# Time Series Analysis and Forecasting

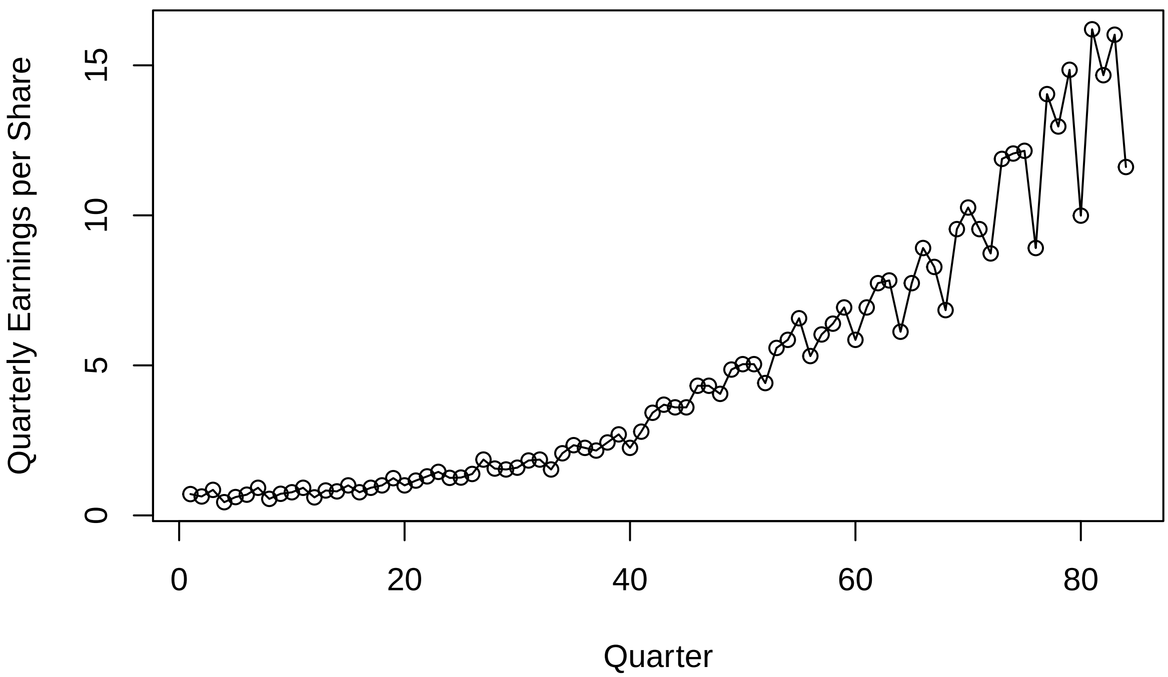
*Warlon Zeng*

*10/12/16*

## Question 1

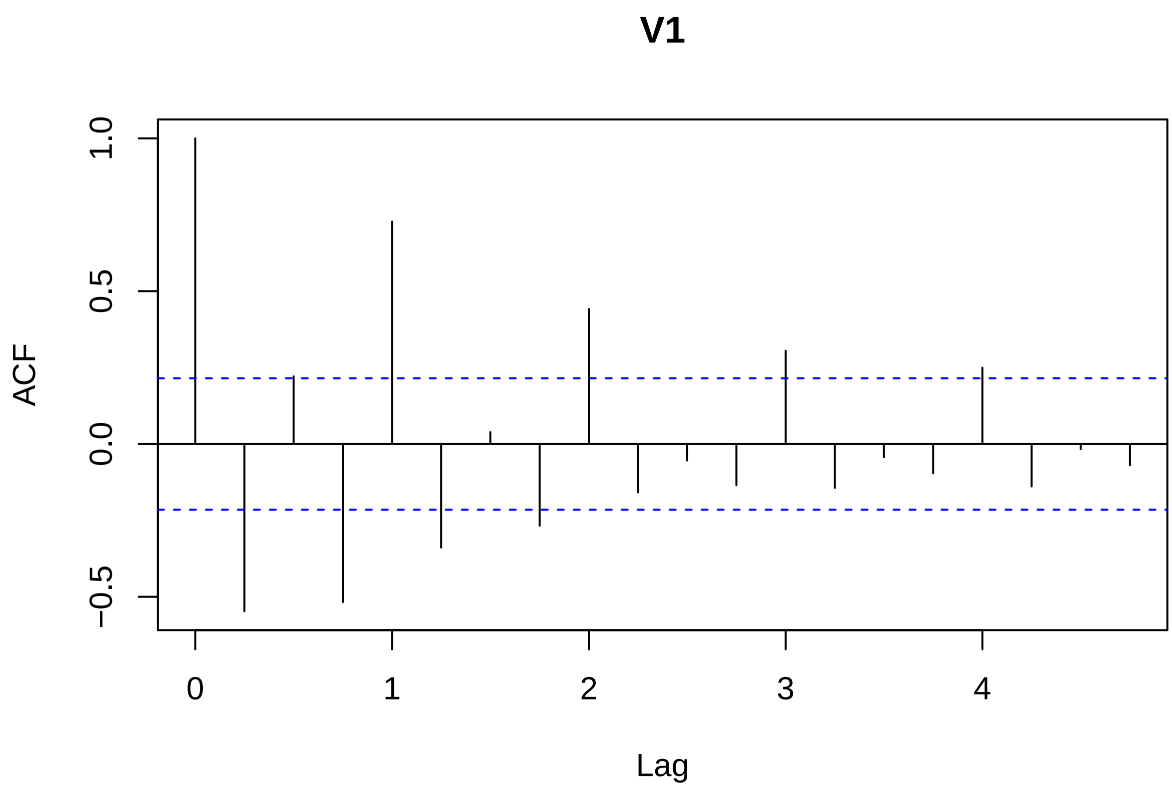
### Part A

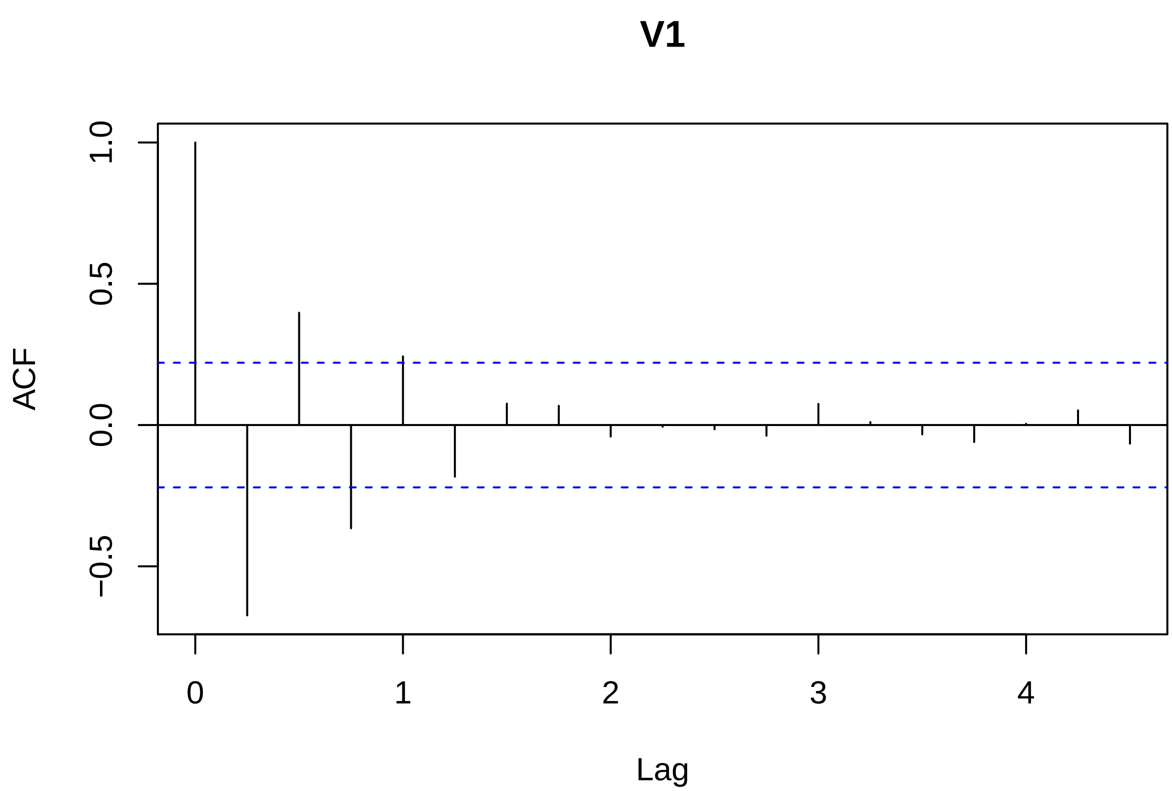
The plot below clearly shows a upward curve trend. Data starts off slow and rises in the last bits of the year. Length is 84 and frequency is 4.

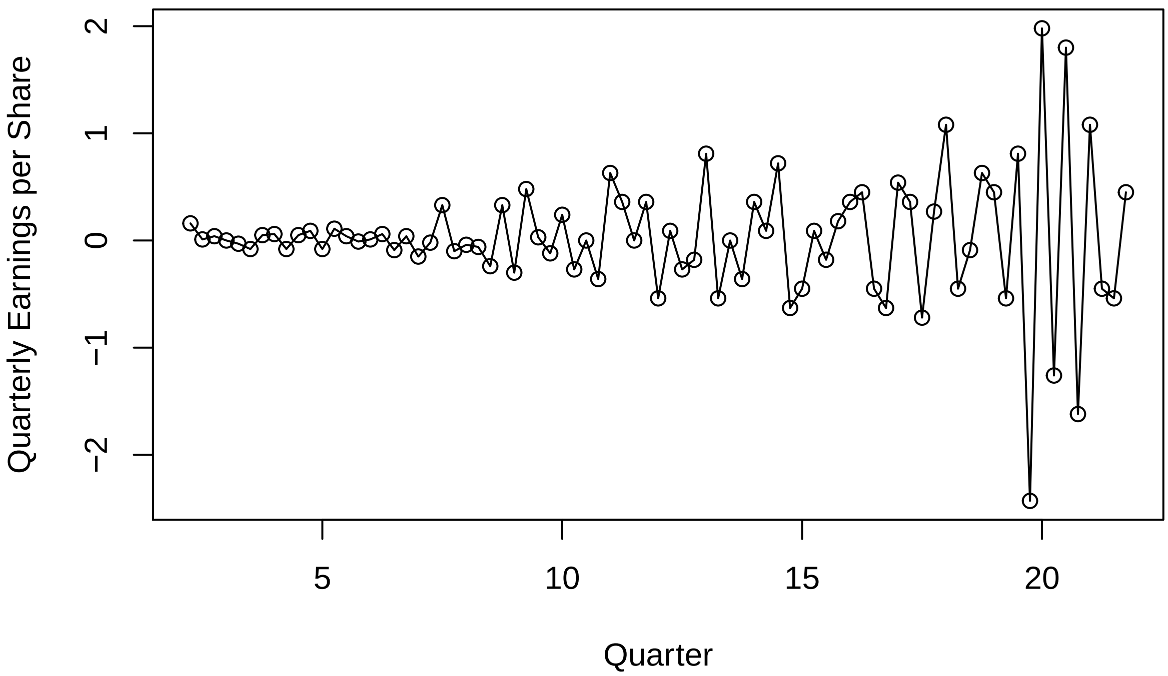



### Part B

Question is asking us to remove trend and seasonal component to make the data stationary. We will do this quickly by assessing autocorrelation (lag) through acf. We see that after lag = 4 periods, seasonal components can be differenced to the point where the data can be said stationary, i.e., meets a certain threshold (sarima, Box-Ljung test) with p value being significant enough. Personally I think this is overkill as p-value $< 0.05$ already.

