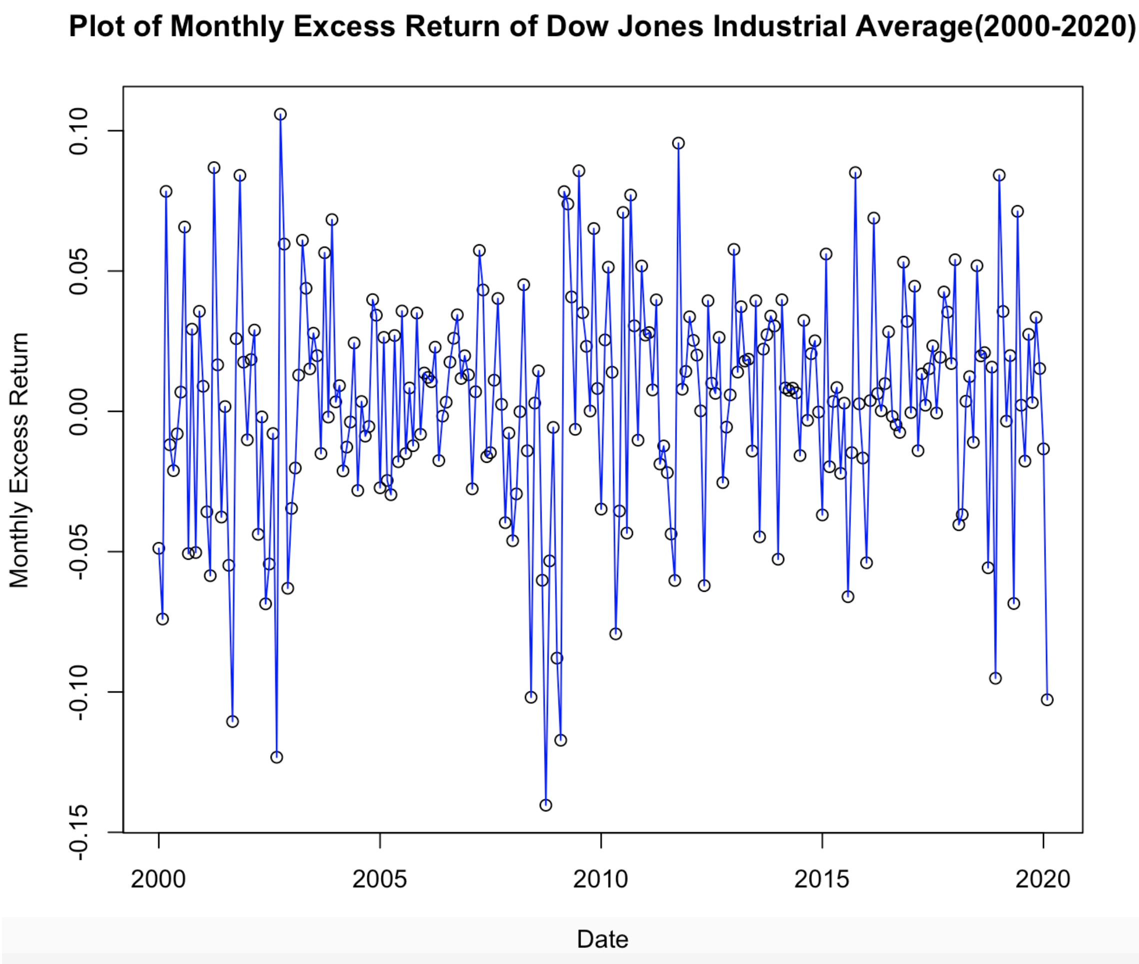
Forecasting the Monthly Excess Return of March and April, 2020

In this report, I want to discuss and forecast Dow Jones Industrial Average monthly excess return of March and April, 2020. This forecaster is based on the data from Jan, 2020 to Feb, 2020.

**Check for the seasonal trend or other trend:**

Firstly, time series data of monthly excess return from 2000, Jan to 2020, February was collected to plot a time series diagram.




From the plot, this time series model seems to be random fluctuations over time. Thus, this model has no trend over time, since there is no increase or decrease trend over time that we can see. No need for removing trend.

