

IE MBD 2020

Forecasting Time Series

Homework 2

11 October 2020

Team F:

Allison Black

Deborah Cheng

Guillermo Germade

Eun Suk Hong

Vasileios Sagiannos

Roberto Picon

# Introduction

Using the Box-Jenkins methodology, we identified seven linear time series models for the quarterly earnings per share of the Coca-Cola Company from the first quarter of 1983 to the third quarter of 2009. We identified our models using the entire available sample. Out of the seven models identified, we chose two (highlighted in green) for our forecasting. We left out the last 24 real values to compare both of the models in terms of forecasting. Because this is quarterly data, we used seasonal lags with values of 4, 8, 12, and so on.

# Model 1

1. Ensuring stationarity

1. We plotted the data (see Figure 1). We saw that the data was not stationary in the mean and therefore required transformations.
2. We chose to experiment with 20, 40, and 60 lags.
3. We plotted the ACF and PACF. At 20 lags and 40 lags, we did not see the sinusoidal pattern (the cycle) in the ACF; we only saw this pattern at 60 lags. Since the formal test will NOT understand the cyclical pattern (it will tell us we need a difference because of the slow decay to zero in the ACF), we can try modeling *without* taking this difference.
4. After we performed the formal tests, we saw that we needed 1 regular difference and 1 seasonal difference. We started by only taking the seasonal difference and *not* the regular difference (because of the sinusoidal pattern).
5. We modeled the data according to the formal test. Taking 1 seasonal difference we plotted the residuals. We still did not have stationarity in the mean (see Figure 2). After computing the number of regular and seasonal differences again, the formal test told us that we did not need to take any more differences, but we took 1 regular difference to see if the residuals appear to be more stationary. After including the regular difference, we achieved stationarity in the mean (see Figure 3). We did not take any more differences because is proved in econometrics, that with no more than 2 differences, we can capture stochastic trends.

2. Identification

1. Now that our data was stationary, we checked the ACF and PACF of the residuals (see Figure 4). Looking at our PACF, we saw the first and fourth lags are out of bounds, so we identified p = 1 (regular AR) and P = 1 (seasonal AR).

3. Estimation

1. Option 1: P = 1: (0,1,0)x(1,1,0) s=4
   1. The coefficient is significant with a confidence interval of (-0.6604, -0.3079).
2. Option 2: (1,1,0) x (0,1,0) s=4.
3. Option 3: (0,1,0)x(1,1,0) s=4 with log

4. Diagnostic checking

1. We then plotted the ACF and PACF of the residuals for (0,1,0)x(1,1,0) s=4 (see Figure 5). Since lags 2 and 10 are just barely out of bounds, we visually considered this model white noise. After performing the Box test, we can verify that this model is white noise. However, performing the Shapiro test, we see that our data is *not* normally distributed.
2. The model (1,1,0) x (0,1,0) s=4, although significant, displays many lags out of bounds on both residuals plots of the ACF and PACF, indicating we do not have white noise and will therefore not choose this model.

5. Forecasting

1. We ran our predictions (see Figure 6) and did out of sample forecasting for the model. We took 24 real values out so that we are forecasting 83 observations. Because with 60 lags there is a break in the data (the 1998 dip), we are chose to use a rolling scheme for this model.
2. For the first period ahead, we got an MSFE value of 0.00168 and a MAPE value of 7.599%. For the second period ahead, we got an MSFE value of 0.00289 and a MAPE value of 8.943%.

