Seventh meeting notes

Daniel L.

4/01/2022

To prepare for this analysis:

- Step 1: Order of differencing
  - TS plot of residuals from ARIMA(0,0,0) w/ constant in not exactly stationary -> d = 1
- Step 2: AR() or MA()

1. Obtain residual from current model: ARIMA(0,1,0) w/ constant
2. Plot PACF of the residual

- Step 3: Seasonality()

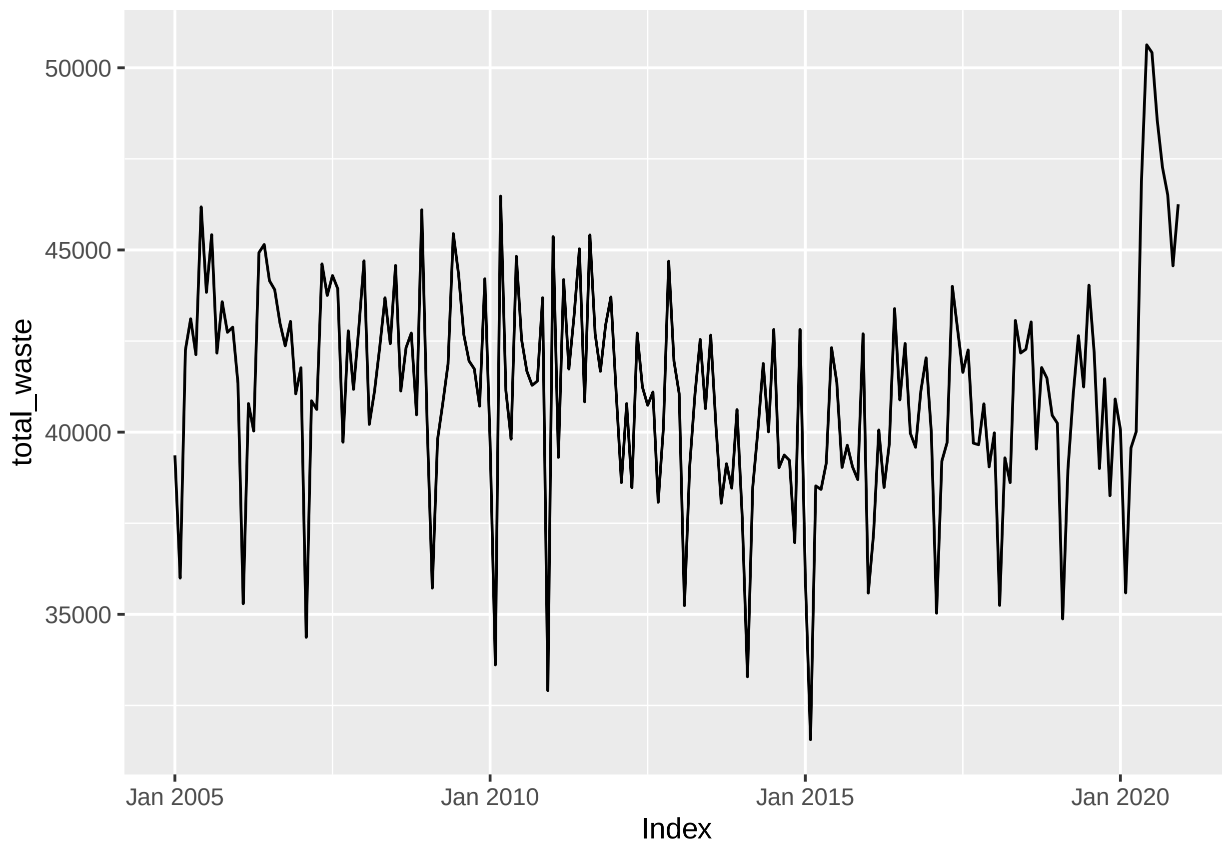
1. Look at the PACF to determine AR() or MA() terms

Creating models with zoo() and the arima package from stats()

Here we create the ts and create the autoplot of the ts using the zoo package

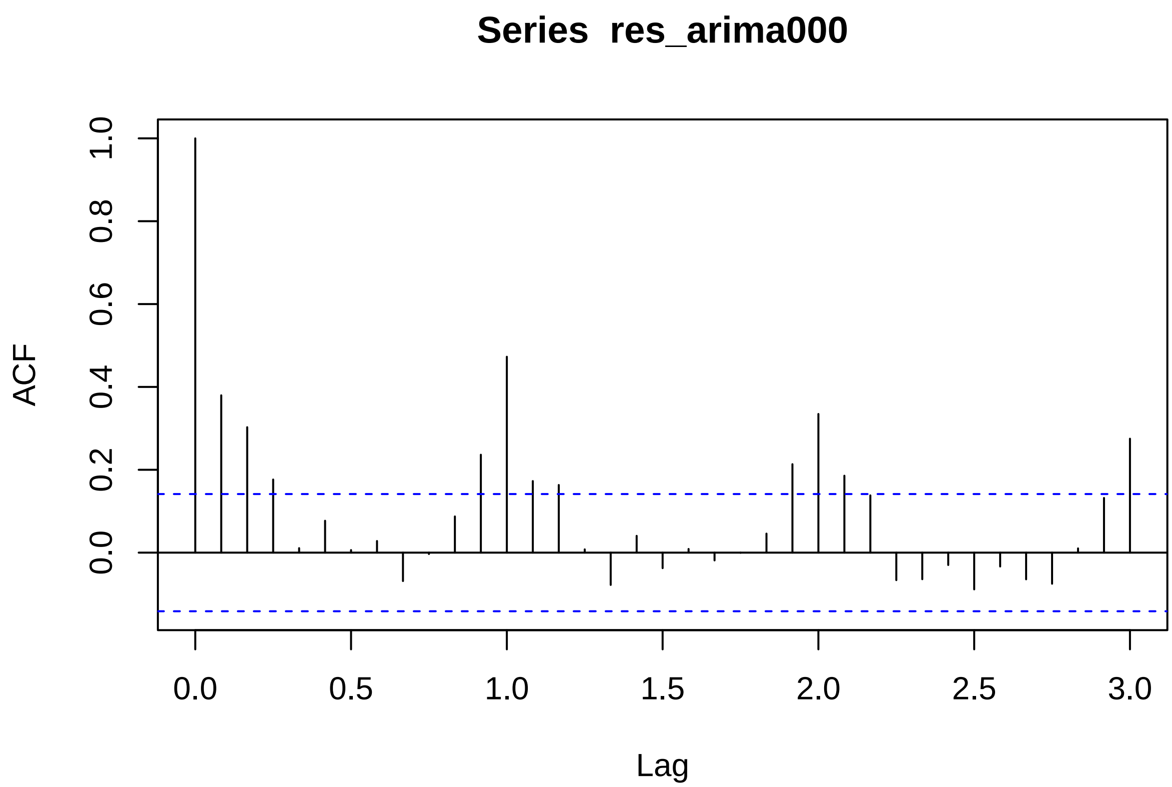
```
DSNY_BX_zoo_ts <- ts(DSNY_third_bronx[,-1],
                     start = as.yearmon(DSNY_third_bronx$month)[1],
                     frequency = 12)

autoplot(as.zoo(DSNY_BX_zoo_ts))
```

