Eighth meeting notes

Daniel L.

4/08/2022

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()

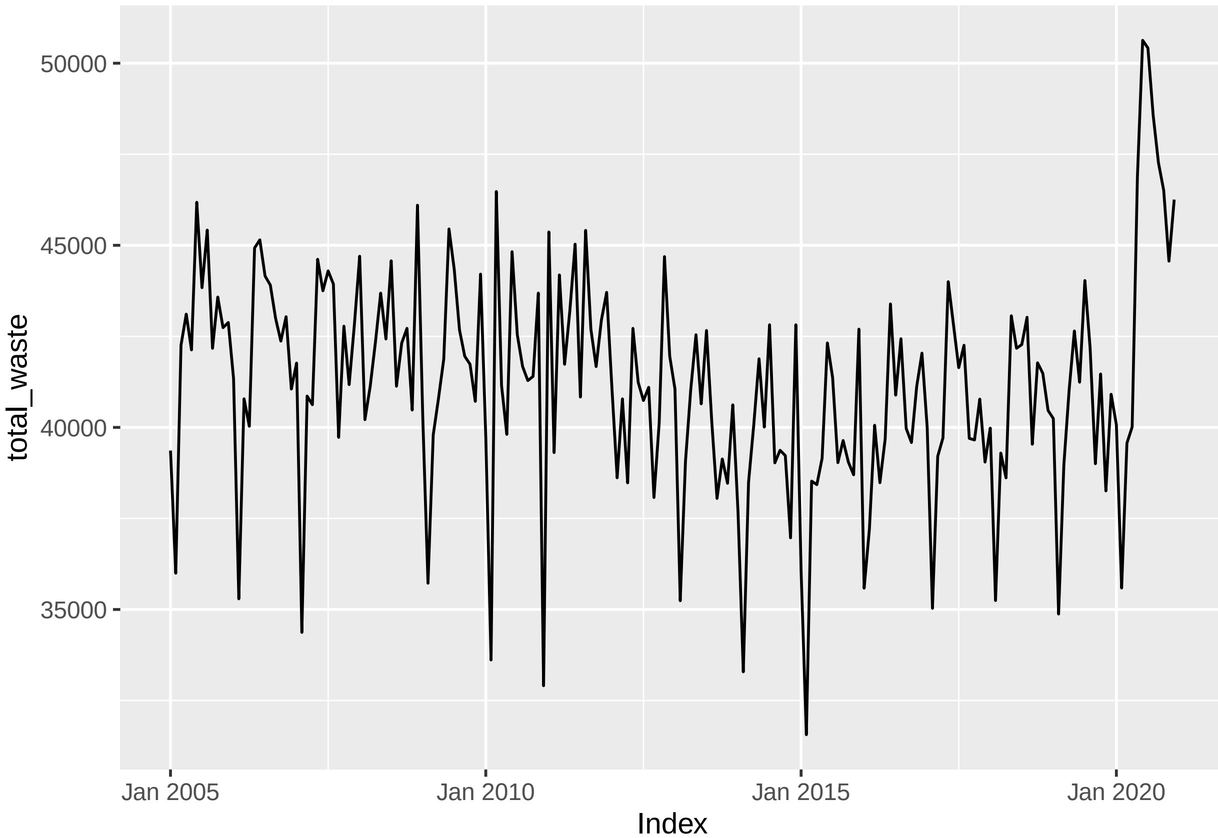
1. Look at the PACF to determine AR() or MA() terms

- To return the RMSE, I had to use the accuracy function from the fabletools package. Which meant that I had to reuse or create the ARIMA models in the same manner shown in the fpp3 book.

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

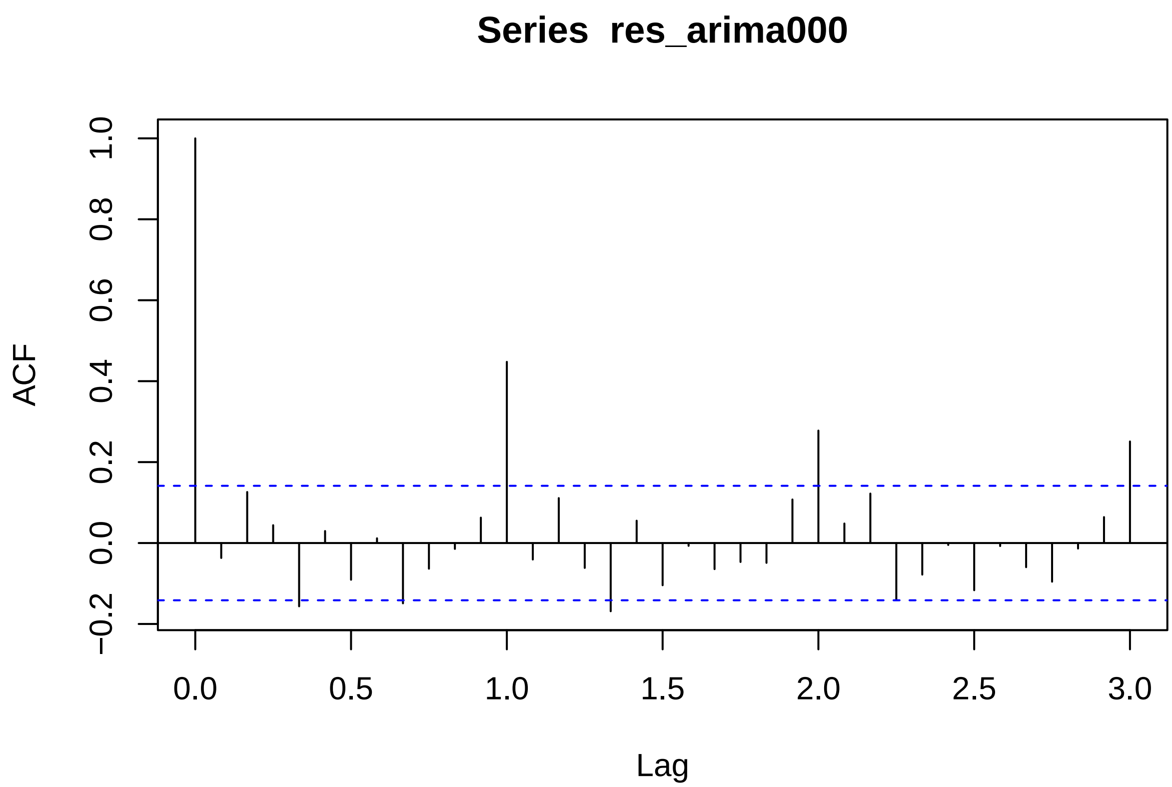
Here we create the ts and create the autoplot of the ts using the zoo package

```
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))
```

