US-EARNINGS

1. Analysis objective

The 'US_Earnings' dataset contains data on the earnings generated by listed American companies, on a quarterly basis, from 1950 to the third quarter of 2021.

The purpose of this paper is to provide an estimate as accurate as possible about the value that these profits will have in the near future, with the aim of understanding whether or not there will be a growth in earnings. The approach adopted broadly follows the Box Jenkins Modeling Approach.

2. Description of the dataset

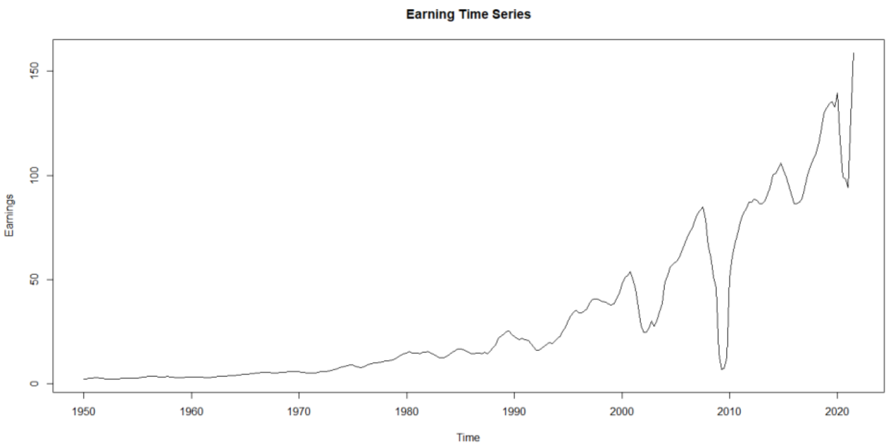

Figure 1- Historical earnings series

The preliminary exploratory analysis of the data confirms that earnings have been growing over time, with cycles of instability starting in the early 2000s and a sharp decrease in the 2008-2010 crisis period, then confirming themselves to new highs in the following decade. A notable decline in earnings was then seen to coincide with the start of the 2019-2020 pandemic period. The average earnings considering the entire period is equal to 32 billion dollars, with a minimum of 2.32 and a maximum reached of 175 billion dollars.

Once the time series has been created using the 'tseries' package, the approach followed suggests a stationarity analysis of the series which will be performed here using a unit root test, the most famous of which is the ADF-test.

Even if the series does not seem stationary as it clearly shows the presence of an increasing trend over time, with the ADF-test we confirm this result, affirming in fact the non-stationarity with a p-value of 0.95, as shown in figure 2.

```
        Augmented Dickey-Fuller Test

data:  US_Earnings.TS
Dickey-Fuller = -0.86534, Lag order = 6, p-value = 0.955
alternative hypothesis: stationary
```

Figure 2- Augmented Dickey-Fuller Test

The next step is to differentiate the series by a delay (d = 1) and repeat the ADF test to check if with a delay the series is made stationary. In this case, a delay is more than sufficient, the p-value shown by the test is significant so we can reject the null hypothesis and validate the alternative hypothesis, effectively making the series stationary, as shown in Figure 3.

```
        Augmented Dickey-Fuller Test

data:  ts_diff
Dickey-Fuller = -7.6868, Lag order = 6, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(ts_diff) : p-value smaller than printed p-value
```

Figure 3- Augmented Dickey-Fuller Test with one lag

Subsequently, a graphic analysis of seasonality allows me to generally affirm the absence of the latter in the series itself, or at least, if it is present, in any case in a slight form. In any case, the seasonal adjustment was performed using the moving average method.

3. Analysis process

The next step is to select the appropriate ARIMA model, which means finding the most consistent values of p and q for an ARIMA model (p, d, q). It is usually necessary to examine the correlogram and partial correlogram of stationary time series in graphic form.

