Flight Delay Predictions

GROUP 2

Mary Doerries
Markus Olson
Varsha Sajja
Maggie Wolff

DePaul University
DSC 425: Time Series Analysis
Dr. John McDonald
June 10, 2022

Non-Technical Summary

In this project we wanted to discover the best way to predict how long a flight would be delayed if the weather was bad. More specifically, how does the amount of precipitation affect the length of the flight delay.

Our analysis used 2 years of data from two different sources. For the flight data, we used the Bureau of Transportation Statistics Carrier On-Time Performance data, and filtered it to Chicago O’Hare Airport. For the precipitation data we used the U.S. Local Climatological Data from the National Oceanic and Atmospheric Administration, and filtered that data to the O’Hare zip code.

The weather data was at the hourly level, and the flight data was by individual flight, so we aggregated the data to daily and weekly levels. For the daily level, we summed the precipitation on a daily level and averaged the flight delays for the day. For the weekly data, we summed both the precipitation and the total flight delay times for the week. The data sets were joined on the date column.

From here we looked at a variety of models to try to determine what the best way to model our forecast would be. Through this analysis we determined the best way would be to use an x-reg ARIMA model as that produced the best type of model to use and it gave us something to model, and we got pretty good fits for our models.

When looking at the results to determine what model, the daily or weekly, gave us the best predictions we used the mean absolute percentage error and tried to minimize that as much as possible. The daily model had a MAPE of 54% while the weekly model had a MAPE of 37%. The results were that the weekly model performed better than the daily model.

While in this project we only looked at one airport, one zip code, and only two years of data, there is still a lot more we could look at to target a better forecast model. If we were to continue

this analysis we would want to look at more airports than just Chicago O'Hare to see broader flight delay data. There are also other things that can cause flight delays such as mechanical issues or delays from incoming flights. We would also look at a larger time frame as that could provide some better accuracy in the forecast as 2 years is a pretty short time period.

In conclusion we were pretty happy with the results of our small data set and just looking at one airport and one type of delay, but we only scratched the surface of the types of analysis that could be done to forecast flight delays. We would be very interested to see the other types of analysis that could be done and see how it improves our model, or even if things other than weather have a large impact on flight delays.

Technical Summary

Exploratory Analysis of Data

Data Cleaning

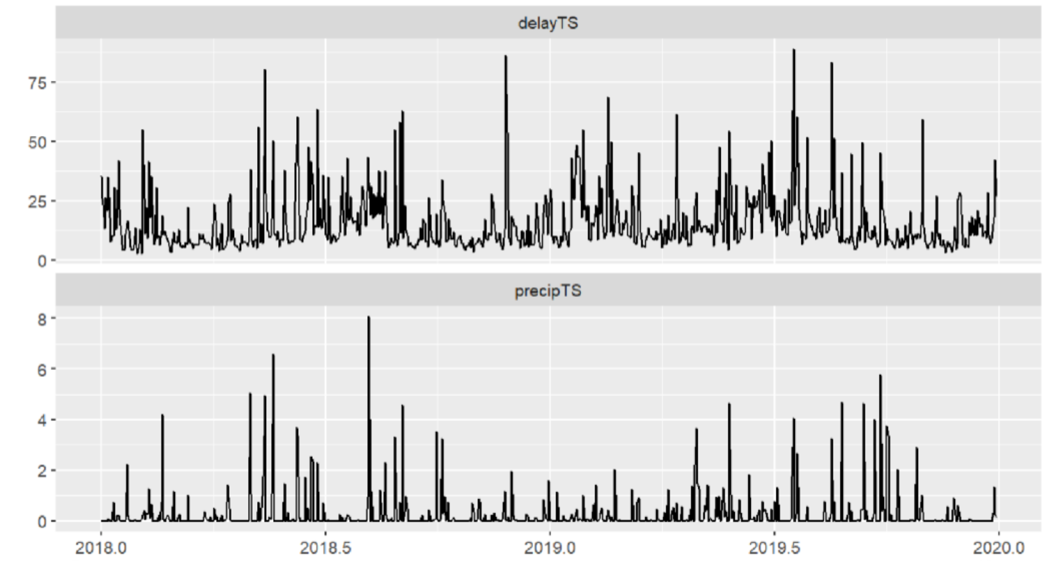
For the project we used data from the Bureau of Transportation Statistics Carrier On-Time Performance data and National Oceanic and Atmospheric Administration U.S. Local Climatological Data. To get the data into a usable format, some data cleansing was performed. For the weather data, we used the precipitation (in inches) and date columns. Any precipitation field that was blank or "NA" was filled in with a 0. There was also a "T" in this field that stood for trace amount; this was replaced with the value of 0.001, which was smaller than the smallest recorded amount in the dataset. Some precipitation values had an "s" on the end, which was stripped so only the numeric value remained. For the flight data, we used the date and the weather delay (in minutes) column. Any blanks for the delay time were replaced with a 0.

The weather delay data was at an hourly level and was summed at both daily and weekly levels. The flight data was for each individual flight and was aggregated at a daily (average) and weekly (sum) level. Finally the two datasets were joined together on the date.

Data Exploration

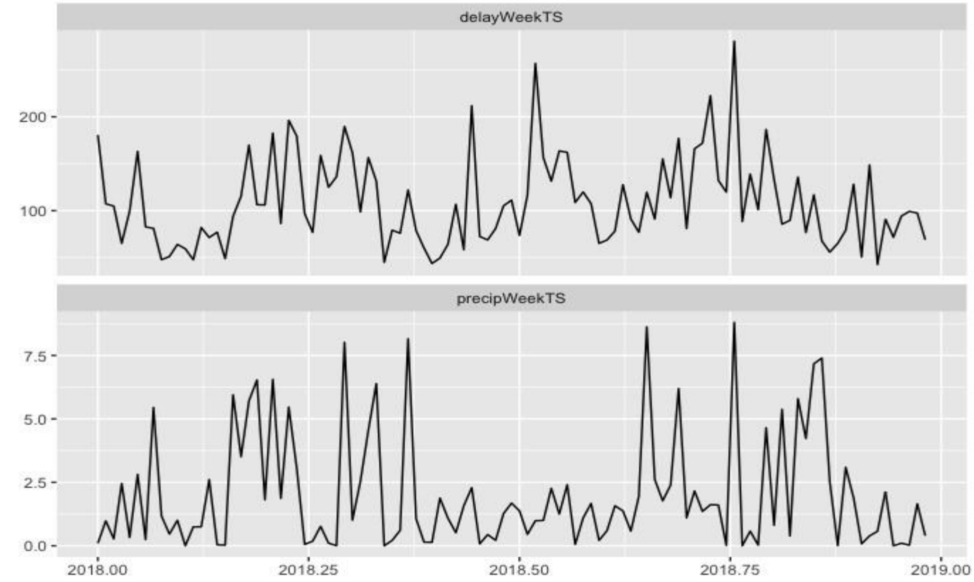
54% of days had measurable precipitation, the median amount of 0.02 inches, the mean was 0.3 inches, and the maximum was 8.088 inches of precipitation in a single day.

100% of days had some amount of flight delays due to weather. The minimum average flight delay in a single day was 2.9 minutes, the median was 11.1 minutes, the mean was 15.6 minutes, and the maximum was 89.1 minutes. Note that flights that left early had a delay listed with a negative amount, which could offset these numbers.