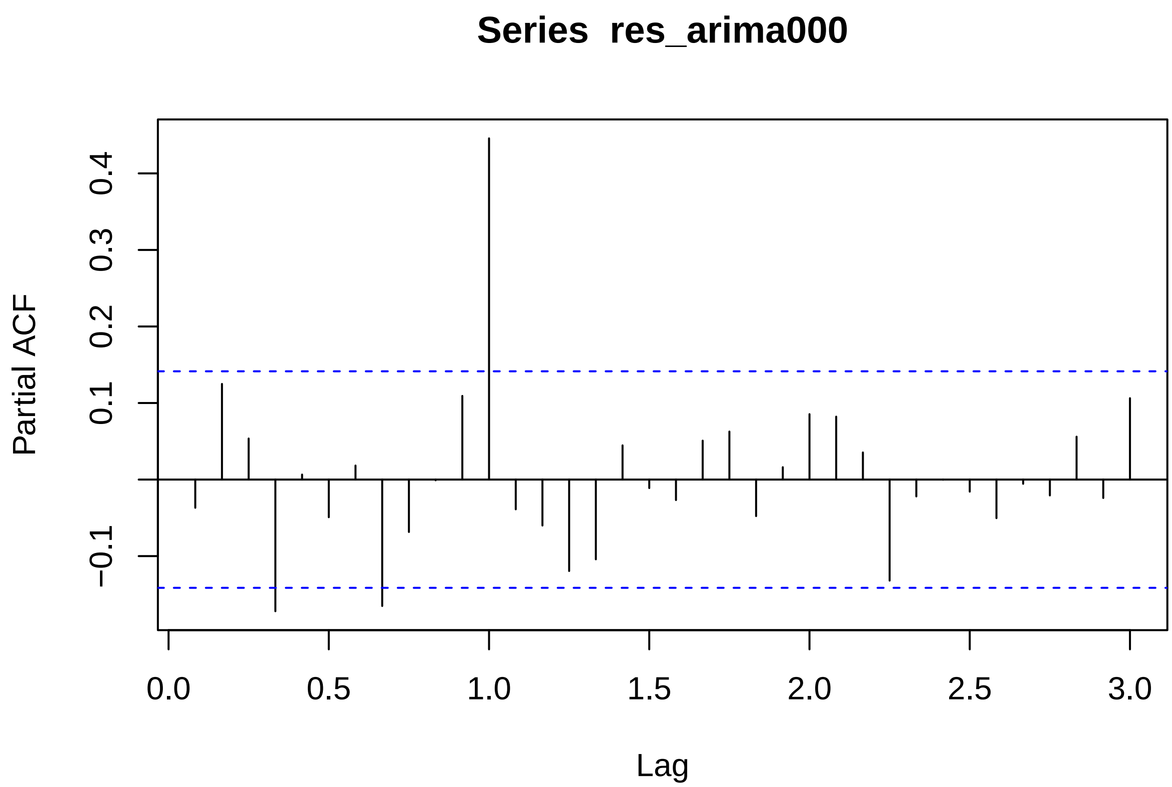


ARIMA(0,0,0) with constant

```
zoo_arima000_fit <- arima(DSNY_BX_zoo_ts, order = 1 + 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)
```




```
accuracy(bx_arima000_fit)[4]
```

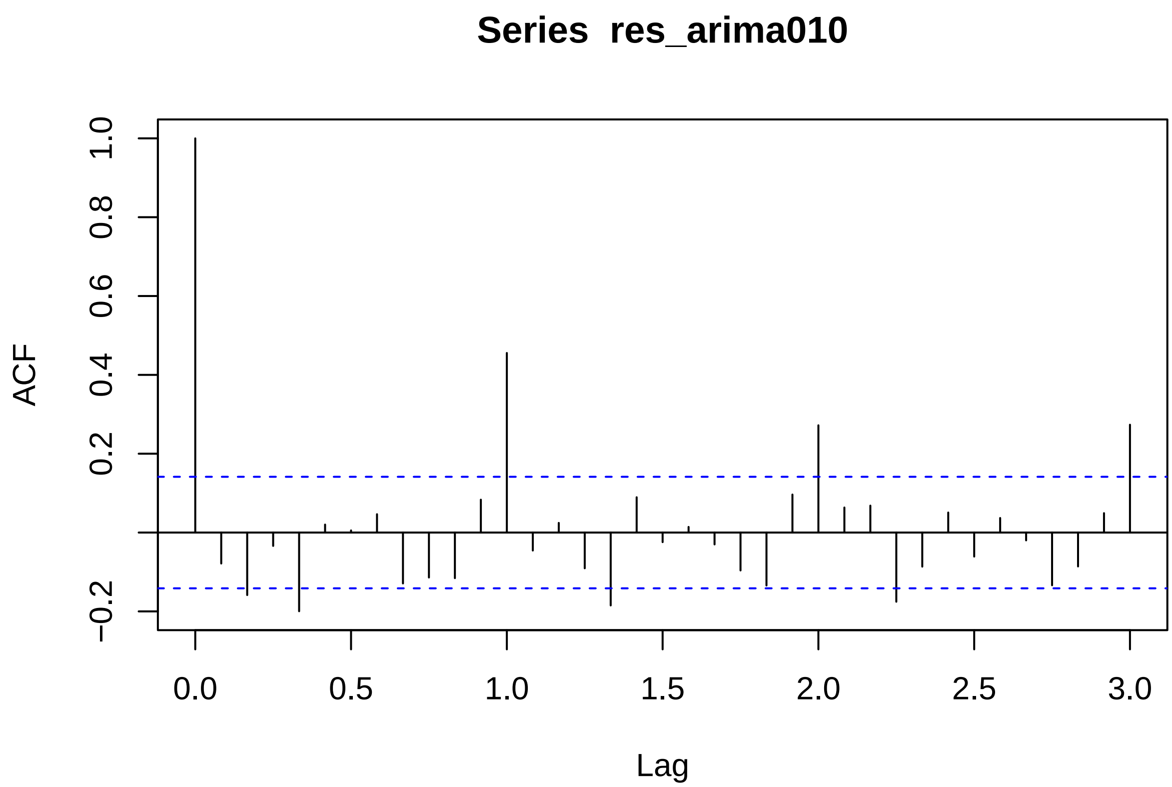
```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 2998.
```

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.