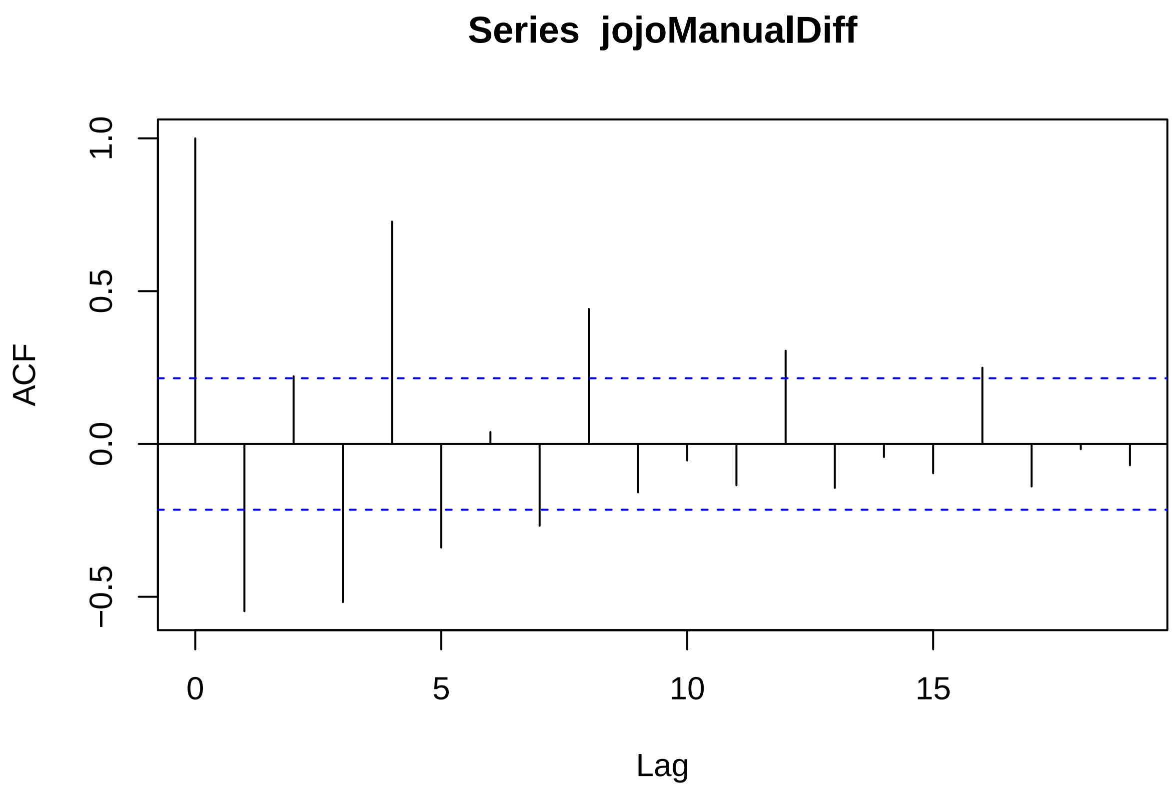

V1
ACF
Lag
1.0
0.5
0.0
−0.5
0
1
2
3
4

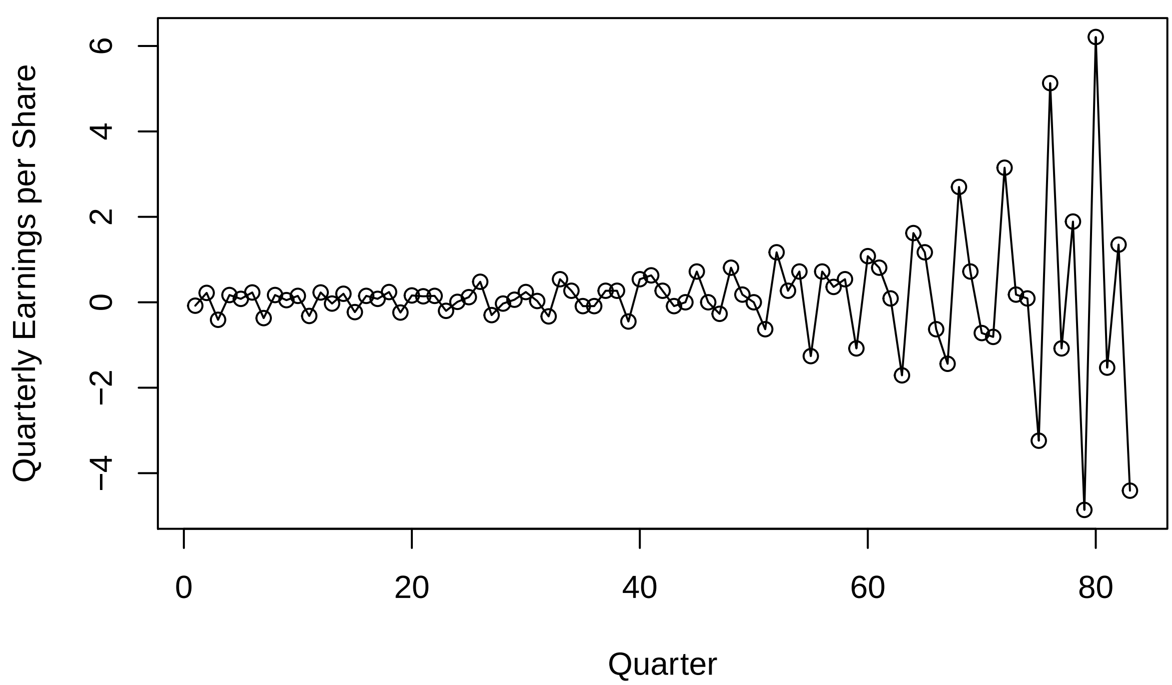

V1
ACF
Lag
1.0
0.5
0.0
−0.5
0
1
2
3
4




## Part C

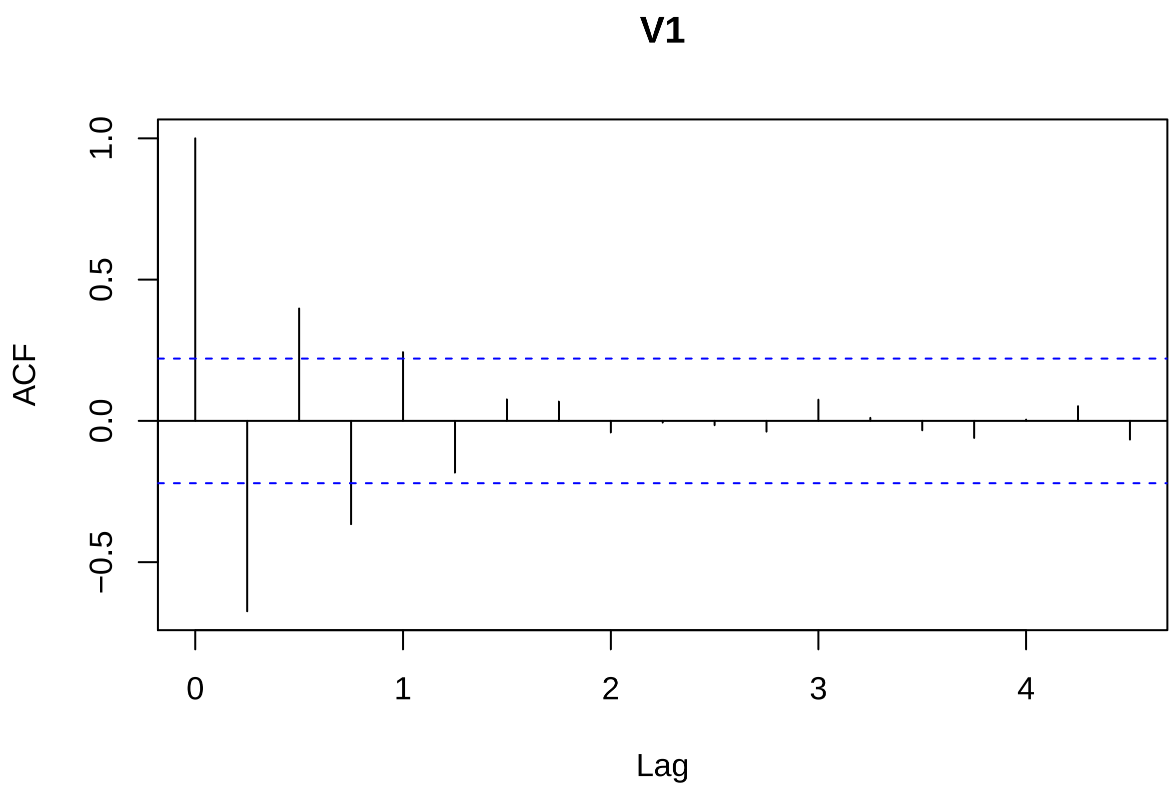
Creating a way to remove trend and seasonal components manually. We will use m = 1 in y’$t = y$_$t$ - $y$(t-m).