Top: Time Series of daily average flight delays

Bottom: Time Series of daily precipitation



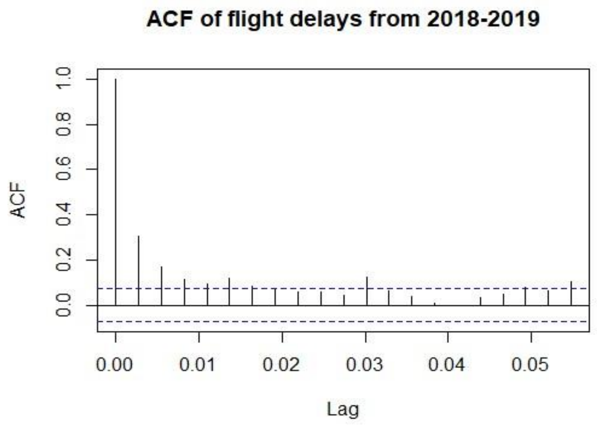
Top: Time Series of weekly aggregated flight delays

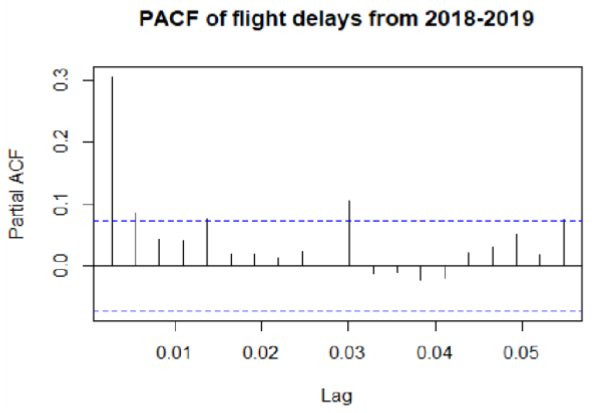
Bottom: Time Series of weekly precipitation

Model Fitting

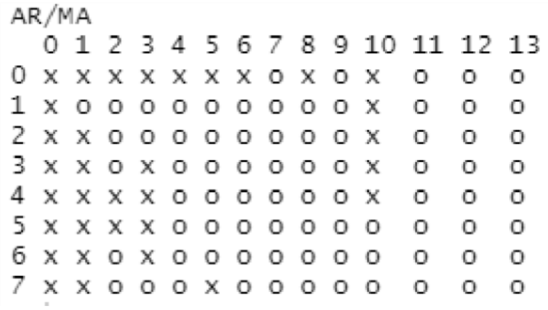
Daily Flight Delay Models

We started fitting the models using the ACF, PACF, EACF to determine potential ARIMA models. From the ACF we can see that there is some potential MA behavoir, because lags 1-6 are above the confidence interval. Lags 3-6 are just barely above the confidence interval so it is unlikely that they were needed in the model.





The PACF showed that there was some AR behavoir in the series. We can see some activity definitely at lag1 and maybe lag2 as well, however it is very close to the line.

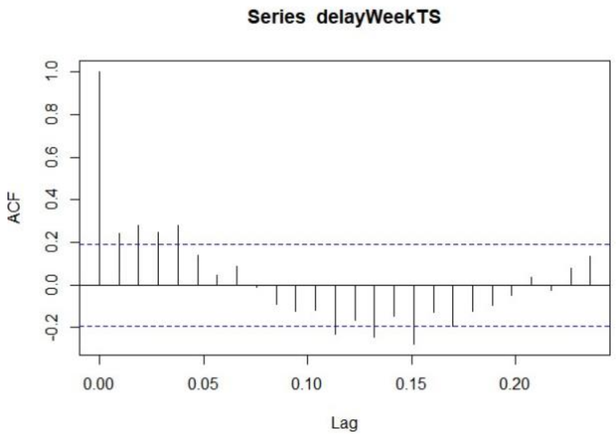


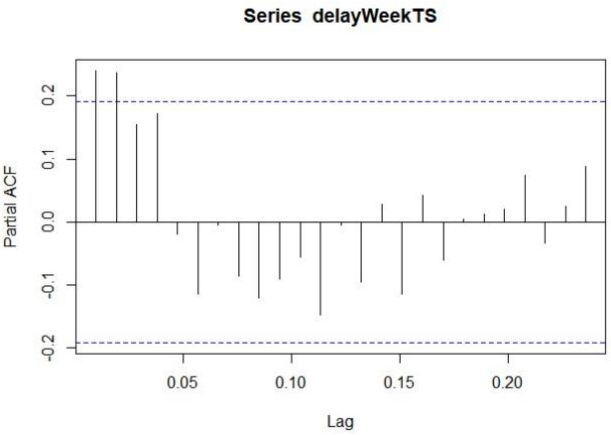
```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x x x o x o x  o  o  o
1 x o o o o o o o o o x  o  o  o
2 x x o o o o o o o o x  o  o  o
3 x x o x o o o o o o x  o  o  o
4 x x x x o o o o o o x  o  o  o
5 x x x x o o o o o o o  o  o  o
6 x x o x o o o o o o o  o  o  o
7 x x o o o x o o o o o  o  o  o
```

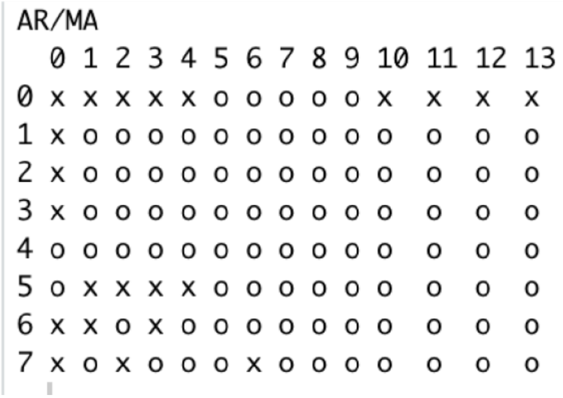
From the EACF, we can tell that we will want to include at least an MA(1) and AR(1) in the model. There is also a long line of X's at the top of the EACF. We tried differencing the data because of this, however the differenced ACF and PACF showed clear signs of over differencing with the first couple of lags switching from positive to negative.

Weekly Flight Delay Models

Using the weekly aggregated data, we started fitting the models using the ACF, PACF, EACF to determine potential ARIMA models. From the ACF, we can see that there are AR and MA behaviors, because lags 1-5 are above the confidence interval. Lags 6-12 are just barely between the confidence interval, so it is unlikely that they were needed in the model. Thus, we should include at least one AR and one MA. From the PACF, we can interpret that model can be AR(2) but the coefficients seem to be insignificant to be included into the model. From EACF plot, we can clearly say that ARIMA(1,0,1) model can be considered for further analysis.





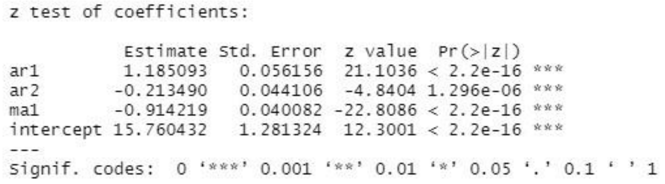


```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x x x x o o o o o x  x  x  x
1 x o o o o o o o o o o  o  o  o
2 x o o o o o o o o o o  o  o  o
3 x o o o o o o o o o o  o  o  o
4 o o o o o o o o o o o  o  o  o
5 o x x x x o o o o o o  o  o  o
6 x x o x o o o o o o o  o  o  o
7 x o x o o o x o o o o  o  o  o
```

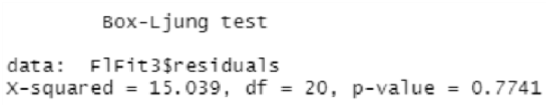
Residual analysis and model diagnostics

Daily Model Residual Analysis and Model Diagnostics

We tested several different models and decided on an ARIMA (2,0,1) model with a non-zero mean. Model equation $Y_t = 15.76 + 1.2Y_{t-1} - 0.2Y_{t-2} - 0.91a_{t-1} + a_{t.}$ The Ljung Box test of the residuals for this model has a p-value of over 0.05 meaning that we cannot reject white noise for the residuals. The coefficient test also showed that all the coefficients in the model are significant.