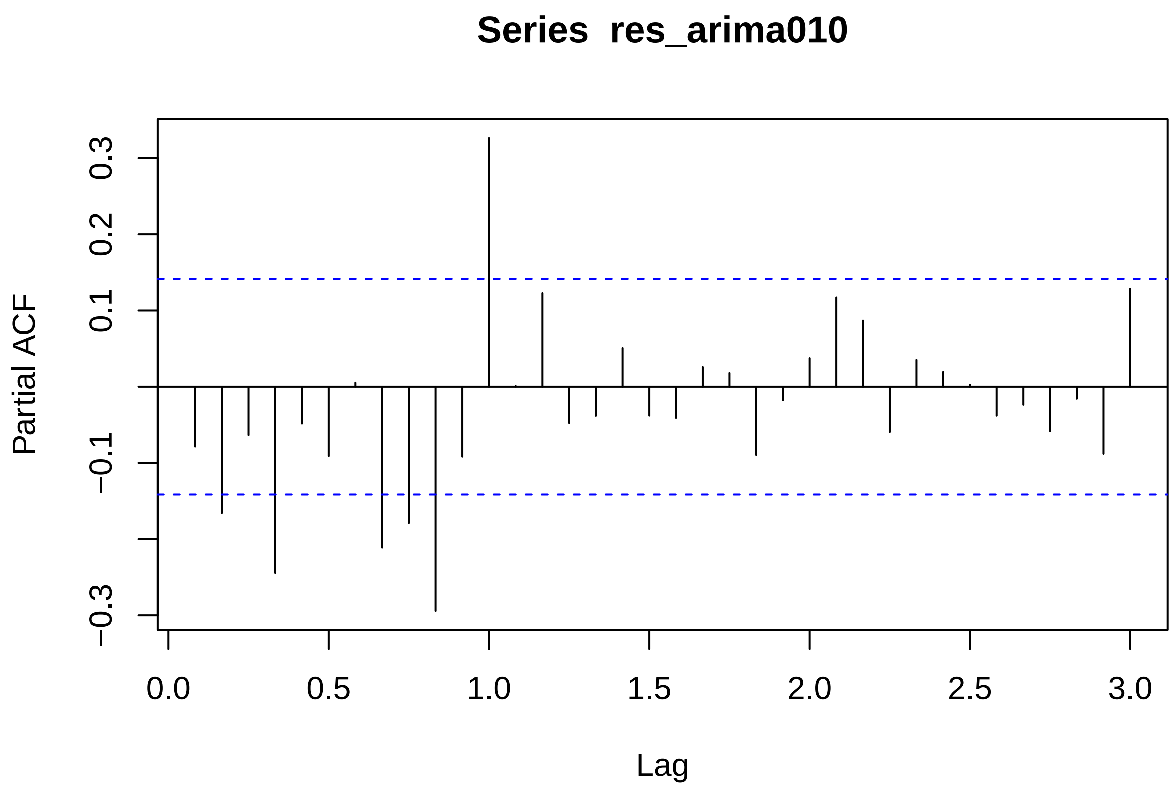
For ARIMA models with differencing, the differenced series follows a zero-mean ARMA model. Documentation by DataCamp.

ARIMA(0,1,0) with 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)
```



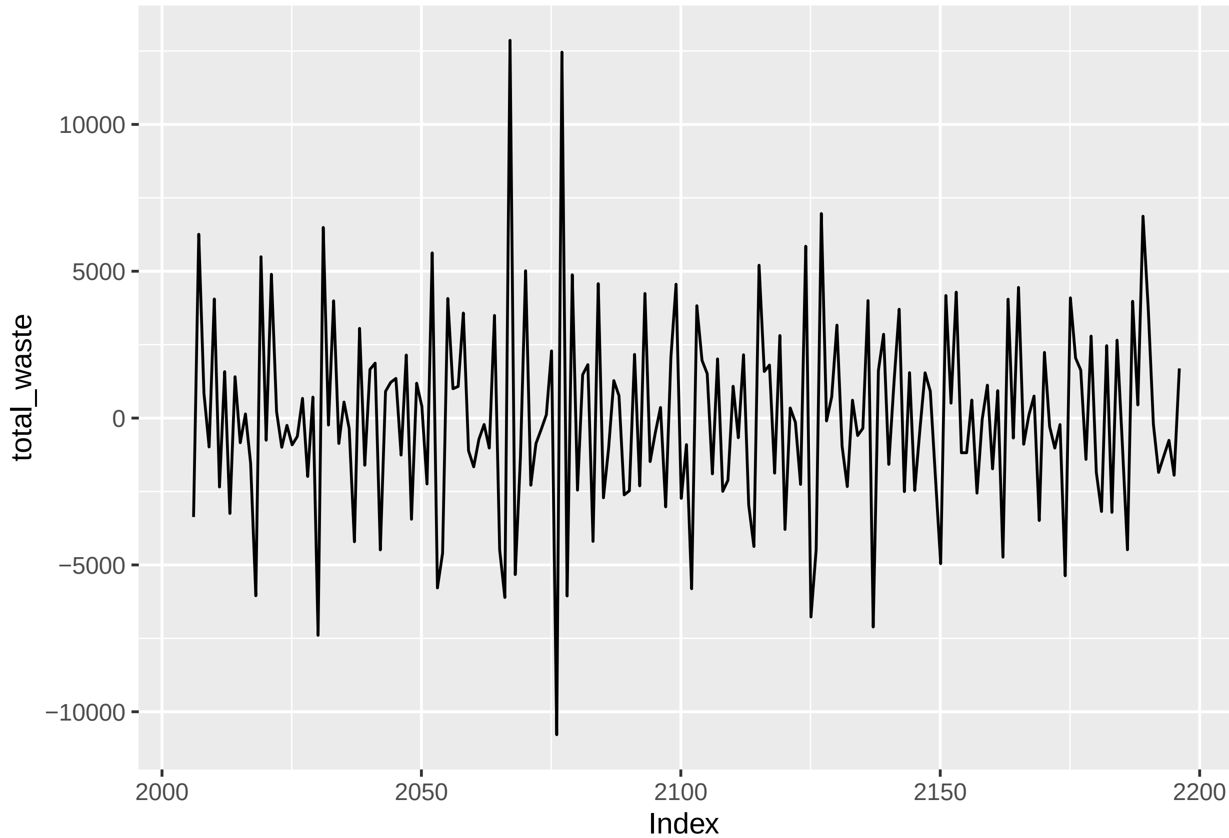

