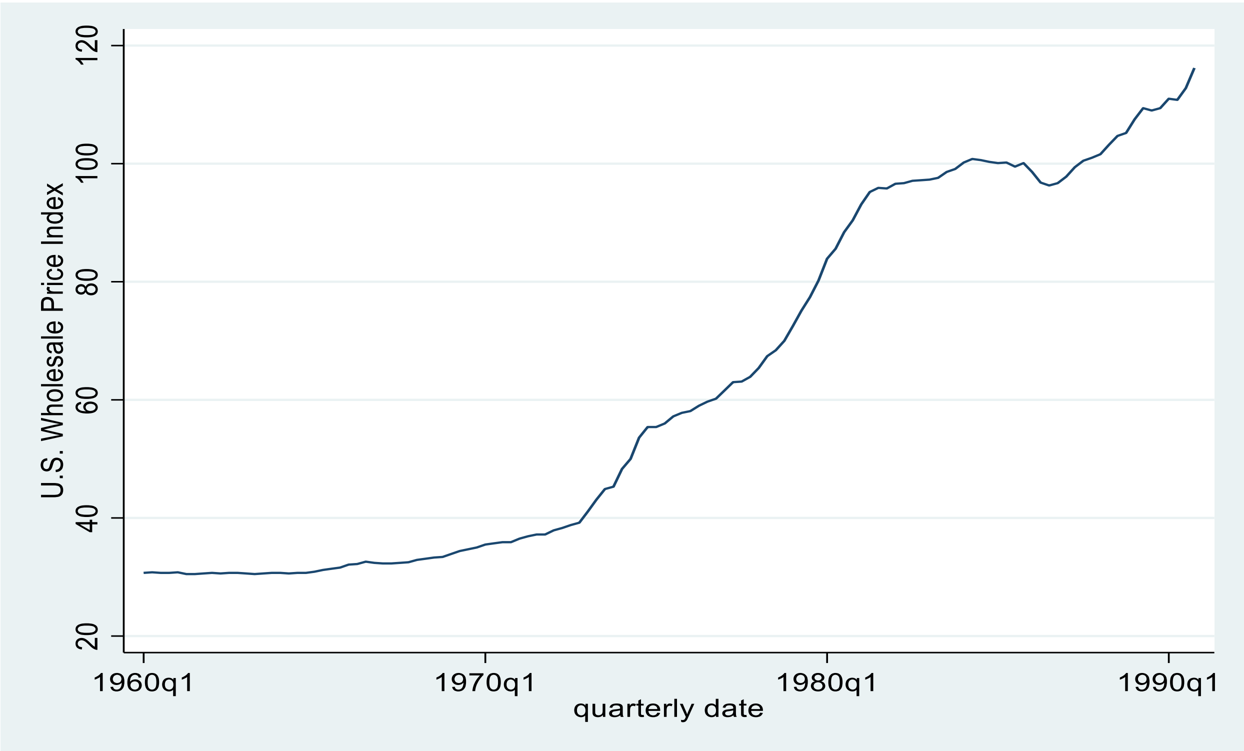
Data: wpi1.dta

Data Series Information: wpi is the U.S Wholesale Price Index; the data runs quarterly.

PLOTS AND STATA COMMANDS:

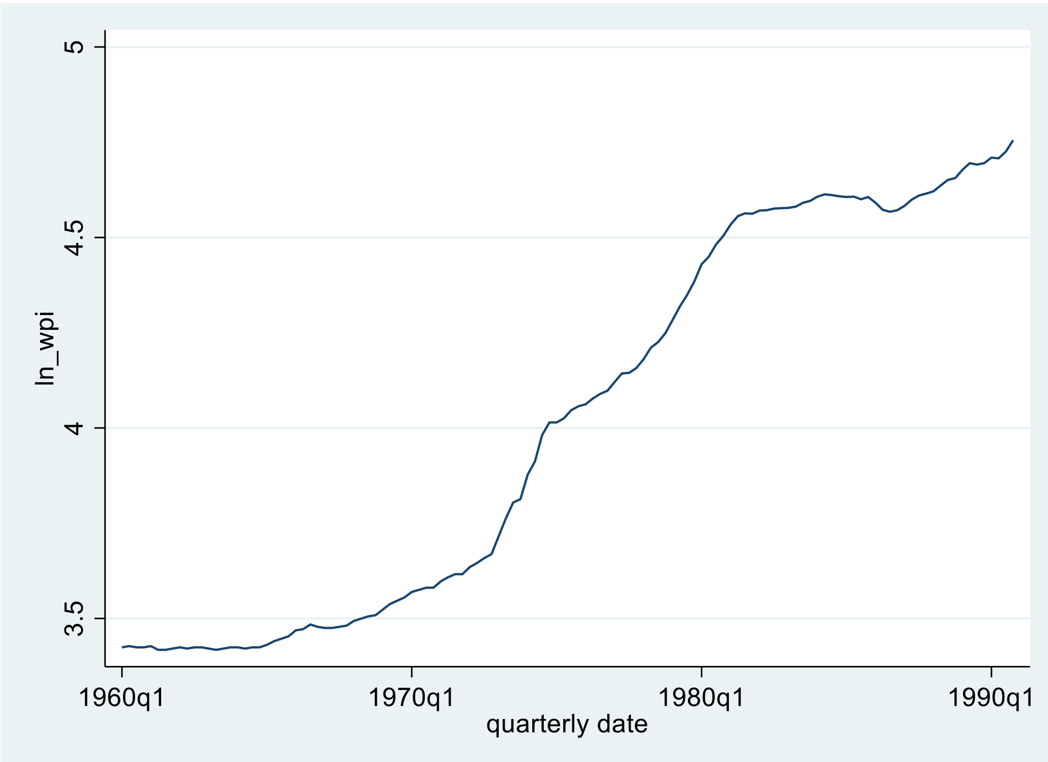
tsset t, quarterly

twoway (tsline wpi)



The line graph of the U.S Wholesale Price Index is exhibiting a time trend. To perform our Time series tests we will need stationary data. To take this trend away we need to difference our data. We can do so by taking the natural logarithm of our data.