## Model 1 plots

Figure 1: Original data

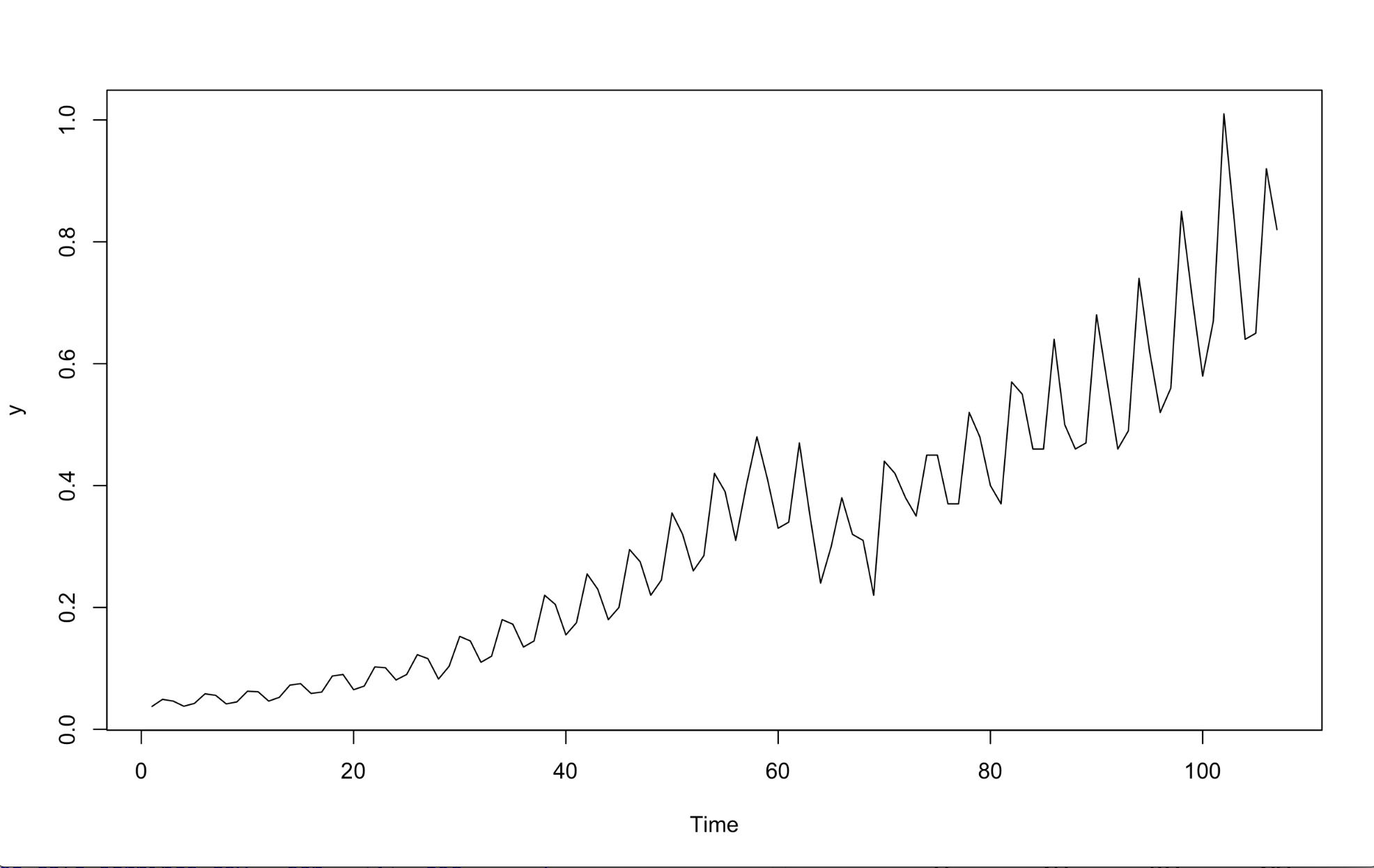


Figure 2: Plot of residuals after taking 1 seasonal difference