**Test for stationarity:**

We use Ljung-Box test for testing stationarity, then we get our p-value is 0.3639. According to our result of Ljung-Box test, we have no evidence to reject null hypothesis, since the p-value is larger than 0.05. Namely, it means that there has no significant evidence for non-zero correlation at lags 1-20.

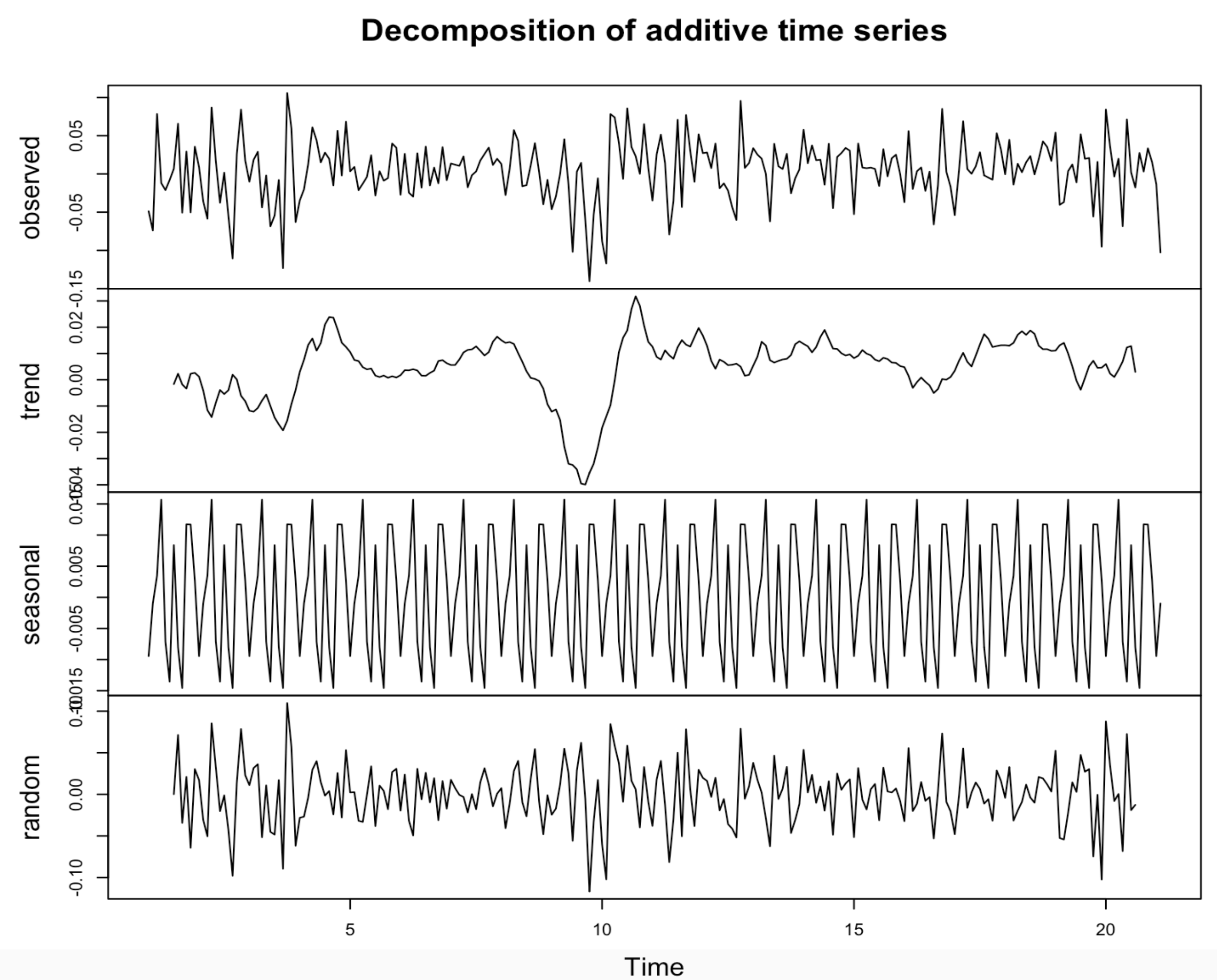
```
> adf.test(ex_r)

	Augmented Dickey-Fuller Test

data:  ex_r
Dickey-Fuller = -5.6338, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary
```

we now using the Augmented Dickey-Fuller Test for testing the stationarity. The null hypothesis is saying that if p-value is larger than 0.05 then it indicates non-stationarity, otherwise it means that the model is stationarity. So, in this model, it gives us a p-value is equal to 0.01, which is less than 0.05, so we can conclude that our model is stationary. However, we need to decompose our time series to make sure that there is no seasonal, because if it is seasonal, it still could be stationary.