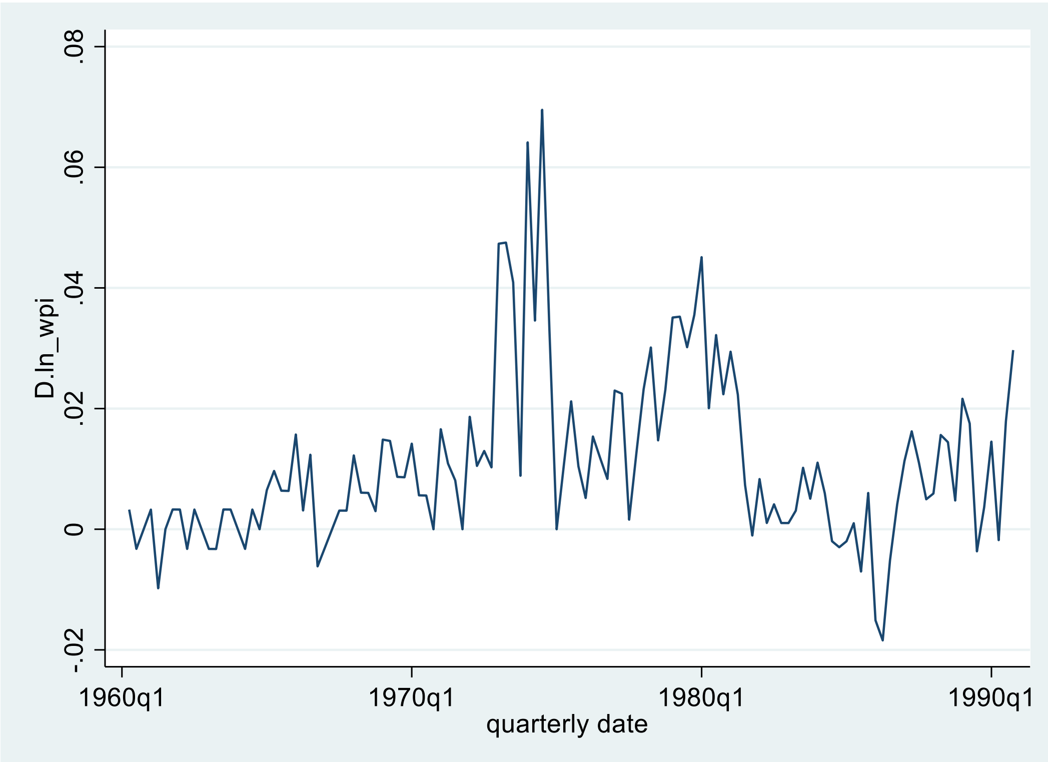
twoway (tsline ln_wpi)



We observe that even when we take the natural logarithm of our data, the time trend does not go away. Hence, the next step would be to try and take the first difference of the natural logarithm of our wpi variable.

NEW PLOTS AFTER REMOVING SEASONALITY:

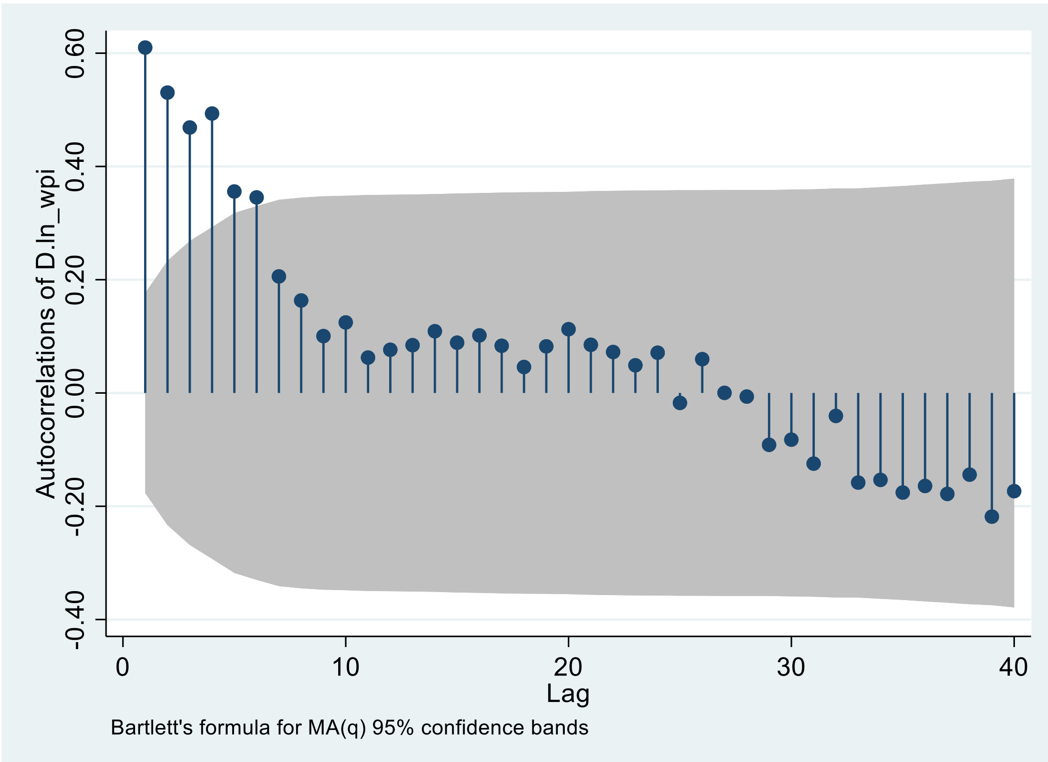
twoway (tsline D.ln_wpi)