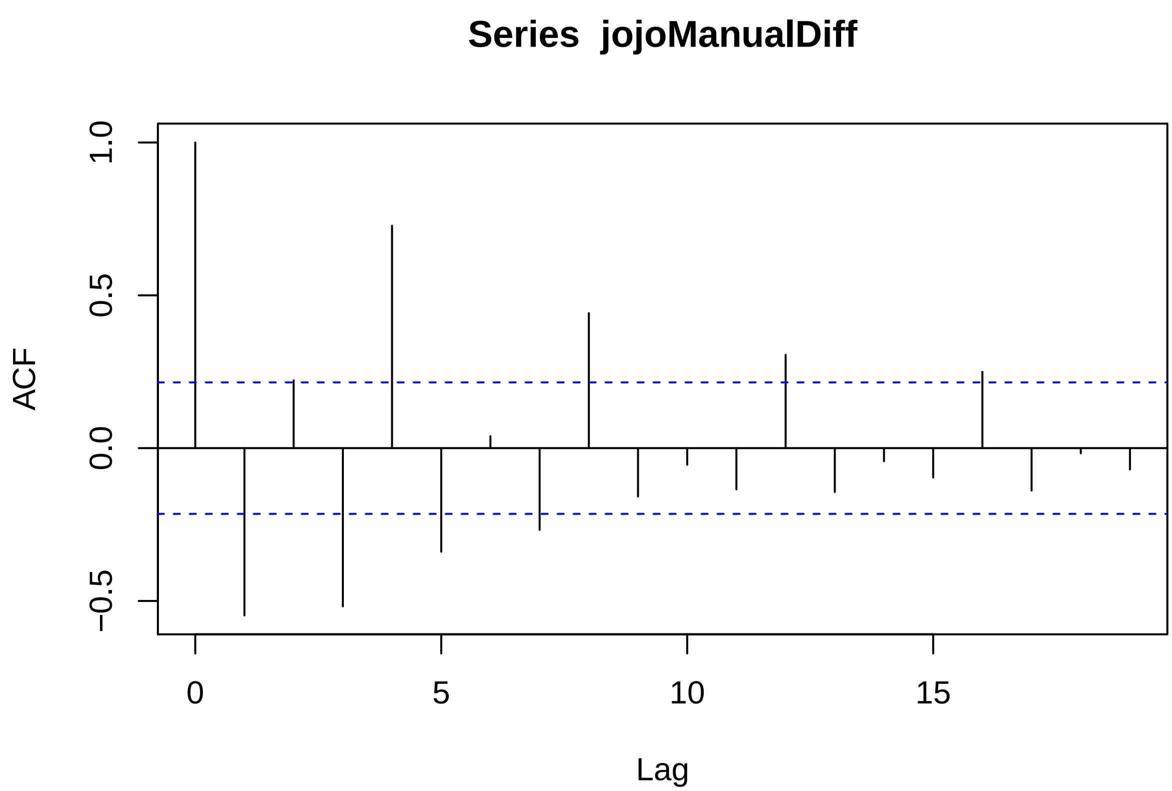

Series jojoManualDiff
ACF
1.0
0.5
0.0
−0.5
0
5
10
15
Lag


Quarterly Earnings per Share
6
4
2
0
−2
−4
0
20
40
60
80
Quarter

## Part D

Assessing if the data is stationary or not. We will use acf and pcf to test Part B and Part C. I believe the data is stationary for Part B because the ACF drops drastically at the start.


V1
ACF
1.0
0.5
0.0
−0.5
0
1
2
3
4
Lag


Series jojoManualDiff
ACF
1.0
0.5
0.0
−0.5
0
5
10
15
Lag

## Part E

Confirming if the data is stationary or not using Augmented Dickey-Fuller (ADF) t-statistic test. Data confirms both p-values $< 0.05$.

```
## Warning in adf.test(jojoDiff): p-value smaller than printed p-value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  jojoDiff
## Dickey-Fuller = -5.9559, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  jojoManualDiff
## Dickey-Fuller = -3.9886, Lag order = 4, p-value = 0.01421
## alternative hypothesis: stationary
```

# Question 2

## Part A

Applying ses() function to multiple extreme low alpha values appear to decrease forecast, lo, and hi for each lower alpha value we go to. Applying ses() function to multiple extreme low alpha values appear to decrease forecast, lo, and hi for each higher alpha value we go to.

```
##    Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 85       3.605153 -2.713464 9.923771 -6.058338 13.26864
```

```
##    Point Forecast    Lo 80 Hi 80    Lo 95    Hi 95
## 85       9.226236 5.542472 12.91 3.592404 14.86007
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95  Hi 95
## 85       13.39684 11.90004 14.89364 11.10768 15.686
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 85       11.82728 10.06149 13.59307 9.126731 14.52783
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 85       11.65397 9.851896 13.45604 8.897938 14.40999
```

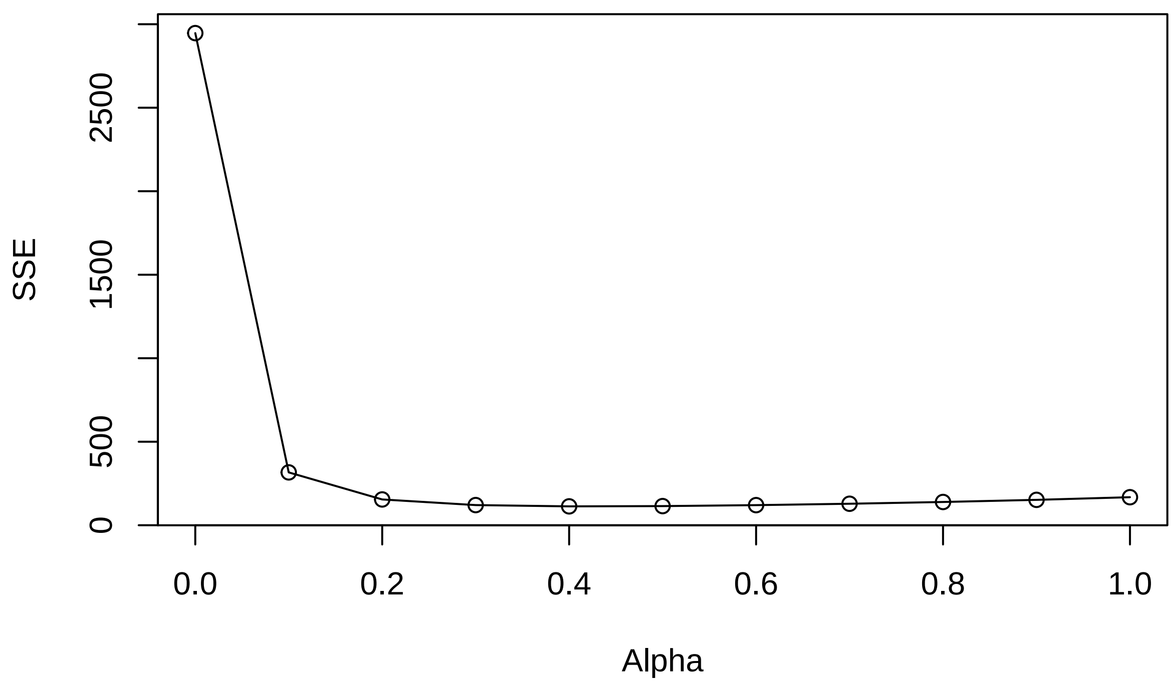
## Part B

Calculating sse from one-step-ahead within-sample forecasts. Base formula uses sum(y_t - yhat_t)^2, used tools to ease calculation…

```
## [1] 289.5153

## [1] 170.3247

## [1] 25.37368

## [1] 11.7024

## [1] 11.05451
```

### Part C

The SSE vs Alpha graph plotted below indicates that SSE is high when alpha = 0.00, and generally decreases until alpha = 0.4 or 0.5. The effect of Alpha is a measure of accuracy for forecasts. An optimal alpha will minimize SSE, meaning the predictions will get closer to our actuals.