```
pacf(res_arima010, lag.max = 36)
```




```
accuracy(bx_arima010_fit)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 3317.
```

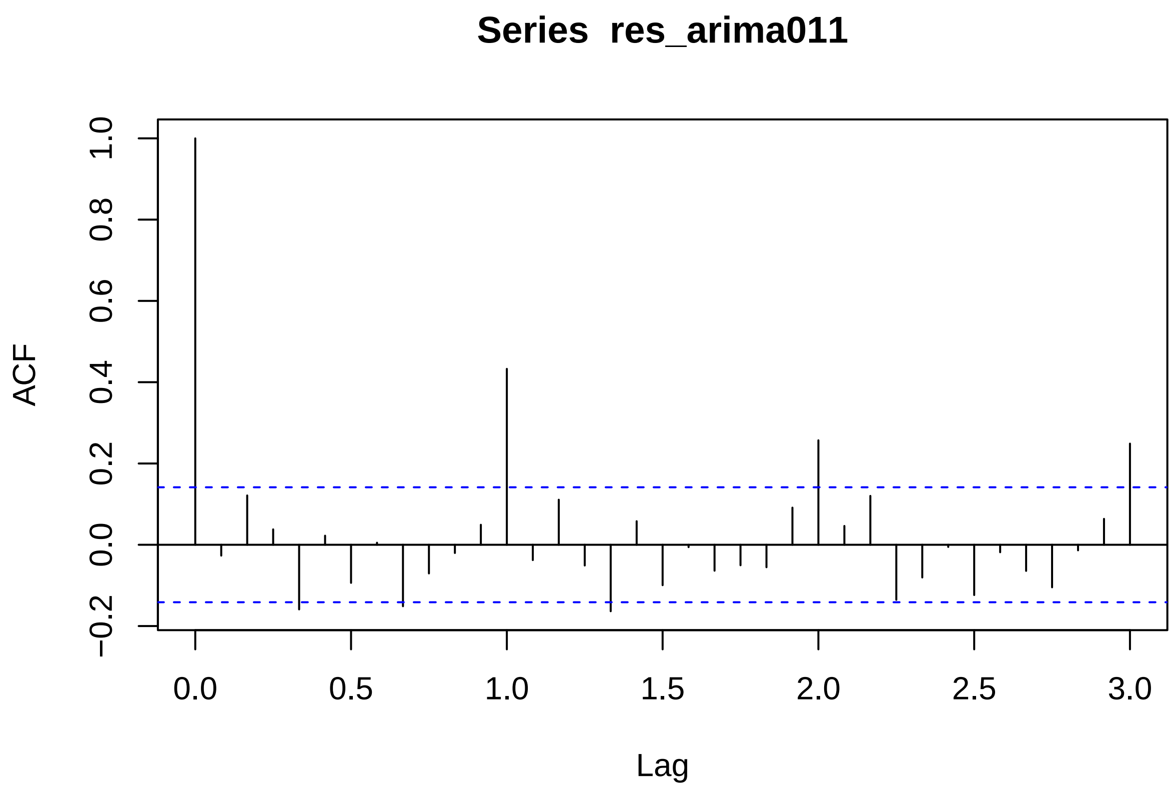
```
autoplot(as.zoo(difference(DSNY_BX_zoo_ts),1))
```



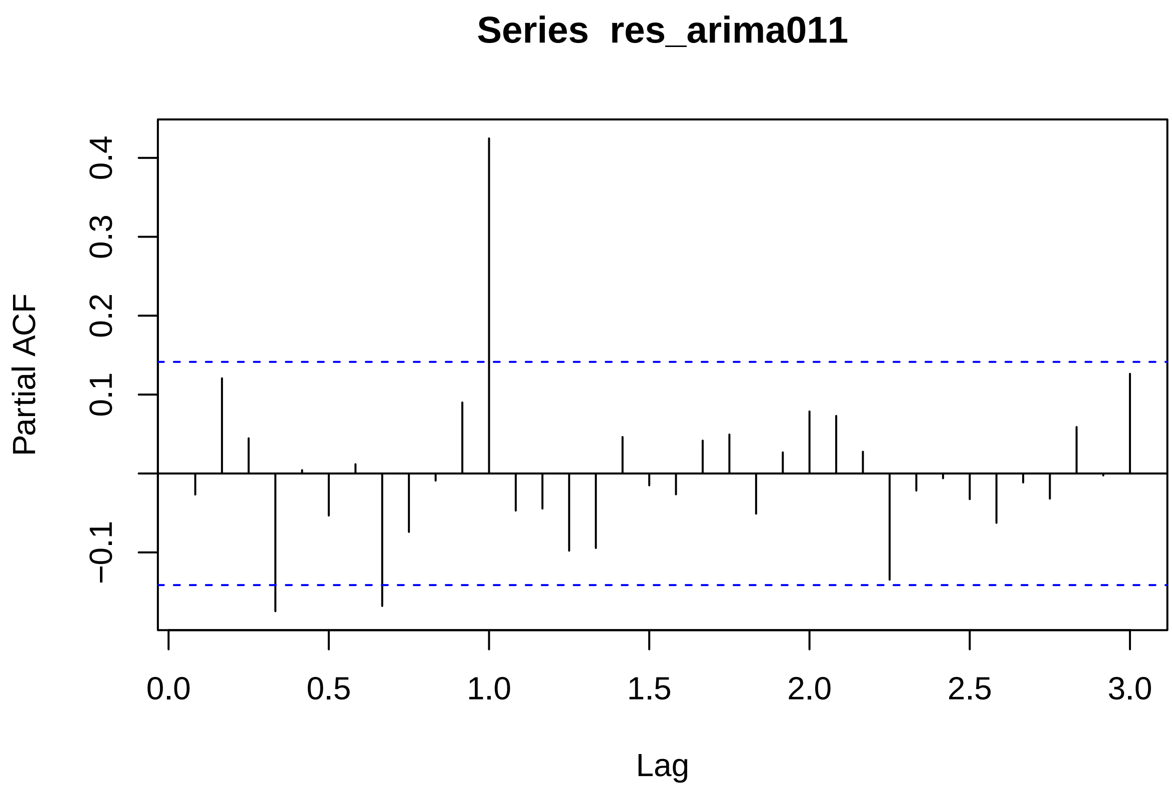
A constant term has been added with '1 + c(0,1,0)'. As discussed before, a negative ACF value is indicative of the need to add an MA() parameter. Here we see that lag 3 is no longer significant, perhaps we try to add MA(1) and MA(2) arguments to our **differenced** time series.

ARIMA(0,1,1) with a constant