We now see something closer to stationarity. Since the first order differencing seems to have solved our problem of non-stationarity, there is no need for us to use higher order differencing.

ACF COMMAND AND GRAPH:

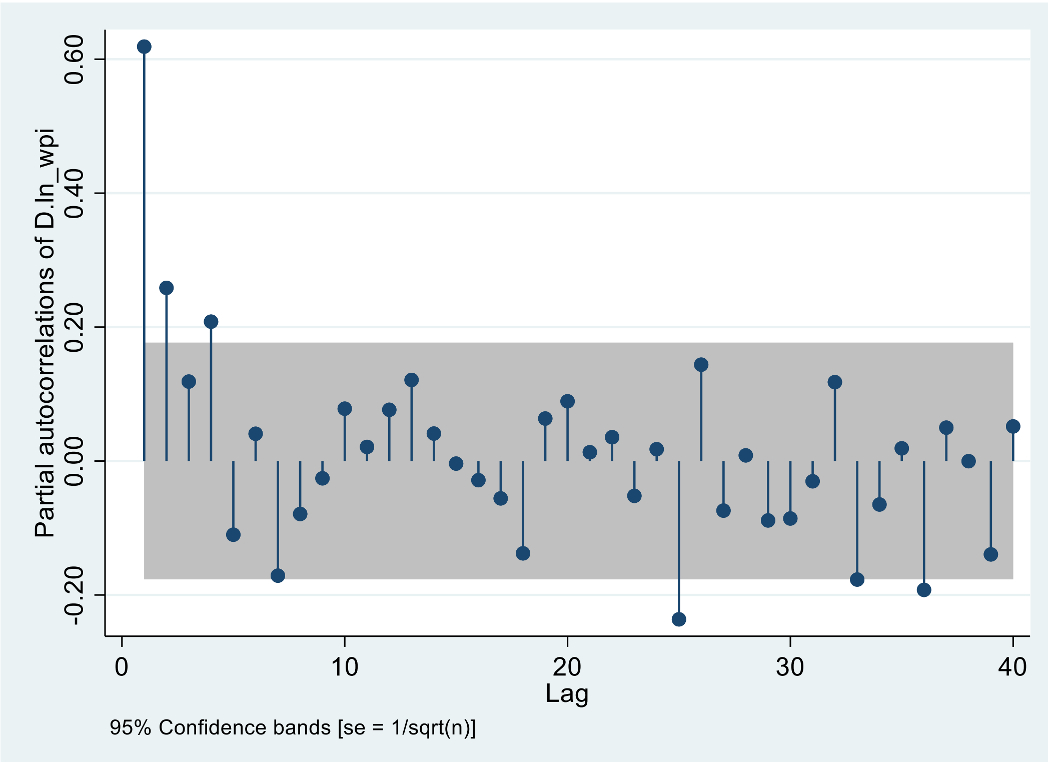
ac D.ln_wpi



Correlogram (ac) for our first order differencing. We see that we need a Moving-average order of 4. That is, there are 4 autocorrelations that are statistically significant from zero. The other two autocorrelations outside of the confidence band are not significantly outside of the band for us to include in our model.

PAC COMMAND AND GRAPH:

pac D.ln_wpi



The first order partial autocorrelation is well outside of the confidence band. This suggests that we have at least p=1, possibly p=2. Hence, we need to set our Austoregressive order either equal to 1 or equal to 2.

OUTPUT:

arima D.ln_wpi, arima(1,1,4)

The model that best fits this time series data is that of an ARIMA model with autoregressive order of 1, Integrated difference of order 1, and a moving-average order of 4.