### Part D

We set alpha=NULL so that the ses will estimate the optimal alpha by itself. According to Part B, I had an alpha=0.5 and alpha=NULL predictions are not very far from alpha=0.5. This suggests that alpha=0.4 or alpha=0.5 is indeed optimal.

```
##    Point Forecast     Lo 80    Hi 80     Lo 95    Hi 95
## 85       3.605153 -2.713464 9.923771 -6.058338 13.26864
```

```
##    Point Forecast    Lo 80 Hi 80    Lo 95    Hi 95
## 85       9.226236 5.542472 12.91 3.592404 14.86007
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95  Hi 95
## 85       13.39684 11.90004 14.89364 11.10768 15.686
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 85       11.82728 10.06149 13.59307 9.126731 14.52783
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 85       11.65397 9.851896 13.45604 8.897938 14.40999
```

```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 85       13.52969 12.04394 15.01544 11.25743 15.80195
## 86       13.52969 11.91629 15.14309 11.06221 15.99717
## 87       13.52969 11.79803 15.26136 10.88134 16.17804
## 88       13.52969 11.68734 15.37205 10.71205 16.34733
```

### Part E

Does not make much difference. Both are very accurate and only differs by 0.00003 in forecast.

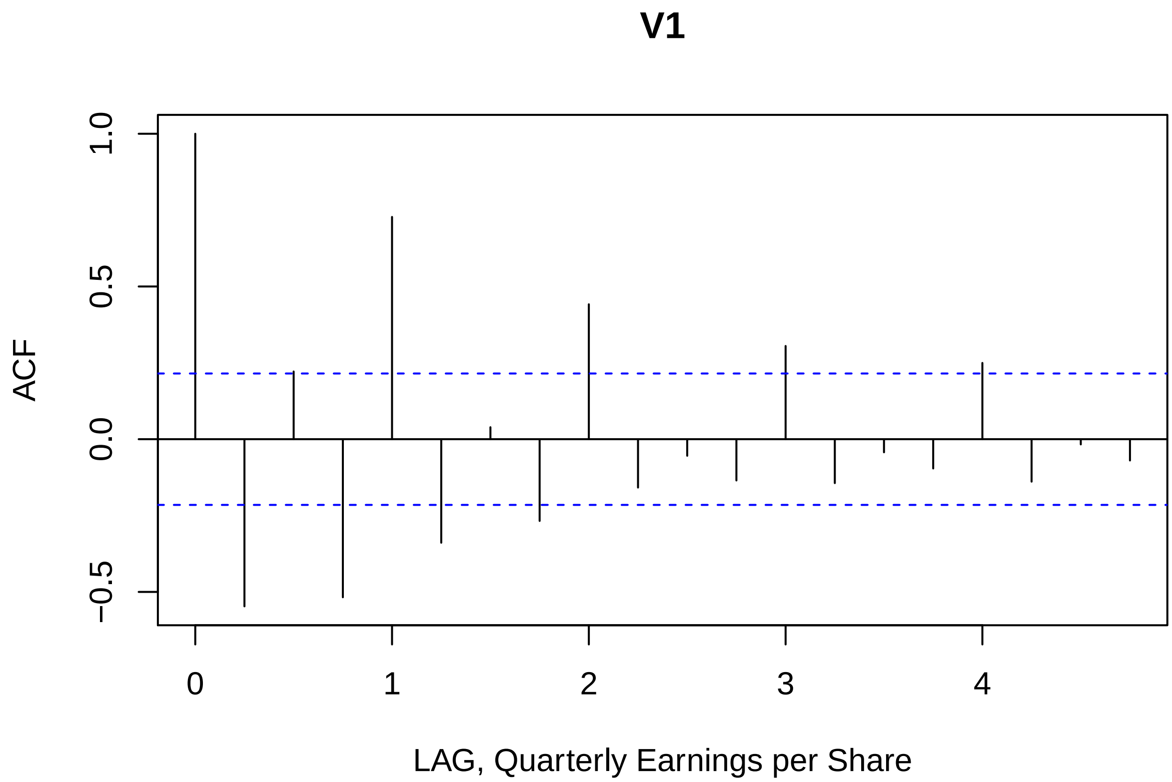
```
##    Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 85       13.52969 12.04394 15.01544 11.25743 15.80195
## 86       13.52969 11.91629 15.14309 11.06221 15.99717
## 87       13.52969 11.79803 15.26136 10.88134 16.17804
## 88       13.52969 11.68734 15.37205 10.71205 16.34733
```

```
##       Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## 22 Q1       13.52972 12.04397 15.01546 11.25747 15.80197
## 22 Q2       13.52972 11.91634 15.14310 11.06226 15.99717
## 22 Q3       13.52972 11.79808 15.26135 10.88141 16.17802
## 22 Q4       13.52972 11.68740 15.37203 10.71214 16.34729
```

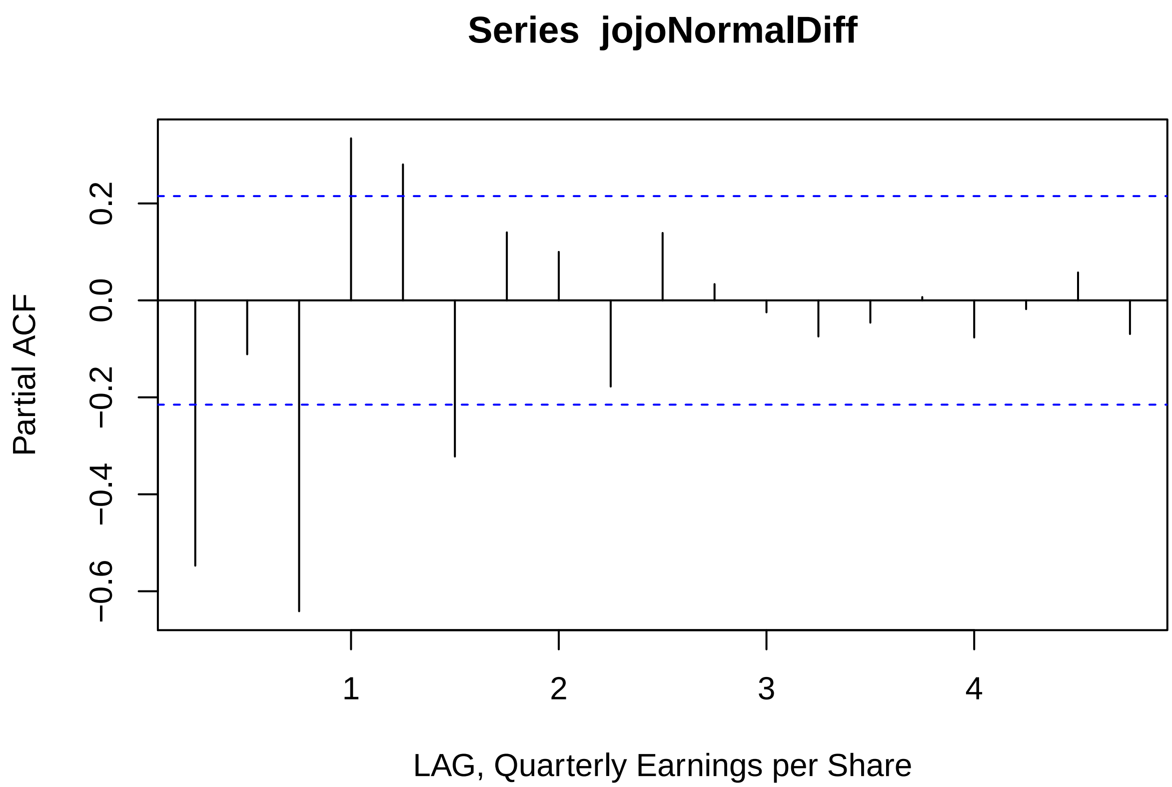
## Question 3

### Part A

Plotted acf and pacf. I can also plot acf2 to see the a common scale on the same plot sheet. LAG are in units of Quarterly Earnings per Share. Based on this model a MA(2) is approriate because after 2nd lag it is clear the model is decaying. AR(1) is approriate because after the 1st lag in the pacf, the model spiked to near zero.