```
bx_arima011_fit_cons <- bx_ts %>%
  model(arima011_constant = ARIMA(total_waste ~ 1 + pdq(0,1,1)))
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)
```



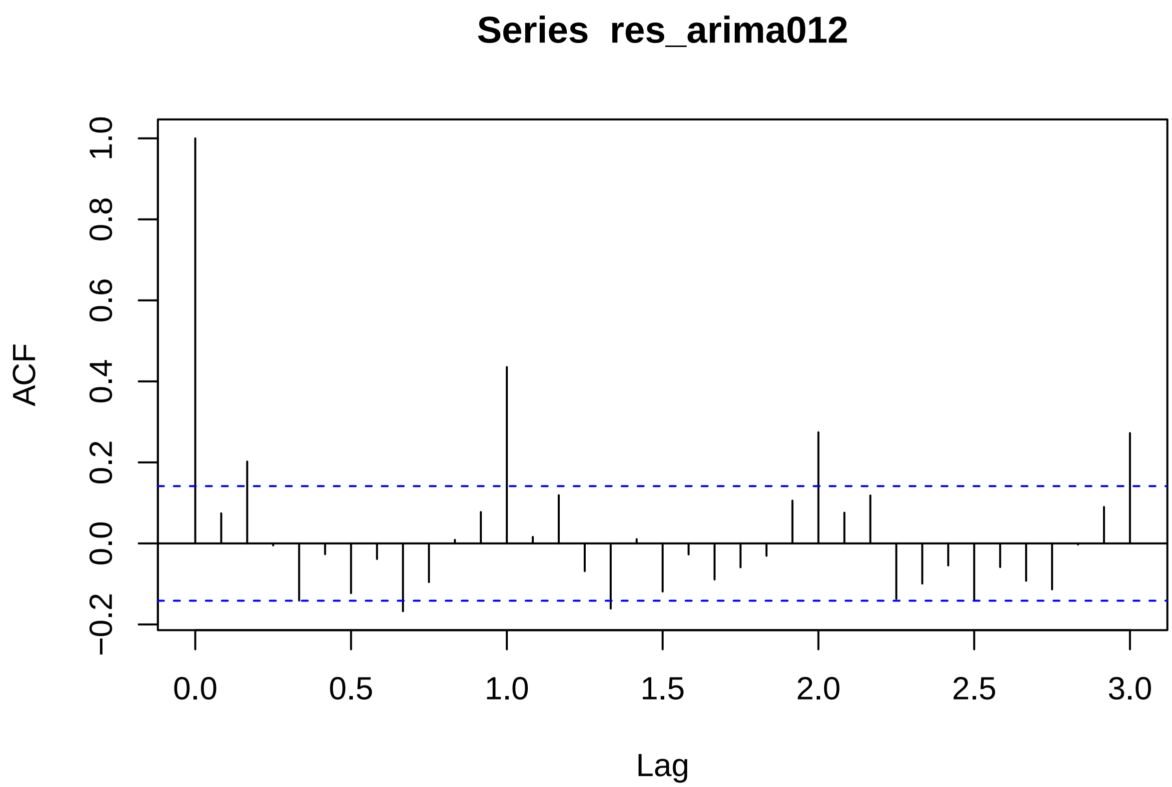

```
accuracy(bx_arima011_fit_cons)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 2818.
```

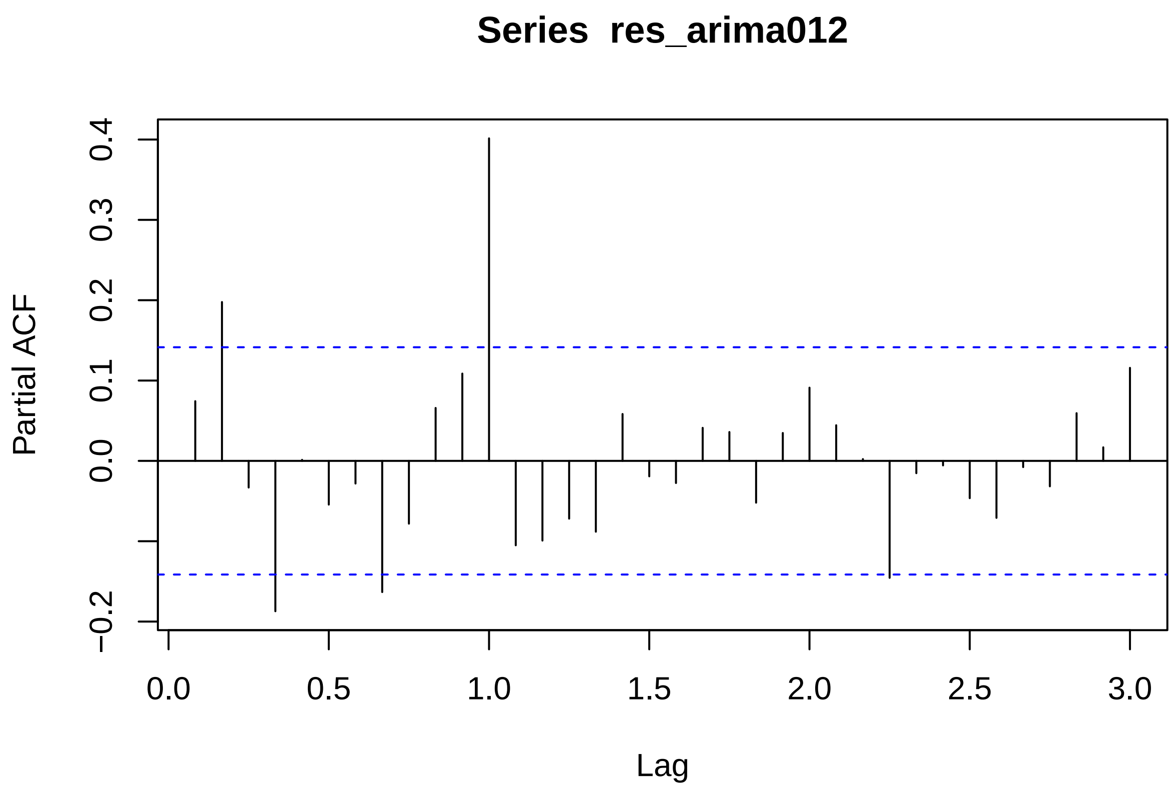
We can see that the RMSE of this model decreased, when compared to the ARIMA model with d = 1, and no other argument. There definitely should be a seasonal argument, as we see a seasonal pattern of significant lags at each integer lag.

ARIMA(0,1,2) with a constant

```
bx_arima012_cons_fit <- bx_ts %>%
  model(arima012_cons = ARIMA(total_waste ~ 1 + pdq(0,1,2)))