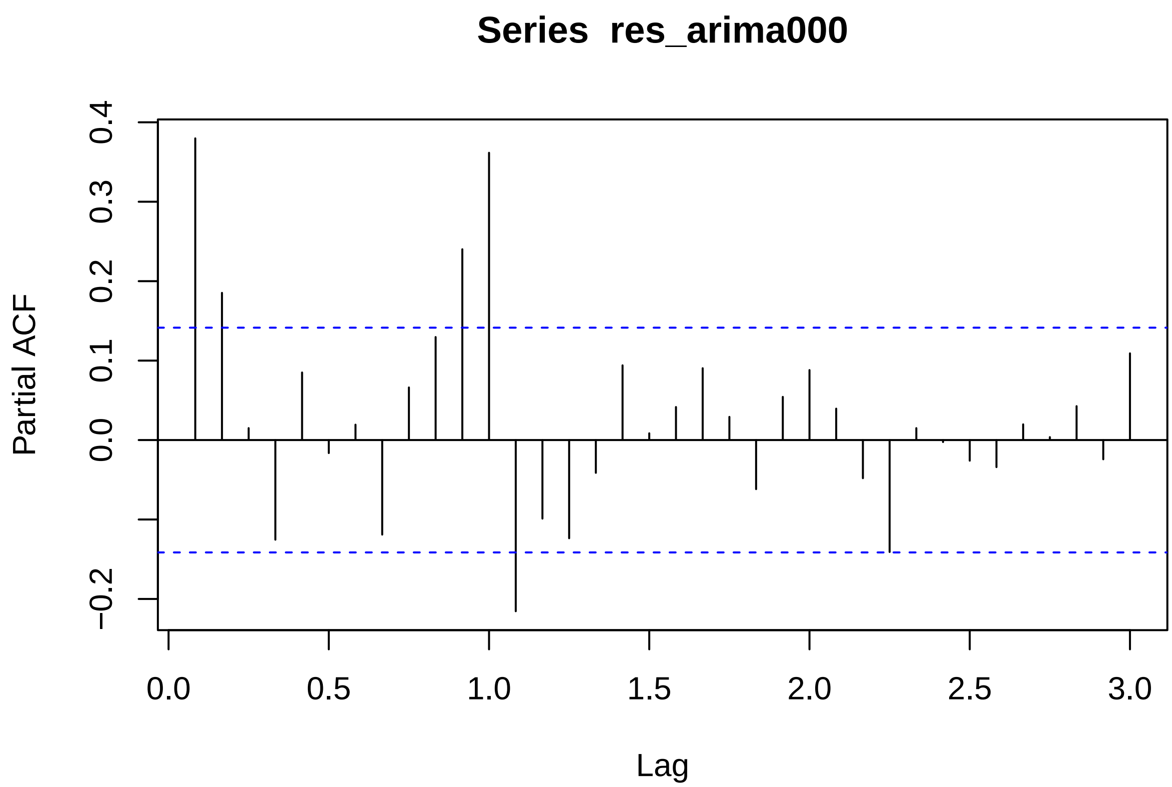


ARIMA(0,0,0) with no constant

```
zoo_arima000_fit <- arima(DSNY_BX_zoo_ts, order = c(0,0,0))
#names(zoo_arima000_fit)
res_arima000 <- zoo_arima000_fit$residuals
acf(res_arima000, lag.max = 36)
```




```
pacf(res_arima000, lag.max = 36)
```

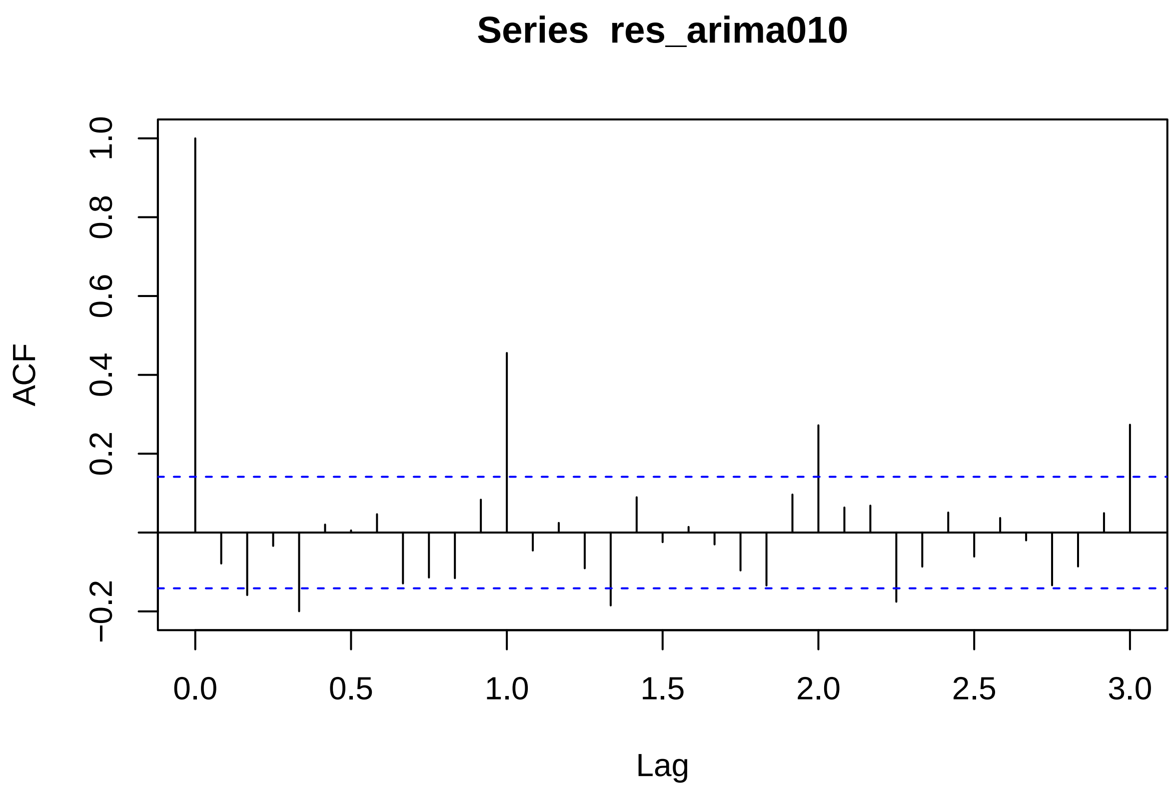


The lags are in decimal format. With the frequency defined as 12, I believe Lag 1.0 = 12, Lag 2.0 = 24, Lag 3.0 = 36.

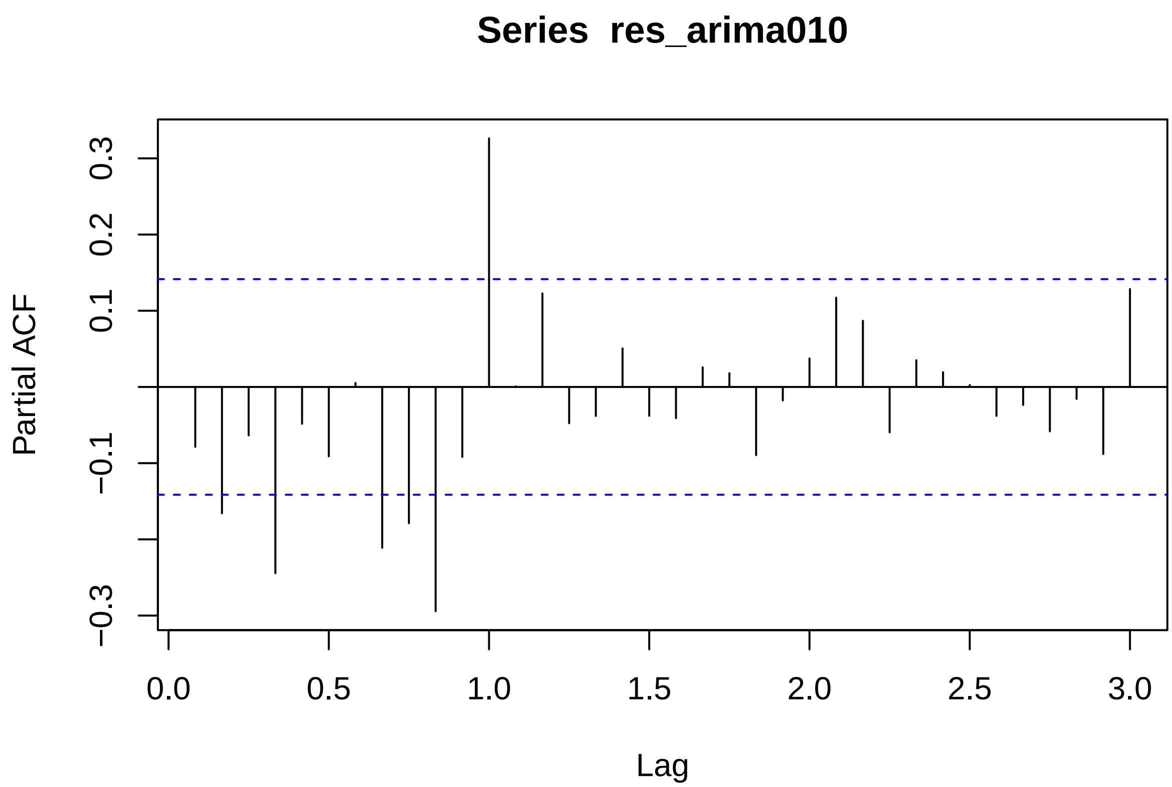
For ARIMA models with differencing, the differenced series follows a zero-mean ARMA model. Documentation by DataCamp.