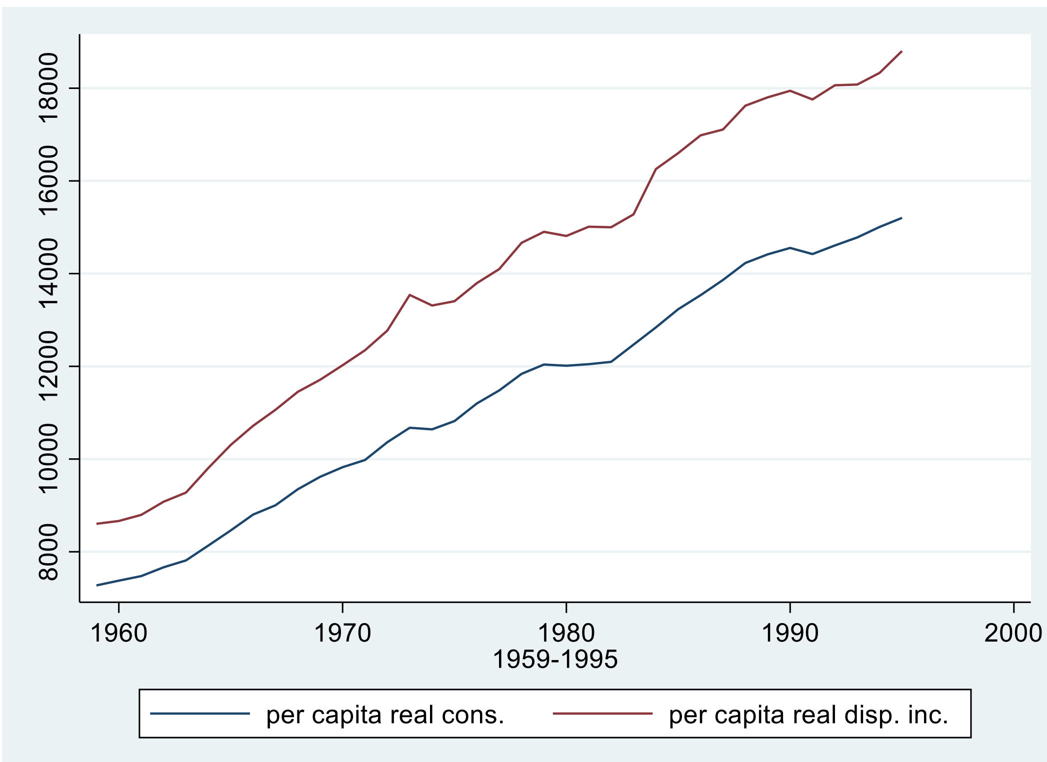
QUESTION 2.

a. **reg c y, r**

A one dollar increase in income results in a 0.7794 dollar increase in consumption. It would be difficult to argue that this result is reliable given that running a simple regression we are ignoring a crucial components of time series. Although the significance level of our income coefficient, the high R-square value as well as the F-test indicate that the relationship between the consumption and income variables is crucial, we won't be able to get reliable time series results without first determining whether our data has a trend or whether it is stationary.

b. **tsset year, yearly**

twoway (tsline c) (tsline y)



The two time series have an upward trend.

c. **gen t=_n**

reg c y t, r

Including a time trend our results become even more significant. The coefficient of income on consumption decreases slightly to 0.5081. Our R-squared coefficient increases even more, while our F-test suggests that our results become less significantly different from zero. The goodness of fit of the data increases. A time trend captures the path of the variable over time, providing forecasts of our variable. It also captures the effect of relevant variables in the regression equation that change over time and are not directly measurable.

d. **reg c t, r**

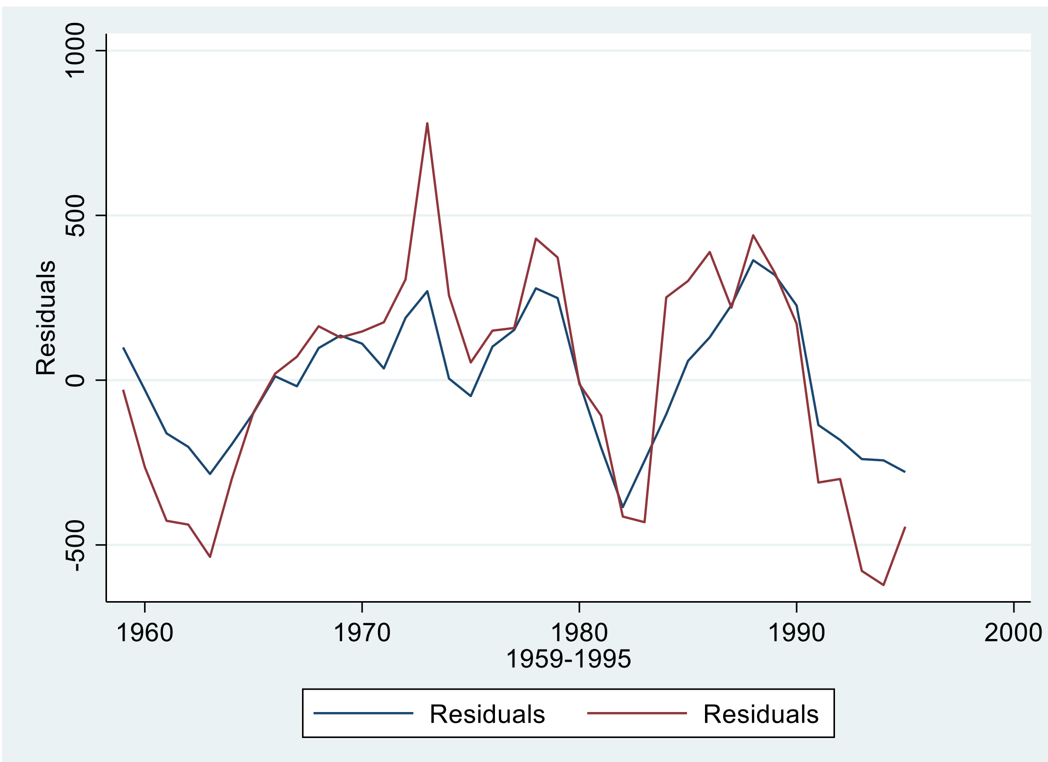
predict c_detrended, resid

reg y t,r

predict y_detrended, resid

reg c_detrended y_detrended,r

twoway (tsline c_detrended) (tsline y_detrended)



The y coefficient is the same as that in c. However, the goodness of fit slightly decreases when we de-trend the series since the R-squares decreases from 0.99 to 0.78. On the other hand, according to the F-test, our coefficient becomes more significant with respect to c.

