

Modeling & Forecasting Atmospheric Carbon Concetration

Citlally Reynoso

1/19/2021

Overview

Carbon dioxide is a greenhouse gas, it causes heat to remain trapped within the atmosphere of Earth. It is a naturally occurring gas, but human activities, such as the combustion of fossil fuels, disrupt the natural carbon cycle. Due to its greenhouse effect and the magnitude with which humans produce it, carbon dioxide is a crucial component in understanding climate change.

About the Data

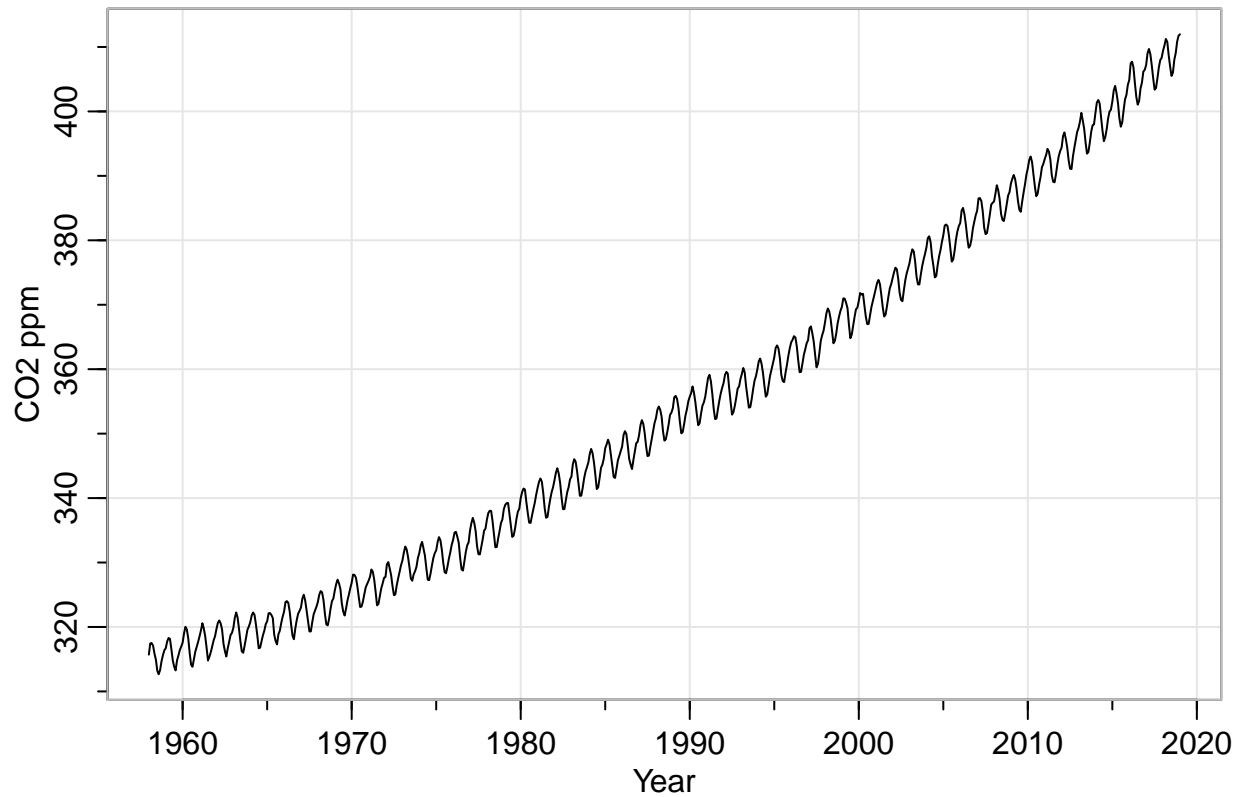
The data set has the carbon dioxide monthly means measured at Mauna Loa, Hawaii from 1958 to April 2020. Throughout this project I model and analyze the trend and seasonality of the CO2 monthly means throughout the years. I use this information to fit a Seasonal Autoregressive Integrated Moving Average (SARIMA) model to the data. I then forecasted CO2 levels using the SARIMA model to validate it and test its accuracy. Below is a short description of the columns in the data:

column name	description
date	Year when carbon dioxide (CO2) reading was taken
season corr	Month when the reading was taken
decimal	Specific date -in decimal format- when reading was taken
average	Average CO2 parts per million (ppm) for the month (-99.99 means there is no data for that month)
interpolated	Average CO2 ppm for the month. Data was interpolated,predicted using a model, for months where reading was not available
trend	Trends observed in the data
days	Number of days for which we have data for that month. Until mid-1974 this column has -1 observations because the data for these months came from another lab.

Vizualizing the Data

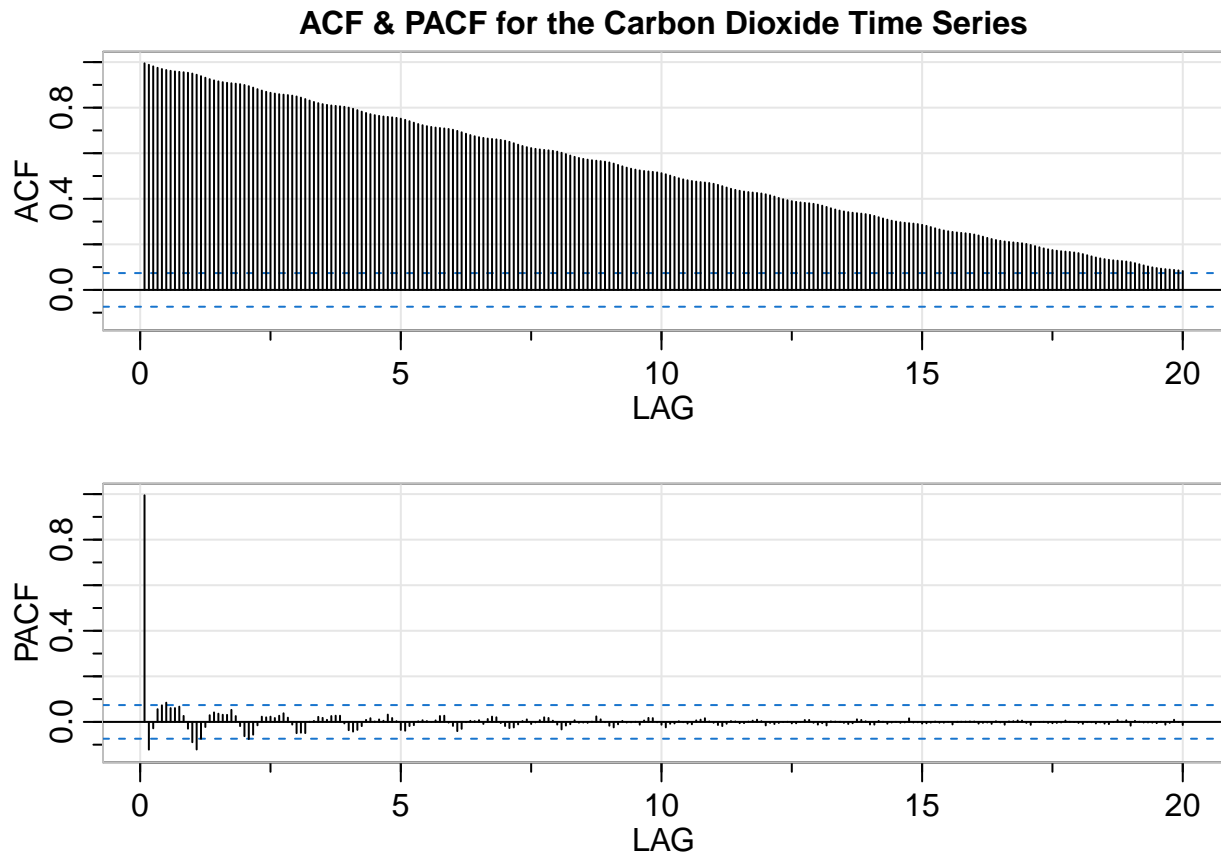
```
##  date season corr  decimal average interpolated  trend days
## 1 1958          3 1958.208  315.71      315.71 314.62   -1
## 2 1958          4 1958.292  317.45      317.45 315.29   -1
## 3 1958          5 1958.375  317.50      317.50 314.71   -1
## 4 1958          6 1958.458  -99.99      317.10 314.85   -1
## 5 1958          7 1958.542  315.86      315.86 314.98   -1
## 6 1958          8 1958.625  314.93      314.93 315.94   -1
```

Carbon Dioxide Monthly Means



Modeling

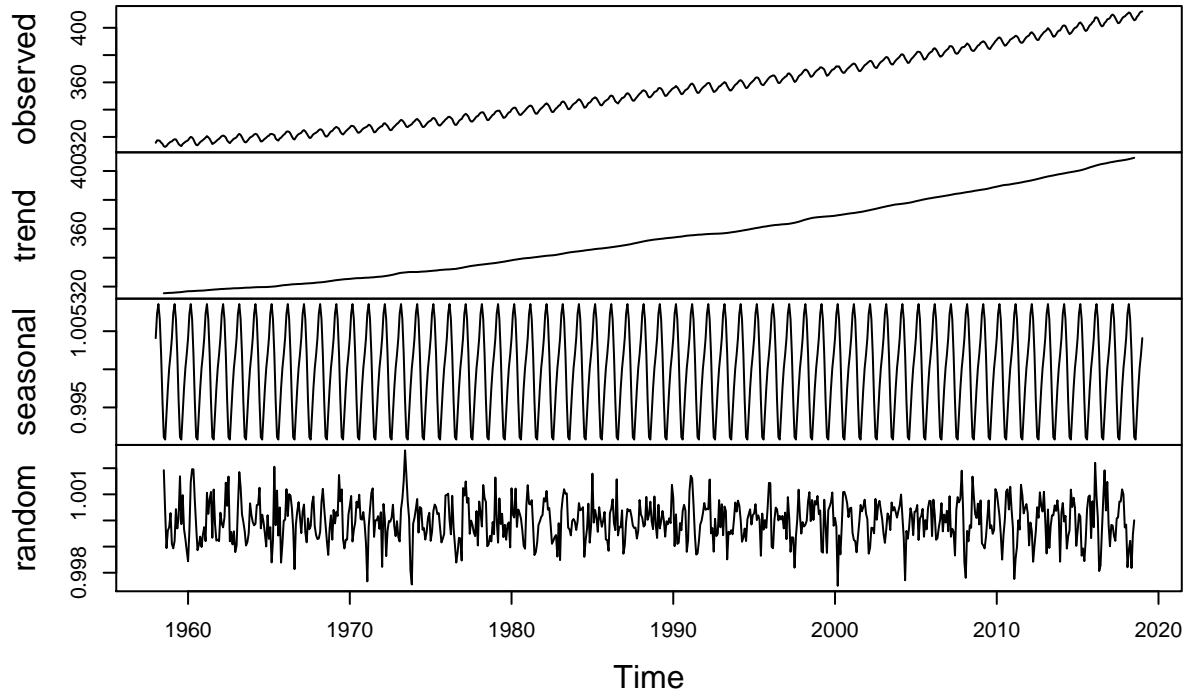
Autocorrelation and Partial Autocorrelation



- The Autocorrelation Function (ACF) of the time series is significant until a lag of 20 years. The autocorrelation decays slowly, which means that the time series is not stationary.
- The Partial Autocorrelation Function (PACF), on the other hand, seems to fall within the insignificant range before 2.5 lags. In order to explore this further, we can decompose the data into trend and seasonality.