V1
ACF
1.0
0.5
0.0
−0.5
0
1
2
3
4
LAG, Quarterly Earnings per Share




### Part B

Fitted model of p=1, d=1, q=2. Also tried other possible values of p's and q's. It appears p=1, q=2 has the lowest s.e.. Also, p=1, q=2 seems to have the lowest aic of all the possible combinations tried.

```
##
## Call:
## arima(x = jojoTS, order = c(1, 1, 2))
##
## Coefficients:
##          ar1      ma1     ma2
##       0.5431  -1.6986  1.0000
## s.e.  0.0999   0.0564  0.0607
##
## sigma^2 estimated as 0.807:  log likelihood = -112.46,  aic = 232.92
```

```
##
## Call:
## arima(x = jojoTS, order = c(1, 1, 1))
##
## Coefficients:
##           ar1      ma1
##       -0.3277  -0.4313
## s.e.   0.1403   0.1050
##
```

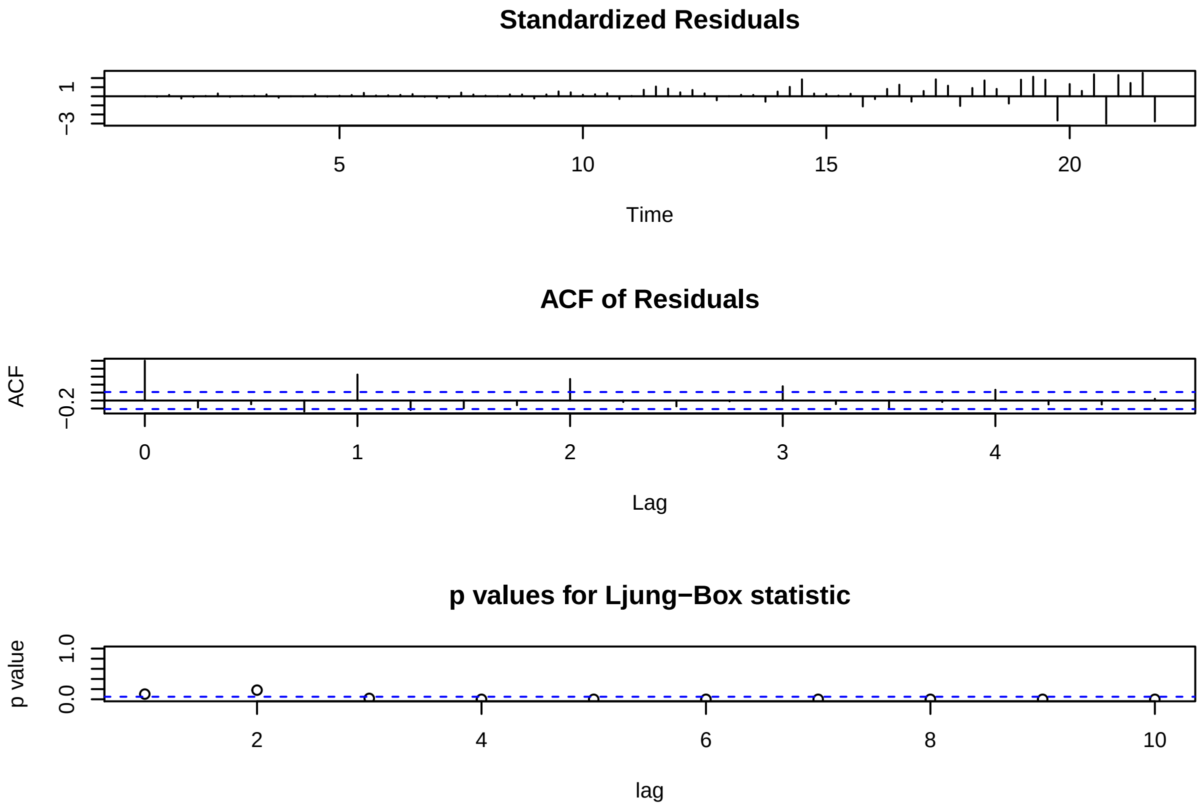
```
## sigma^2 estimated as 1.282:  log likelihood = -128.37,  aic = 262.74
```

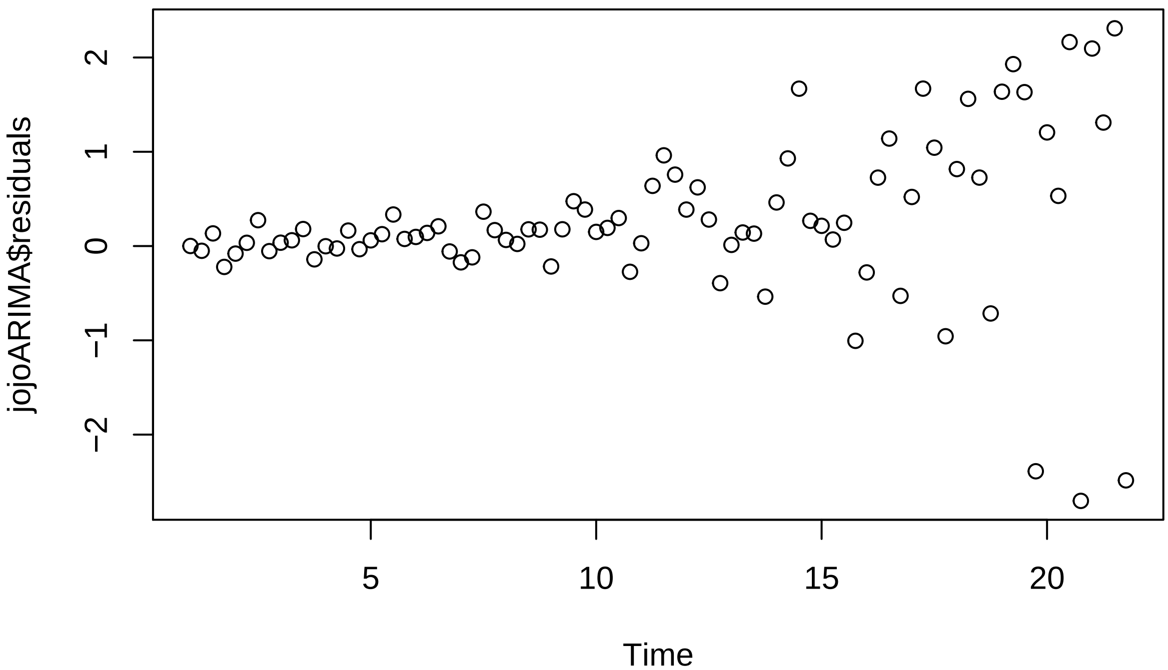
```
##
## Call:
## arima(x = jojoTS, order = c(2, 1, 1))
##
## Coefficients:
##           ar1      ar2      ma1
##       -0.3447  -0.0221  -0.4185
## s.e.   0.1794   0.1472   0.1353
##
## sigma^2 estimated as 1.281:  log likelihood = -128.36,  aic = 264.72
```