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)
```




```
accuracy(bx_arima012_cons_fit)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 2765.
```

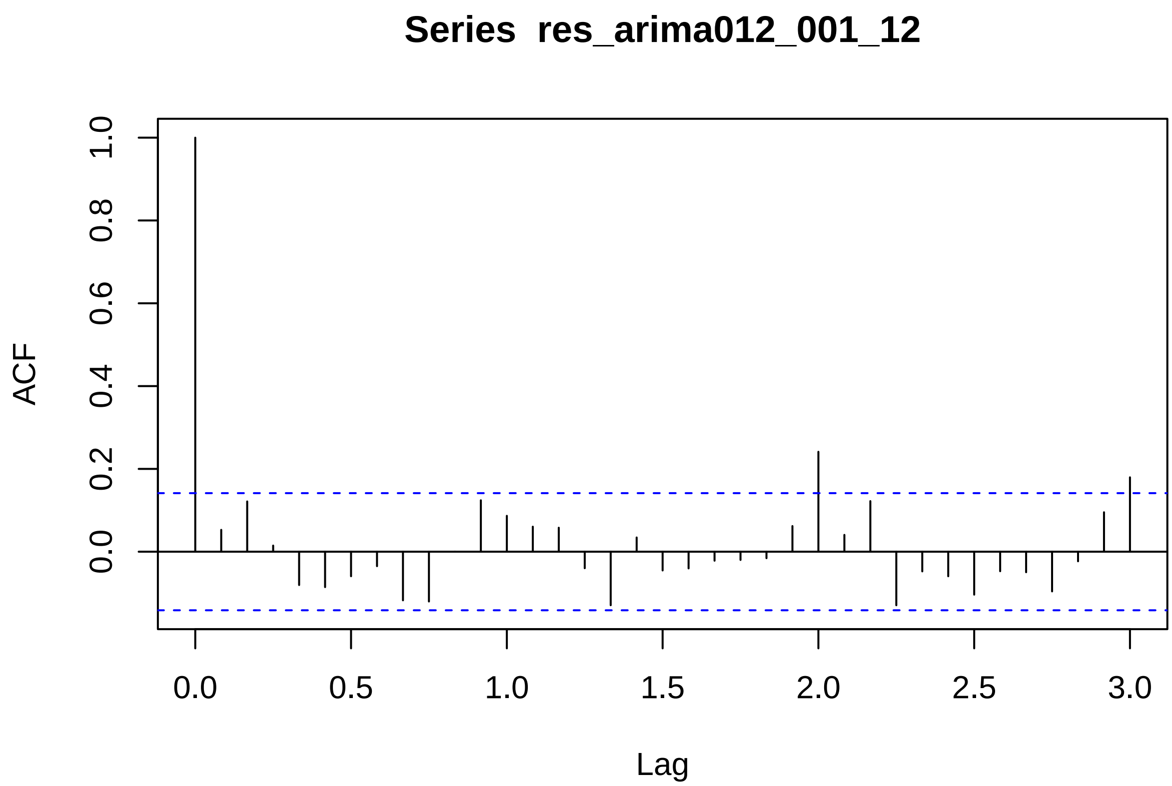
The RMSE continues to decrease. Lags (2,4,8) are significant in the PACF(). I am not sure if the lag 2 value of the PACF() is encouraging us to add an AR(2) argument. I would believe that this would remove any progress we have made. For now, I will attempt an MA(1) model. I also think it is best to address the seasonal lags.

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

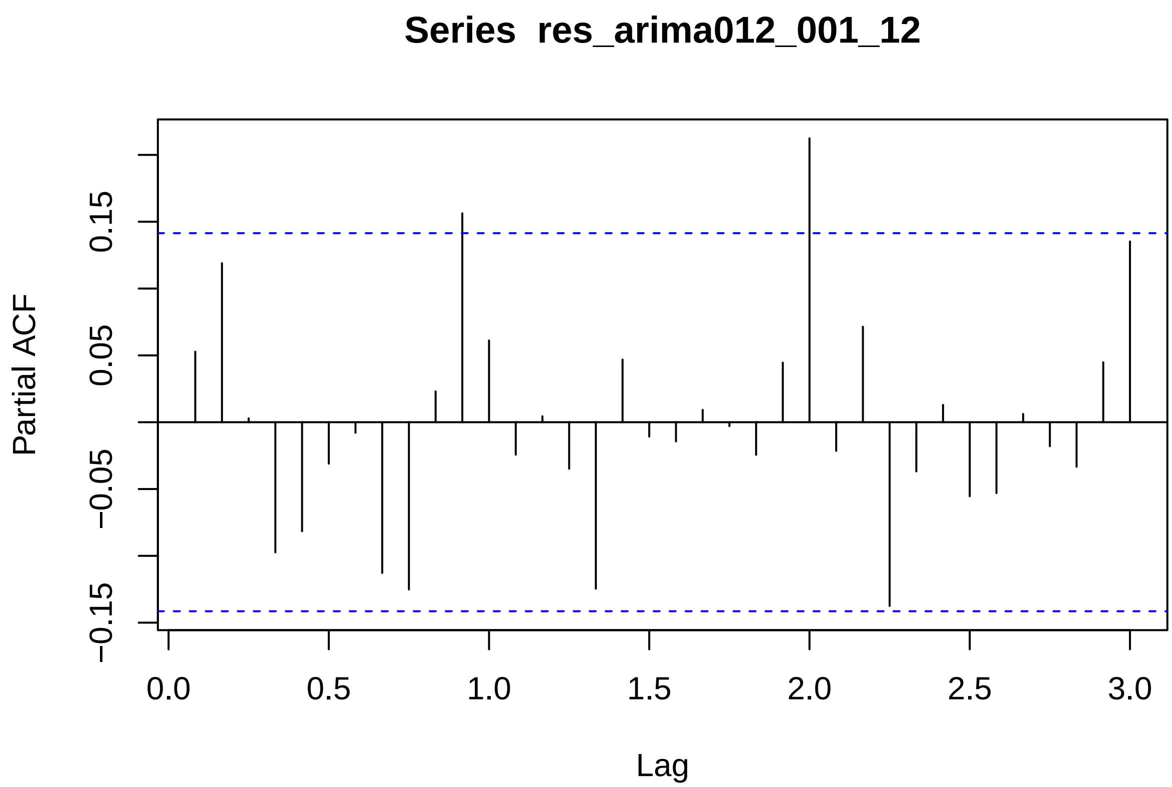
ARIMA(0,1,2)$(0,0,1)_{12}$

```
bx_arima012_cons_001_12 <- bx_ts %>%
  model(bx_arima012_001_12 = ARIMA(total_waste ~ 1 +
                                    pdq(0,1,2) +
                                    PDQ(0,0,1, period = 12)))

zoo_arima012_001_12fit <- arima(DSNY_BX_zoo_ts,
                                order = 1 + c(0,1,2),
                                seasonal = list(order = c(0L,0L,1),
```

```r
                                                    period = 12))

res_arima012_001_12 <- zoo_arima012_001_12fit$residuals
acf(res_arima012_001_12, lag.max = 36)
```




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