Decomposing the data into Trend and Seasonality

Decomposition of multiplicative time series



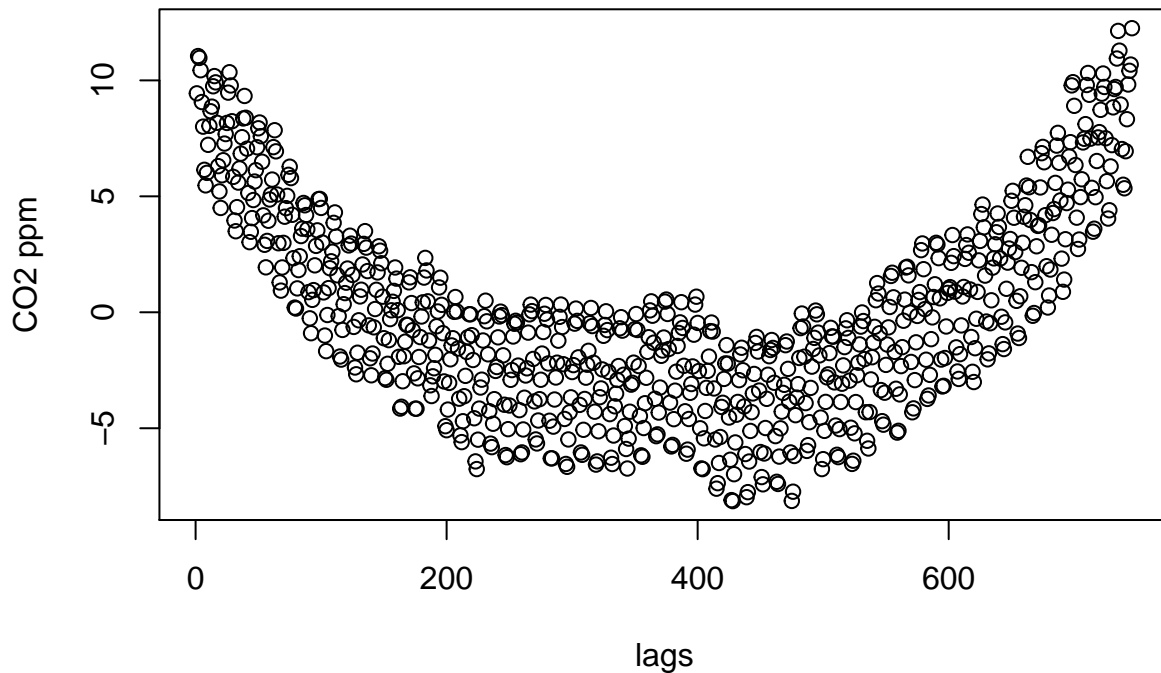
When we deconstruct the time series into trend and seasonality, we are able to visualize the influence that each one has on the data.

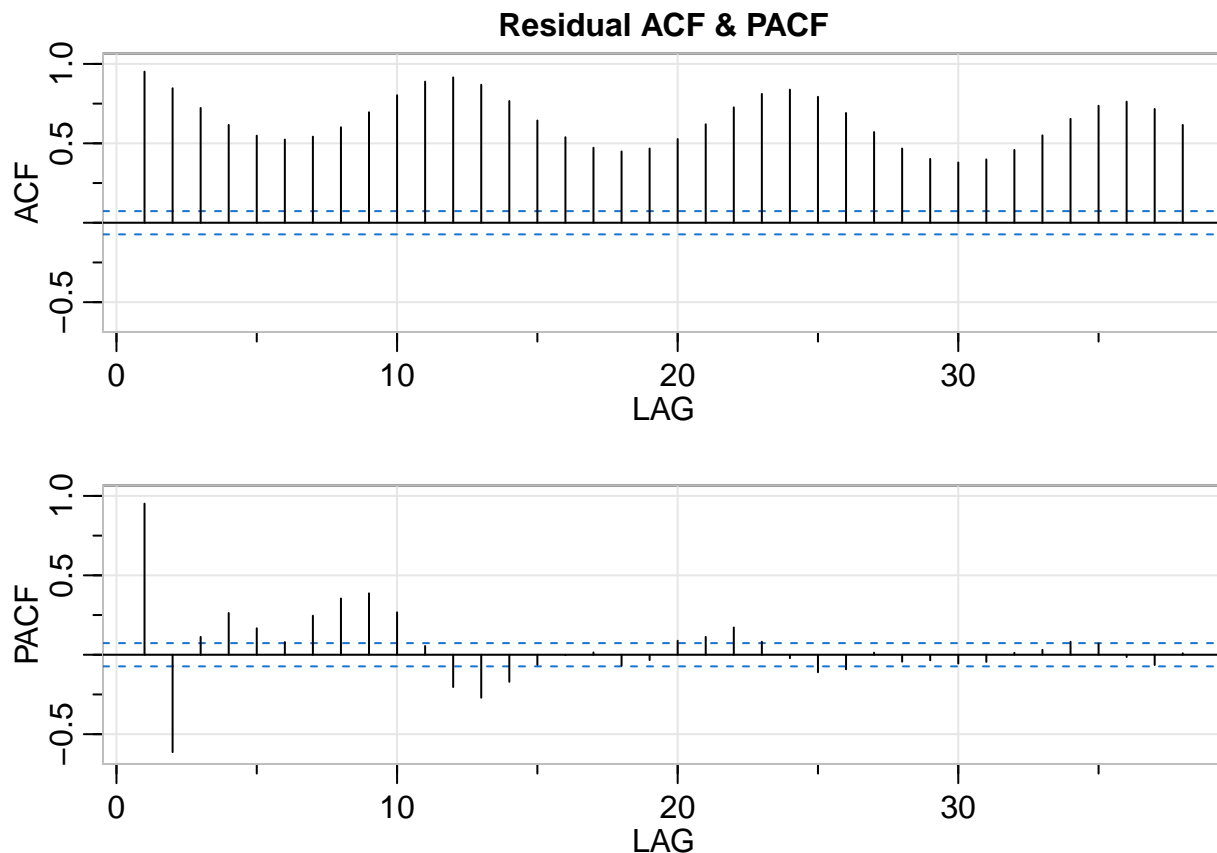
1. **Observed:** This is a visualization of the CO2 monthly means as they exist.
2. **Trend:** shows us there is a positive trend over time in the **carbon dioxide parts per million (CO2 ppm)** present in our atmosphere. As the years have gone by, there is a constant increase in the CO2 ppm.
3. **Seasonal:** graph shows an oscillating pattern in the CO2 monthly means. This indicates that the CO2 ppm are following a seasonal pattern throughout the year. During certain months, CO2 is generally higher and during others it tends to be lower.
4. **Random:** This is the variation in CO2 monthly means that is left over when trend and seasonality are accounted for. As you can see in the graph, when trend and seasonality are accounted for there is no discernible pattern left over in the data. This is what we call *white noise*.

Fitting trend into the model

```
##  
## Call:  
## lm(formula = interpolated ~ decimal, data = co2_mm)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -8.141 -3.071 -0.629  2.847 12.250   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) -2.775e+03  1.762e+01  -157.5  <2e-16 ***  
## decimal      1.574e+00  8.856e-03   177.7  <2e-16 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 4.341 on 744 degrees of freedom  
## Multiple R-squared:  0.977, Adjusted R-squared:  0.9769  
## F-statistic: 3.157e+04 on 1 and 744 DF, p-value: < 2.2e-16
```