ARIMA(0,1,0) with no constant/mean

```
zoo_arima010_fit <- arima(DSNY_BX_zoo_ts,
                          order = 1 + c(0,1,0)) #the PACF plot return the same values
                          #with or without the include.mean = parameter
res_arima010 <- zoo_arima010_fit$residuals
acf(res_arima010, lag.max = 36)
```



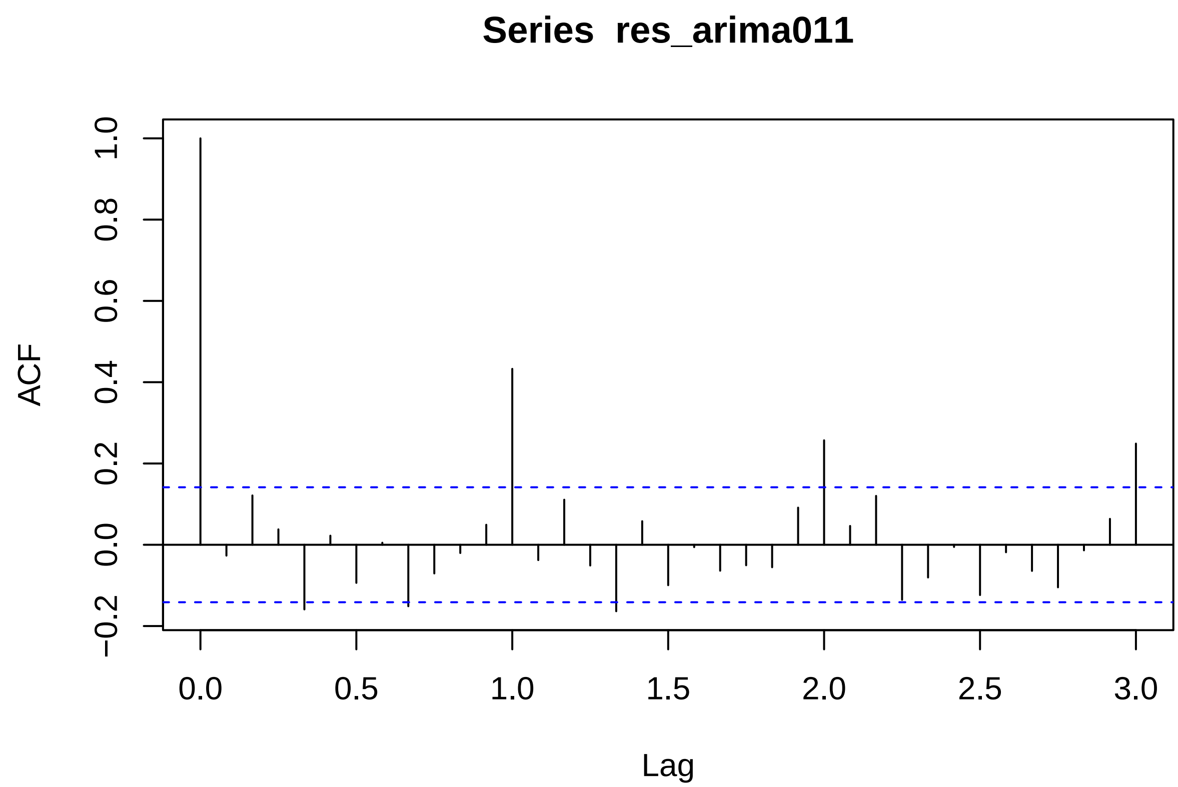

```
pacf(res_arima010, lag.max = 36)
```



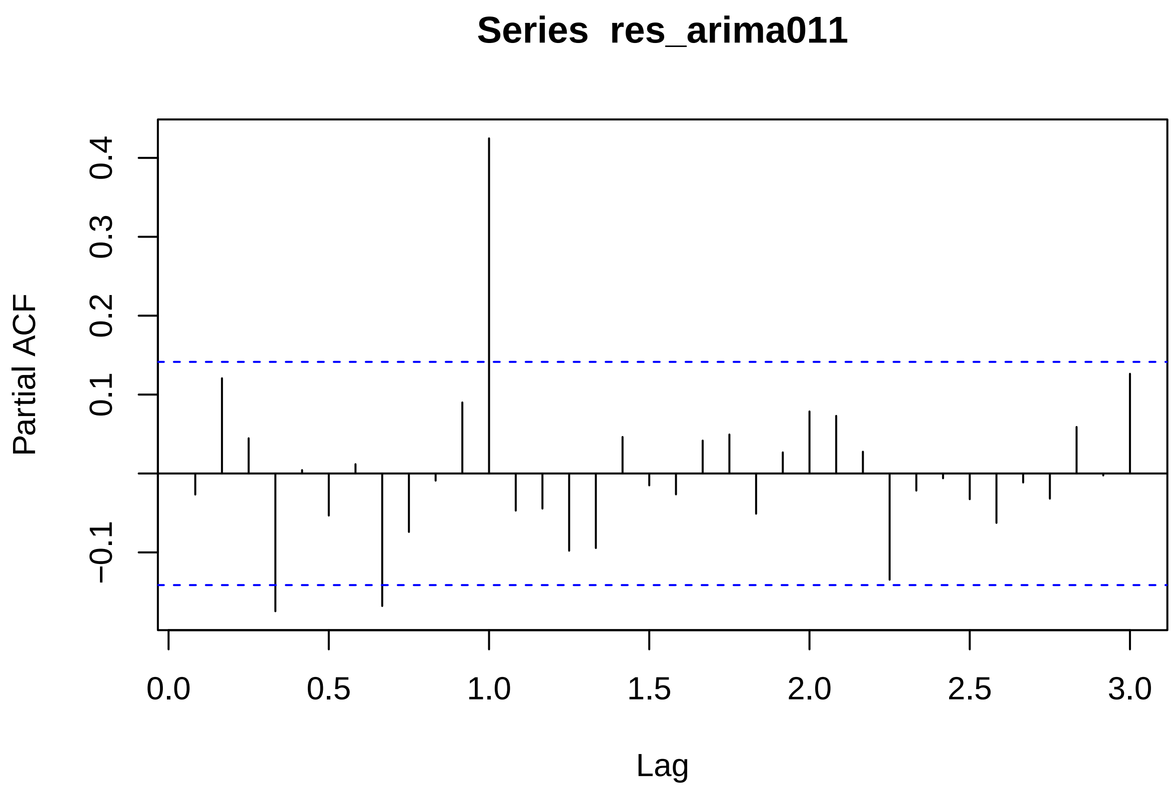
A constant term has been added with '1 + c(0,1,0)'. As discussed before, a negative ACF value is indicative of the need to add an MA() parameter. Here we see that lag 3 is no longer significant, perhaps we can try to add MA(1) and MA(2) arguments to our differenced time series.