```
accuracy(bx_arima012_cons_001_12)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 2471.
```

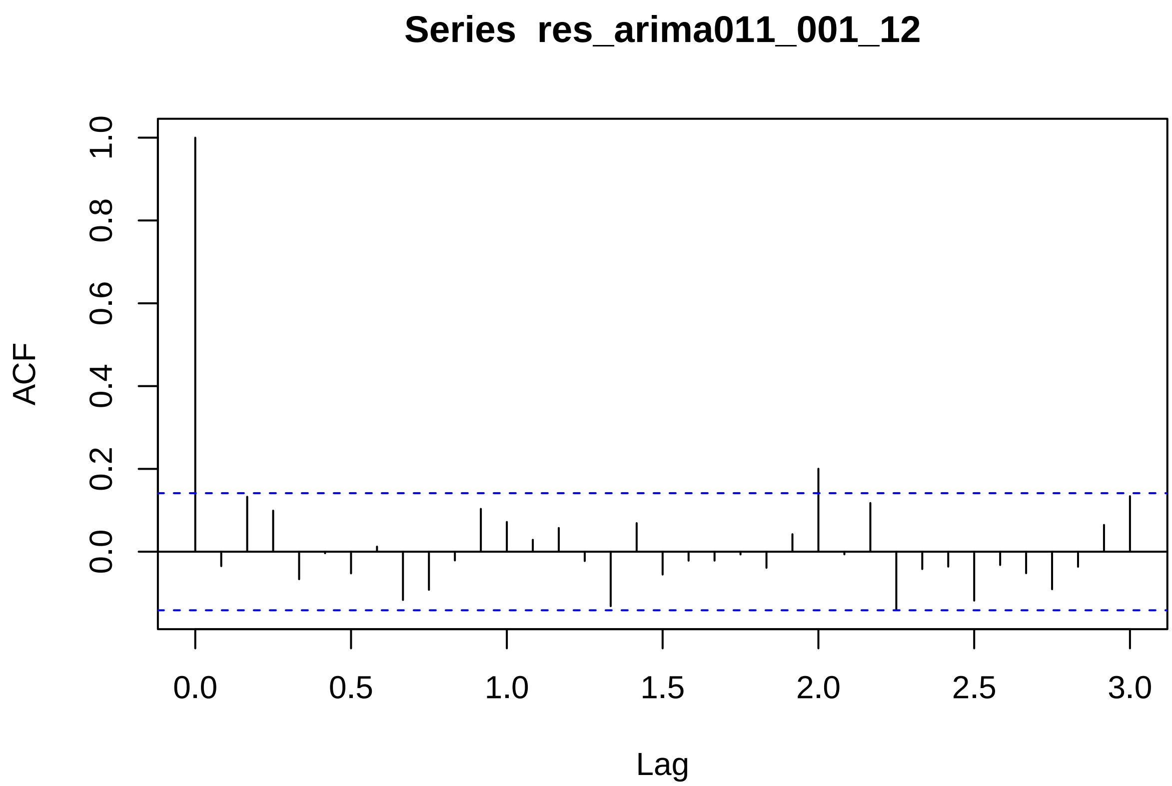
It appears that the seasonal lags in the PACF() plot appear to still be significant. It also appears that we have bounded the PACF values between (-0.15, 0.2). The RMSE is now the lowest of them all. From meeting 6, I believe that this was the best ARIMA() fitting of them all. But I will try to reduce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

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

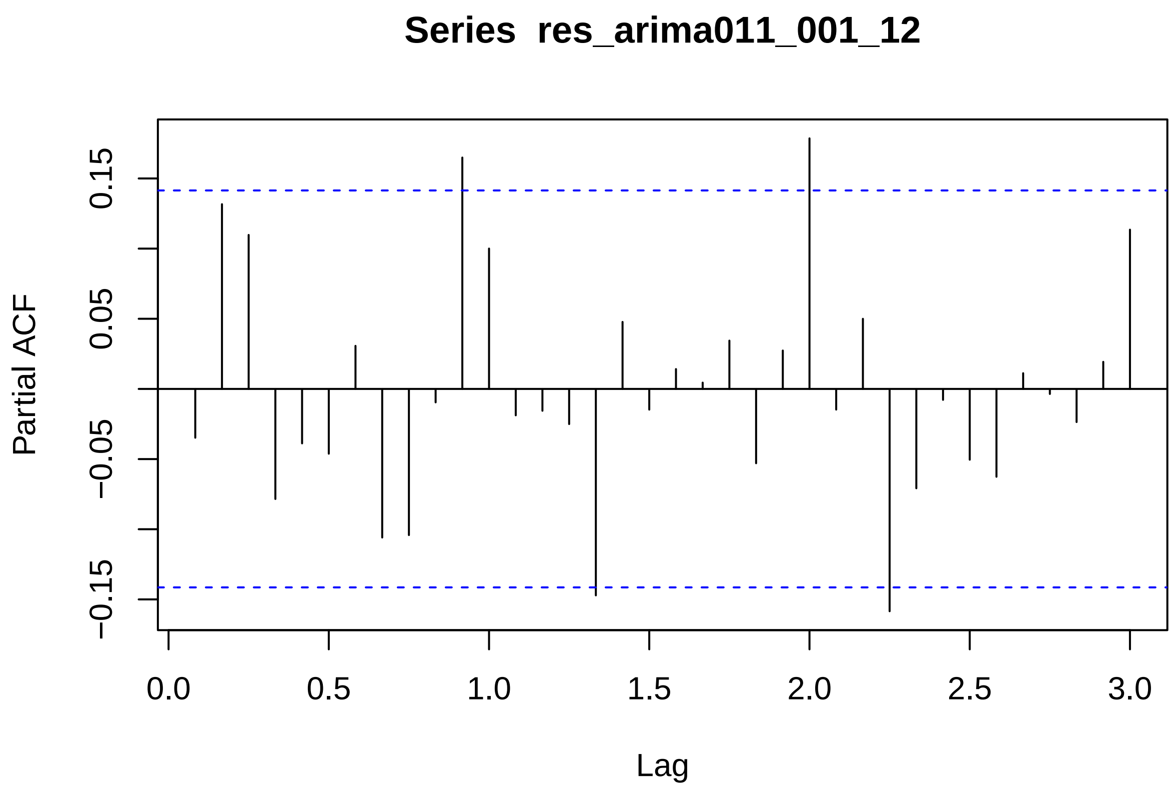
ARIMA(0,1,1)(0,0,1)$_{12}$

```
bx_arima011_cons_001_12 <- bx_ts %>%
  model(bx_arima011_cons_001_12 = ARIMA(total_waste ~ 1 +
                                        pdq(0,1,1) +
                                        PDQ(0,0,1, period = 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)
```