Residuals for CO2 ppms

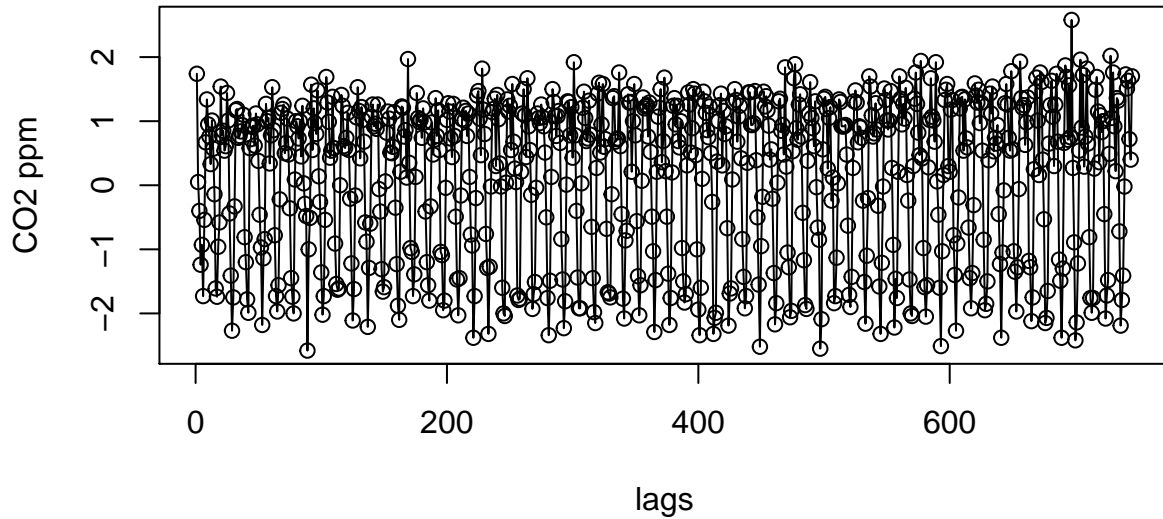




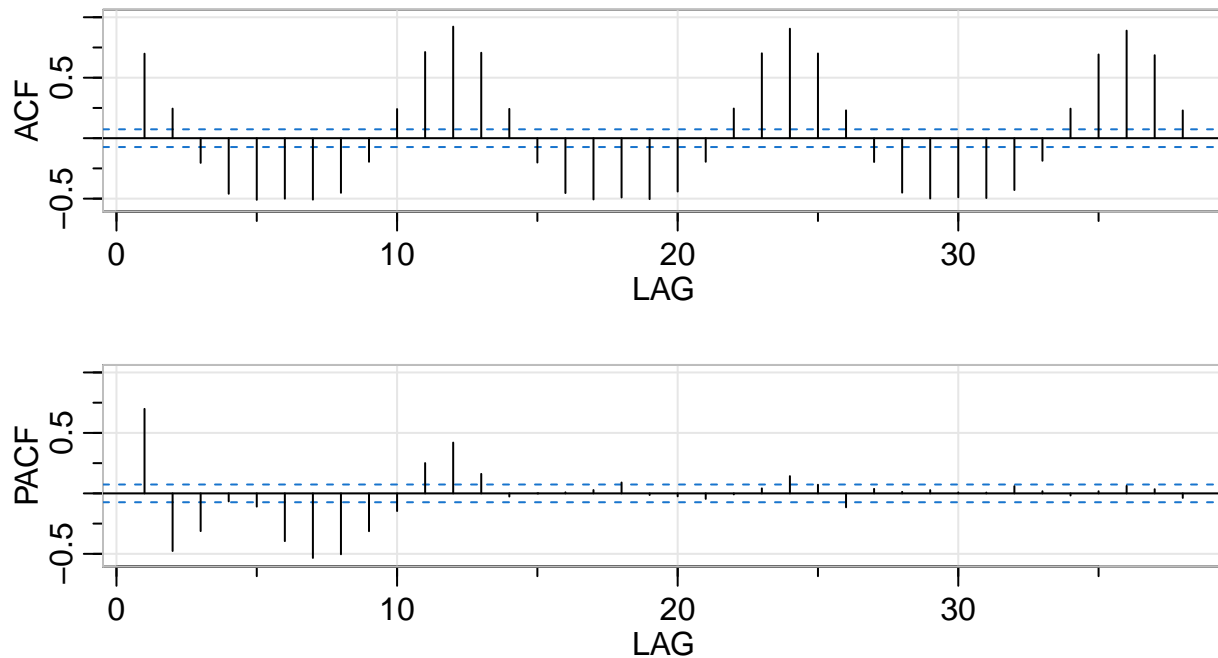
- The variable *decimal*, gives the exact date in decimal form for when the CO2 reading was taken. Including this variable in the model helps us account for the positive trend in the CO2 ppm.
- This predictor alone, gives us an **R-squared of 0.977**.
- Residuals allow us to see what kind of variance we still have not captured:
- **Residual Plot:** When looking at the residual plot, it is apparent that there is still a lot of variance unaccounted for. Specifically, there seems to be a quadratic term which is heavily influencing the data that we have not yet accounted for in this model.
- **ACF & PACF:** The ACF is still significant after a lot of lags, but now it moves up and down over the zero line. The PACF seems to have significance for more lags than before adding in the time factor.

Differentiating the Data Once:

Residuals after Differnciating CO2 Monthly Means Once



ACF & PACF after Differnciating CO2 Monthly Means Once

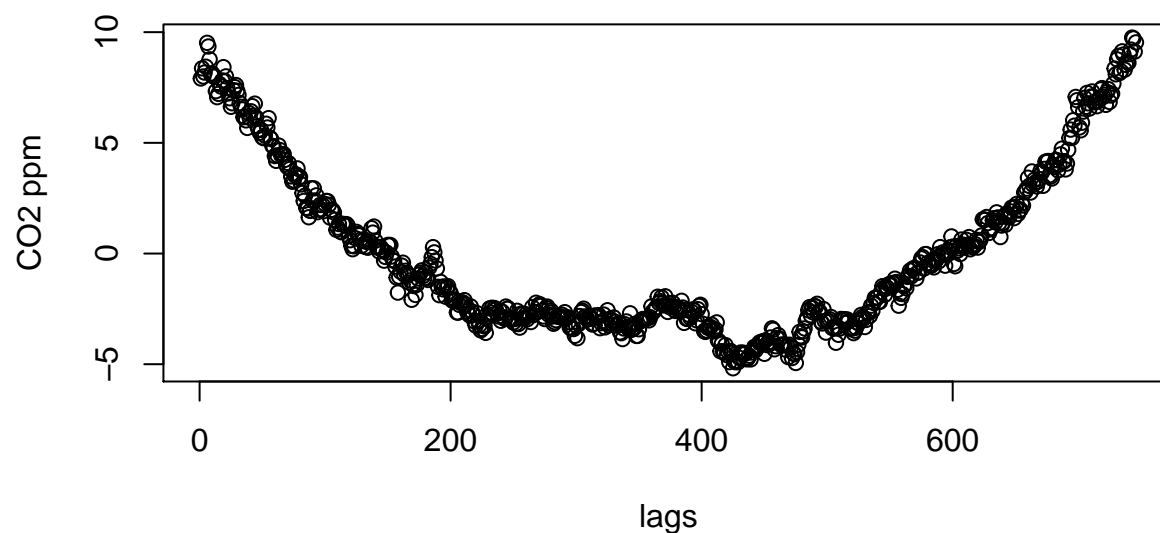


- When I differentiate the data once, the data loses its trend and the mean becomes centered at zero. When I fit the SARIMA model to the data, differentiate the data once will likely be enough to make it stationary.
- Seasonality is still visible in the plot. This tells us that both seasonality and trend are sources of variation in the data, and should both be included into our SARIMA model.

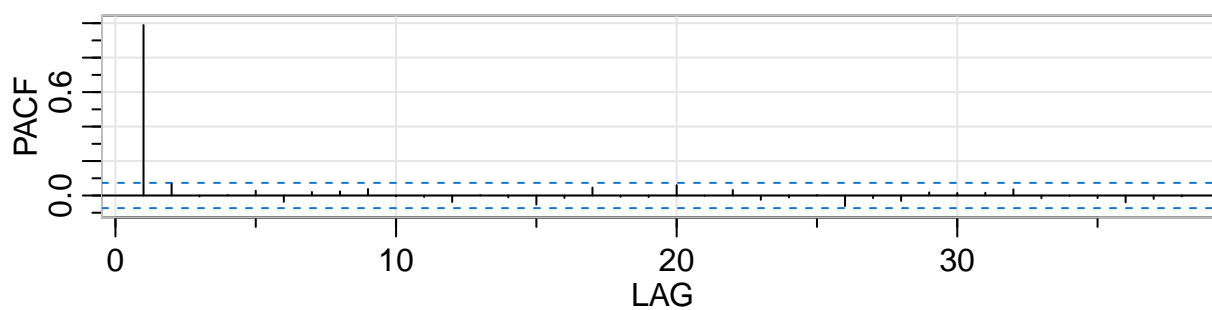
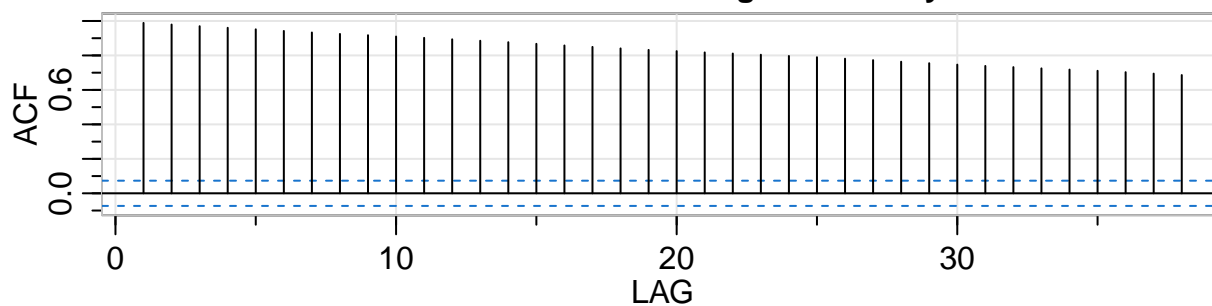
Fitting Seasonality into the data

```
##
## Call:
## lm(formula = interpolated ~ decimal + factor(`season corr`),
##     data = co2_mm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.183 -2.954 -1.397  2.185  9.754
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.776e+03  1.553e+01 -178.723  < 2e-16 ***
## decimal        1.574e+00  7.803e-03  201.719  < 2e-16 ***
## factor(`season corr`)2  6.289e-01  6.869e-01   0.916  0.360210
## factor(`season corr`)3  1.494e+00  6.842e-01   2.184  0.029284 *
## factor(`season corr`)4  2.643e+00  6.842e-01   3.863  0.000122 ***
## factor(`season corr`)5  2.931e+00  6.869e-01   4.267  2.24e-05 ***
## factor(`season corr`)6  2.223e+00  6.869e-01   3.236  0.001266 **
## factor(`season corr`)7  5.880e-01  6.869e-01   0.856  0.392286
## factor(`season corr`)8 -1.553e+00  6.869e-01  -2.261  0.024044 *
## factor(`season corr`)9 -3.248e+00  6.869e-01  -4.728  2.72e-06 ***
## factor(`season corr`)10 -3.340e+00  6.869e-01  -4.863  1.42e-06 ***
## factor(`season corr`)11 -2.134e+00  6.869e-01  -3.107  0.001962 **
## factor(`season corr`)12 -9.555e-01  6.869e-01  -1.391  0.164620
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.824 on 733 degrees of freedom
## Multiple R-squared:  0.9824, Adjusted R-squared:  0.9821
## F-statistic: 3408 on 12 and 733 DF, p-value: < 2.2e-16
```