ARIMA(0,1,1) with a constant

```
zoo_arima011_fit <- arima(DSNY_BX_zoo_ts,
                          order = 1 + c(0,1,1))
res_arima011 <- zoo_arima011_fit$residuals
acf(res_arima011, lag.max = 36)
```



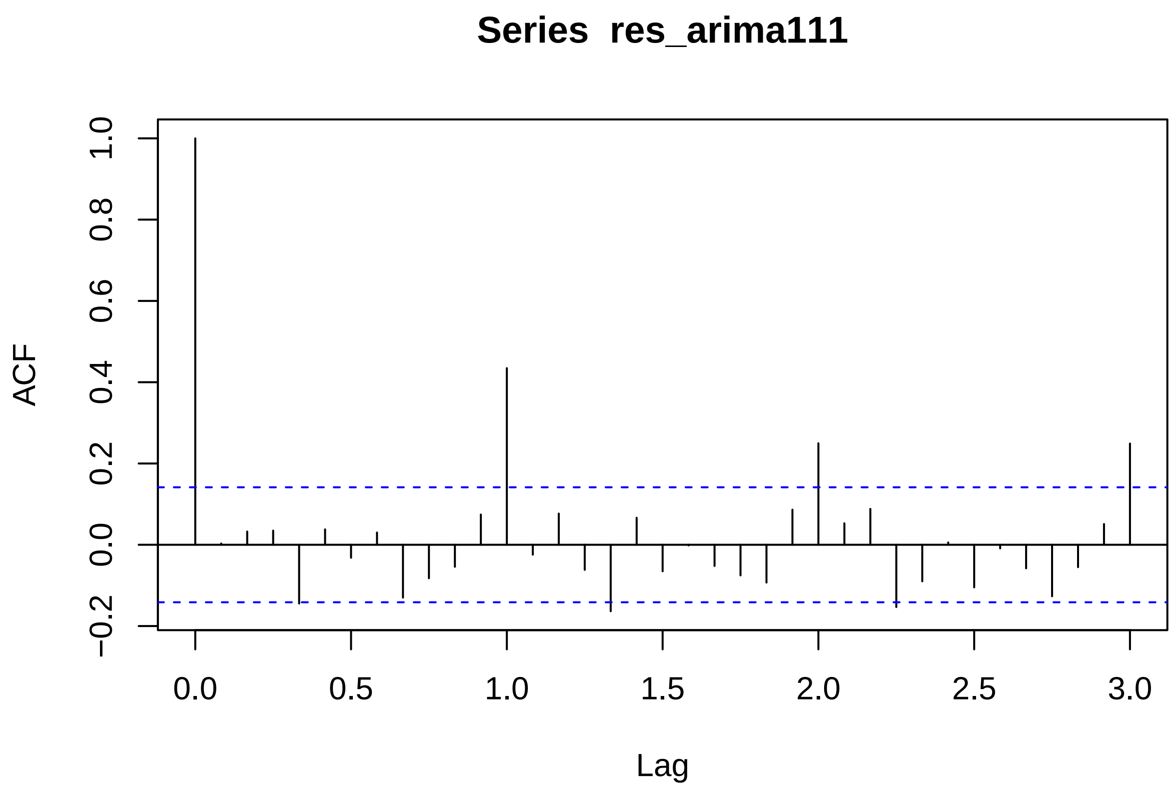

```
pacf(res_arima011, lag.max = 36)
```



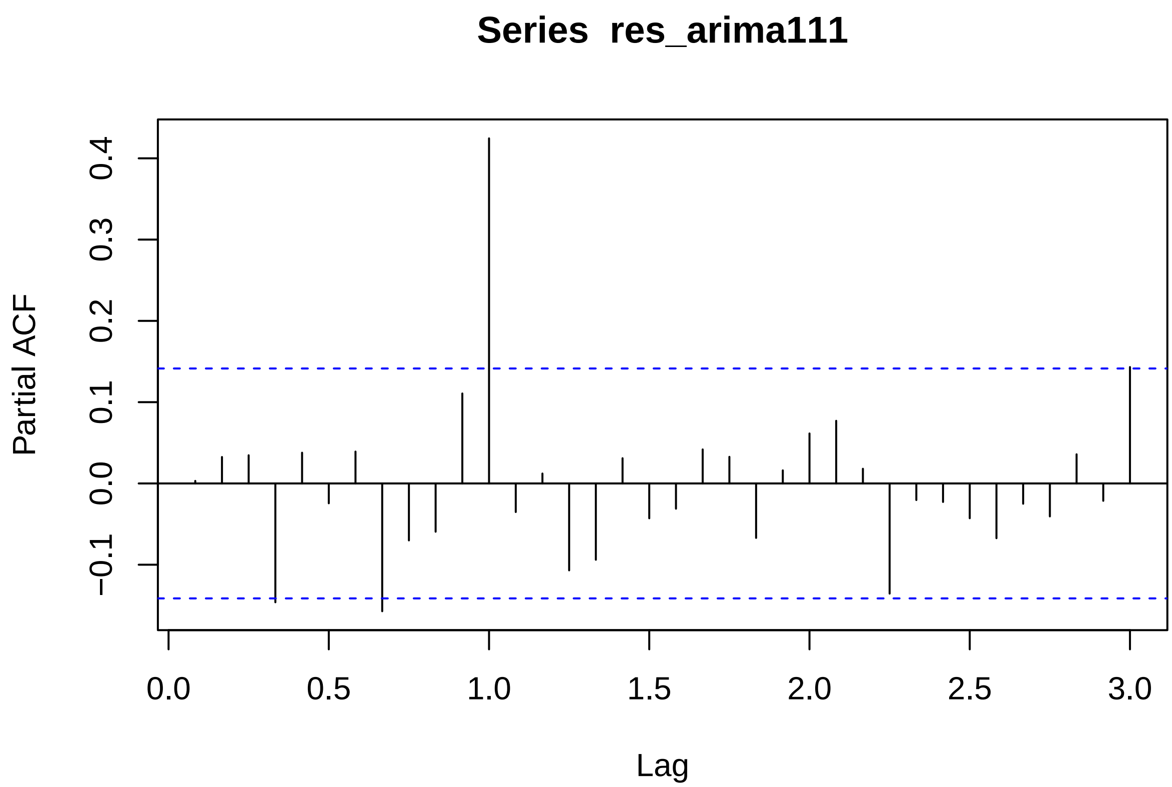
The lag one value of the ACF plot is 1.0 and positive. With a seasonal pattern of significant lags at each whole lag. There definitely should be a seasonal argument.

ARIMA(1,1,1) with constant

```
zoo_arima111_fit <- arima(DSNY_BX_zoo_ts,
                          order = 1+ c(1,1,1))
res_arima111 <- zoo_arima111_fit$residuals
acf(res_arima111, lag.max = 36)
```



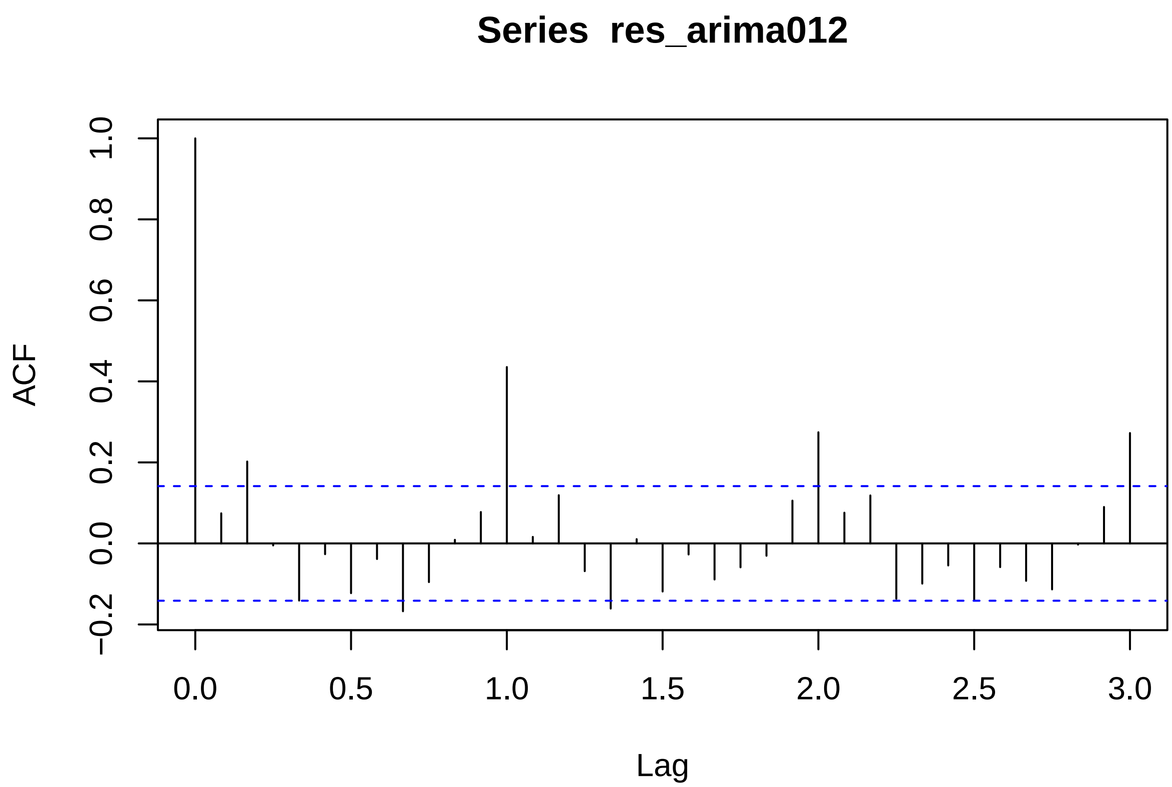

```r
pacf(res_arima111, lag.max = 36)
```