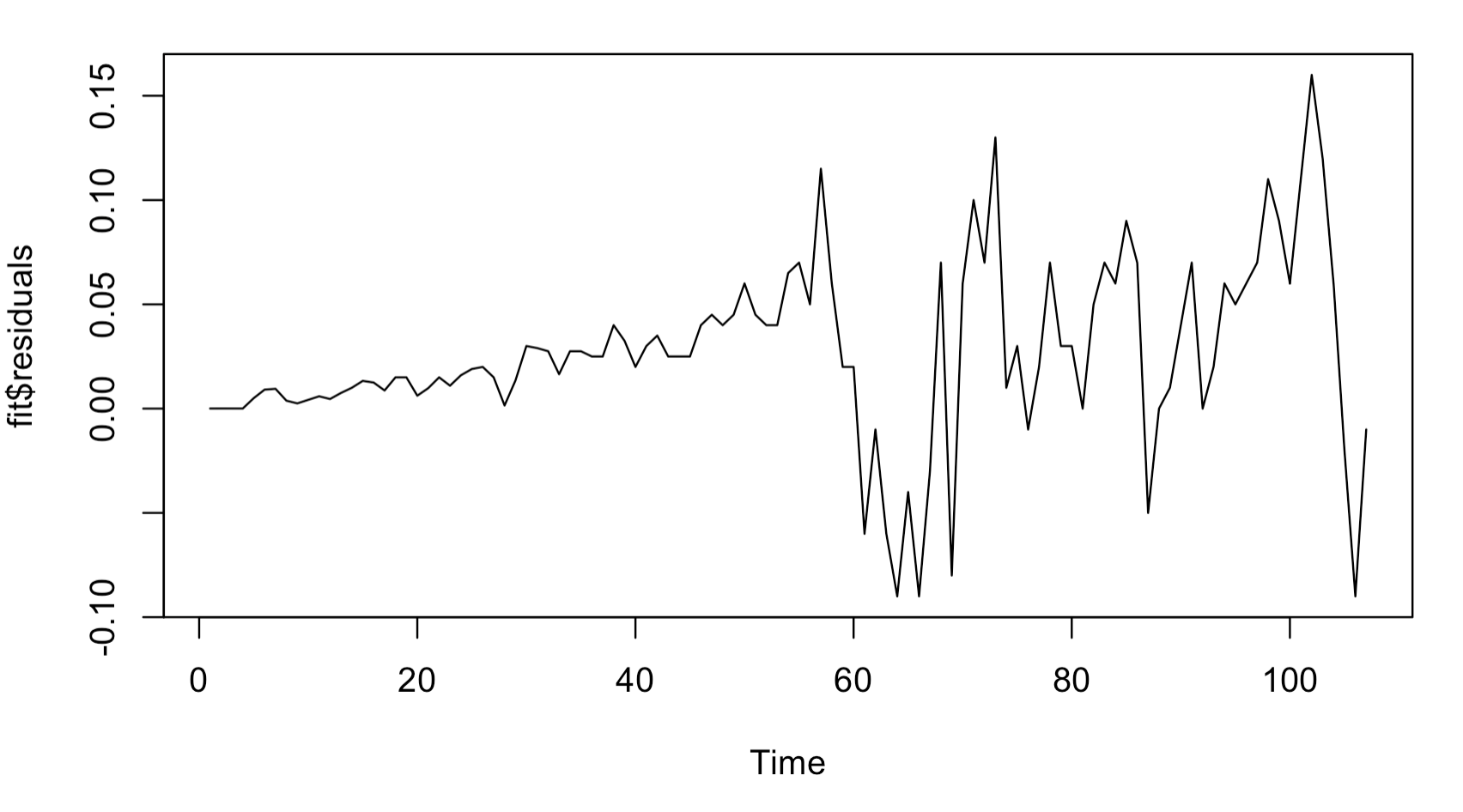


Figure 3: Plot residuals after taking 1 regular difference and 1 seasonal difference.