e. **gen ln_c=log(c)**

gen growth_ratec=(ln_c[_n]-ln_c[_n-1])/ln_c[_n-1]

gen ln_y=log(y)

gen growth_ratey=(ln_y[_n]-ln_y[_n-1])/ln_y[_n-1]

reg growth_ratec growth_ratey,r

Running a regression in growth rates results in a slightly higher coefficient of income on consumption, that of 0.58. Considering the overtime effect is a crucial component of time series and it is one more representative of the real world. The goodness of fit, on the other hand, has slightly decreased with an R-squared value of 0.69.

f. ***CONTEMPORANOUS EFFECT:

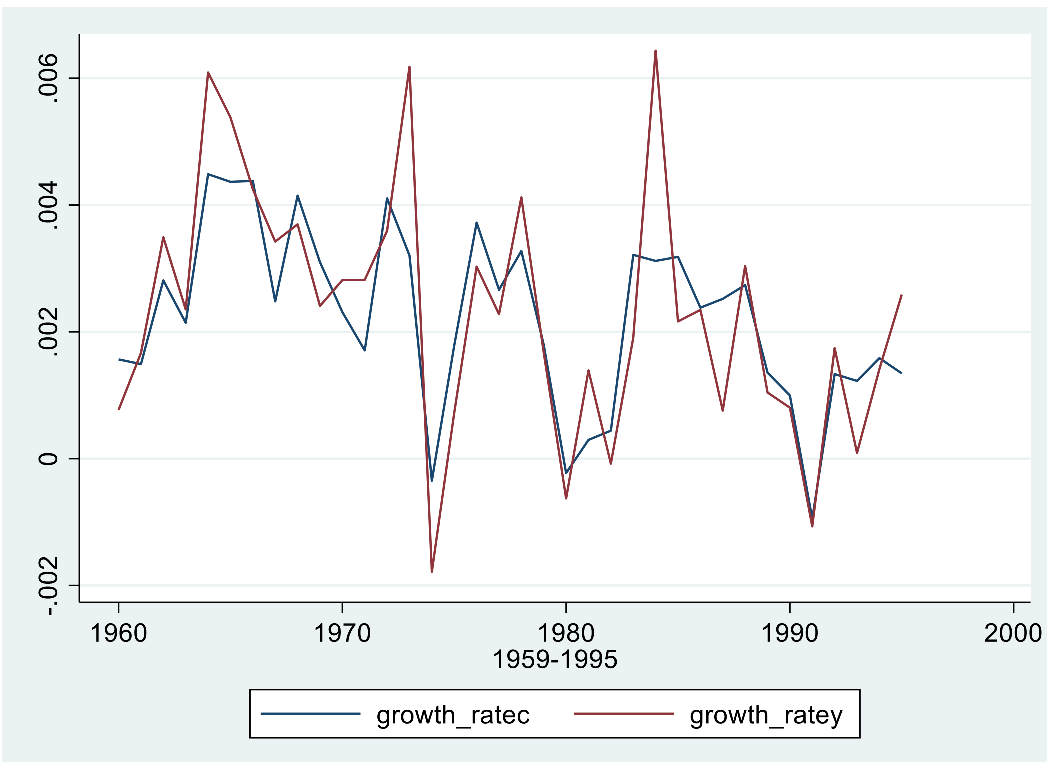
reg growth_ratec growth_ratey,r

***LONG-RUN EFFECT

reg growth_ratec L(0/4).D.growth_ratey,r

tsset year, yearly

twoway (tsline growth_ratec) (tsline growth_ratey)



FINITE SAMPLE ASSUMPTION OF OLS PROPERTIES:

The first Finite Sample Assumption of the OLS Properties is the linearity of the OLS parameters. We think of this assumption as being essentially the same as the first assumption on cross-sectional data, with the only difference being that we are not specifying a model with time series data.

The second Finite Sample Assumption of the OLS Properties is 'No Perfect Collinearity'. This assumption implies that in a time series process, no independent variable is constant nor a perfect linear combination of the others. This assumption allows the independent variables to be correlated, but it rules out perfect correlation in the sample.

The third Finite Sample assumption has to do with 'Zero Conditional Mean'. This assumption implies that the error at time t is uncorrelated with each explanatory variable in every time period. Since in a time series context random sampling is almost never appropriate, we must explicitly assume that the expected value of the error term at time t is not related to the explanatory variables in any time period. Hence, for OLS to be unbiased, it is not sufficient to assume contemporaneous exogeneity[1] in addition to the first two assumptions. We must also assume strict exogeneity[2].

If all three assumptions stated above are satisfied (including strict exogeneity), then the OLS are thought to be unbiased estimators.

The fourth Finite Sample assumption is 'Homoskedasticity'. This assumptions states that conditional on X, the variance of the error term at time t is the same for all t. This implies that the variance of the error term cannot depend on X. Hence ut and X must be independent, and the variance of the error term is constant over time.

The fifth Finite Sample assumption is 'No serial Correlation'. This assumption implies that conditional on X, the errors in two different time periods are uncorrelated. In cross sectional data we did not have to make such an assumption. This is due to the random sampling assumption. Hence the potential serial correlation problem is only relevant for time series data.

The sixth Finite Sample assumption is 'Normality'. This assumption states that the errors ut are independent of X and are independent and identically distributed as normal (0, var).

Strict exogeneity, homoskedasticity and no serial correlation assumptions are very strict and are often violated by time series data. However, these assumptions are much less of a problem if our sample size is large enough.