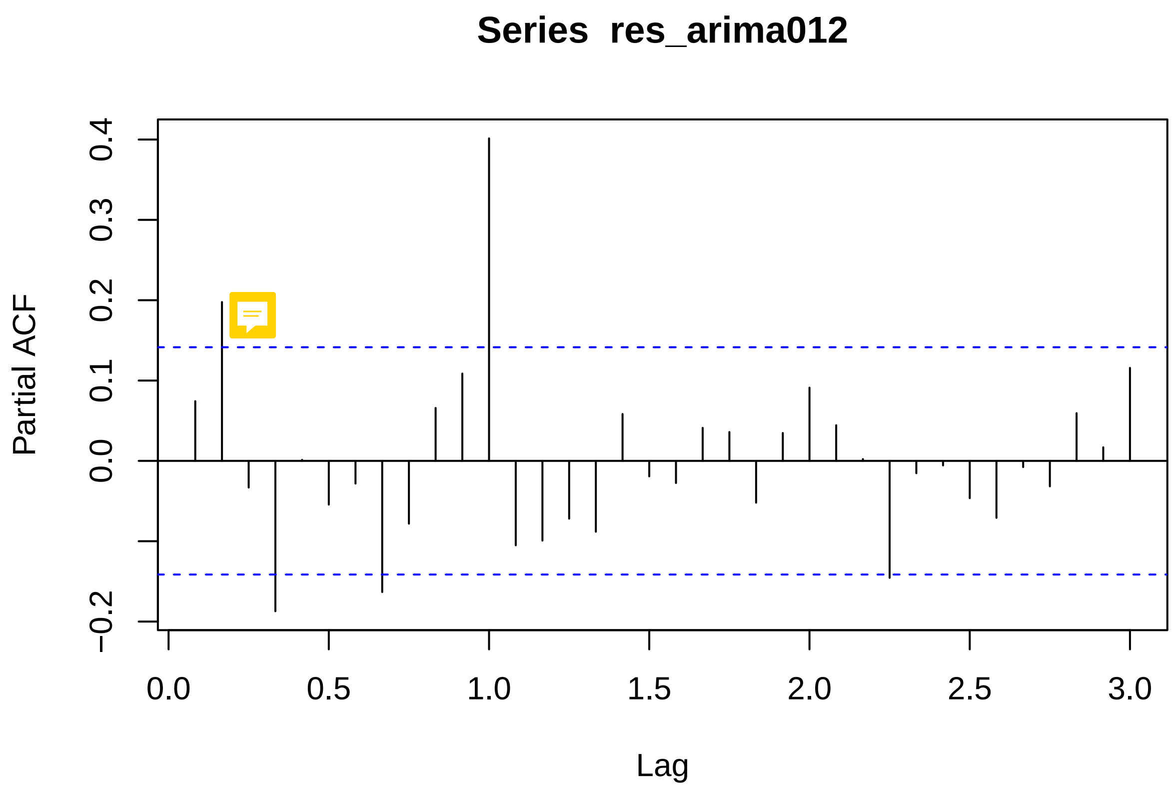
We can see that in the ACF(), the seasonal lags are not being contained within the range. The PACF() is showing promising signs. Lags 4, and 8 are negative and significant and lag 12 is positive and significant. Before committing to a single MA() paramter, I would like to test an MA(2) model on the differenced data.

ARIMA(0,1,2) with a constant

```
zoo_arima012_fit <- arima(DSNY_BX_zoo_ts,
                          order = 1+ c(0,1,2))
res_arima012 <- zoo_arima012_fit$residuals
acf(res_arima012, lag.max = 36)
```



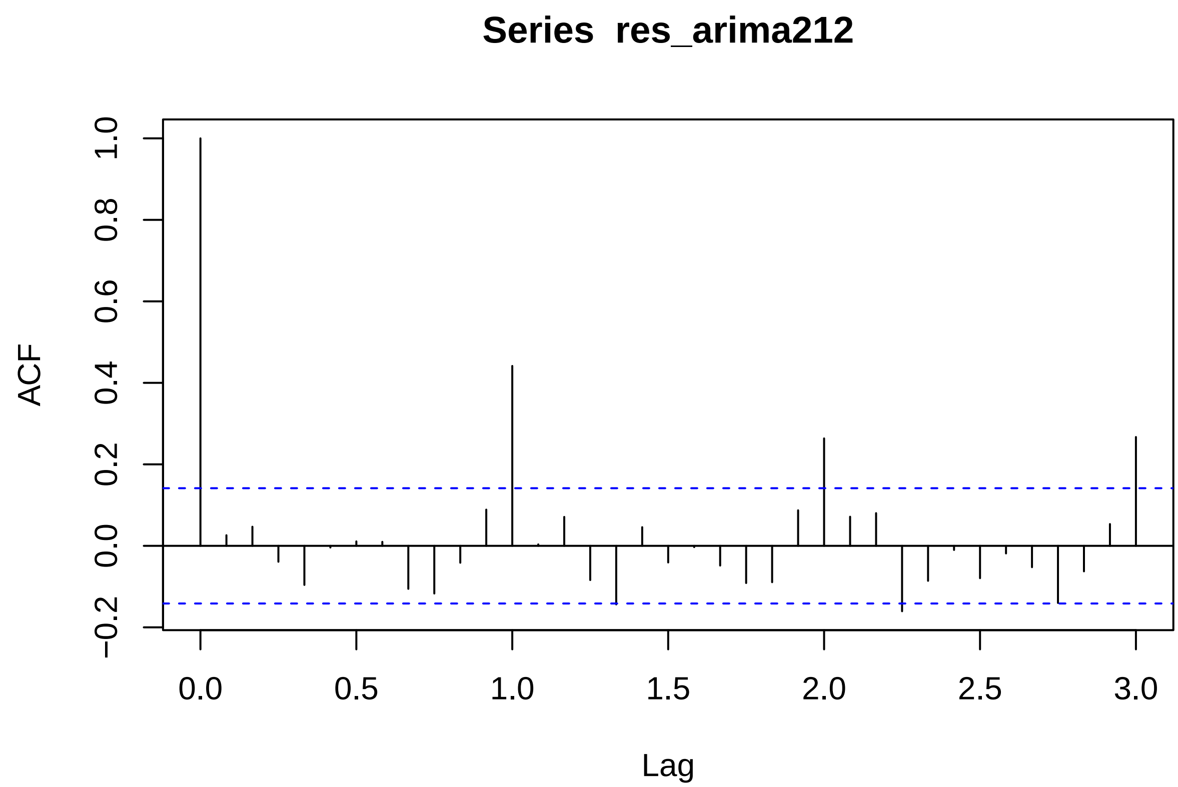

```
pacf(res_arima012, lag.max = 36)
```