Residuals after Fitting Seasonality



ACF & PACF after Fitting Seasonality

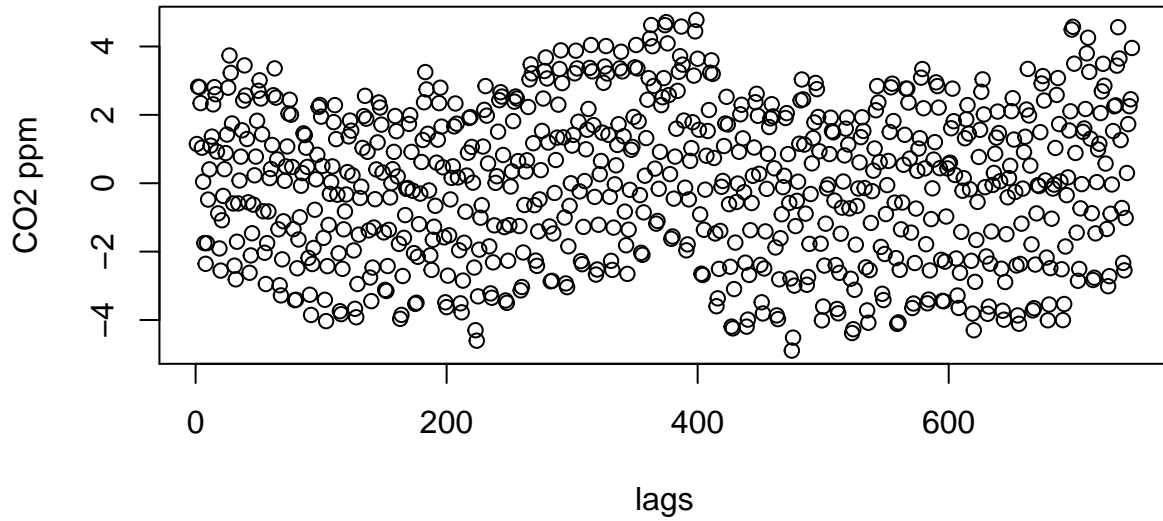


```
## Analysis of Variance Table
##
## Model 1: interpolated ~ decimal
## Model 2: interpolated ~ decimal + factor(`season corr`)
##   Res.Df  RSS Df Sum of Sq    F    Pr(>F)
## 1     744 14019
## 2     733 10721  11    3297.7 20.497 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

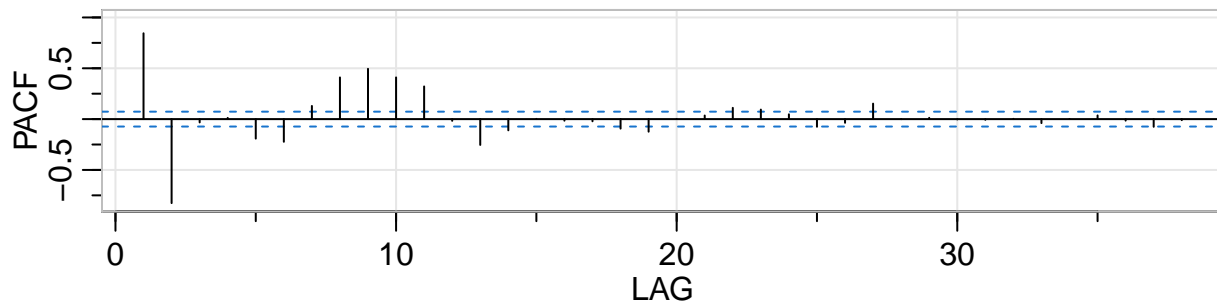
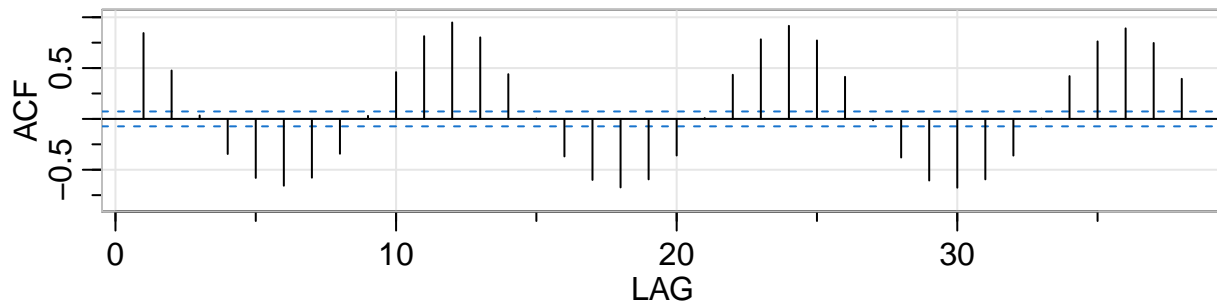
- *season corr* is the month in which the CO2 measurement was taken. The variable allows us to explore the role of seasonality plays in CO2 concentration.
- The months are in number format, and I converted them into factors so they would be analyzed as a categorical variable. They do not really behave in the same way that the numerical scale does. January and December are the farthest in the numerical scale, but probably have more in common than perhaps June and January.
- This model actually has a higher R squared than the one with just the date, coming in at a **0.982**. As you are able to see in the anova results above, the added predictor is actually very significant, and helps to capture a lot more information on the data.
- The residuals still show the quadratic trend, so that will be the next thing I add to the model.
- The ACF no longer has the sinusoidal behavior, but it remains significant for many lags. However, the PACF falls within the significance boundaries within one lag.

Fitting a Quadratic term into the Model

Residuals after Fitting a Quadratic Term



ACF & PACF after Fitting a Quadratic Term



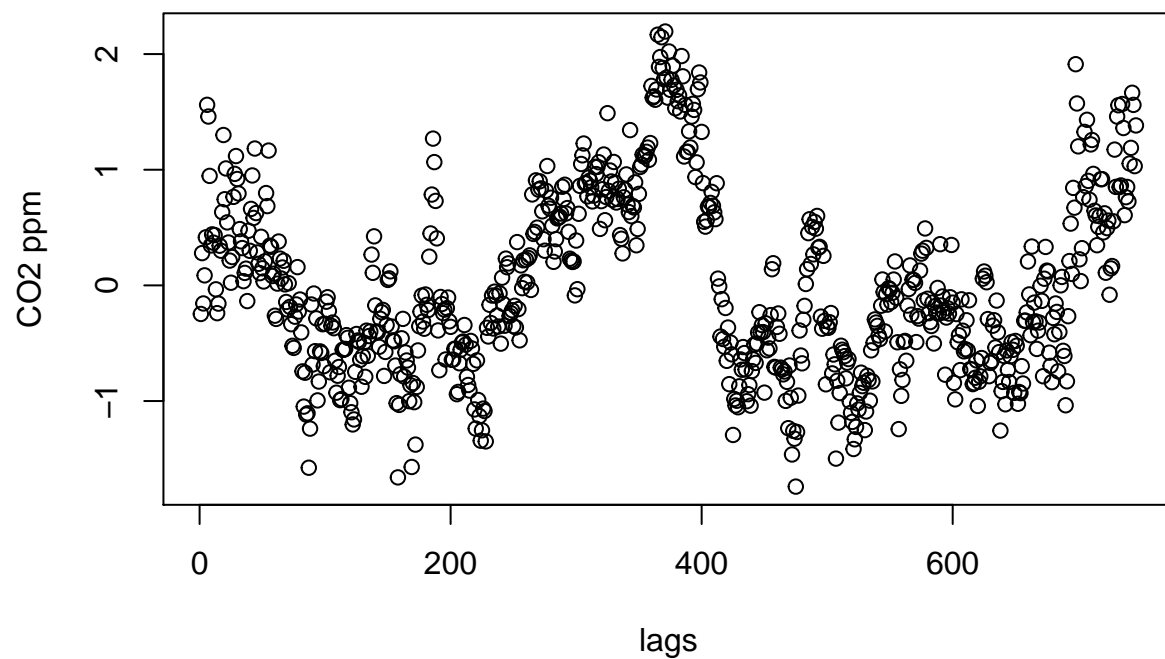
```
## Analysis of Variance Table
##
## Model 1: interpolated ~ decimal + factor(`season corr`)
## Model 2: interpolated ~ decimal + I(decimal^2)
##   Res.Df    RSS Df Sum of Sq F Pr(>F)
## 1     733 10721.1
## 2     743  3674.5 -10    7046.6
```

- We have finally come around to fitting a quadratic term into the model! I have chosen to add this predictor on its own before adding in seasonality in order to see how it behaves individually. Our R squared has again improved to a **0.994**.
- The residuals are constantly distributed around the zero line and there seems to be no apparent trend.
- As expected, when we remove seasonality from the model, the ACF returned to the sinusoidal behavior, and the PACF seems to lose significance at about 20 lags.

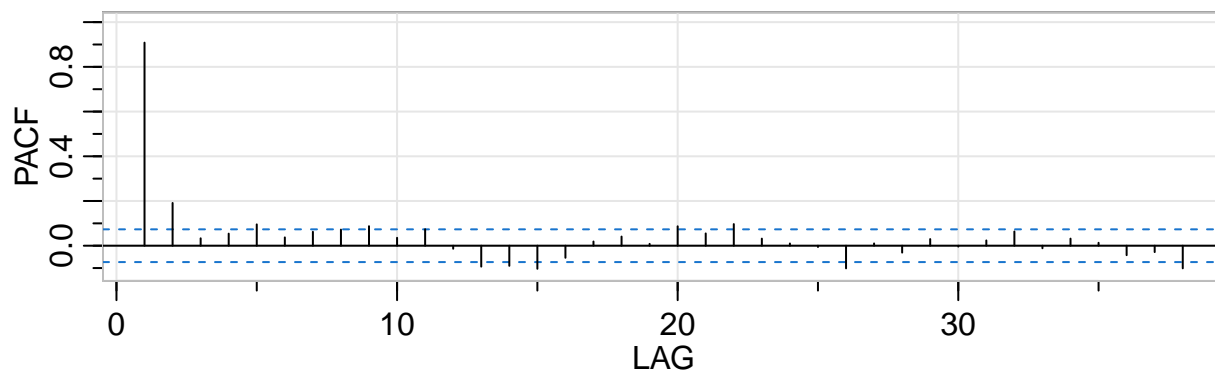
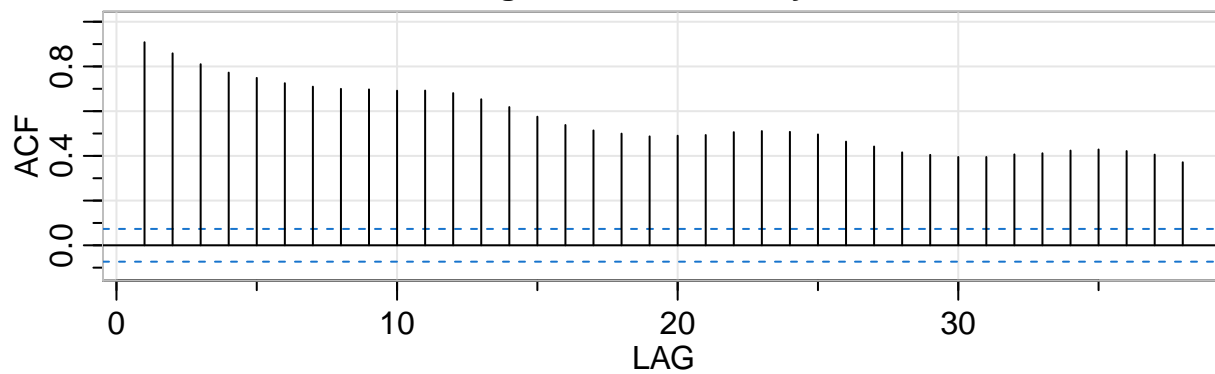
Combining Trend, Seasonality, and the Quadratic Term into one Model