---

[1]Contemporaneous Exogeneity is the condition that the expected value of the error term at time t, given the explanatory variable at time t, is equal to zero.

[2] Strict Exogeneity is the condition that the expected value of the error term at time t, given the explanatory variable at all past and future values, is equal to zero.

ASYMPTOTIC ASSUMPTION OF OLS PROPERTIES

A crutial requirement for asymptotic analysis of time series is that the time series we are working with are stationary and weakly dependent[3].

First Asymptotic assumption of the OLS Properties is 'Linearity and Weak Dependence'. Hence, the assumption here is the same as than in the first Finite Sample Assumption, but it adds the condition of Stationarity and Weak dependence.

The second Asymptotic assumption of the OLS Properties is 'No Perfect Collinearity. This assumption is the same as the second Finite Sample Assumption of OLS.

The third Asymptotic assumption of the OLS Properties is 'Zero Conditional Mean'. This assumption is different from that of the Finite Sample OLS assumption in that it only required Contemporaneous Exogeneity. It is hence a less restrictive assumption since it does not require Strict Exogeneity.

Under the first three assumptions OLS estimators are **consistent** as the probability limit of our sample beta equals our population beta.

The fourth Asymptotic assumption of the OLS Properties is 'Homoskedasticity'. This assumption requires that the errors are contemporaneously heteroskedastic, such that the variance of the error at time t is independent with the explanatory variable at time t, but not with past and future values of t.

The fifth Asymptotic assumption of the OLS Properties is 'No Serial Correlation'. This assumption holds for AR(1) models. As long as only one lag belongs to our regression equation, the errors will be serially uncorrelated.

Under the 5 assumptions stated above, the OLS estimators are asymptotically normally distributed.

To conclude, the reason why we might need to rely on asymptotic assumptions in Time series data compared to Cross Sectional Data is the fact that Time Series deals with all kinds of dynamic effect which are restricted under the assumption of Strict Exogeneity, strict Homoskedasticity and "No serial Correlation". We need something that allows for the feedback effect of the dependent variable on the future values of the explanatory variables.

---

[3] Weakly Dependent Time Series means that the correlation between xt and xt+h goes to zero" sufficiently quickly" as h goes to infinity. This essentially relaces the assumption of random sampling in implying that Law of Large Numbers and Central Limit Theorem hold.

ACF OF AR(2) :

AR (2) : $Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + U_t$

We assume the innovation is serially uncorrelated, with mean zero and constant variance σ^2 (white noise)

ACF : $\rho(s) = \dfrac{\gamma(s)}{\gamma(0)}$

$\gamma(s)$ = autocovariance
$\gamma(0)$ = variance

$$\gamma(s) = E\left[(Y_t - E[Y_t])(Y_{t-s} - E[Y_{t-s}])\right]$$

by stationarity, $E[Y_t] = 0$, so

$$\gamma(s) = E[Y_t Y_{t-s}]$$

Multiplying both sides of our AR(2) process by Y_{t-s}, we get

$$\gamma(s) = E[Y_t Y_{t-s}] = \beta_1 E[Y_{t-1} Y_{t-s}] + \beta_2 E[Y_{t-2} Y_{t-s}]$$

For s=0

$$\gamma(0) = E[(Y_t)^2] = \beta_1 E[Y_t Y_{t-1}] + \beta_2 E[Y_t Y_{t-2}] + E[U_t Y_t]$$

$$\gamma(0) = \beta_1 \gamma(1) + \beta_2 \gamma(2)$$

For s=1

$$\gamma(1) = E[Y_t Y_{t-1}] = \beta_1 E[Y_{t-1}{}^2] + \beta_2 E[Y_{t-1} Y_{t-2}] + E[U_t Y_{t-1}]$$

$$\gamma(1) = \beta_1 \gamma(0) + \beta_2 \gamma(1)$$

For s=2

$$\gamma(2) = E[Y_t Y_{t-2}] = \beta_1 E[Y_{t-1} Y_{t-2}] + \beta_2 E[Y_{t-2} Y_{t-2}] + E[U_t Y_{t-2}]$$

$$\gamma(2) = \beta_1 \gamma(1) + \beta_2 \gamma(0)$$

ACF : $\rho(s) = \dfrac{\gamma(s)}{\gamma(0)}$

$\rho(0) = \dfrac{\gamma(0)}{\gamma(0)} = 1$; $\rho(1) = \dfrac{\gamma(1)}{\gamma(0)} = \dfrac{\beta_1 \gamma(0) + \beta_2 (\gamma(1))}{\beta_1 \gamma(1) + \beta_2 \gamma(2)} = \dfrac{\gamma(0) + \gamma(1)}{\gamma(1) + \gamma(2)}$

$\rho(2) = \dfrac{\gamma(2)}{\gamma(0)} = \dfrac{\gamma(1) + \gamma(0)}{\gamma(1) + \gamma(2)}$.