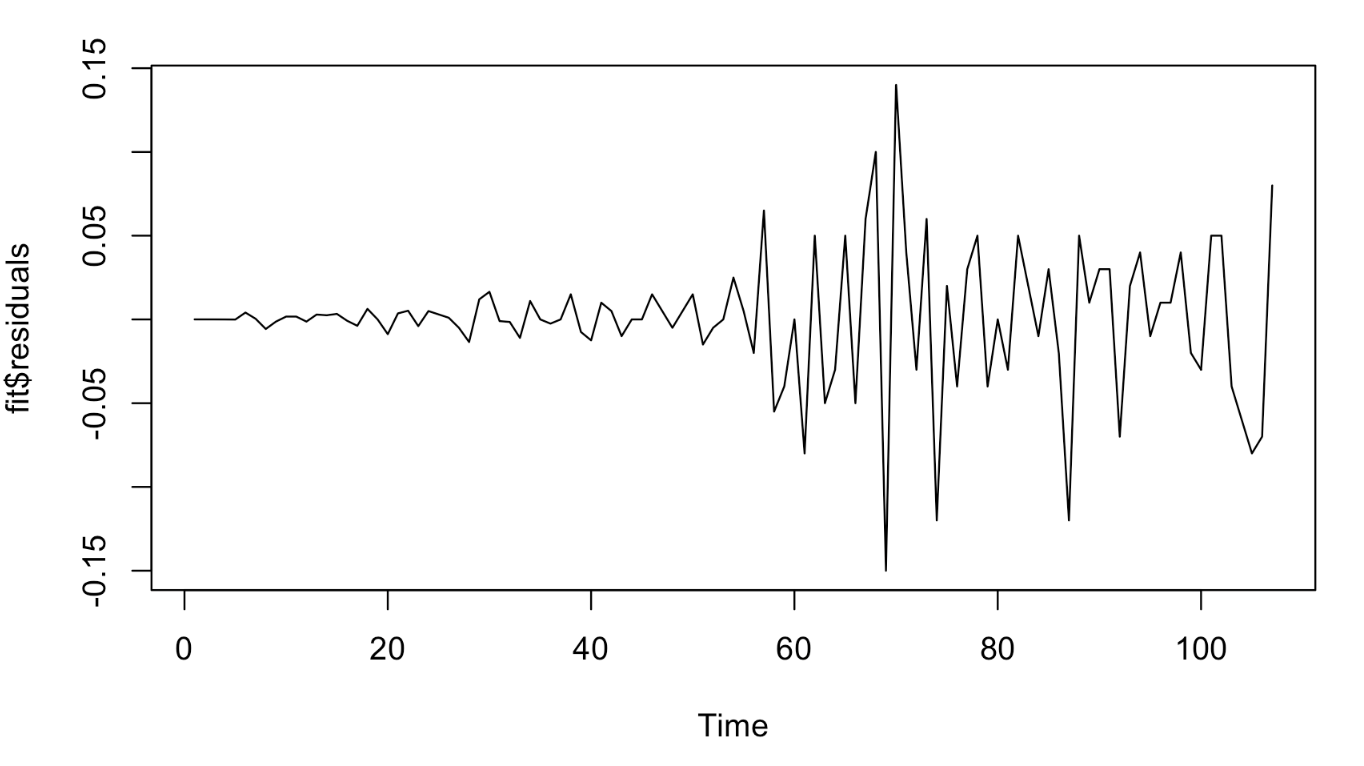


Figure 4: ACF and PACF of residuals for model (0,1,0) x (0,1,0) s=4