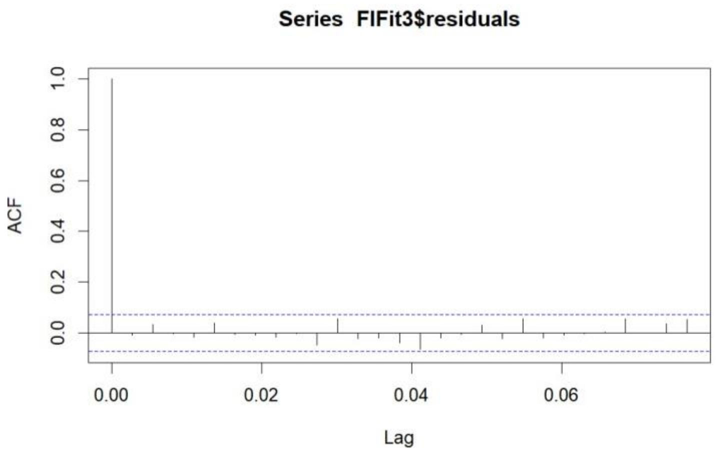


```
z test of coefficients:

            Estimate Std. Error  z value  Pr(>|z|)
ar1         1.185093   0.056156  21.1036 < 2.2e-16 ***
ar2        -0.213490   0.044106  -4.8404 1.296e-06 ***
ma1        -0.914219   0.040082 -22.8086 < 2.2e-16 ***
intercept 15.760432   1.281324  12.3001 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

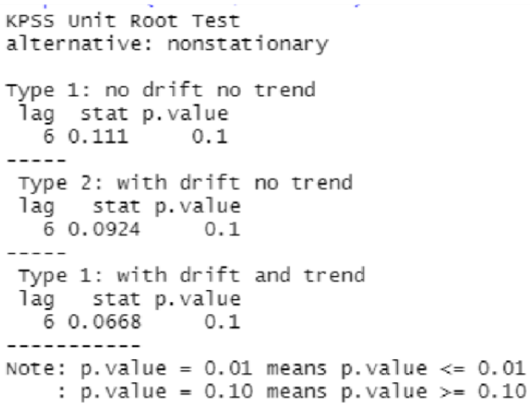


```
        Box-Ljung test

data:  FlFit3$residuals
X-squared = 15.039, df = 20, p-value = 0.7741
```

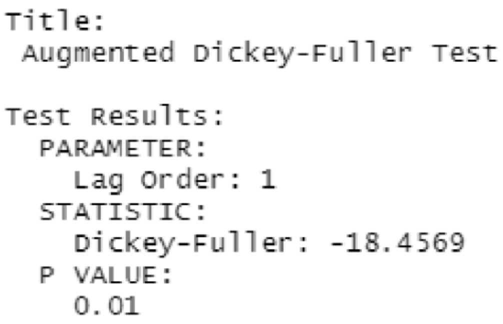
The autocorrelation of the residuals was insignificant after lag 0.



Looking at the Augmented Dickey-Fuller (using a constant) and the KPSS test of the residuals for this model, both tests agree that the residuals are stationary.

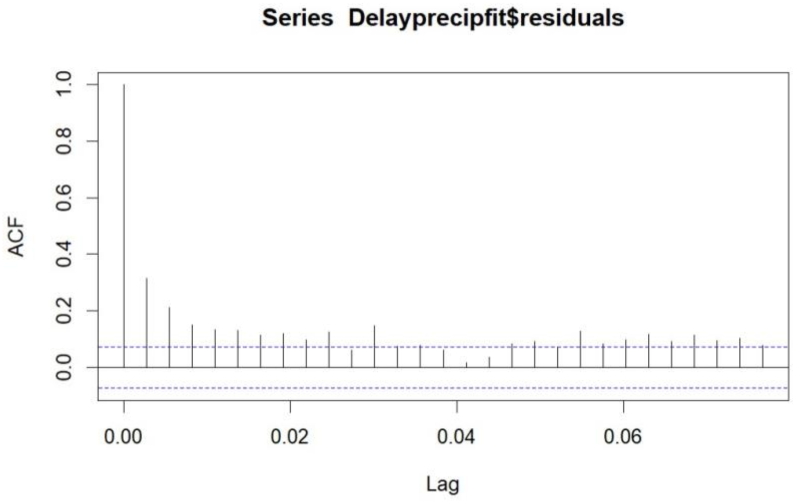


```
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag  stat p.value
   6 0.111     0.1
-----
 Type 2: with drift no trend
 lag   stat p.value
   6 0.0924     0.1
-----
 Type 1: with drift and trend
 lag   stat p.value
   6 0.0668     0.1
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```

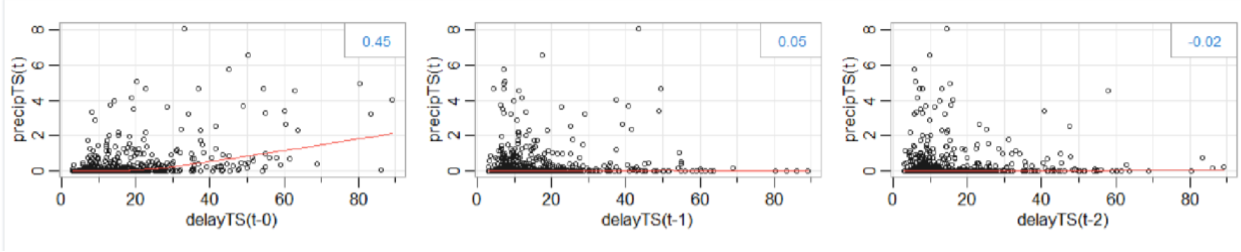


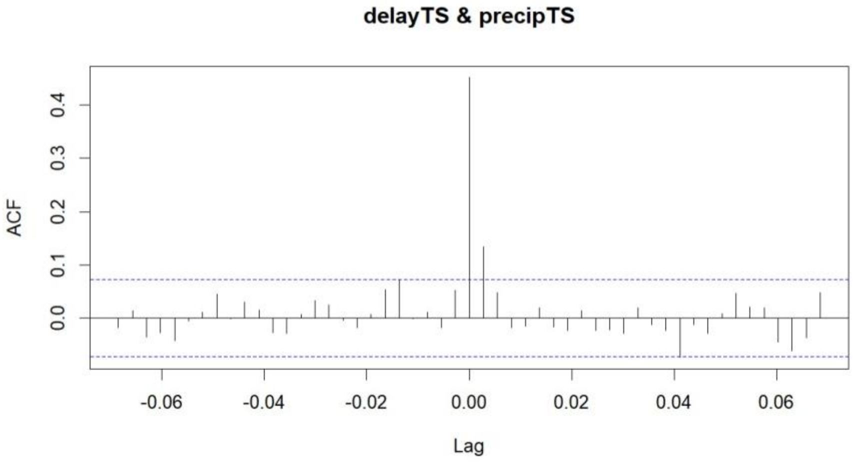
```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -18.4569
  P VALUE:
    0.01
```

We also checked a linear regression of the flight delays on the precipitation data. The residuals showed some autocorrelation.