```
##
## Call:
## lm(formula = interpolated ~ decimal + I(decimal^2) + factor(`season corr`),
##     data = co2_mm)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.7401 -0.5702 -0.1565  0.5624  2.1951
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.820e+04  3.934e+02  122.523 < 2e-16 ***
## decimal       -4.968e+01  3.956e-01 -125.603 < 2e-16 ***
## I(decimal^2)    1.288e-02  9.943e-05  129.584 < 2e-16 ***
## factor(`season corr`)2  6.282e-01  1.405e-01   4.471 9.00e-06 ***
## factor(`season corr`)3  1.361e+00  1.399e-01   9.727 < 2e-16 ***
## factor(`season corr`)4  2.510e+00  1.399e-01  17.936 < 2e-16 ***
## factor(`season corr`)5  2.931e+00  1.405e-01  20.860 < 2e-16 ***
## factor(`season corr`)6  2.223e+00  1.405e-01  15.823 < 2e-16 ***
## factor(`season corr`)7  5.885e-01  1.405e-01   4.189 3.14e-05 ***
## factor(`season corr`)8 -1.552e+00  1.405e-01 -11.049 < 2e-16 ***
## factor(`season corr`)9 -3.246e+00  1.405e-01 -23.109 < 2e-16 ***
## factor(`season corr`)10 -3.339e+00  1.405e-01 -23.770 < 2e-16 ***
## factor(`season corr`)11 -2.133e+00  1.405e-01 -15.186 < 2e-16 ***
## factor(`season corr`)12 -9.550e-01  1.405e-01  -6.798 2.20e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7822 on 732 degrees of freedom
## Multiple R-squared:  0.9993, Adjusted R-squared:  0.9993
## F-statistic: 7.65e+04 on 13 and 732 DF, p-value: < 2.2e-16
## Analysis of Variance Table
##
## Model 1: interpolated ~ decimal + I(decimal^2)
## Model 2: interpolated ~ decimal + I(decimal^2) + factor(`season corr`)
##   Res.Df    RSS Df Sum of Sq    F    Pr(>F)
## 1      743 3674.5
## 2      732  447.8 11    3226.7 479.46 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

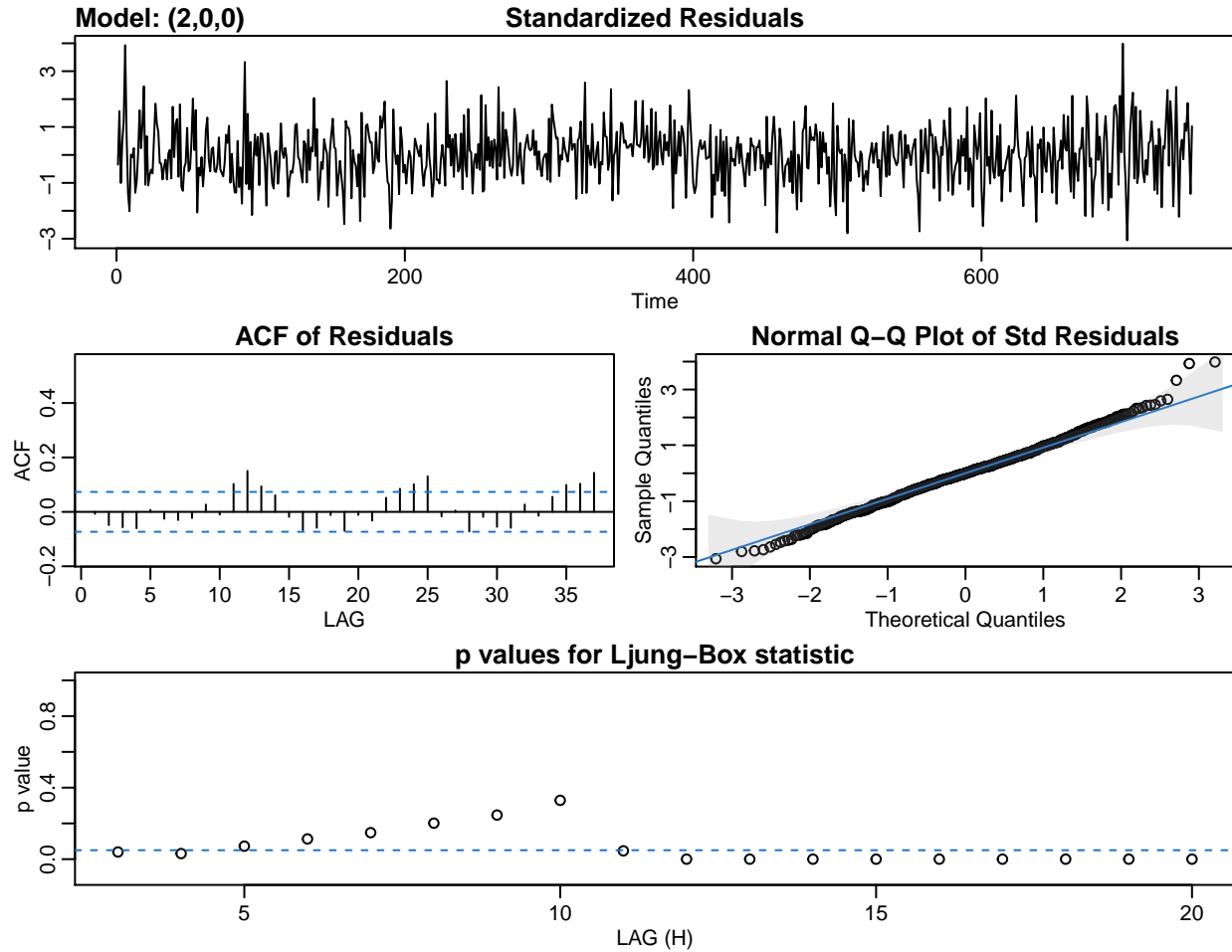
Residuals after Fitting Trend, Seasonality, and the Quadratic Term



ACF & PACF after Fitting Trend, Seasonality, and the Quadratic Term



```
## initial value -0.253963
## iter 2 value -0.340034
## iter 3 value -1.111815
## iter 4 value -1.129517
## iter 5 value -1.153599
## iter 6 value -1.156448
## iter 7 value -1.156464
## iter 8 value -1.156464
## iter 9 value -1.156464
## iter 10 value -1.156467
## iter 11 value -1.156478
## iter 12 value -1.156478
## iter 12 value -1.156478
## iter 12 value -1.156478
## final value -1.156478
## converged
## initial value -1.154837
## iter 2 value -1.154838
## iter 3 value -1.154845
## iter 4 value -1.154845
## iter 4 value -1.154845
## iter 4 value -1.154845
## final value -1.154845
## converged
```



- Here we combine all the different elements that we have found to be significant into one model. The resulting model has an **R squared of 0.9992**, and all the predictors included in the model are significant.
- When we compare this model with our previous one, the ANOVA table shows that this is a significantly better model, as our p-value is $< 2.2e-16$.

Components of a SARIMA Model

component	Description
p	number of autoregressive terms
d	number of nonseasonal differences needed for stationarity
p	number of moving average terms
P	number of seasonal autoregressive terms
D	number of seasonal differences
Q	number of seasonal moving averages
m	number of periods per season

Fitting a SARIMA Model to our Data

```
##
## Fitting models using approximations to speed things up...
##
## ARIMA(2,1,2)(1,1,1)[12] : 419.7603
## ARIMA(0,1,0)(0,1,0)[12] : 878.2927
## ARIMA(1,1,0)(1,1,0)[12] : 608.7397
## ARIMA(0,1,1)(0,1,1)[12] : 439.8985
## ARIMA(2,1,2)(0,1,1)[12] : 430.2981
## ARIMA(2,1,2)(1,1,0)[12] : 592.9529
## ARIMA(2,1,2)(2,1,1)[12] : 401.5694
## ARIMA(2,1,2)(2,1,0)[12] : 499.2808
## ARIMA(2,1,2)(2,1,2)[12] : 399.6999
## ARIMA(2,1,2)(1,1,2)[12] : 411.4671
## ARIMA(1,1,2)(2,1,2)[12] : 391.1464
## ARIMA(1,1,2)(1,1,2)[12] : 407.8446
## ARIMA(1,1,2)(2,1,1)[12] : 393.62
## ARIMA(1,1,2)(1,1,1)[12] : 413.037
## ARIMA(0,1,2)(2,1,2)[12] : 393.9551
## ARIMA(1,1,1)(2,1,2)[12] : 390.1578
## ARIMA(1,1,1)(1,1,2)[12] : 406.2688
## ARIMA(1,1,1)(2,1,1)[12] : 392.4911
## ARIMA(1,1,1)(1,1,1)[12] : 411.0685
## ARIMA(0,1,1)(2,1,2)[12] : 393.8133
## ARIMA(1,1,0)(2,1,2)[12] : 411.0226
## ARIMA(2,1,1)(2,1,2)[12] : 398.9569
## ARIMA(0,1,0)(2,1,2)[12] : Inf
## ARIMA(2,1,0)(2,1,2)[12] : 406.1716
##
## Now re-fitting the best model(s) without approximations...
##
## ARIMA(1,1,1)(2,1,2)[12] : 389.2503
##
## Best model: ARIMA(1,1,1)(2,1,2)[12]
##
## Series: co_ts
## ARIMA(1,1,1)(2,1,2)[12]
##
## Coefficients:
##          ar1      ma1      sar1      sar2      sma1      sma2
##          0.1858 -0.5551 -0.3729 -0.0337 -0.4951 -0.3114
```

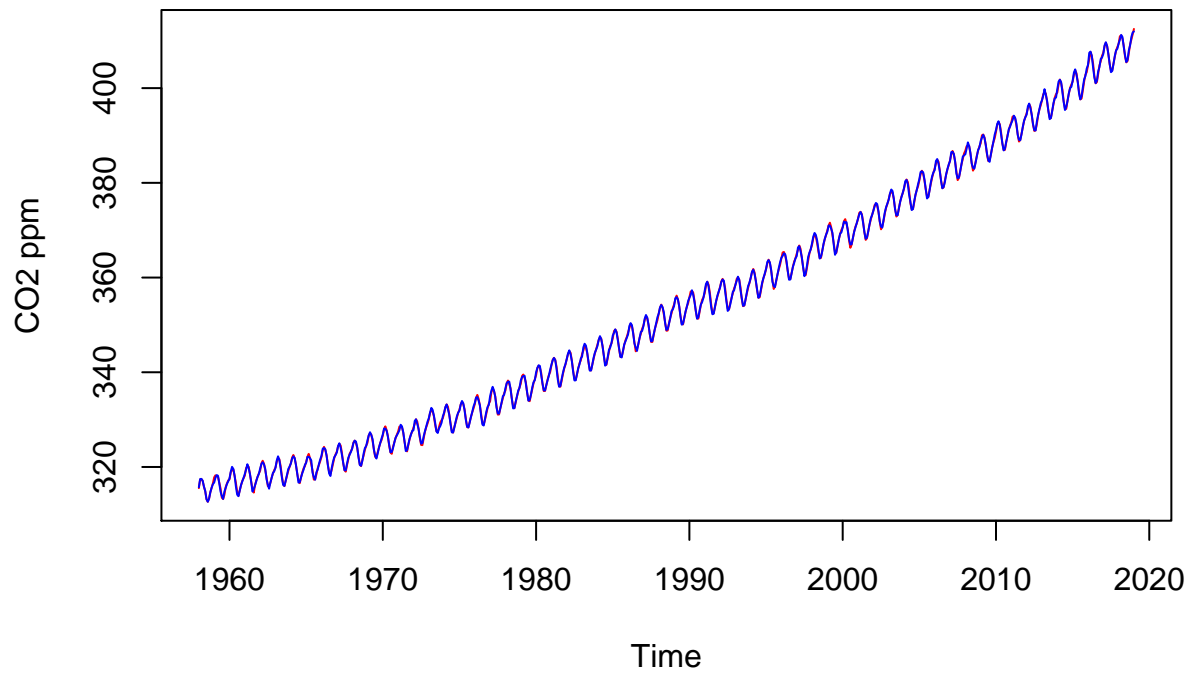


```
## s.e.  0.0947   0.0808   0.7338   0.0432   0.7335   0.6343
##
## sigma^2 estimated as 0.09891:  log likelihood=-187.55
## AIC=389.09   AICc=389.25   BIC=421.15
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.02510703 0.3103991 0.2403723 0.006817949 0.06800183 0.1535048
##              ACF1
## Training set -0.002163006
```