```
accuracy(bx_arima011_cons_001_12)[4]
```

```
## # A tibble: 1 x 1
##    RMSE
##   <dbl>
## 1 2486.
```

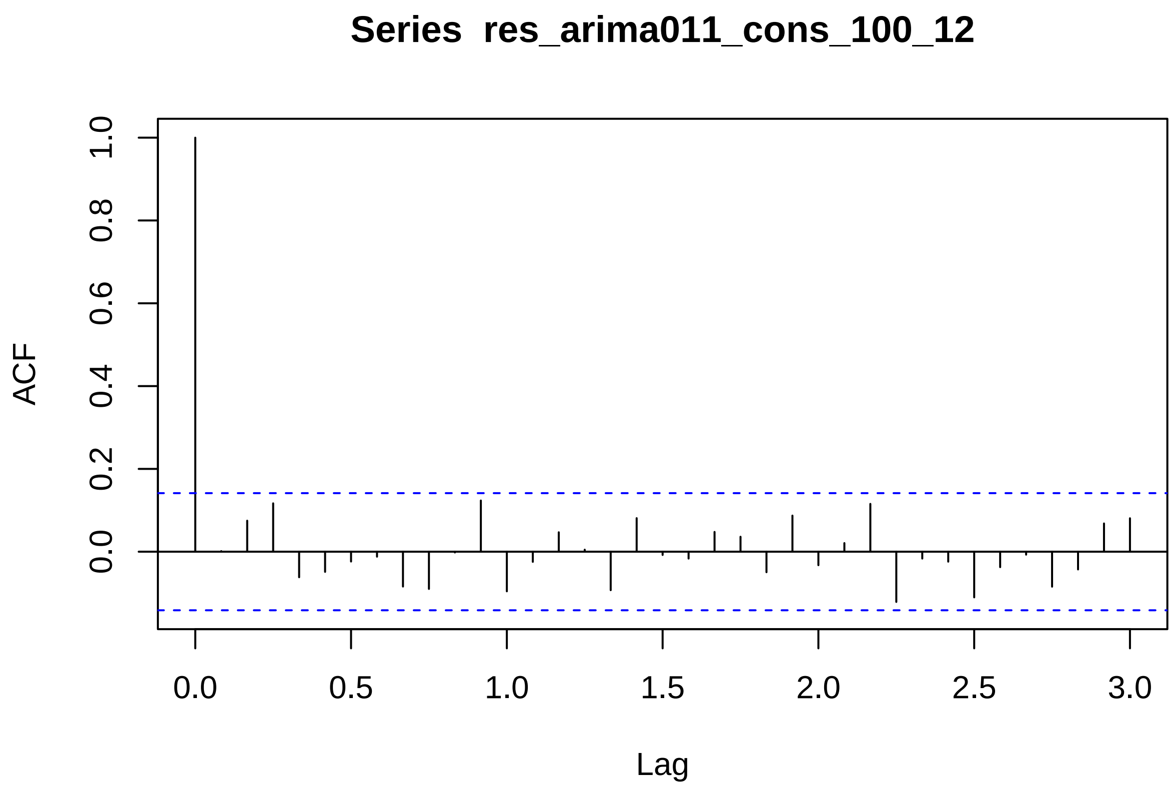
In the ACF Plot(), the seasonal lags continue to be significant, as in they pass the 95% threshold. A similar story with the PACF(), but the non-seasonal lags appear to be close to the 95% confidence levels. The PACF values are small as well, so it would not be a major deal breaker. With our period = 12, and seasonal lags are positive. This indicates that it is best to add an AR() seasonal argument.

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

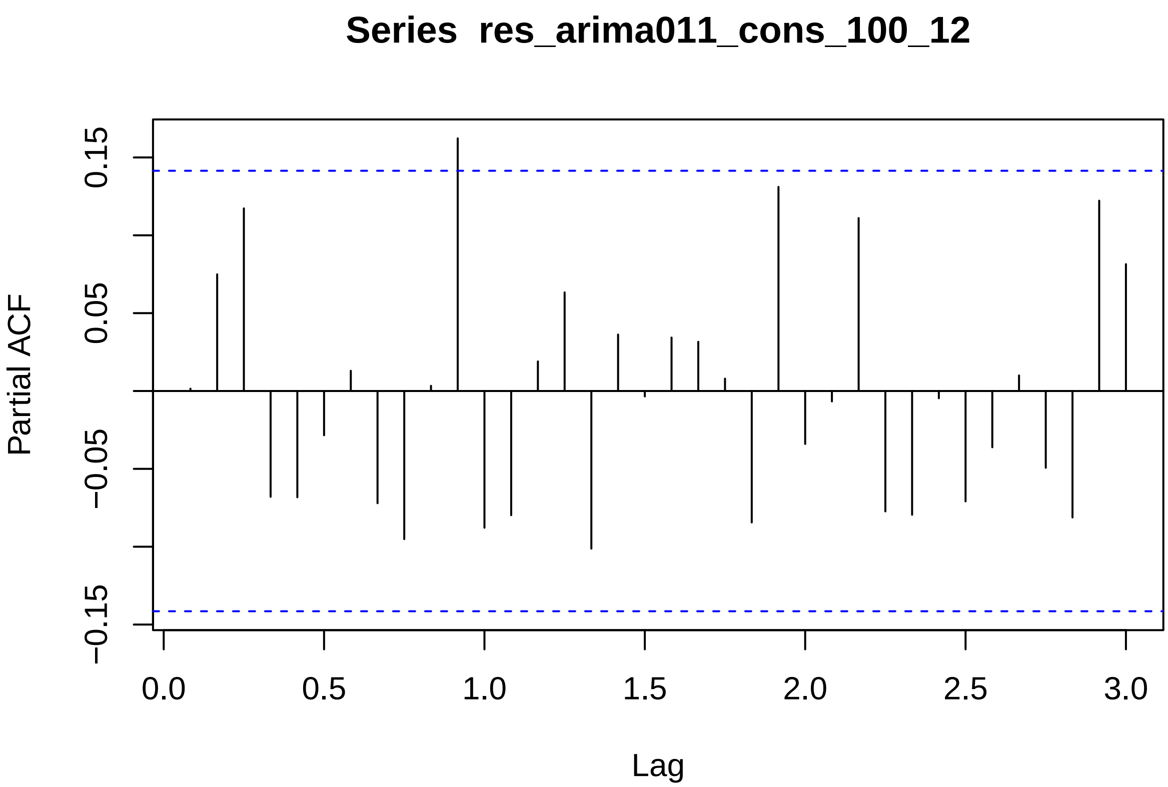
ARIMA(0,1,1)$(1,0,0)_{12}$

```
bx_arima011_cons_100_12 <- bx_ts %>%
  model(bx_arima011_cons_100_12 = ARIMA(total_waste ~ 1 +
                                        pdq(0,1,1) +
                                        PDQ(1,0,0, period = 12)))

zoo_arima011_cons_100_12fit <- arima(DSNY_BX_zoo_ts,
                                order = 1 + c(0,1,1),
                                seasonal = list(order = c(1,0L,0L),
```

```
                                              period = 12))

res_arima011_cons_100_12 <- zoo_arima011_cons_100_12fit$residuals
acf(res_arima011_cons_100_12, lag.max = 36)
```




```
pacf(res_arima011_cons_100_12, lag.max = 36)
```




```
accuracy(bx_arima011_cons_100_12)[4]
```

```
## # A tibble: 1 x 1
##     RMSE
##    <dbl>
## 1 2341.
```

This model returns the lowest RMSE. The seasonal lags in th PACF are no longer significant. All non-seasonal lags, except for lag = 11, are insignificant.

In conclusion

The models with the most promising ACF and PACF plots, and RMSE are ARIMA(0,1,1)(1,0,0)$_{12}$, and ARIMA(0,1,1)(0,0,1)$_{12}$.