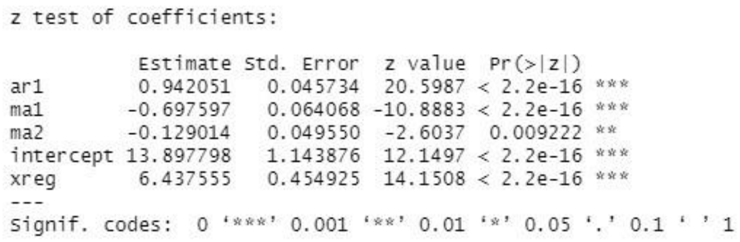


Next we added an xreg of the precipitation time series to the ARIMA model. Based on the lag and cross-correlation we did not need to account for any lagged correlations between the two values.



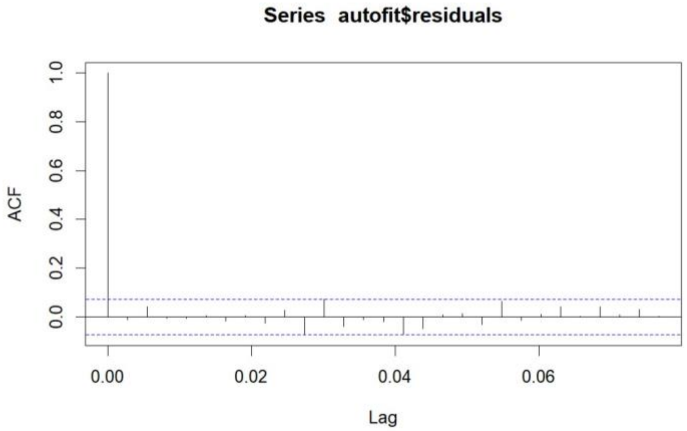


The coeftest of the model plus the xreg shows that all the variables including the added regression are all significant. The MA(2) variable is a little less significant than without the regression but it is still very significant.

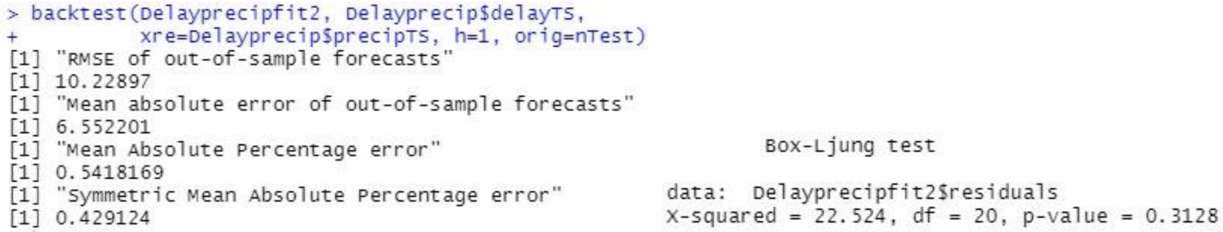


```
z test of coefficients:

           Estimate Std. Error  z value  Pr(>|z|)    
ar1        0.942051   0.045734  20.5987 < 2.2e-16 ***
ma1       -0.697597   0.064068 -10.8883 < 2.2e-16 ***
ma2       -0.129014   0.049550  -2.6037  0.009222 ** 
intercept 13.897798   1.143876  12.1497 < 2.2e-16 ***
xreg       6.437555   0.454925  14.1508 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The autocorrelation of the residuals shows insignificant values:

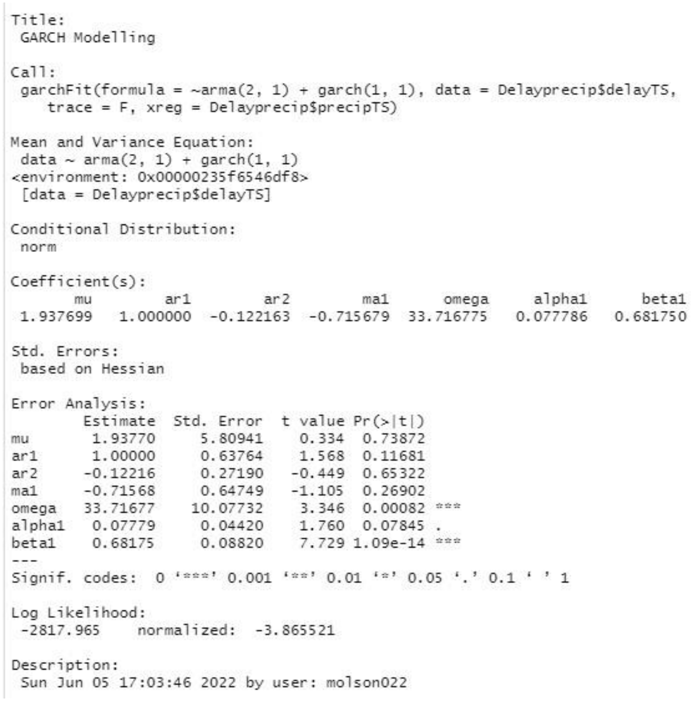


Once again, the Ljung Box test of the residuals for this model has a p-value of over 0.05 meaning that we can not reject white noise for the residuals. Diving further into the model analysis we can see from the backtest that the MAPE is around 54%, meaning that the model will be off about 54% of the time for future predictions.



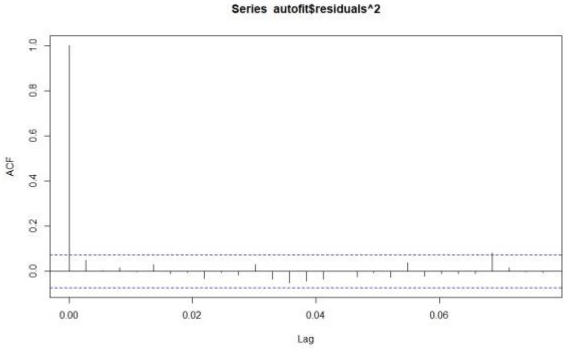
```
> backtest(Delayprecipfit2, Delayprecip$delayTS,
+          xre=Delayprecip$precipTS, h=1, orig=nTest)
[1] "RMSE of out-of-sample forecasts"
[1] 10.22897
[1] "Mean absolute error of out-of-sample forecasts"
[1] 6.552201
[1] "Mean Absolute Percentage error"                      Box-Ljung test
[1] 0.5418169
[1] "Symmetric Mean Absolute Percentage error"        data:  Delayprecipfit2$residuals
[1] 0.429124                                          X-squared = 22.524, df = 20, p-value = 0.3128
```

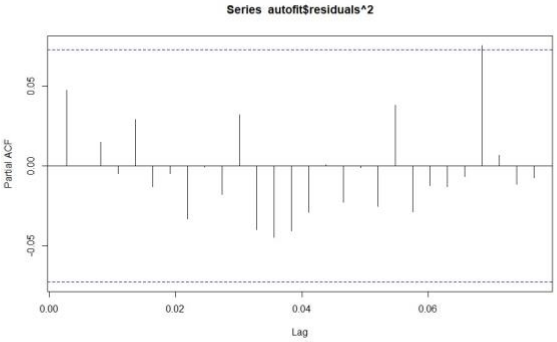
We did try GARCH modelling and saw a small improvement on the residuals, but not enough to want to use the GARCH model going forward. This is because our MA and AR coefficients were not significant anymore and the alpha is barely significant. Below you can see our coefficients test along with the before and after ACF and PACF residuals squared.



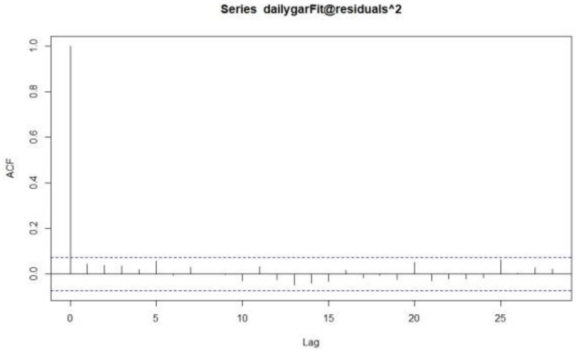
```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(2, 1) + garch(1, 1), data = Delayprecip$delayTS,
    trace = F, xreg = Delayprecip$precipTS)