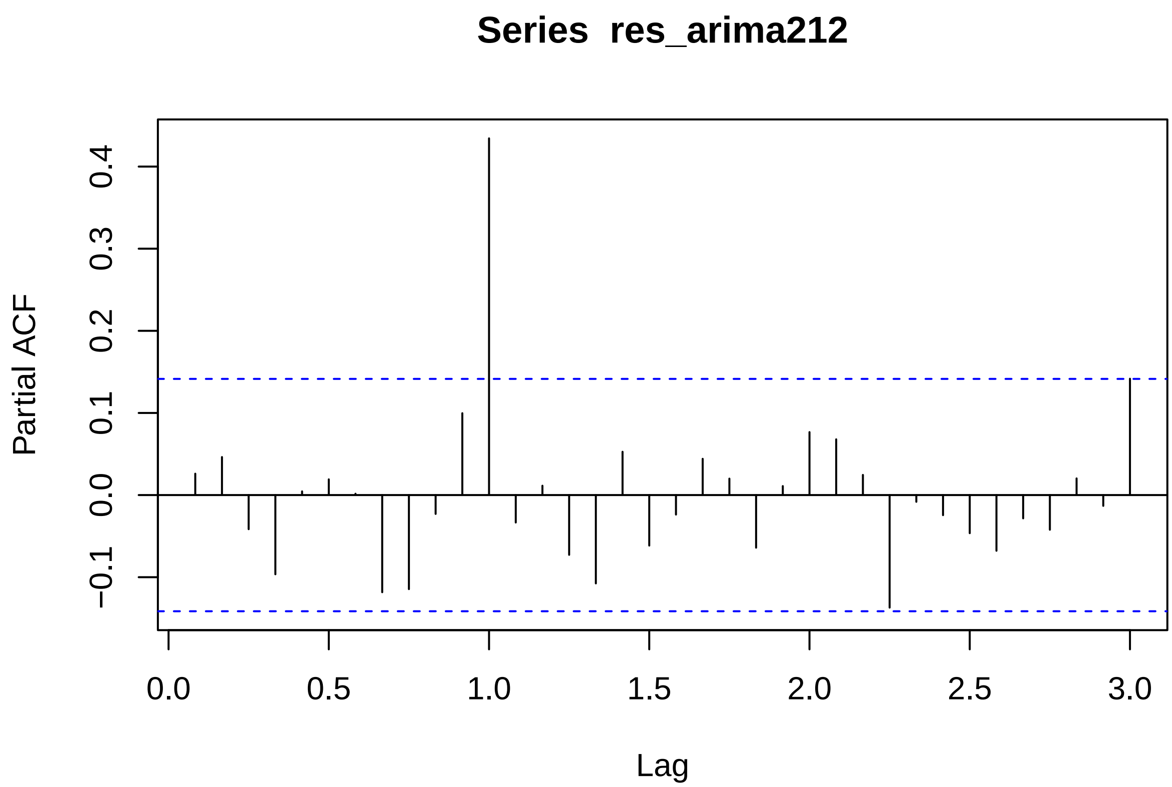
Lags (2,4,8) are significant in the PACF(). I am not sure if the lag 2 value of the PACF() is encouraging us to add an AR(2) argument. I would believe that this would remove any progress we have made. For now, I think it is best to address the seasonal lags.

ARIMA(2,1,2) with a constant

```
zoo_arima212_fit <- arima(DSNY_BX_zoo_ts,
                          order = 1+ c(2,1,2))
res_arima212 <- zoo_arima212_fit$residuals
acf(res_arima212, lag.max = 36)
```




```
pacf(res_arima212, lag.max = 36)
```

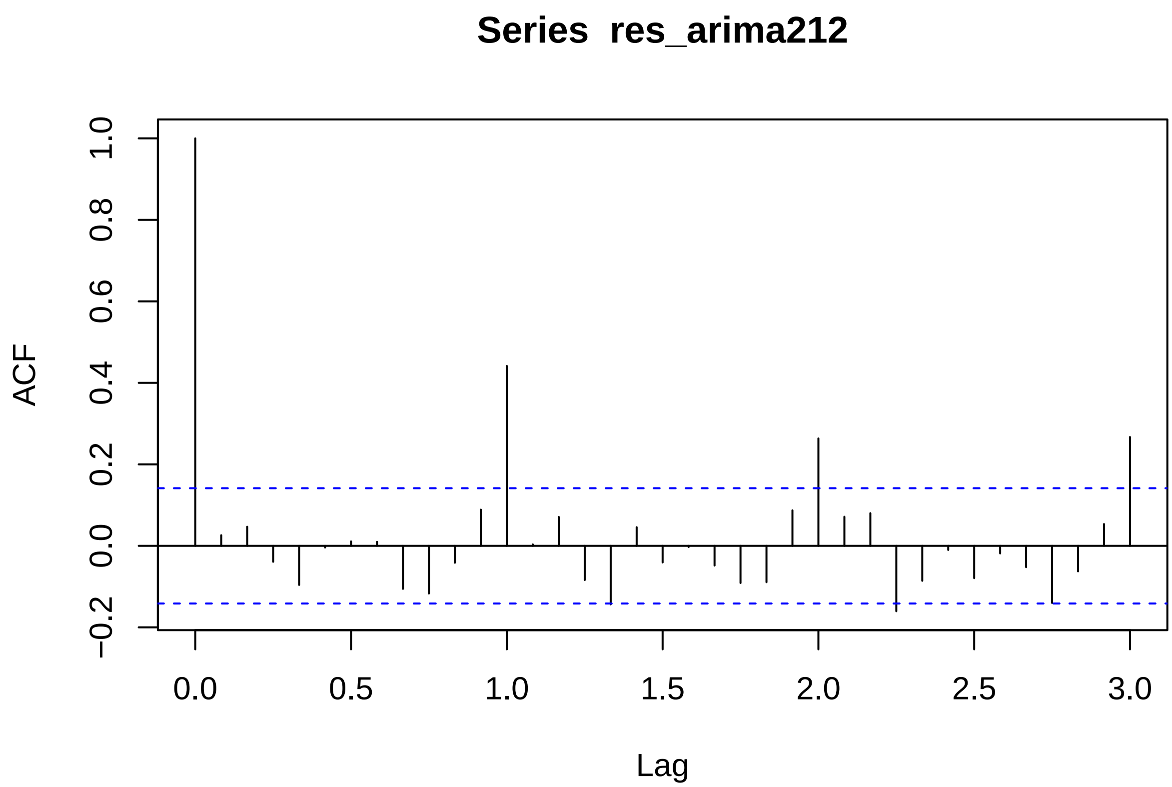


I have attempted an ARIMA(2,1,2) and the PACF() plot is showing promising signs. The non-seasonal lags are contained within range, and only lag 12 is significant. However, I am worried that this model maybe over-fitted. The addition of both AR and MA arguments will make this model hard to interpret.