ACF OF AR(3) :

$AR(3)$: $Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + U_t$

$$\gamma(s) = E[(Y_t - E[Y_t])(Y_{t-s} - E[Y_{t-s}])]$$

by stationarity, $E[Y_t] = 0$

$$\gamma(s) = E[Y_t Y_{t-s}]$$

Multiplying both sides of our AR(3) process, we get

$$\gamma(s) = E[Y_t Y_{t-s}] = \beta_1 E[Y_{t-1} Y_{t-s}] + \beta_2 E[Y_{t-2} Y_{t-s}] + \beta_3 E[Y_{t-3} Y_{t-s}]$$

For $s=0$
$$\gamma(0) = \beta_1 E[Y_t Y_{t-1}] + \beta_2 E[Y_t Y_{t-2}] + \beta_3 E[Y_t Y_{t-3}] + E[U_t Y_t]$$
$$\gamma(0) = \beta_1 \gamma(1) + \beta_2 \gamma(2) + \beta_3 \gamma(3)$$

For $s=1$
$$\gamma(1) = E[Y_t Y_{t-1}] = \beta_1 E[Y_{t-1}^2] + \beta_2 E[Y_{t-1} Y_{t-2}] + \beta_3 E[Y_{t-1} Y_{t-3}] + E[U_t Y_{t-1}]$$
$$\gamma(1) = \beta_1 \gamma(0) + \beta_2 \gamma(1) + \beta_3 \gamma(2)$$

For $s=2$
$$\gamma(2) = E[Y_t Y_{t-1}] = \beta_1 E[Y_{t-1} Y_{t-2}] + \beta_2 [Y_{t-2}^2] + \beta_3 [Y_{t-2} Y_{t-3}] + E[U_t Y_{t-2}]$$
$$\gamma(2) = \beta_1 \gamma(1) + \beta_2 \gamma(0) + \beta_3 \gamma(1)$$

For $s=3$
$$\gamma(3) = \beta_1 E[Y_{t-1} Y_{t-3}] + \beta_2 E[Y_{t-2} Y_{t-3}] + \beta_3 [Y_{t-3}^2] + E[U_t Y_{t-3}]$$
$$\gamma(3) = \beta_1 \gamma(2) + \beta_2 \gamma(1) + \beta_3 \gamma(0)$$

ACF : $\rho(s) = \dfrac{\gamma(s)}{\gamma(0)}$

$\rho(0) = 1$; $\rho(1) = \dfrac{\gamma(1)}{\gamma(0)} = \dfrac{\gamma(0) + \gamma(1) + \gamma(2)}{\gamma(1) + \gamma(2) + \gamma(3)}$

$$\rho(2) = \frac{\gamma(2)}{\gamma(0)} = \frac{\gamma(1) + \gamma(0) + \gamma(1)}{\gamma(1) + \gamma(2) + \gamma(2)}$$

$$\rho(3) = \frac{\gamma(3)}{\gamma(0)} = \frac{\gamma(2) + \gamma(1) + \gamma(0)}{\gamma(1) + \gamma(2) + \gamma(3)}$$

ACF OF MA(2)

$$MA(2) = Y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2}$$

ACF : $\rho(s) = \dfrac{\gamma(s)}{\gamma(0)}$ where $\gamma(s)$ = autocovariance of Y_t.
$\gamma(0)$ = variance of Y_t

$$\gamma(s) = E\left[(Y_t - E[Y_t])(Y_{t-s} - E[Y_{t-s}])\right]$$

$$= E\left[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2})(\varepsilon_t + \alpha_1 \varepsilon_{t-s-1} + \alpha_2 \varepsilon_{t-s-2})\right]$$

For $s=0$
$$\gamma(0) = E\left[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2})^2\right]$$
$$= E\Big[\varepsilon_t^2 + \alpha_1^2 \varepsilon_{t-1}^2 + \alpha_2^2 \varepsilon_{t-2}^2 + 2\alpha_1 \underbrace{\varepsilon_t \varepsilon_{t-1}}_{=0} + 2\alpha_1\alpha_2 \underbrace{\varepsilon_{t-1}\varepsilon_{t-2}}_{=0} + 2\alpha_2 \underbrace{\varepsilon_t \varepsilon_{t-2}}_{=0}\Big]$$
$$= (1+\alpha_1^2+\alpha_2^2)\, E[\varepsilon_t^2] = (1+\alpha_1^2+\alpha_2^2)\sigma^2.$$

Cancellations are possible because all covariances of the ε_t's with their own lagged realizations are zero by assumption.

For $s=1$
$$\gamma(1) = E\left[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2})(\varepsilon_{t-1} + \alpha_1 \varepsilon_{t-2} + \alpha_2 \varepsilon_{t-3})\right]$$
$$= \alpha_1 E[\varepsilon_{t-1}^2] + \alpha_1\alpha_2 E[\varepsilon_{t-2}^2] = \alpha_1(1-\alpha_2)\sigma^2$$

For $s=2$
$$\gamma(2) = E\left[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2})(\varepsilon_{t-2} + \alpha_1 \varepsilon_{t-3} + \alpha_2 \varepsilon_{t-4})\right]$$
$$= \alpha_2 E[\varepsilon_{t-2}^2] = \underline{\underline{\alpha_2 \sigma^2.}}$$

From then on, the autocovariance function remains inactive :

$$\gamma(s) = 0 \quad \text{for} \quad s \geq 3$$

ACF: $\rho(0) = \dfrac{\gamma(0)}{\gamma(0)} = 1$; $\rho(1) = \dfrac{\gamma(1)}{\gamma(0)} = \dfrac{\alpha_1(1+\alpha_2)}{1+\alpha_1^2+\alpha_2^2}$

$$\rho(2) = \frac{\gamma(2)}{\gamma(0)} = \frac{\alpha_2}{1+\alpha_1^2+\alpha_2^2}$$

ACF OF MA(3) :