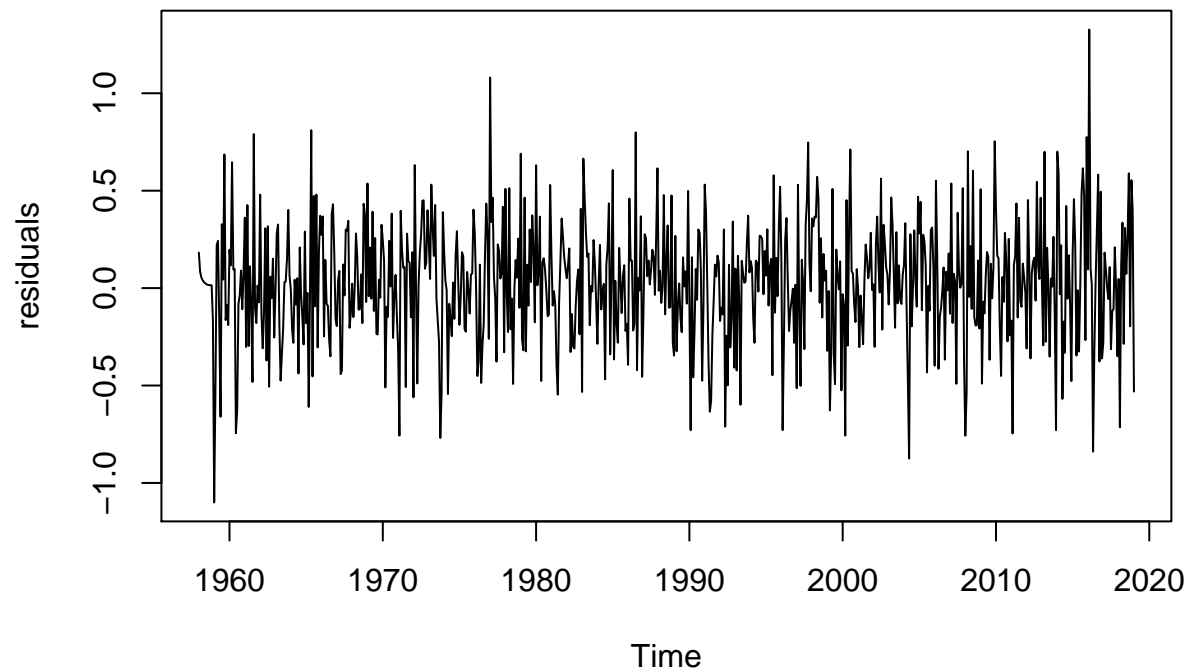
- I used the `auto.arima()` function in order to help me find the best SARIMA model for the data. This function uses either AIC, AICc or BIC value to find the best model, and the results that are returned are much like the ones that I expected from exploring the data.
- p & P: While investigating the data, I suspected that the number of autoregressive lags in the data was going to be two, and here we see that the small $p = 1$. When we look at the autoregressive lags associated to seasonality, this does in fact turn out to be equal to two.
- d & D: As shown during the exploration part of the analysis, one difference was enough to make the time series stationary. This is reflected by the model that R choses, as both d & $D = 1$.
- q & Q: The moving averages chosen for this model were 2. This was expected based on the fact that every model that we tested on this data set returned ACF plots which remained significant for many lags.

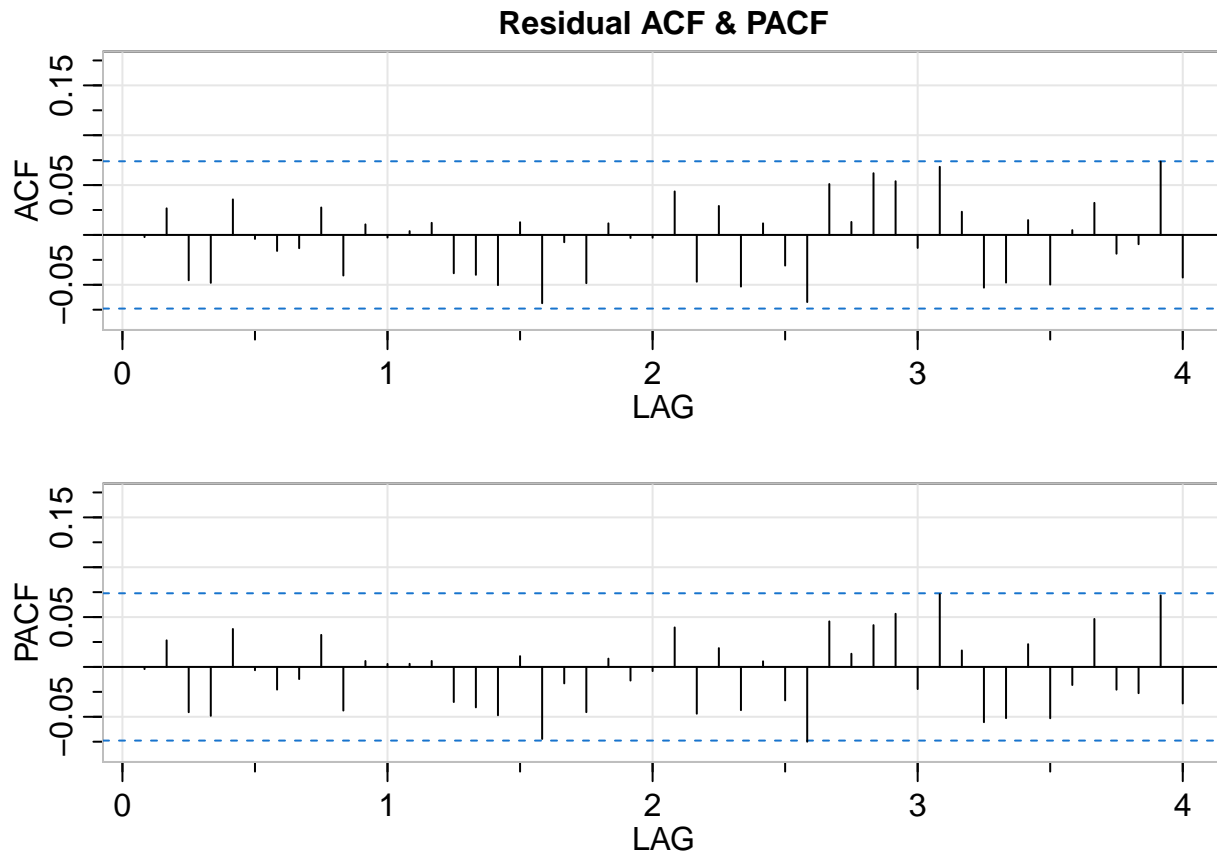
Validating the Model

Predicted v Actual CO2 ppm Values



Residuals for the ARIMA Model

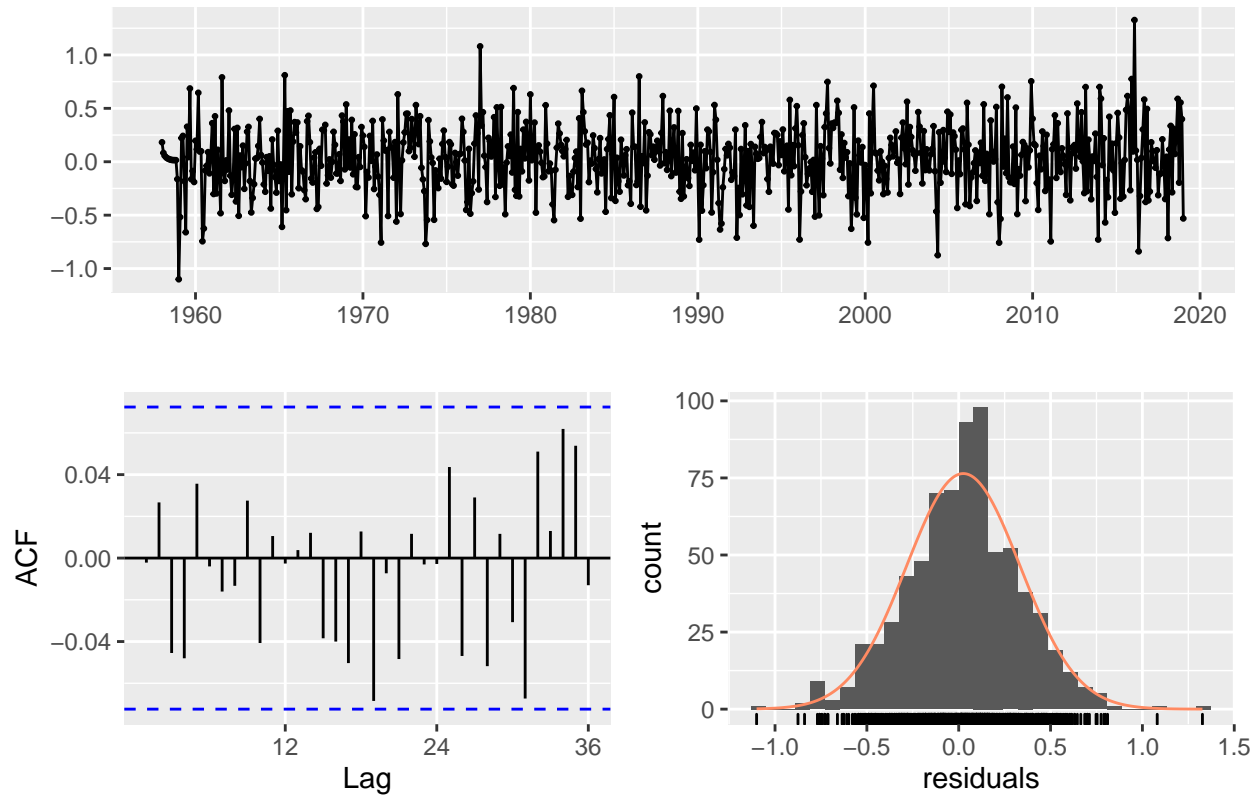




- When we plot the predicted values against the actual data, the two are virtually indistinguishable which means that our model is a good fit for our data.
- The residuals are centered around zero and have a constant variance so it seems like nothing other than white noise is left unaccounted for.
- The ACF & PACF both fall within the significance limits!