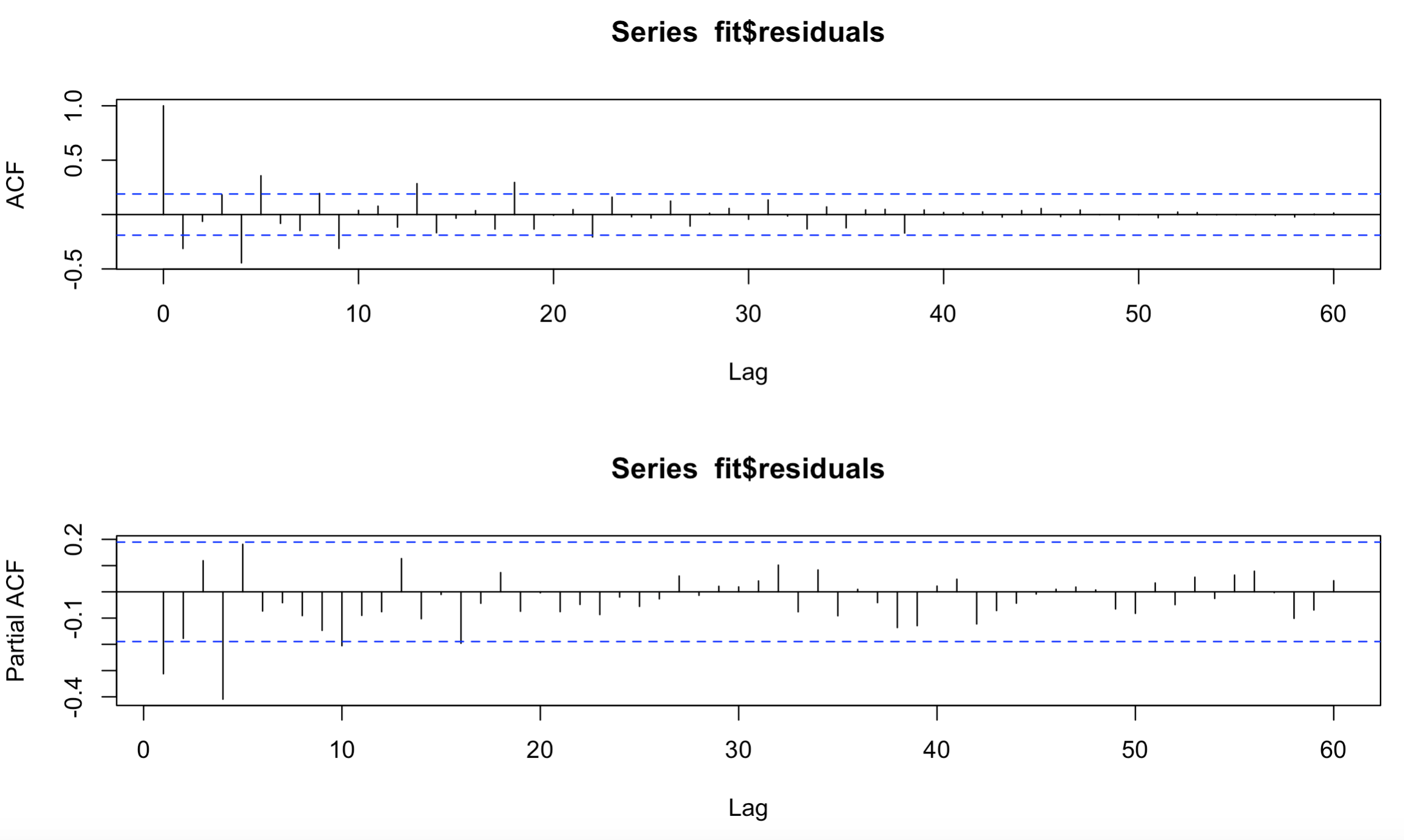




Figure 5: ACF and PACF of residuals for model (0,1,0) x (1,1,0) s=4

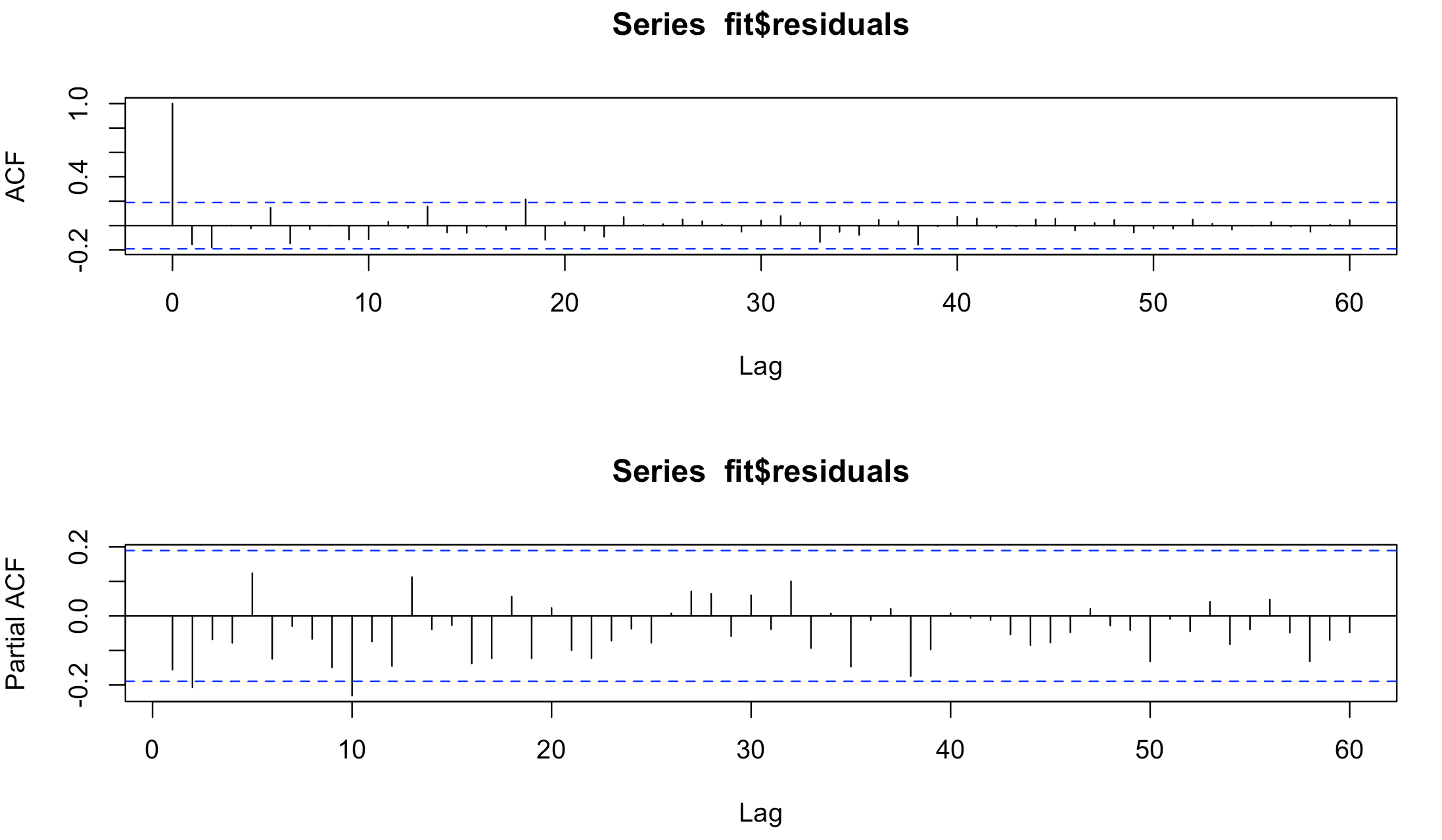
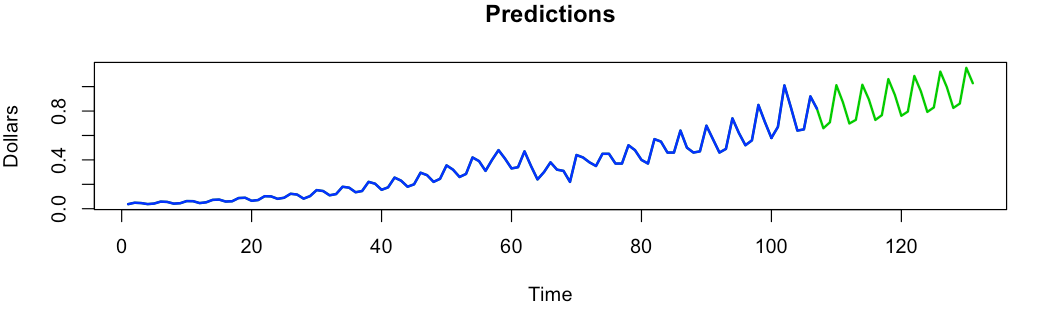