MA(3) : $Y_t = \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3}$

ACF : $\rho(s) = \dfrac{\gamma(s)}{\gamma(0)}$

$\gamma(s)$ = autocovariance of Y_t
$\gamma(0)$ = variance of Y_t

$$\gamma(s) = E[(Y_t - E[Y_t])(Y_{t-s} - E[Y_{t-s}])]$$
$$= E[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3})(\varepsilon_t + \alpha_1 \varepsilon_{t-s-1} + \alpha_2 \varepsilon_{t-s-2} + \alpha_3 \varepsilon_{t-s-3})]$$

For s=0

$$\gamma(0) = E[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3})^2]$$
$$E[\varepsilon_t^2](1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2) = \sigma^2(1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2)$$

For s=1

$$\gamma(1) = E[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3})(\varepsilon_{t-1} + \alpha_1 \varepsilon_{t-2} + \alpha_2 \varepsilon_{t-3} + \alpha_3 \varepsilon_{t-4})]$$
$$= E[\alpha_1 \varepsilon_{t-1}^2 + \alpha_1 \alpha_2 \varepsilon_{t-2}^2 + \alpha_2 \alpha_3 \varepsilon_{t-3}^2]$$
$$= \alpha_1 E[\varepsilon_{t-1}^2] + \alpha_1 \alpha_2 E[\varepsilon_{t-2}^2] + \alpha_2 \alpha_3 E[\varepsilon_{t-3}^2]$$
$$= \sigma^2(\alpha_1 + \alpha_1 \alpha_2 + \alpha_2 \alpha_3).$$

For s=2

$$\gamma(2) = E[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3})(\varepsilon_{t-2} + \alpha_1 \varepsilon_{t-3} + \alpha_2 \varepsilon_{t-4} + \alpha_3 \varepsilon_{t-5})]$$
$$= E[\alpha_2 \varepsilon_{t-2}^2 + \alpha_1 \alpha_3 \varepsilon_{t-3}^2]$$
$$= \alpha_2 E[\varepsilon_{t-2}^2] + \alpha_1 \alpha_3 E[\varepsilon_{t-3}^2]$$
$$= \sigma^2(\alpha_2 + \alpha_1 \alpha_3).$$

For s=3

$$\gamma(3) = E[(\varepsilon_t + \alpha_1 \varepsilon_{t-1} + \alpha_2 \varepsilon_{t-2} + \alpha_3 \varepsilon_{t-3})(\varepsilon_{t-3} + \alpha_1 \varepsilon_{t-4} + \alpha_2 \varepsilon_{t-5} + \alpha_3 \varepsilon_{t-6})]$$
$$= E[\alpha_3 \varepsilon_{t-3}^2] = \underline{\underline{\alpha_3 \sigma^2}}.$$

From then on, the autocovariance function remains inactive.

ACF: $\rho(0) = \dfrac{\gamma(0)}{\gamma(0)} = 1$; $\rho(1) = \dfrac{\gamma(1)}{\gamma(0)} = \dfrac{\alpha_1 + \alpha_1 \alpha_2 + \alpha_2 \alpha_3}{1 + \alpha_1^2 + \alpha_2^2 + \alpha_3^2}$

PACF of AR(2) :

$$\frac{\text{Covariance}\,(x_t\,,\,x_{t-2}\,|\,x_{t-1})}{\sqrt{\text{Variance}\,(x_t\,|\,x_{t-1})\,\text{Variance}\,(x_{t-2}\,|\,x_{t-1})}}$$

The formula above is the correlation between values two time periods apart conditional on knowledge of the value in between.
Hence, the partial autocorrelation between x_t and x_{t-2} is the conditional correlation between x_t and x_{t-2}, given x_{t-1}.

Similarly,
PACF of AR(3):

$$\frac{\text{Covariance}\,(x_t\,,\,x_{t-3}\,|\,x_{t-1}\,,\,x_{t-2})}{\sqrt{\text{Variance}\,(x_t\,|\,x_{t-1}\,,\,x_{t-2})\,\text{Variance}\,(x_{t-3}\,|\,x_{t-1}\,,\,x_{t-2})}}$$

PACF OF MA(2) :

In general : $\text{PACF}\,(\phi_{hh}) = \text{Corr}\,(X_h - X_h^{h-1}\,,\,X_0 - X_0^{h-1})$ For $h = 2, 3, \ldots$

PACF of an invertible MA(q) :

$$X_t = \sum_{i=1}^{q} \theta_i W_{t-i} + W_t\,, \qquad X_t = -\sum_{i=1}^{\infty} \pi\, X_{t-i} + W_t$$

where $X_{n+1}^{n} = P\,(X_{n+1}\,|\,X_1, \ldots, X_n)$

$$= P\left(-\sum_{i=1}^{\infty} \pi_i X_{n+1-i}\,|\,X_1, \ldots, X_n\right)$$

$$= -\sum_{i=1}^{n} \pi_i X_{n+1-i} - \sum_{i=n+1}^{\infty} \pi_i\, P\,(X_{n+1-i}\,|\,X_1, \ldots, X_n)$$

In general, $\phi_{nn} \neq 0$

The pattern that I see with ACF & PACF functions of AR(2), AR(3), MA(2) and MA(3) processes is that the ACF function of an AR process results in a geometric decay as the number of lags increases, while the PACF function of an AR process results in lags which are significant till p lags. Namely for AR(2), there will be two significant lags, while for an AR(3) there will be three significant lags.

The results of MA processes are flipped when it comes to the ACF & PACF functions. That is, for the MA(2) and MA(3) processes the ACF will tend to show significant lags till lag 2 and 3, respectively. On the other hand, for these two processes the pacf will convey a geometric decay.