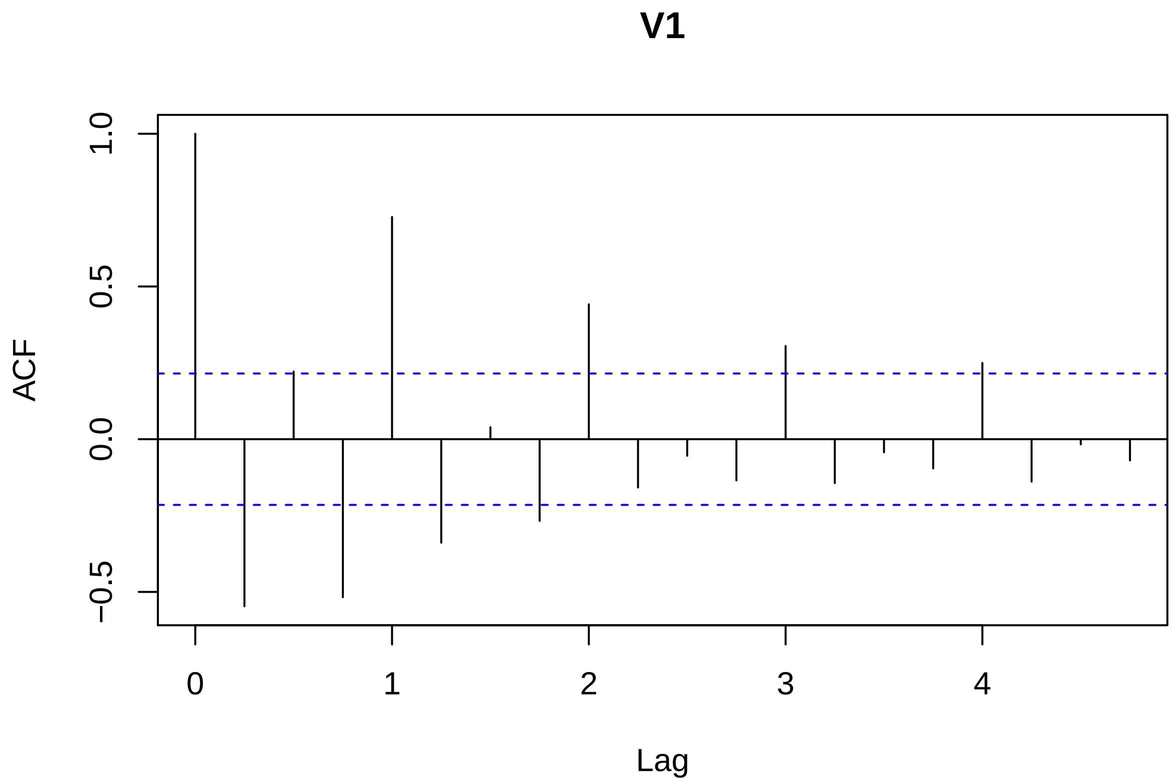
```
##
## Call:
## arima(x = jojoTS, order = c(2, 1, 2))
##
## Coefficients:
##          ar1      ar2      ma1     ma2
##       0.1717  -0.2602  -1.3093  0.9219
## s.e.  0.1508   0.1576   0.1466  0.1014
##
## sigma^2 estimated as 0.8452:  log likelihood = -112.82,  aic = 235.64
```