Ljung-Box test

Residuals from ARIMA(1,1,1)(2,1,2)[12]

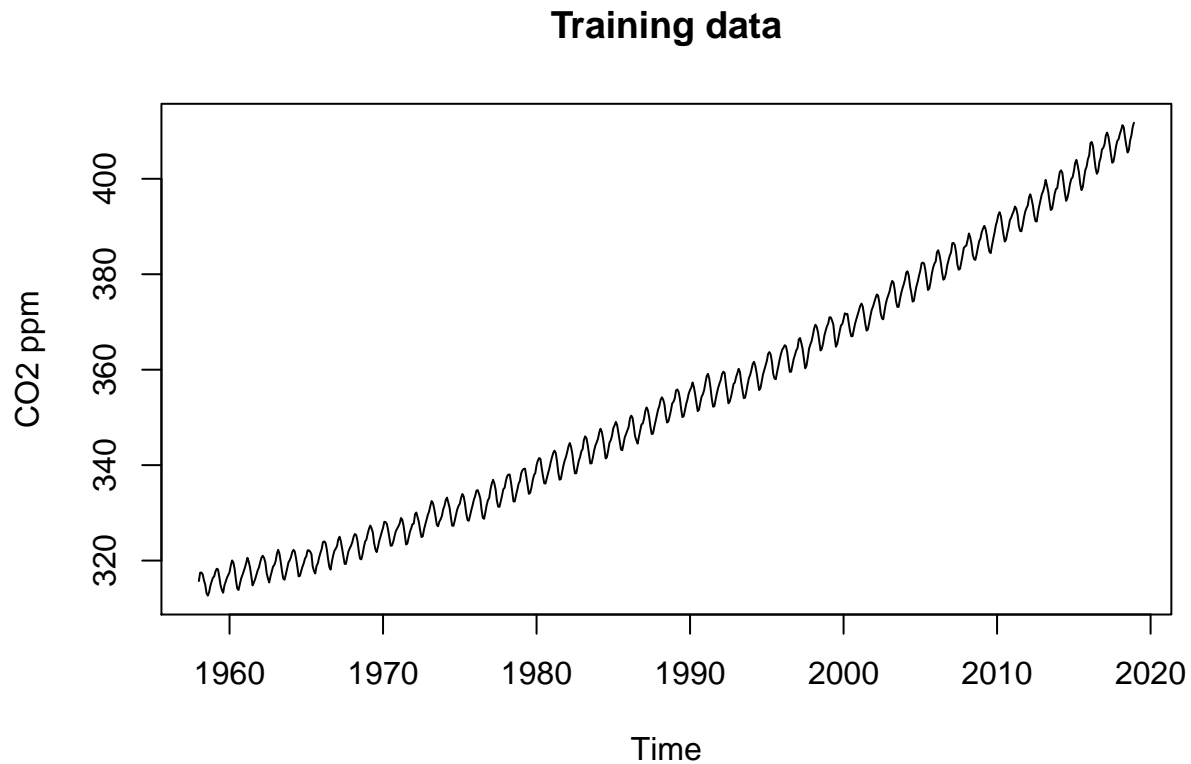


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(2,1,2)[12]
## Q* = 6.9202, df = 6, p-value = 0.3283
##
## Model df: 6.    Total lags used: 12
```

- The Ljung-Box test tests whether the residuals of a model are actually white noise, or if there's some kind of trend that we did not account for. If the remaining residuals are in fact white noise, then we know that the model that was selected is actually a good fit for the data. The null hypothesis of the test is that the model properly fits the data, which means that if our resulting p-value is significant we do not have a good model. Fortunately, our **p-value = 0.3246** which means that our model passed the Ljung-Box test and I can confidently use it to make a forecast.

Forecasting

Creating Training and Testing Data set:



- Data omitted: The training data is composed of the monthly CO2 ppm means for the years 1958 to the end of 2017. Therefore, there is an omission of a little over a year.
- Test data: The test data is from the beginning of 2019 through month 4 of 2020.

Creating the SARIMA Model

```
##
## Fitting models using approximations to speed things up...
##
## ARIMA(2,1,2)(1,1,1)[12] : 417.3212
## ARIMA(0,1,0)(0,1,0)[12] : 874.2091
## ARIMA(1,1,0)(1,1,0)[12] : 606.937
## ARIMA(0,1,1)(0,1,1)[12] : 437.5698
## ARIMA(2,1,2)(0,1,1)[12] : 428.1664
## ARIMA(2,1,2)(1,1,0)[12] : 591.773
## ARIMA(2,1,2)(2,1,1)[12] : 398.8025
## ARIMA(2,1,2)(2,1,0)[12] : 497.5467
## ARIMA(2,1,2)(2,1,2)[12] : 396.9926
## ARIMA(2,1,2)(1,1,2)[12] : 408.9982
## ARIMA(1,1,2)(2,1,2)[12] : 388.5148
## ARIMA(1,1,2)(1,1,2)[12] : 405.4638
## ARIMA(1,1,2)(2,1,1)[12] : 390.8945
## ARIMA(1,1,2)(1,1,1)[12] : 410.6434
```