**Decompose a Time Series:**
We decompose time series to explore seasonality.

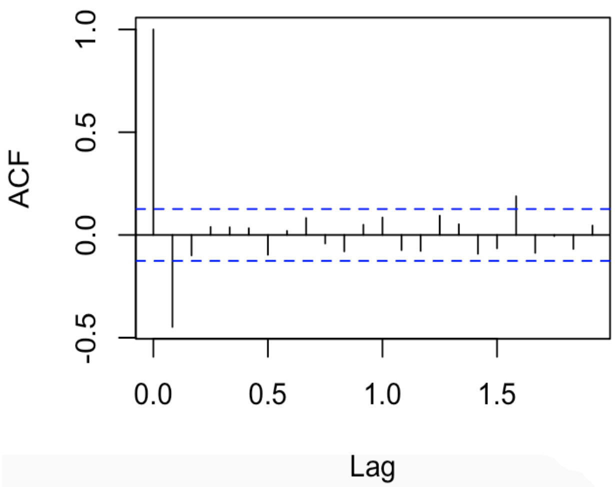


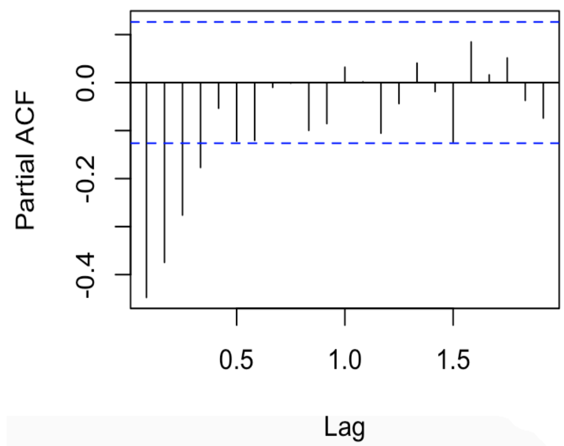

Check from the plots, there may be some seasonality exists in our time series object. Therefore, we need to transform our time series object.

**Transforming Time Series Object:**
We need to do the differencing across its consecutive values to make a stationary time series. Differencing can be used to stabilize the mean, so that we can make the time series stationary.

**Diagnosing the ACF and PACF plots:**

Since we got the stationary time series, we need to exploit ACF and PACF plots to check two problems 1. Is it this AR or MA process. 2. What order of AR or MA process do we need to use? So, in order to figure out these two problem, I need to draw ACF and PACF plots. Which means that we are looking for the lags for AR and MA series.




From the PACF, we see that it is a geometric. From the ACF plot, I saw that there are three significant lags. So use the model MA(2). Since we use the differenced time series, I need to use the combined model ARIMA, USE the function auto.arima in r, I got that the model is ARIMA(2,0,2)(2,0,1)[12]

```
Coefficients:
          ar1      ar2     ma1     ma2     sar1   sar2    sma1    mean
      -0.1145  -0.9496  0.1157  0.8865  -0.7475  0.057  0.7446  0.0040
s.e.   0.0613   0.0452  0.0922  0.0654   0.2916  0.077  0.2885  0.0026

sigma^2 estimated as 0.001649:  log likelihood=435.67
AIC=-853.34   AICc=-852.57   BIC=-821.94
```