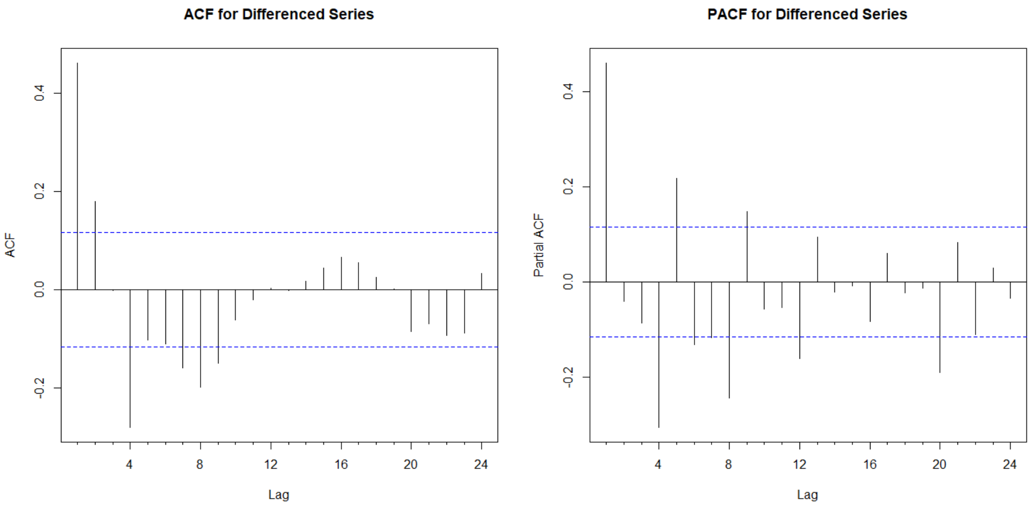

Figure 4- ACF and PACF

In light of the results obtained from the graphs just mentioned, as shown in figure 4 it would seem that the most suitable values are p = 1 and q = 3.

The ARIMA model (1,1,3) using the maximum likelihood method does not offer the lowest value of the AIC information criterion (equal to 1530) nor of the BIC (equal to 1549). Furthermore, the Ljung-Box test confirms that there is a slight presence of serial autocorrelation, given the observed value of the p-value.

The choice then moved to an ARIMA model (1,1,4) which is better from the point of view of AIC and BIC, results respectively 1516 and 1538. Furthermore, the Ljung-Box test confirms the absence of autocorrelation residual serial, with a p-value equal to 0.1823.

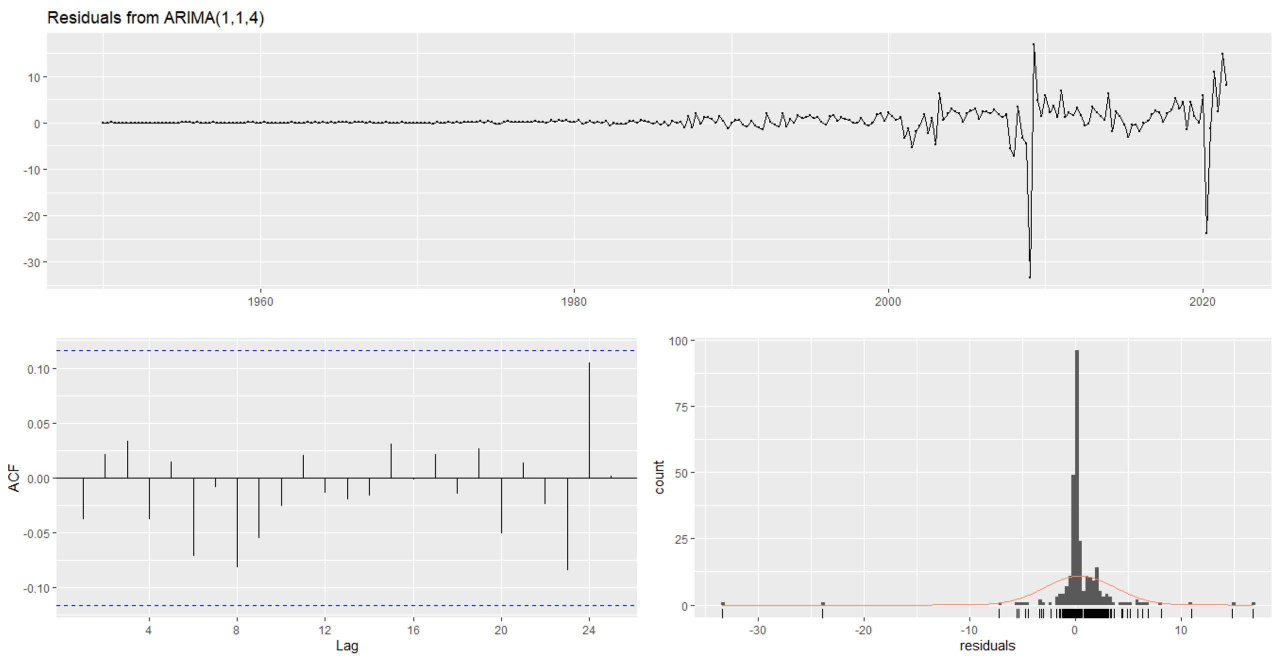

Figure 5- Arima model (1,1,4)

Other ARIMA models (p, d, q) have been tested but with less satisfactory results than the latter shown above, so the final choice falls on values of p = 1 and q = 4. Other values tested for these parameters mostly generate serial autocorrelation of the residuals, as well as a worsening of the values of the information criteria obtained.