```

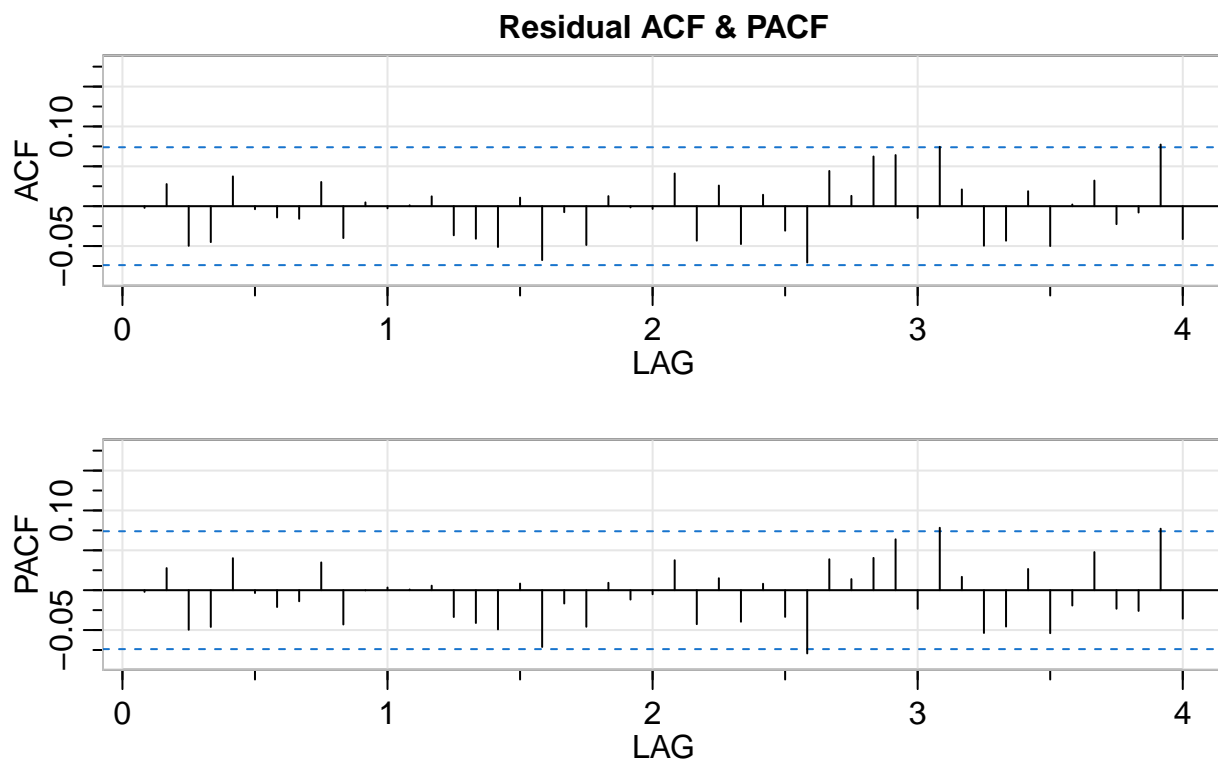
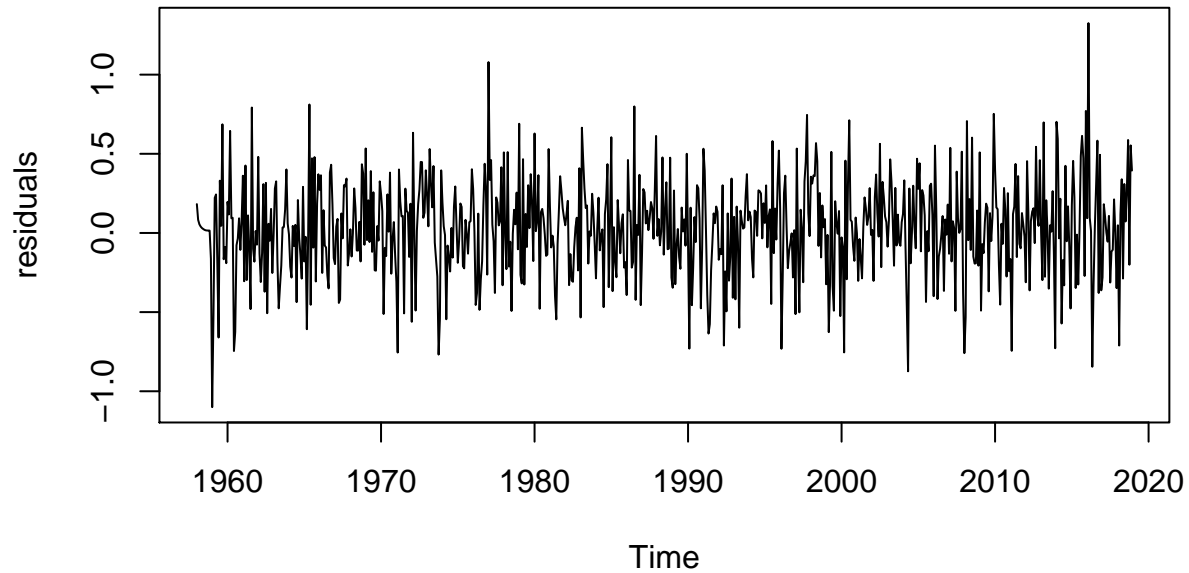
## ARIMA(0,1,2)(2,1,2)[12] : 391.3334
## ARIMA(1,1,1)(2,1,2)[12] : 387.5738
## ARIMA(1,1,1)(1,1,2)[12] : 403.8803
## ARIMA(1,1,1)(2,1,1)[12] : 389.8328
## ARIMA(1,1,1)(1,1,1)[12] : 408.6687
## ARIMA(0,1,1)(2,1,2)[12] : 390.9683
## ARIMA(1,1,0)(2,1,2)[12] : 407.071
## ARIMA(2,1,1)(2,1,2)[12] : 396.3372
## ARIMA(0,1,0)(2,1,2)[12] : Inf
## ARIMA(2,1,0)(2,1,2)[12] : 403.1876
##
## Now re-fitting the best model(s) without approximations...
##
## ARIMA(1,1,1)(2,1,2)[12] : 386.7956
##
## Best model: ARIMA(1,1,1)(2,1,2)[12]
## Series: co2_training
## ARIMA(1,1,1)(2,1,2)[12]
##
## Coefficients:
##          ar1      ma1      sar1      sar2      sma1      sma2
##          0.1806 -0.5469 -0.4106 -0.0357 -0.4571 -0.3439
## s.e. 0.0961 0.0827 0.6973 0.0438 0.6968 0.6037
##
## sigma^2 estimated as 0.09865: log likelihood=-186.32
## AIC=386.64 AICc=386.8 BIC=418.68
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.02554781 0.3099845 0.2399844 0.006916028 0.0679221 0.1533923
##              ACF1
## Training set -0.002077693

```

- Only the testing data was used to create the SARIMA model in order to use the remaining data as a way to cross validate our model through forecasting. The resulting model is an ARIMA(1,1,1)(2,1,2)12, which is the same one that resulted from the complete data set.

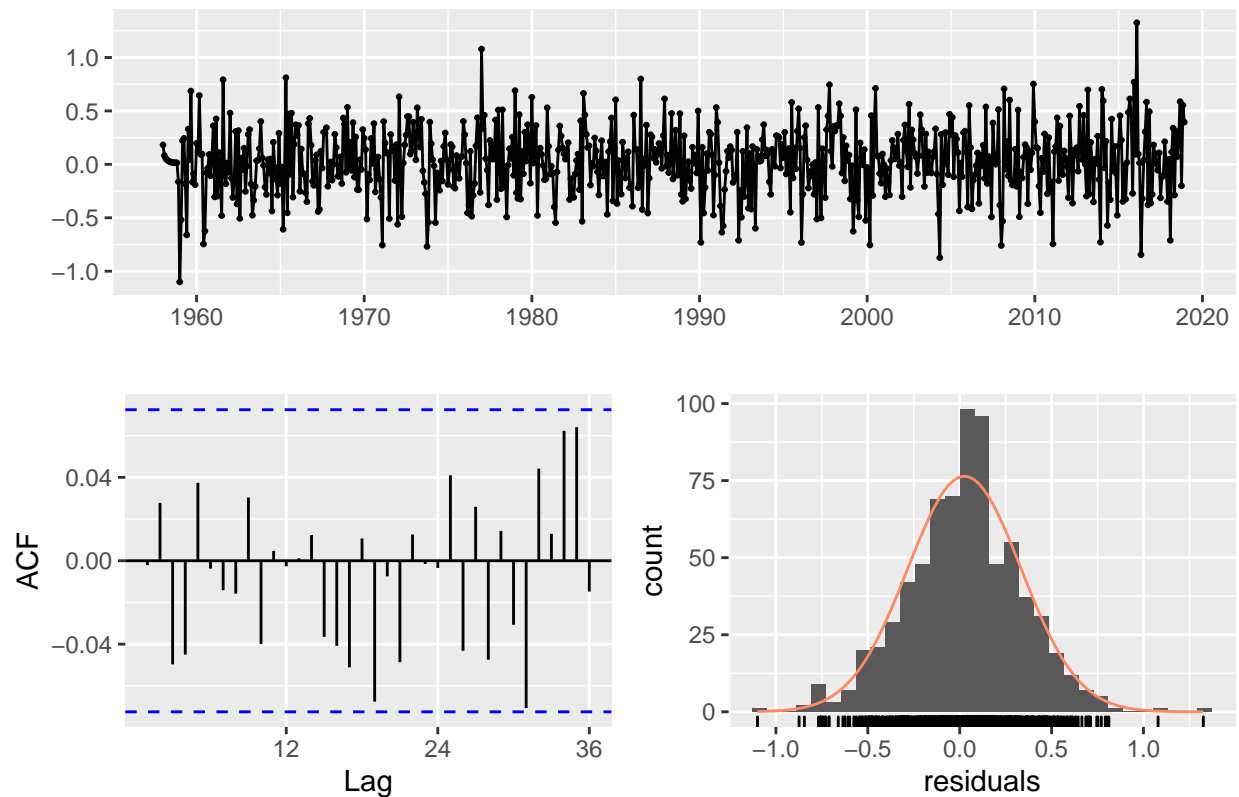
Validating the Forecasting Model

Residuals for the Forecasting Model: ARIMA(1,1,1)(2,1,2)₁₂



- Just by looking at the remaining residuals from our forecasting model, we can see that the errors are white noise. The ACF and PACF fall within the significance limits; these two plots are strong indicators that the model will pass the Ljung-Box test.

Residuals from ARIMA(1,1,1)(2,1,2)[12]

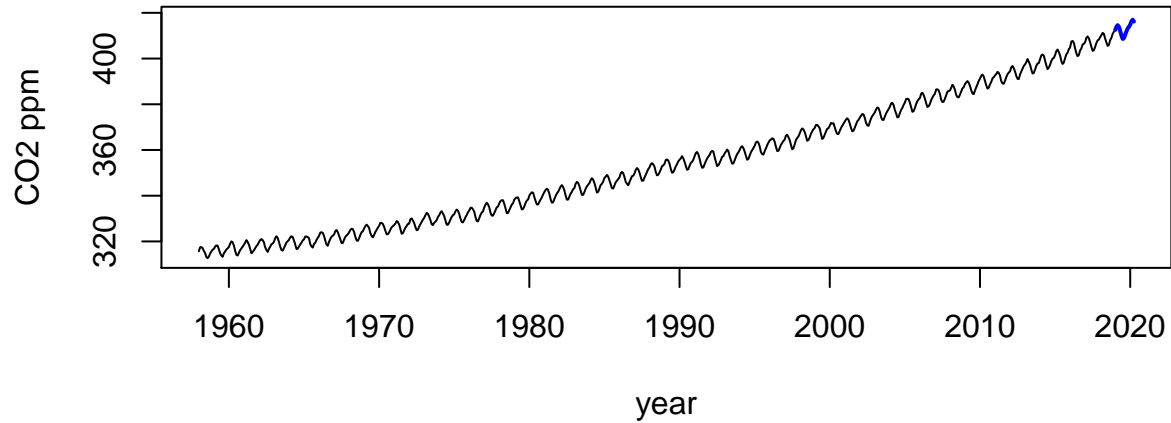


```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)(2,1,2)[12]
## Q* = 16.971, df = 18, p-value = 0.5251
##
## Model df: 6.    Total lags used: 24
```

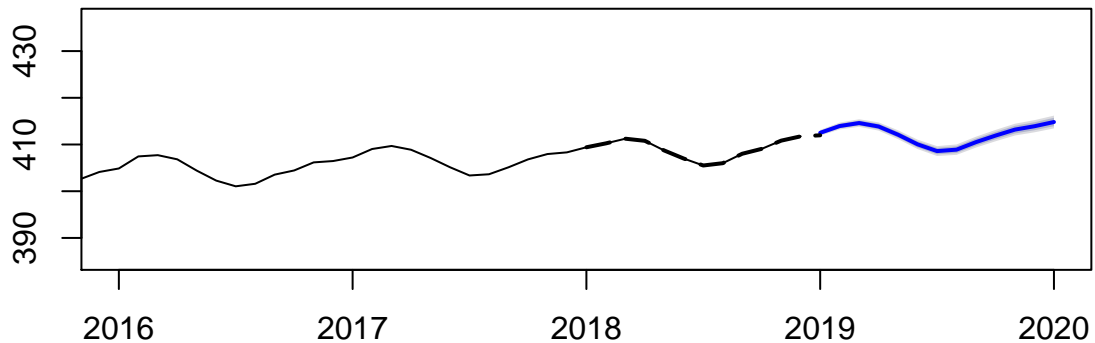
- As expected, or **p-value = 0.4137** so we know that our model is a good fit for our data. I can confidently use the model to create some forecasts.

Forecasts

Forecasts from ARIMA(1,1,1)(2,1,2)[12]



Forecasts from ARIMA(1,1,1)(2,1,2)[12]



```
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set  0.02554781 0.3099845 0.2399844 0.006916028 0.0679221 0.1533923
## Test set     -0.53558910 0.5355891 0.5355891 -0.130006820 0.1300068 0.3423358
##               ACF1
## Training set -0.002077693
## Test set     NA
```

- The blue line on the first plot shows the forecasts made by the model. Everything previous to that is part of the data used to train the model.
- The second plot zooms in on the observations leading up to the forecast and the forecasts the model made. The lightly shaded blue area around the dark blue line demonstrates the 95% confidence interval. As you can see, the true observations for this time period fall within the confidence interval of the forecasting model, so it actually did a good job forecasting those values.

Results and Insights

The model used to forecast carbon dioxide concentration in the atmosphere was a Seasonal ARIMA(1,1,1)(2,1,2)₁₂

1. **Seasonality** is an important component to explaining the variation that is present in the CO2 ppm. This means that the CO2 ppm levels naturally fluctuate up and down during certain times of the year.
 - CO2 levels are **significantly higher** during the months of **February - July** compared to the levels seen during January.
 - February has the highest average CO2 levels.
 - CO2 levels are **significantly lower** during the months of **August - December** compared to the typical levels experienced during January.
 - December has the lowest average CO2 levels.
2. **Trend** is a crucial component to modeling and understanding the variation in the CO2 atmospheric concentration. As we have seen throughout the data visualization and the final SARIMA model, there is a positive trend in the level of CO2 present in our atmosphere. As the years have gone by there is a constant, predictable increase in the CO2 concentration.
3. **Forecasting & Forecasting Accuracy** I tested the accuracy of the model by using forecasting.
 - I used data from 1958 to 2018 train the model
 - The forecasts were made for CO2 values from January 2019 to April 2020.
 - Forecasting Accuracy can be gauged by analyzing forecasting errors. These values are in the same scale of the data, and the CO2 ppm values range from 312.66 to 416.21ppm. Error terms that are close to zero, either positive or negative, indicate that the forecasts are unbiased and accurate. The forecast errors for the SARIMA model I created, are printed below: **the errors are close to zero, indicating that my model is unbiased and accurately predicted the CO2 ppm values.**

Forecast errors for the SARIMA model

	ME	RMSE	MAE	MPE	MAPE	MASE
## Training set	0.02554781	0.3099845	0.2399844	0.006916028	0.0679221	0.1533923
## Test set	-0.53558910	0.5355891	0.5355891	-0.130006820	0.1300068	0.3423358
##	ACF1					
## Training set	-0.002077693					
## Test set	NA					