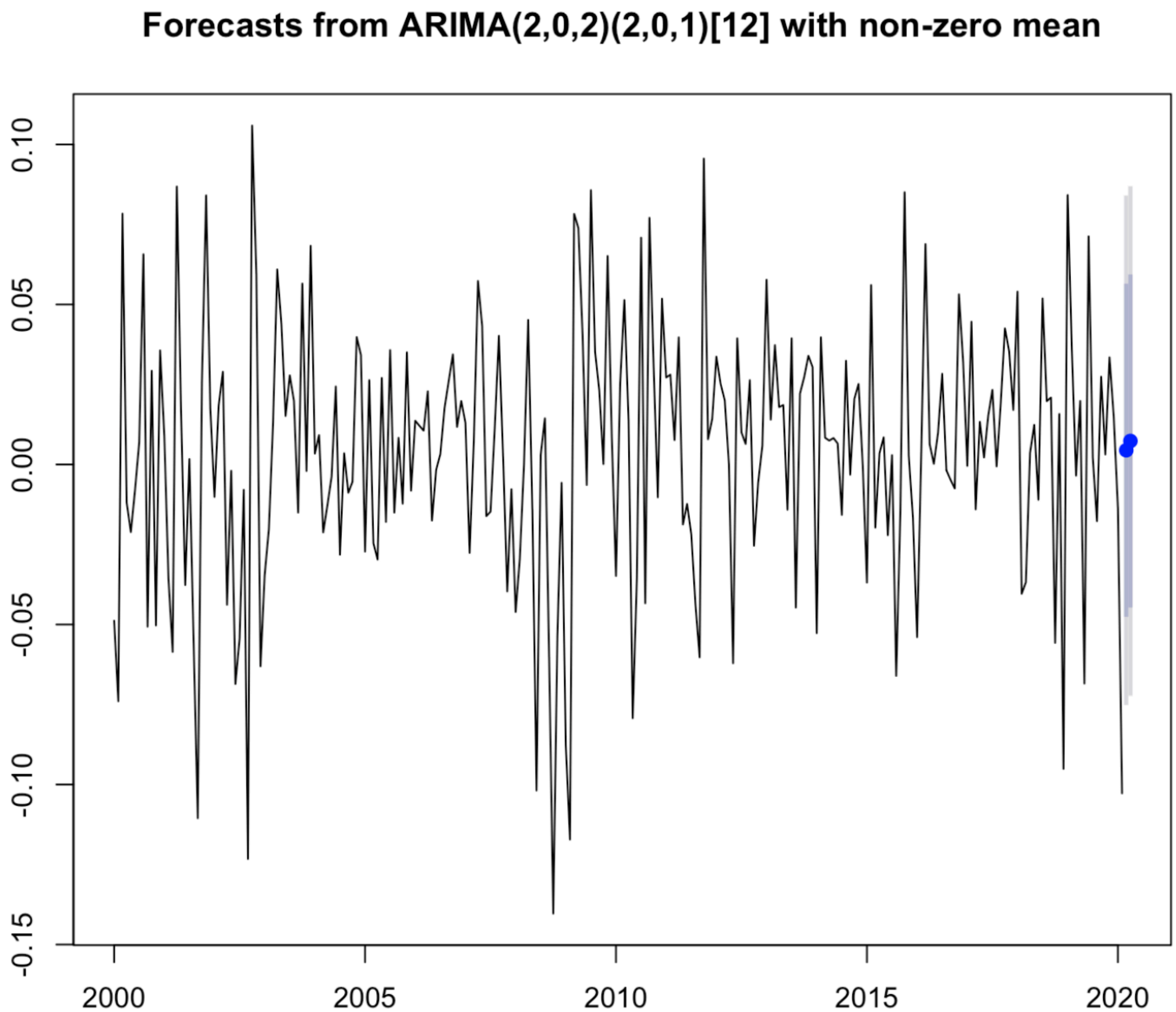
which means our time series object is general seasonal ARIMA model.

It forms as ARIMA(p,d,q)x(P,D,Q),
where P=number of seasonal autoregressive terms
D=number of seasonal differences
Q=number of seasonal moving average terms

Forecast from ARIMA(2,0,2)(2,0,1)[12]:
To visualize the predictions for the March and April 2020:

```
         Point Forecast       Lo 80      Hi 80       Lo 95      Hi 95
Mar 2020    0.004442619 -0.04759096 0.05647620 -0.07513588 0.08402111
Apr 2020    0.007334888 -0.04469873 0.05936851 -0.07224367 0.08691344
```

Forecasting the Monthly Excess Return of March and April, 2020




The blue points are the forecasting points of March and April 2020 with the 80% prediction intervals as a light blue shaded area, and the 95% prediction intervals as a grey shaded area.

## Bibliography:

Eulogio, Raul, January 30, 2018, https://blogs.oracle.com/datascience/performing-a-time-series-analysis-on-the-sandp-500-stock-index