Figure 6 : Model 1 Predictions



# Model 2

1. Ensuring stationarity

1. Starting from the beginning, we saw that in order to achieve stationarity, we need to take differences. We changed the number of lags to 40 took 1 regular difference and 1 seasonal difference (same as model 1) and we got the same result: no more differences needed. However, we wanted to see what happened when we took 2 seasonal differences and no regular differences as well as 2 regular differences and no seasonal differences.
2. Taking 2 regular differences and no seasonal differences: the plot looked stationary in the mean (see Figure 7), but the formal test told us we still needed to take 1 seasonal difference. Since we did not want to take more than 2 differences (as mentioned in step 1.e for Model 1), we disregarded this.
3. Taking 2 seasonal differences and no regular differences: the plot seemed stationary in the mean (see Figure 8), but we performed the formal test to check. The formal test told us we did not need further transformation.

2. Identification

1. Now that our data was stationary, we checked the ACF and PACF of the residuals (see Figure 9). Looking at the ACF, the first, fourth, and ninth lags were out of bounds, so we identified q = 1, 4, 9 (regular MA) and Q = 1 (seasonal MA).
2. Looking at the PACF, the first, fourth, fifth, sixth, and eighth lags are out of bounds, so we identified p = 1, 4, 6, 8 (regular AR) and P = 1, 2 (seasonal AR).

3. Estimation

1. Option 1: We went for simple: P = 1: (0,0,0) x (1,2,0) s=4
   1. Coefficients are significant (-0.65554, -0.25766)
2. Option 2: To previous model added Q = 1: (0,0,0) x (1,2,1) s=4
   1. Coefficients are significant (-1.148, -0.744)
3. Option 3: We kept Q = 1, but removed P = 1: (0,0,0)x(0,2,1)s=4
   1. Coefficients are significant with a confidence interval of (-1.1197, -0.7704).
4. Option 3: We kept Q = 1, and added p = 5,: (5,0,0) x (0,2,1) s=4
   1. Coefficients are significant with a confidence interval of (0.1407, 0.5503).

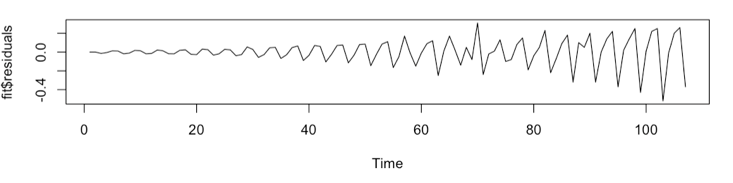
4. Diagnostic checking

1. After checking the residuals for options 1, 2 and 3 (see Figure 10), we saw many lags out of bounds in the ACF and PACF and therefore we did not have white noise. We did not use these models.
2. Option 4 (Figure 11) showed white noise, so this model seemed promising. After performing the Box test to double check, we verified that this model is white noise.
3. We performed the Shapiro test to check for normality and the data is not normally distributed; therefore, we do not have Gausian white noise.

5. Forecasting

1. We ran our predictions (see Figure 12) and we did out of sample forecasting for the model. We took 24 real values out and forecasted 83 observations. With 40 lags we no longer had the break in the data (from the 1998 dip) that we had in model 1. Therefore, we used the recursive scheme for this model.
2. For the first period ahead, we got an MSFE value of 0.0023 and a MAPE value of 8.319%. For the second period ahead, we got an MSFE value of 0.0030 and a MAPE value of 9.693%.