Mean and Variance Equation:
 data ~ arma(2, 1) + garch(1, 1)
<environment: 0x00000235f6546df8>
 [data = Delayprecip$delayTS]

Conditional Distribution:
 norm

Coefficient(s):
       mu        ar1        ar2        ma1      omega     alpha1      beta1
 1.937699   1.000000  -0.122163  -0.715679  33.716775   0.077786   0.681750

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu       1.93770     5.80941    0.334  0.73872
ar1      1.00000     0.63764    1.568  0.11681
ar2     -0.12216     0.27190   -0.449  0.65322
ma1     -0.71568     0.64749   -1.105  0.26902
omega   33.71677    10.07732    3.346  0.00082 ***
alpha1   0.07779     0.04420    1.760  0.07845 .
beta1    0.68175     0.08820    7.729 1.09e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 -2817.965    normalized:  -3.865521

Description:
 Sun Jun 05 17:03:46 2022 by user: molson022
```

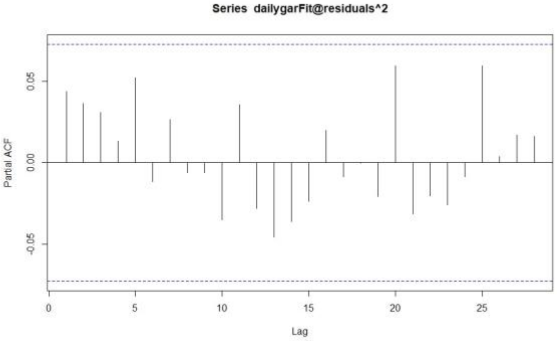
Before GARCH





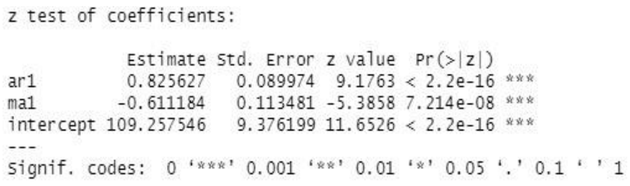
After Garch:



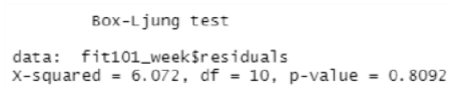


Weekly Model Residual Analysis and Model Diagnostics

Based on ACF/PACF/EACF plots, we tested several different models and decided on an ARIMA (1,0,1) model. Model equation is $Y_t = 109.26 + 0.82Y_{t-1} - 0.61a_{t-1} + a_t$. The Ljung Box test of the residuals for this model has a p-value of over 0.05 meaning that we cannot reject white noise for the residuals. The coefficient test also showed that all the coefficients in the model are significant.

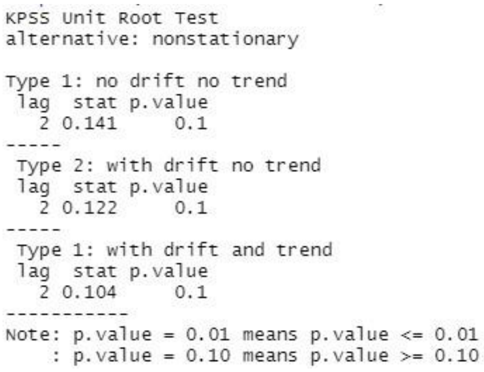


```
z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
ar1         0.825627   0.089974  9.1763 < 2.2e-16 ***
ma1        -0.611184   0.113481 -5.3858 7.214e-08 ***
intercept 109.257546   9.376199 11.6526 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

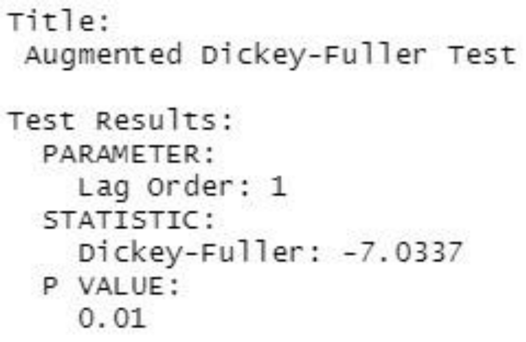


```
        Box-Ljung test

data:  fit101_week$residuals
X-squared = 6.072, df = 10, p-value = 0.8092
```

Looking at the Augmented Dickey-Fuller test and the KPSS test of the residuals, both tests agree that the residuals are non-stationary.