By means of a z-test on the coefficients it has been shown that for the ARIMA model (1,1,4), each single coefficient is significant with a lower p-value of .05, small standard-errors and high z-scores, confirming the goodness of the model itself, as shown in figure 6.

```
z test of coefficients:

     Estimate Std. Error  z value  Pr(>|z|)    
ar1  0.519835   0.084373   6.1612 7.221e-10 ***
ma1  0.147569   0.055443   2.6616  0.007776 ** 
ma2  0.144335   0.055130   2.6181  0.008842 ** 
ma3  0.106265   0.052788   2.0131  0.044109 *  
ma4 -0.817749   0.051287 -15.9445 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 6- z test of coefficients

Also for the ARIMA model (1,1,3) each coefficient is highly significant.

Once the most appropriate model was identified, the values were forecast over a time interval of two years and three years. For the sake of simplicity, we show in figure 7 the forecasts over a period of two years.

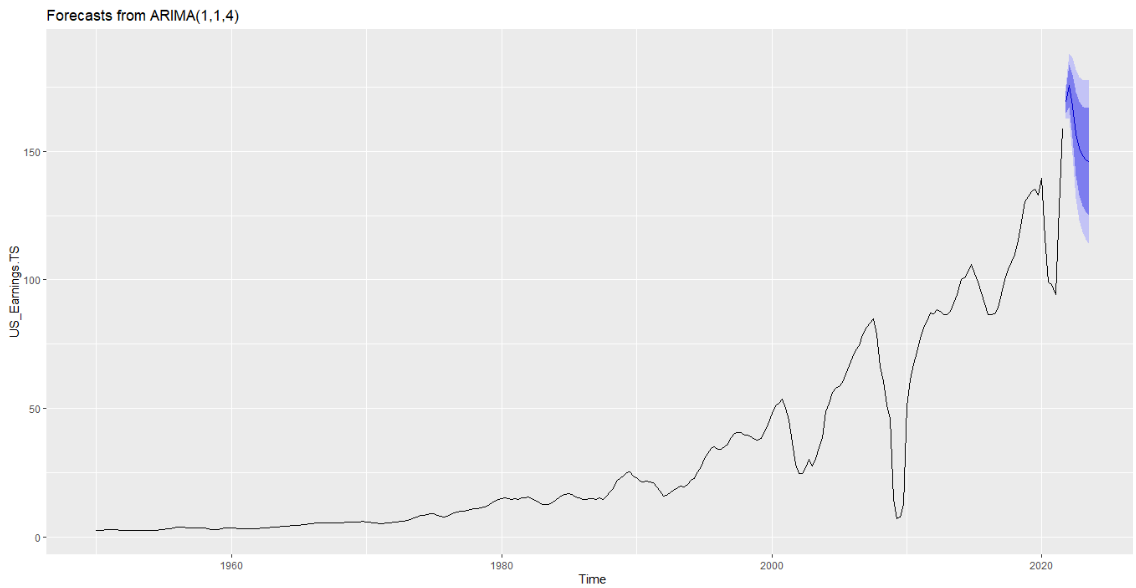

Figure 7- Forecast from Arima (1,1,4)

4. Interpretation of the results

With reference to the results obtained from a two-year forecast, it is possible to state that for the last quarter of 2021 and for the first six months of the following year, the estimated values of the earnings are around historical highs, and then gradually decrease. from the third quarter of 2022 until the end of the estimated time interval, i.e. the third quarter of 2023.

In particular, in the case of Q3 2021, we can most likely estimate a value within a 95% confidence interval that varies between a minimum of 162 and a maximum of $ 175 billion.

Similarly, the peak in the figure would seem to be produced by Q1 of 2022, whose estimate is included in the highest range recorded by the forecast, with a minimum of 163 and a maximum of 188 billion dollars.

According to the estimates produced, 2022 also looks very profitable in the last two quarters of the year, with values around 150 billion dollars and very high confidence intervals.

The quarters of 2023, on the other hand, are characterized by lower-high forecasts (148, 146, 145) and intervals with bearish trends, aimed at perhaps anticipating a possible general decline in profits in the immediate future. The decline could be related to the rather unstable period we have experienced in recent years and which, in some way, could still have its effects for other years, with negative repercussions on the earnings of the companies in question.

Figure 8 shows the forecast values with the relative confidence intervals.

```
Forecasts:
        Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
2021 Q4       169.2900 165.0437 173.5362 162.7959 175.7841
2022 Q1       175.5588 167.3029 183.8148 162.9325 188.1852
2022 Q2       167.5473 155.2113 179.8833 148.6811 186.4135
2022 Q3       156.7591 140.4282 173.0900 131.7832 181.7350
2022 Q4       151.1510 132.9538 169.3482 123.3208 178.9813
2023 Q1       148.2357 128.8658 167.6057 118.6119 177.8596
2023 Q2       146.7203 126.4673 166.9732 115.7461 177.6945
2023 Q3       145.9325 124.9357 166.9293 113.8206 178.0444
> 
```

Figure 8 - Forecast values with confidence intervals

by Ridolfi Lorenzo