## Model 2 plots

Figure 7: Taking 2 regular differences (0,2,0) x (0,0,0) s=4

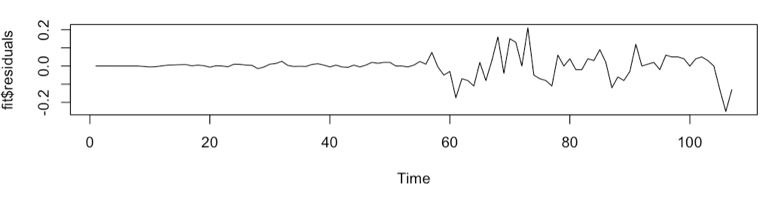
Figure 8: Taking 2 seasonal differences (0,0,0) x (0,2,0) s=4

Figure 9: ACF and PACF of the Residuals

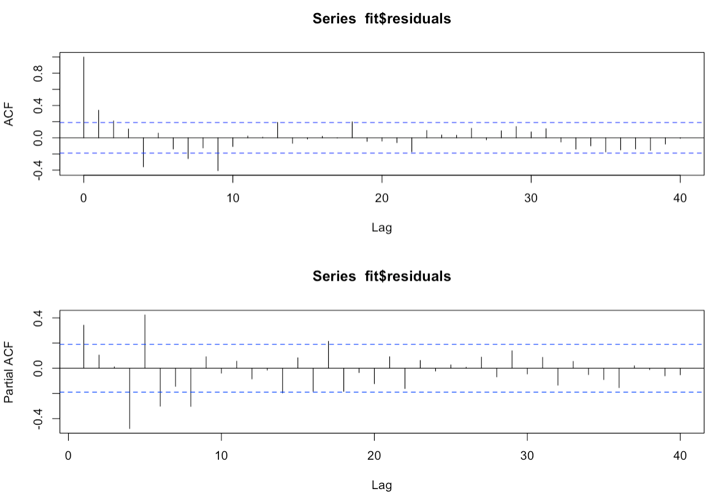


Figure 10: Option 3 ACF and PACF of Residuals (0,0,0)x(0,2,1)s=4

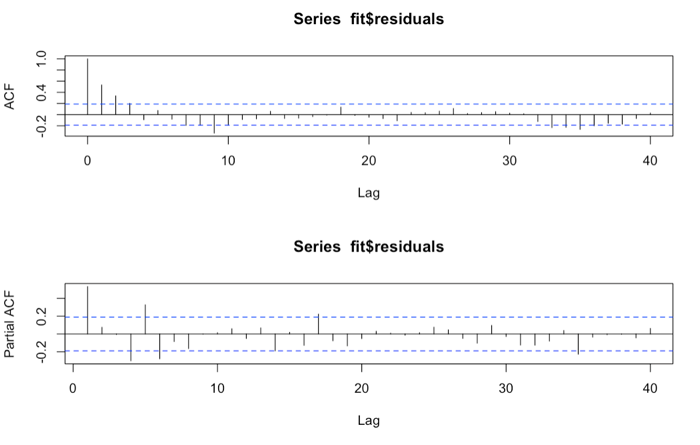
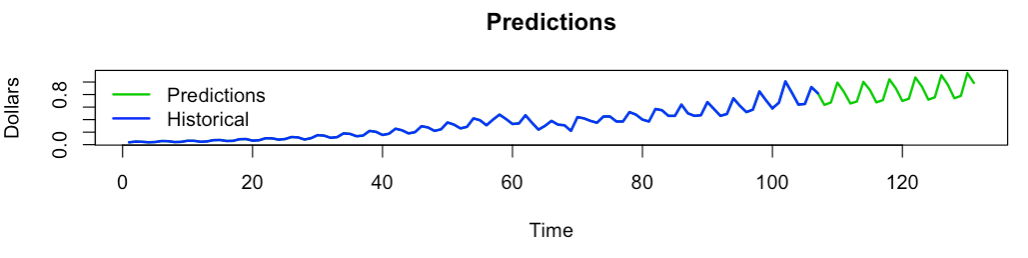


Figure 11: Option 4 ACF and PACF of Residuals (5,0,0)x(0,2,1)s=4

# 

Figure 12: Model 2 Predictions



# Model Comparison and Conclusion

To determine the best model, we looked at the MAPE (mean absolute percentage error) and the MSFE (mean square forecasting error) from doing the out of sample forecasting exercise. Model 1 is the best model because it has the smallest MAPE and MSFE, and therefore the smallest forecasting error.