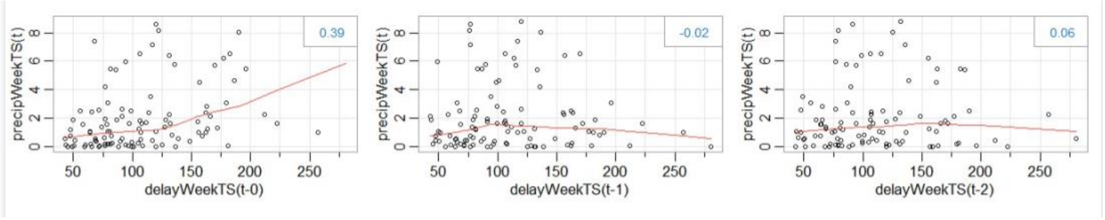


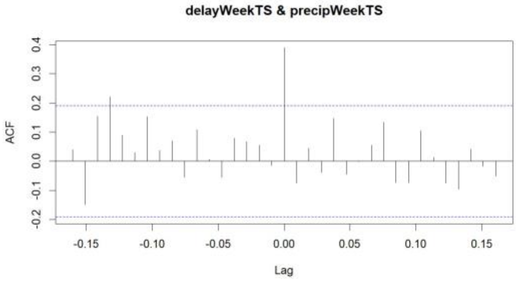
```
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag  stat p.value
   2 0.141     0.1
-----
 Type 2: with drift no trend
 lag  stat p.value
   2 0.122     0.1
-----
 Type 1: with drift and trend
 lag  stat p.value
   2 0.104     0.1
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```



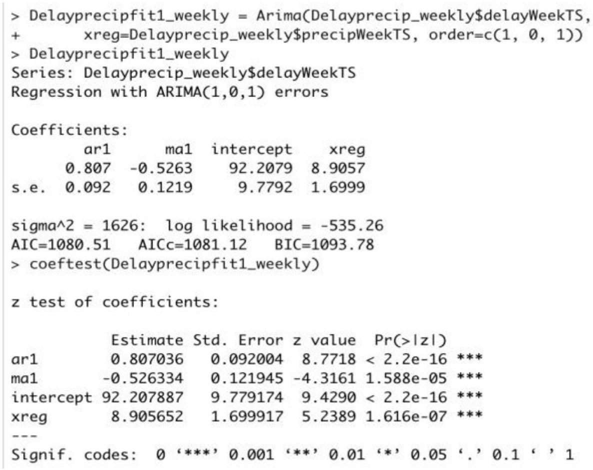
```
Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 1
  STATISTIC:
    Dickey-Fuller: -7.0337
  P VALUE:
    0.01
```

Next, we added an xreg of the precipitation time series to the ARIMA(1,0,1) model. As shown below from the lag and cross-correlations, no lag is needed.

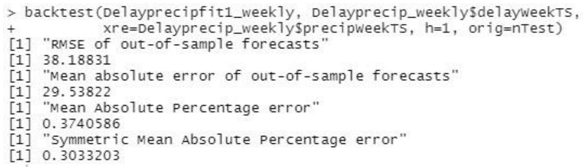




Coeftest shows that all the variables are significant.

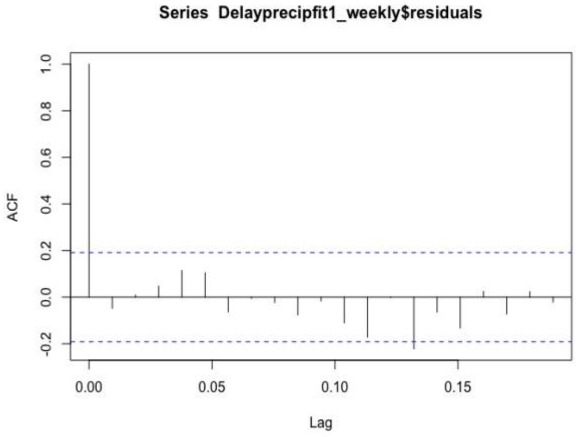


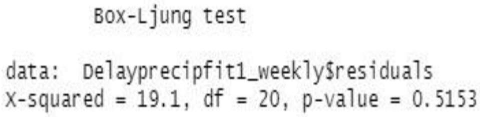
```
> Delayprecipfit1_weekly = Arima(Delayprecip_weekly$delayWeekTS,
+       xreg=Delayprecip_weekly$precipWeekTS, order=c(1, 0, 1))
> Delayprecipfit1_weekly
Series: Delayprecip_weekly$delayWeekTS
Regression with ARIMA(1,0,1) errors

Coefficients:
         ar1      ma1  intercept    xreg
       0.807  -0.5263    92.2079  8.9057
s.e.   0.092   0.1219     9.7792  1.6999

sigma^2 = 1626:  log likelihood = -535.26
AIC=1080.51   AICc=1081.12   BIC=1093.78
> coeftest(Delayprecipfit1_weekly)

z test of coefficients:

           Estimate Std. Error z value  Pr(>|z|)
ar1        0.807036   0.092004  8.7718 < 2.2e-16 ***
ma1       -0.526334   0.121945 -4.3161 1.588e-05 ***
intercept 92.207887   9.779174  9.4290 < 2.2e-16 ***
xreg       8.905652   1.699917  5.2389 1.616e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```



```
> backtest(Delayprecipfit1_weekly, Delayprecip_weekly$delayWeekTS,
+          xre=Delayprecip_weekly$precipWeekTS, h=1, orig=nTest)
[1] "RMSE of out-of-sample forecasts"
[1] 38.18831
[1] "Mean absolute error of out-of-sample forecasts"
[1] 29.53822
[1] "Mean Absolute Percentage error"
[1] 0.3740586
[1] "Symmetric Mean Absolute Percentage error"
[1] 0.3033203
```

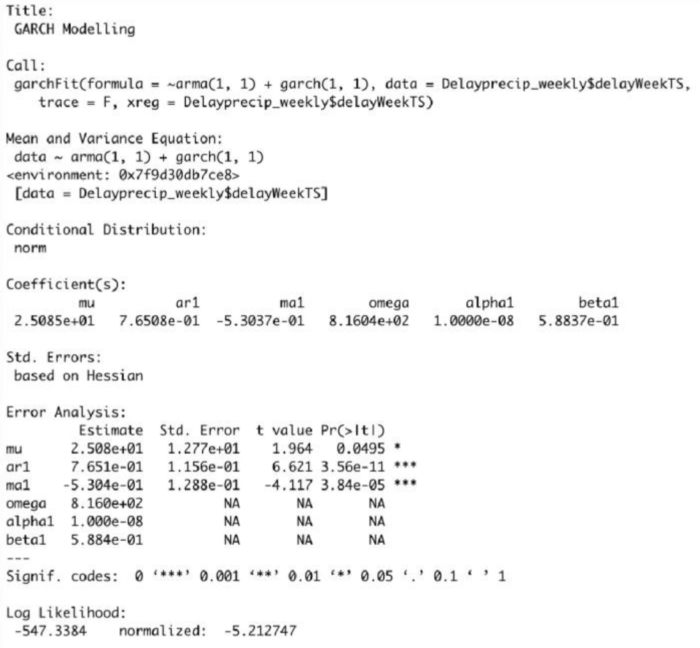
Once again, the Ljung Box test of the residuals for this model has a p-value of over 0.05 meaning that we cannot reject white noise for the residuals. Diving further into model analysis we can see from the backtest that the MAPE is around 37%, meaning that the model will be off about 37% of the time for future predictions.





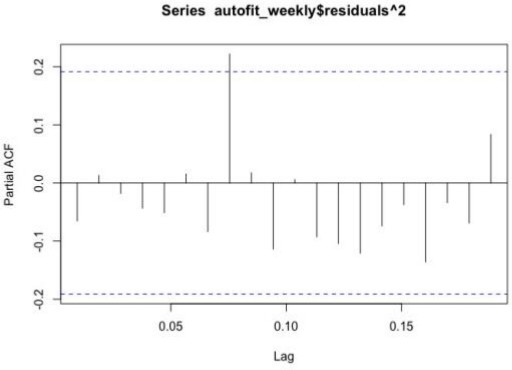
```
                 Box-Ljung test

data:  Delayprecipfit1_weekly$residuals
X-squared = 19.1, df = 20, p-value = 0.5153
```

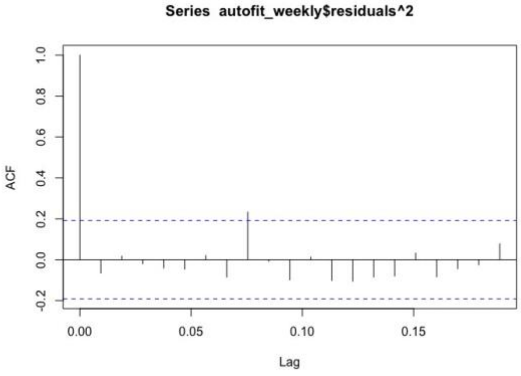
Using GARCH modeling, we saw a small improvement on the residuals with weekly data, but not enough to decide on using the GARCH model going forward. This is because our MA and AR coefficients were not significant any more. Also, the alpha estimate is insignificant. Below you can see our coefficients test along with the before and after ACF/PACF squared residuals.