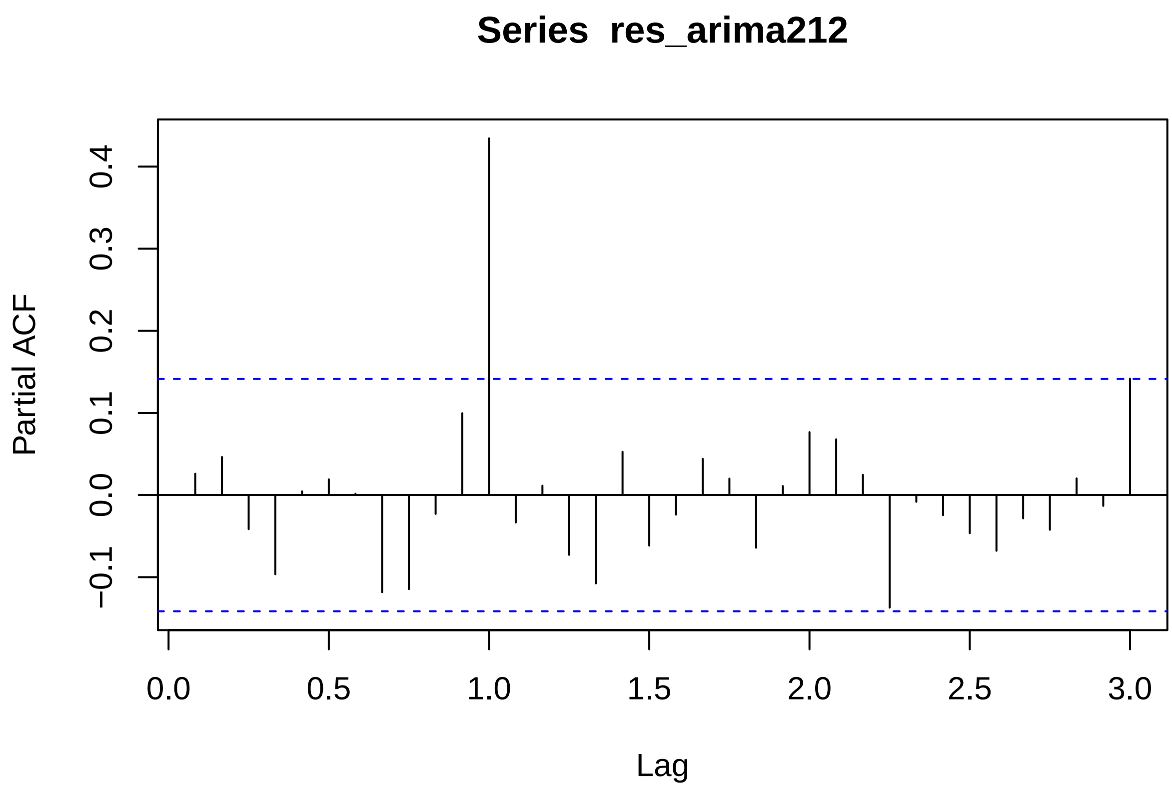
ARIMA(2,1,2) with a constant and seasonal period = 12

ARIMA(2,1,2)(0,0,0)$_{12}$

```
zoo_arima212_000_12fit <- arima(DSNY_BX_zoo_ts,
                                order = 1+ c(2,1,2),
                                seasonal = list(order = c(0L,0L,0L),period = 12))
res_arima212 <- zoo_arima212_000_12fit$residuals
acf(res_arima212, lag.max = 36)
```




```
pacf(res_arima212, lag.max = 36)
```

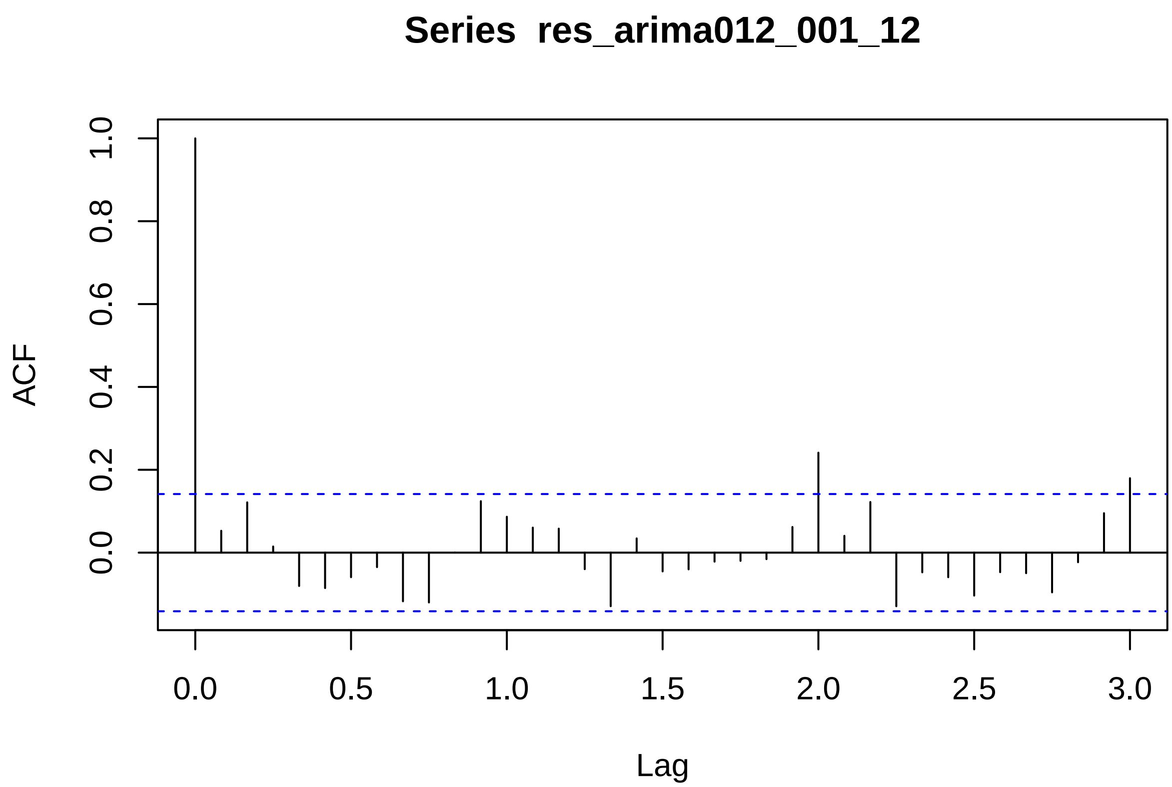


It would appear that this 'overfitted' model can not contain the seasonal lags within bounds. For now, I will not focus too much on this model.

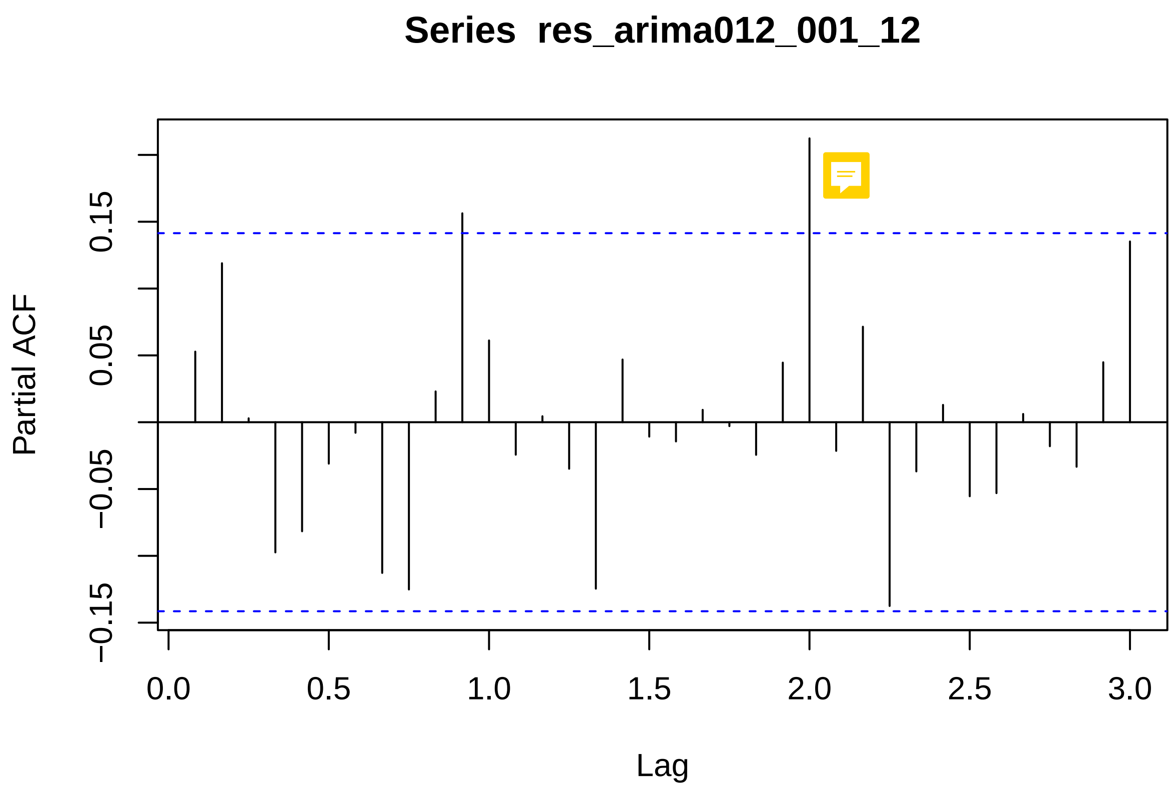
ARIMA(0,1,2)(0,0,1) with a constant and seasonal period = 12