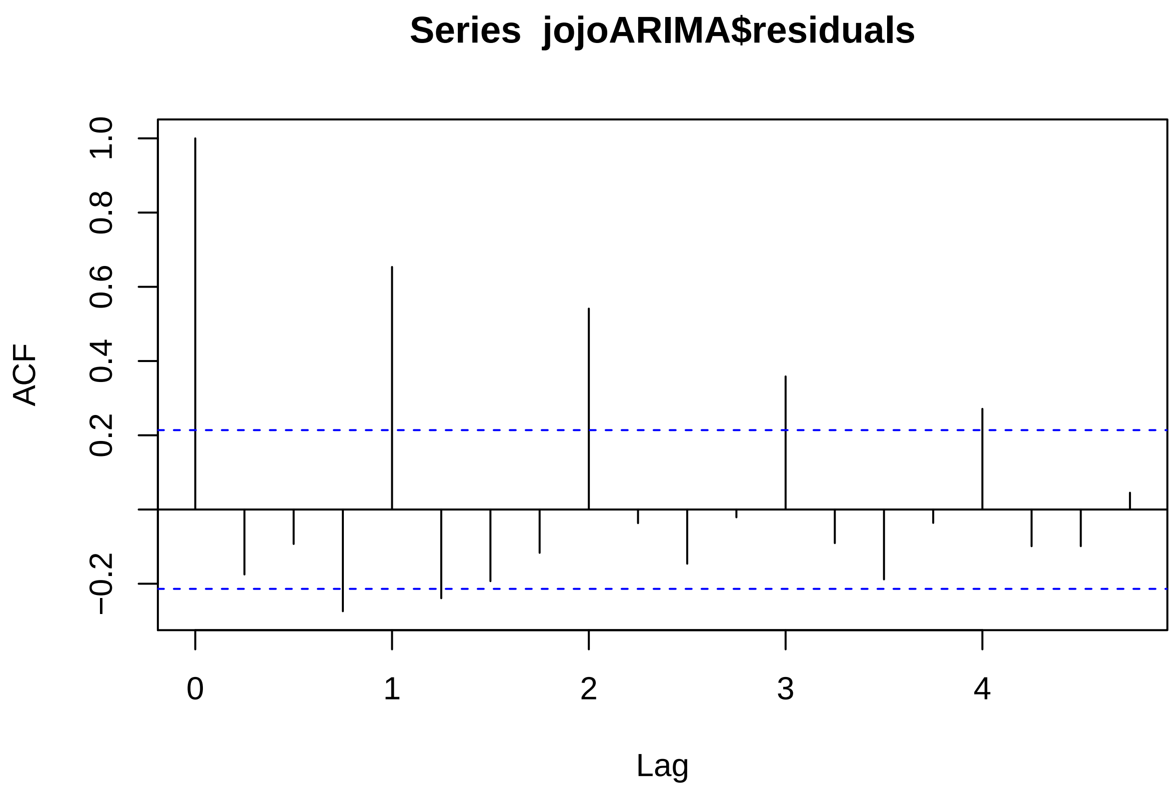
## Part C

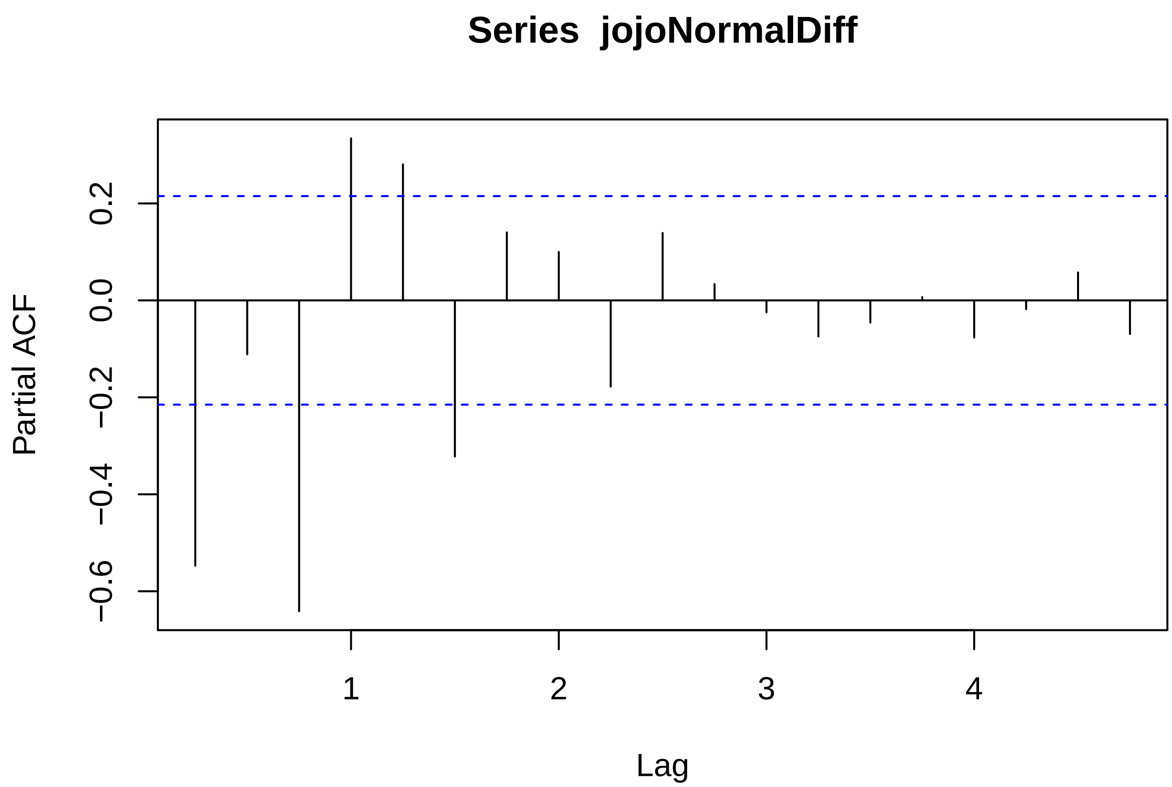
Based on tsdiag(jojoARIMA), which provided a p-values graph for Ljung-Box statistics, the points surpassed 0.05 by a wide margin so therefore the residuals are not stationary; we cannot reject the null hypothesis.

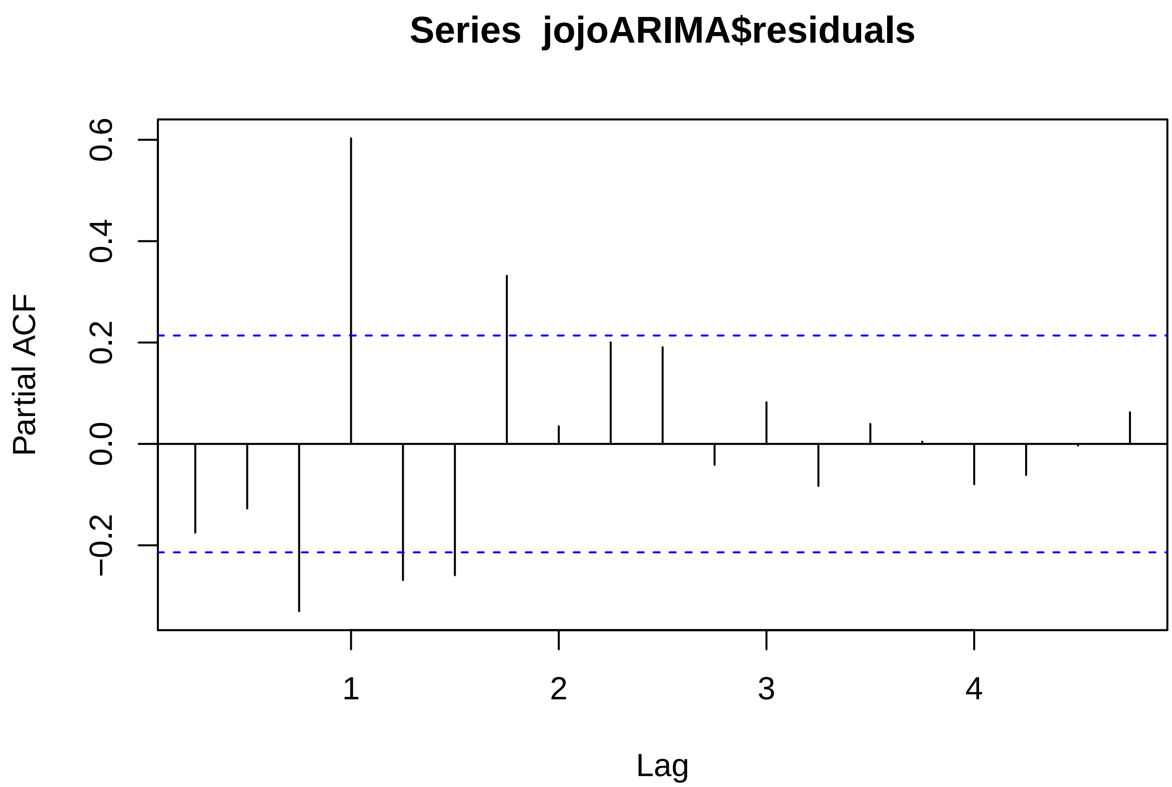

Standardized Residuals
1
−3
5
10
15
20
Time
ACF of Residuals
ACF
−0.2
0
1
2
3
4
Lag
p values for Ljung−Box statistic
p value
1.0
0.0
2
4
6
8
10
lag


jojoARIMA$residuals
Time
2
1
0
−1
−2
5
10
15
20


V1
ACF
1.0
0.5
0.0
−0.5
0
1
2
3
4
Lag


Series jojoARIMA$residuals
ACF
1.0
0.8
0.6
0.4
0.2
−0.2
0
1
2
3
4
Lag


Series jojoNormalDiff
Partial ACF
0.2
0.0
−0.2
−0.4
−0.6
1
2
3
4
Lag




### Part D

Performed predict looking 4 steps into the future. Applied arima to order of p=1, d=0, q=1.

```
## $pred
##        Qtr1     Qtr2     Qtr3     Qtr4
## 22 11.56461 11.56461 11.56461 11.56461
##
## $se
##        Qtr1     Qtr2     Qtr3     Qtr4
## 22 1.421362 2.020448 2.478740 2.864628
```

### Part E

ARIMA(1,0,1) model gave an SSE of 22 to 22.8: 1.4, 2.02, 2.47, 2.86… ses gave an clean cut SSE of 114. ARIMA performed better.