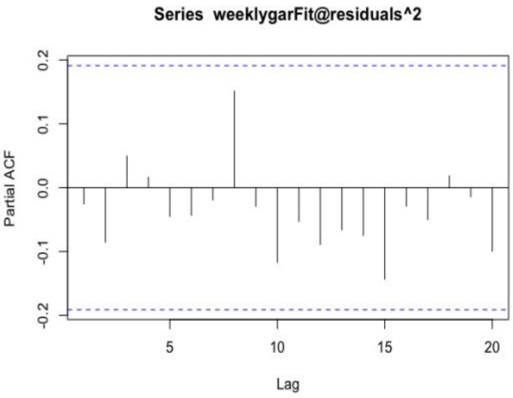
```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~arma(1, 1) + garch(1, 1), data = Delayprecip_weekly$delayWeekTS,
    trace = F, xreg = Delayprecip_weekly$delayWeekTS)

Mean and Variance Equation:
 data ~ arma(1, 1) + garch(1, 1)
<environment: 0x7f9d30db7ce8>
 [data = Delayprecip_weekly$delayWeekTS]

Conditional Distribution:
 norm

Coefficient(s):
        mu          ar1          ma1        omega       alpha1        beta1
 2.5085e+01   7.6508e-01  -5.3037e-01   8.1604e+02   1.0000e-08   5.8837e-01

Std. Errors:
 based on Hessian

Error Analysis:
         Estimate  Std. Error  t value Pr(>|t|)
mu      2.508e+01   1.277e+01    1.964   0.0495 *
ar1     7.651e-01   1.156e-01    6.621 3.56e-11 ***
ma1    -5.304e-01   1.288e-01   -4.117 3.84e-05 ***
omega   8.160e+02          NA       NA       NA
alpha1  1.000e-08          NA       NA       NA
beta1   5.884e-01          NA       NA       NA
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Log Likelihood:
 -547.3384    normalized:  -5.212747
```

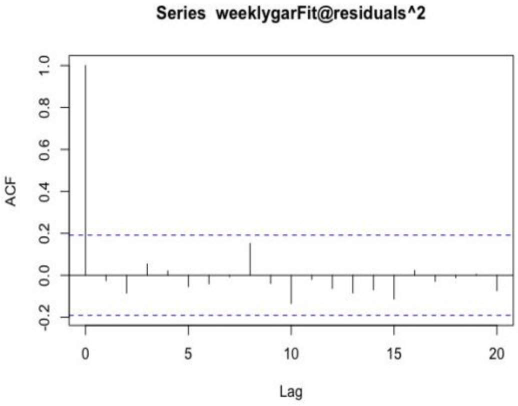
Before GARCH





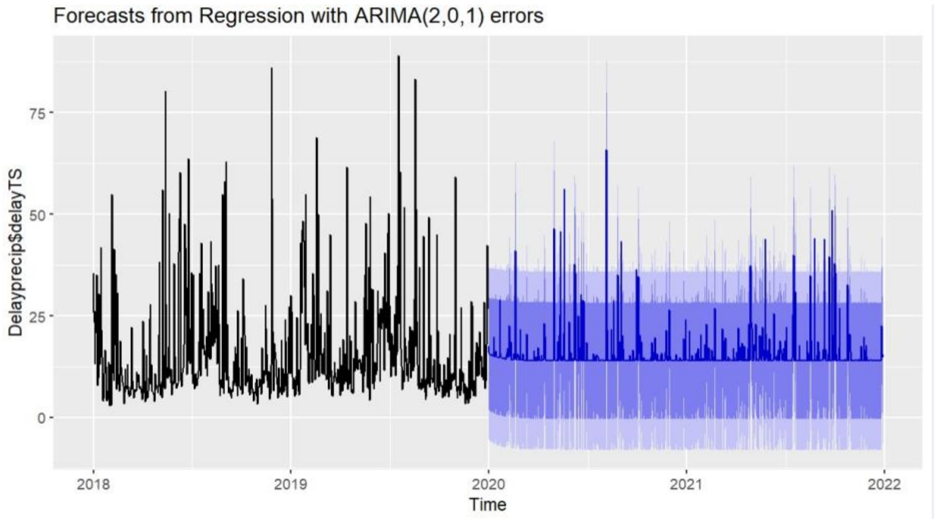
After GARCH

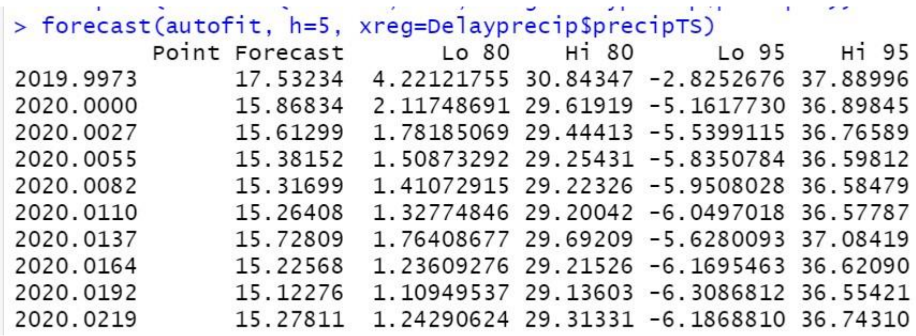




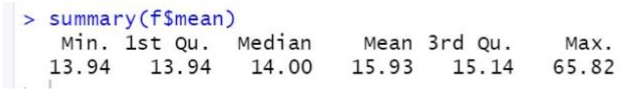
Forecast analysis

Forecasts on Daily data





```
> forecast(autofit, h=5, xreg=Delayprecip$precipTS)
          Point Forecast       Lo 80    Hi 80      Lo 95    Hi 95
2019.9973       17.53234  4.22121755 30.84347 -2.8252676 37.88996
2020.0000       15.86834  2.11748691 29.61919 -5.1617730 36.89845
2020.0027       15.61299  1.78185069 29.44413 -5.5399115 36.76589
2020.0055       15.38152  1.50873292 29.25431 -5.8350784 36.59812
2020.0082       15.31699  1.41072915 29.22326 -5.9508028 36.58479
2020.0110       15.26408  1.32774846 29.20042 -6.0497018 36.57787
2020.0137       15.72809  1.76408677 29.69209 -5.6280093 37.08419
2020.0164       15.22568  1.23609276 29.21526 -6.1695463 36.62090
2020.0192       15.12276  1.10949537 29.13603 -6.3086812 36.55421
2020.0219       15.27811  1.24290624 29.31331 -6.1868810 36.74310
```

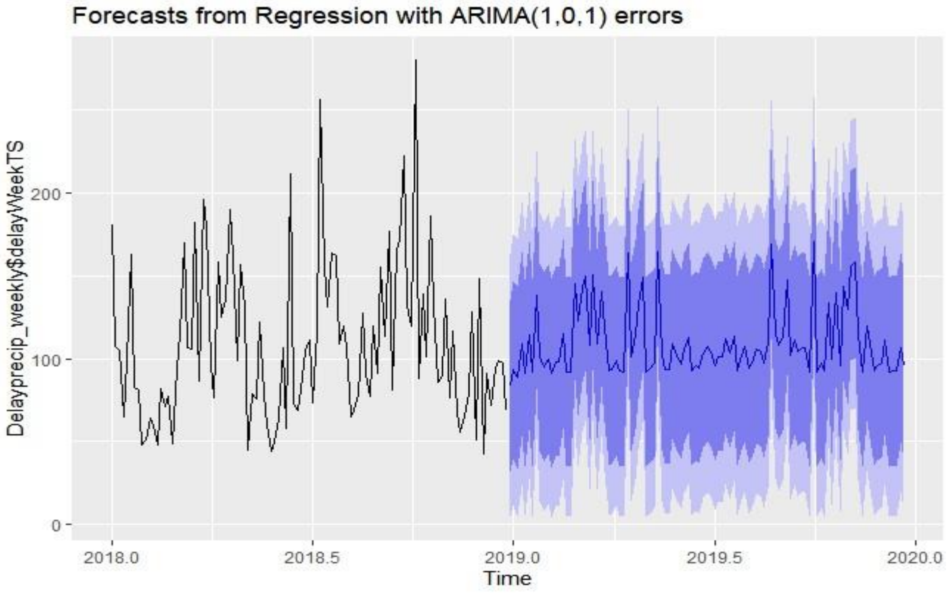


```
> summary(f$mean)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  13.94   13.94   14.00   15.93   15.14   65.82 
```

The mean forecast of daily average flight delays is 15.93 minutes which is close to the mean of the observed data. However, the minimum and maximum (13.94 and 65.82) are closer to the series mean than the observed values.