ARIMA(0,1,2)(0,0,1)$_{12}$

```
zoo_arima012_001_12fit <- arima(DSNY_BX_zoo_ts,
                                order = 1 + c(0,1,2),
                                seasonal = list(order = c(0L,0L,1),
                                                period = 12))
res_arima012_001_12 <- zoo_arima012_001_12fit$residuals
acf(res_arima012_001_12, lag.max = 36)
```




```
pacf(res_arima012_001_12, lag.max = 36)
```

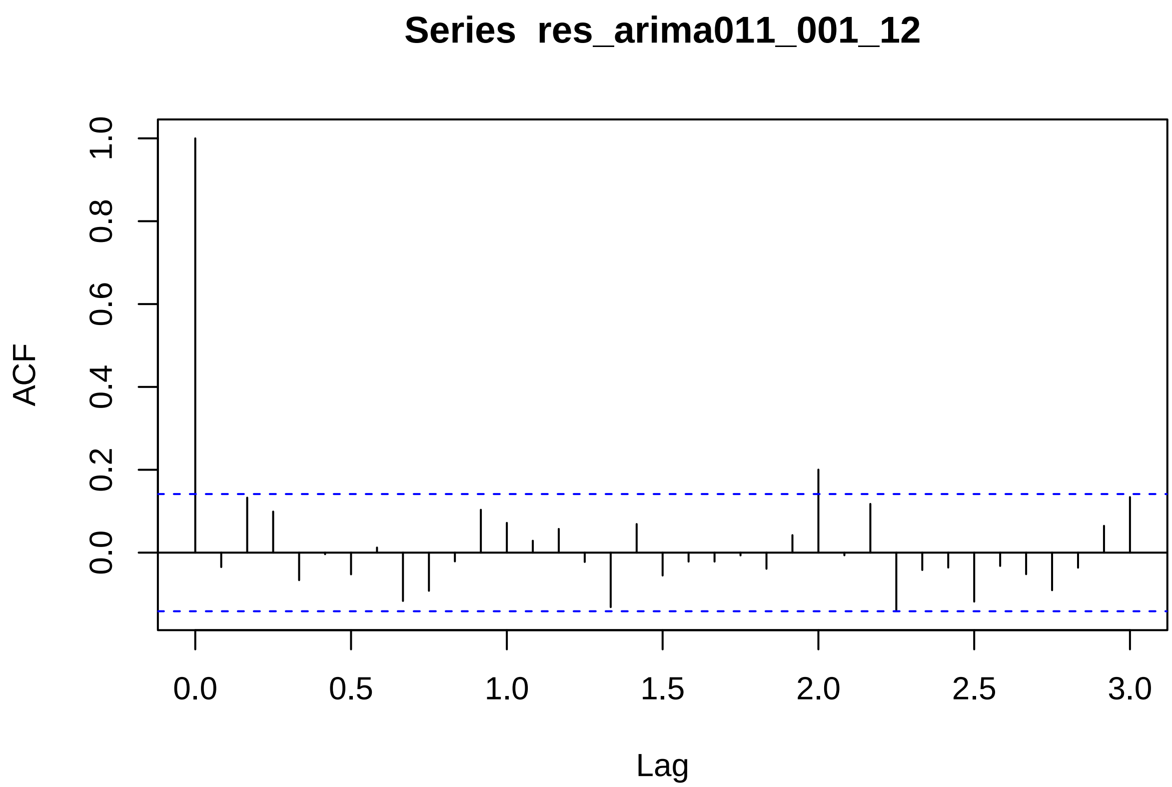


It appears that the seasonal lags in the PACF() plot appear to still be significant. It also appears that we have bounded the PACF values between (-0.15, 0.2). From meeting 6, I believe that this was the best ARIMA() fitting of them all. But I will try to reduce the MA() parameter by one.

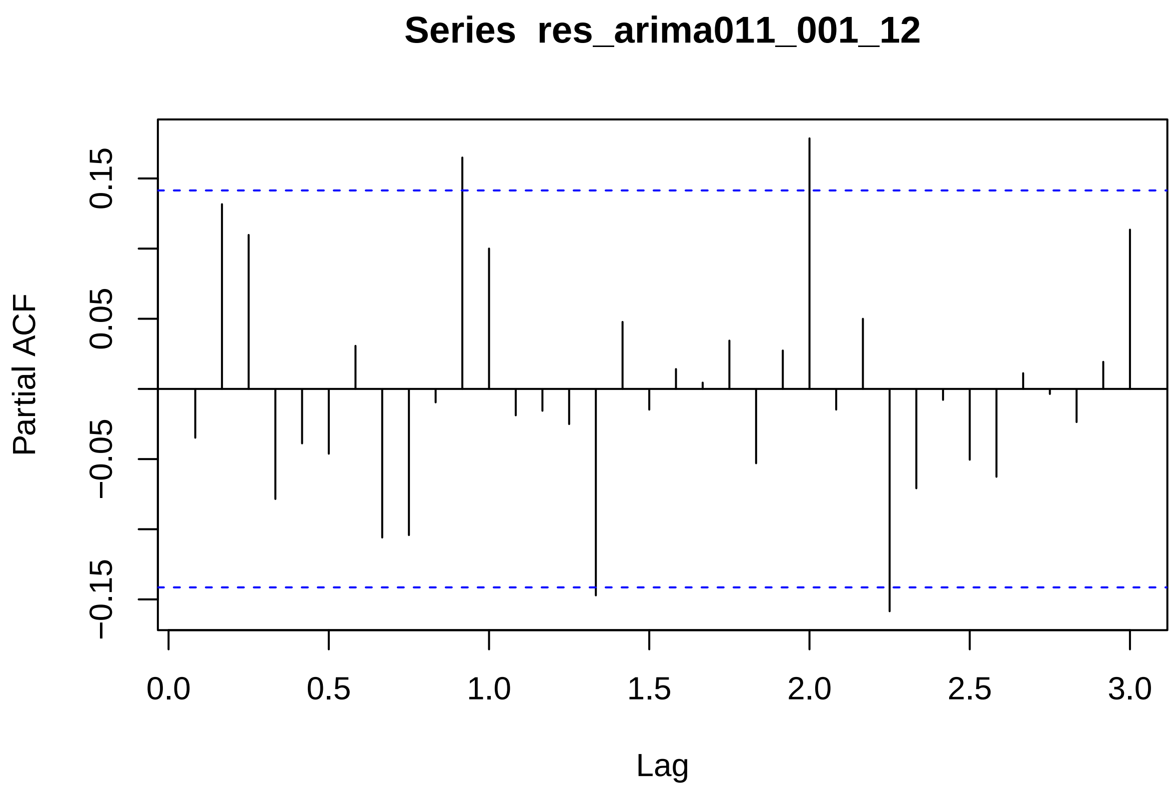
ARIMA(0,1,1)(0,0,1) with a constant and seasonal period = 12

ARIMA(0,1,1)$(0,0,1)_{12}$

```
zoo_arima011_001_12fit <- arima(DSNY_BX_zoo_ts,
                                order = 1 + c(0,1,1),
                                seasonal = list(order = c(0L,0L,1),
                                                period = 12))
res_arima011_001_12 <- zoo_arima011_001_12fit$residuals
acf(res_arima011_001_12, lag.max = 36)
```




```
pacf(res_arima011_001_12, lag.max = 36)
```



In the ACF Plot(), the seasonal lags continue to be significant, as in they pass the 95% threshold. A similar story with the PACF(), but the non-seasonal lags appear to be close to the 95% confidence levels. The PACF values are small as well, so it would not be a major deal breaker.

In conclusion

The models with the most promising ACF and PACF plots are ARIMA(2,1,2)$(0,0,0)_{12}$, ARIMA(0,1,2)$(0,0,1)_{12}$, and ARIMA(0,1,1)$(0,0,1)_{12}$. With my suspicion that the first model is overfitting the timeseries, and the other two models are returning adequate plots.