Forecasts on Weekly data

The forecast shows seasonal patterns including spikes in the forecast from the year 2019 to 2020 as there were delays occurred due to rainy weather or snowfall around Spring and Late Fall/Early Winter in Chicago. Because we summed this data instead of averaging it and because we are looking at weekly values, the mean is much higher than in the daily forecast.



Analysis of the results and discussion/conclusions

Overall, the weekly models performed better in backtesting and forecasting. The daily model had a mean absolute percentage error (MAPE) of 54% while the weekly model had a MAPE of 37%. Surprisingly, seasonality wasn't a factor in model building based on lack of evidence in the autocorrelations. But while a weekly model could certainly give us a better idea of the time of year flight delays are more common, in terms of practicality, forecasting the sum of weekly flight delays doesn't tell us much when it comes to the likelihood of a specific flight being delayed.

For future analysis, looking at data at the hourly level would possibly lead to more accuracy of predicting if a specific flight would be delayed, but the amount of data it would take to make an accurate forecast would require some significant computing power and cloud resources.

Additionally, it would be interesting to look at additional years of data, especially expanding to 2020 when there were significantly fewer flights due to Covid restrictions, so presumably there were fewer factors other than weather that could cause a delay. Additionally, looking at additional airports, especially with different climates, would be a good addition to our analysis.

Appendix

Markus Olson Individual

Through the project I acted as the project manager by organizing meetings, leading discussions, and working with the group to assign different tasks to each team member. I also did a lot of work on the analysis and data cleaning part as well. I was able to aggregate the data from hourly to daily, and cleaned up a little bit to ensure there were no NA’s in our data. Once the data was clean and ready to work with Maggie and I worked together on the daily analysis part of the project. We came up with the R -code the models and the analysis of those models. I found our powerpoint presentation and organized the layout and flow of our final presentation as well as compiling the slides for the intro, data explanation, and with the help of Mary the conclusion. In the actual presentation I spoke in the introduction and conclusion as well as explained the data in our project. Lastly, In the final paper I wrote up the non-technical report and in the technical report I wrote up the data portion.

With this being my last class in the data science program here at Depaul I feel like this class gave me the most tools to use in the real world and learned so much about forecasting , and already have ideas of how to apply these skills. Coming into this class I knew forecasting as an excel exercise with some simple calculations, but here I learned it can be so much more. The amount of techniques you can use and when you should use these techniques. Learning about how you need to evaluate the data before you even start to forecast was really eye opening to me. I now know I need to get the model to be stationary and look at ACF, PACF, EACF to try to determine what type of ARIMA to use. Evaluation using the dicky-fuller and KPSS test to determine if the data is stationary or not. These tangible skills I have already started to be implemented at work for forecasting.

The biggest take away was getting the data to a point where you can forecast using seasonality and how sometimes you have to difference the data to get the best model possible. These were things I would have never thought necessary when I came to forecasting. In my old life if we knew seasonality was going to have a positive or negative effect we would just put in a 5% increase to account for that. I'm very grateful for this class and glad I took it as forecasting now has a whole new meaning for me, and I hope to produce much more accurate forecasts going forward.

Maggie Wolff Individual

For this project, I helped identify the topic and the datasets to use. Once we agreed on a location and a date range, and downloaded the data accordingly, I cleaned the weather data using R. I identified values that needed to be replaced, and cleaned up the data by replacing missing values with 0, and stripping alphabetic characters from numeric columns. I also fixed the date column so it was in a proper format.

With the cleaned data, I explored model building with Markus. We explored models first for the daily precipitation data alone, and then models for the daily aggregated flight delay data regressed on the daily aggregated precipitation data. For modeling the precipitation data alone, we checked the autocorrelation and partial autocorrelation, as well as stationarity tests. We checked differencing as well. We rejected non-stationarity and found that one difference showed over differencing. We explored lag plots and cross-correlations. To determine the proper ARIMA model, we did a linear model of the flight delay on the precipitation and checked the residuals. We tried a few manual ARIMA models, and also ran an auto.arima. We ran backtesting and generated forecasts with that model.

This project has helped me to further understand time series analysis and forecasting. Many of our assumptions were validated but going through the process of model building with a group member also helped to better understand the steps necessary for model building and the appropriate order and tests to check. Additionally, we could help each other troubleshoot the code together if someone wasn't working. Finally, working with "real" data from start to finish, on a topic of our choosing, really helped time series analysis come to life for me.

Varsha Sajja Individual

My initial contribution to the project was deciding upon the datasets with Maggie. Upon proper planning of all the team members and collaboration during each milestones, I was part of data cleaning and getting aggregated data for departure delays (average) to find out daily data analysis on flight delays using the Departure Delay attribute as a prime contributor. Mary and I together did model fitting and residual analysis on the daily flight delay data.

In Milestone 3, I tried ACF, PACF, EACF and unit root tests on flight delay data to check stationary or not. After daily data analysis on flight delays, I came up with the analysis on Weekly data which was aggregated by Mary. I have built an ARIMA model which contributed to the conclusive findings of our data. During this part, I have learnt how to use xreg modelling for our dataset and devised it with the proper ARIMA model. On further analysis, forecasting the data has given me clear insights into the trend and seasonality of our data.

Working on the environmental dataset is absolutely interesting and overwhelming as this is my first time exploring such data out of my comfort zone. This project led me a clear path in identifying seasonal patterns in the dataset. I definitely have advanced learnings from this course (deep diving from DSC 423 & DSC 424) which helped me a lot in this project along with the techniques used by my group members. I would also like to try other models which haven't been used in our project to understand various datasets and their behavior with respect to Time series analysis and forecasting. My important takeaway from this course would be using the statistical methods and interpreting them to further forecast the models upon any typical datasets which I am looking forward to doing.

Mary Doerries Individual

At the start of the project, I helped clean the flight data, I made sure that everything aligned properly (we had some issues with this at first), filtered the data so that only the weather delays showed, filled in missing data (for example some sections where left blank instead of having a zero so I filled in all the blank spaces with zeros). For milestone 3, the initial model submission, I created 4 different models for the flight delay data. I analyzed the ACF, PACF, and EACF as well as the differenced ACF, PACF, and EACFs for the flight delay data. Based on my analysis of these graphs and tests I built two potential models manually and two models using auto.arima and AIC and BIC. For each of these models I used the Ljung-box test on the residuals to check which models had white noise for the residuals and wrote a short analysis for each one.

Next, we further explored our data by further aggregating our data to be weekly to compare to the daily models. For this I aggregated the data to be weekly instead of daily for both the weather and the flight delay data sets. Then for the final presentation I did all the slides for the Daily Data – model exploration, model fitting, forecasting, weekly model exploration, model fitting, forecasting, and half of the conclusion. For each slide on the daily and weekly model exploration, model fitting, and forecasting I wrote a short and succinct analysis for what each of the graphs and different tests showed and what that meant for our models.

The project helped me further understand many of the concepts and principles of time series analysis that we learned in class. Before this class the extent of my forecasting abilities were dragging boxes in excel, so being able to develop more in depth forecasts that take into account seasonality and more nuanced factors will be invaluable.