Text

Description automatically generated

**INTRODUCTION TO ENTERPRISE ANALYTICS**

**ALY 6050, CRN 21596**

**PROFESSOR ROY WADA**

**SUMMARY REPORT & ANALYSIS**

**SUBMITTED BY:**

**RICHA UMESH RAMBHIA**

**NUID: 001523295**

**TABLE OF CONTENTS**

OVERVIEW…………………………………………………………………… 3

PART 1…………………………………………………………………………. 4

PART 2…………………………………………………………………………. 11

PART 3…………………………………………………………………………. 15

ANALYSIS……………………………………………………………………… 32

REFERENCES…………………………………………………………………. 33

**OVERVIEW**

This project deals with *Forecasting Financial Time Series* where the data of the historical stock prices for the following stocks of 252 market days are used for the analysis.

* Coca-Cola (NYSE:KO)
* Costco Wholesale (NASDAQ:COST)

Time series analysis deals with analyzing the past data within a given duration of time to forecast the future which has the components like trend, seasonal variations, cyclic variations and random or irregular movements. Time series forecasting on the other hand is when we associate a time component to that forecast, is when it becomes a time series forecasting and we have the data being called as the time series data.

The dataset of the historical stock prices has closing prices for the two stocks, namely, the COST stock data and the KO stock data and along with the closing price, the other parameters included for analysis are date, period, and volume of the two stock data. With respect to this dataset provided, we need to perform short term forecasting and long term forecasting.

Forecasting is the process of making predictions based on the past and the present data to predict the future data and for analysis. In the short term forecasting, exponential smoothing for the smoothing parameter alpha and adjusted exponential smoothing for the smoothing parameter beta is performed to yield the most accurate forecasts for the COST and KO stocks. Long term forecasting will focus on forecasting the value during specific periods using the 3-period weighted moving averages. Computing these values will help to compare and analyze the forecasted values for periods with their actual close values on the specific days.

Lastly, part 3 of this project will focus on the time series analysis using R where the data for COST and KO is downloaded using R to perform exponential smoothing, fit an AR model using ARIMA function in R or by doing it manually and again using the auto.arima function for forecasting using the dry\_wine.csv data.

**PART 1**

**Short-term Forecasting**

The components of the time series analysis as mentioned above are trend, seasonal variations, cyclic variations and random or irregular movements. The analysis and detecting these patterns and behaviors can be done using a simple line plot of the two time series data. A trend analysis is to check the data for any repeated behavior in the graphical representation and as long as the trend is either increasing or decreasing then it becomes easy for the analysis. (Anish, 2020). But if the time series data has some error, then we need to perform smoothing. Now, with respect to seasonality, it is the repetition of data at a certain period of time interval. (Anish, 2020). Seasonality is measured or calculated by autocorrelation after we subtract the trend from the data and that is one of the most important characteristics of the time series analysis.

The line plot for the two stock data to perform time series analysis based on the components of the time series are as follows:

*Figure 1: Line plot for COST Closing Price with the trend line*

The above graph represents the line plot for the COST stock data closing price, which is varying, and the trend line represented is a linear line plot. Since the trend is not having an increasing or decreasing factor, it means that there is an error due to which there is a need for exponential smoothing to be performed. The increasing or decreasing factor of the trend makes it easier for analysis and so there is a need for the error to be removed. With respect to the seasonality component, we can observe that the closing price has a certain increase during a period and certain decrease during some period and this explains the seasonality component of the time series analysis.

*Figure 2: Line plot for KO Closing Price with the trend line*

The above graph represents the line plot for the KO stock data closing price, which is varying, and the trend line represented is a linear line plot. Since the trend is not having an increasing or decreasing factor, it means that there is an error due to which there is a need for exponential smoothing to be performed for this data too. Here, there is spike of certain decrease of the closing price for the KO stock data during a certain time period and this too explains the seasonality component.

Since we observed that there is an error in the trend analysis component, exponential smoothing and adjusted exponential smoothing is thus performed for both the stock data.

First, exponential smoothing is performed in order to forecast the closing prices for the period 253 for both the stock data and the values of alpha used for the smoothing parameter are as below.

*Table 1: Values for Alpha*

|  |  |
| --- | --- |
| **Sr No.** | **Alpha Values** |
| 1 | 0.15 |
| 2 | 0.35 |
| 3 | 0.55 |
| 4 | 0.75 |

Now, the exponential smoothing forecast which will be generated above will be used to perform adjusted exponential smoothing for the alpha value of 0.55 and the successive beta values used for the same are as below.

*Table 2: Values for Beta*

|  |  |
| --- | --- |
| **Sr No.** | **Beta Values** |
| 1 | 0.15 |
| 2 | 0.25 |
| 3 | 0.45 |
| 4 | 0.85 |

Thus, exponential smoothing for successive values of alpha and adjusted exponential smoothing for successive values of beta for both the stock data are performed. Starting with the exponential smoothing for the COST stock data where alpha is 0.15, forecasting is calculated using the alpha value which further helps in calculating the exponential absolute error which is MAPD i.e., Mean Absolute Percentage Deviation and then using that we get the percentage value for the exponential smoothing that will help in determining which value of alpha has yielded the most accurate forecast for both the stocks.

The MAPD obtained for alpha value of 0.15 is 2.22%, therefore, exponential smoothing for COST where alpha is 0.15 is 2.22%. Similarly, for all the successive values of alpha, the MAPD is computed for both the stock data and the output for which is shown below.

*Table 3: MAPD for successive values of alpha for the COST stock data*

|  |  |
| --- | --- |
| **Alpha Value** | **MAPD** |
| 0.15 | **2.22%** |
| 0.35 | 1.36% |
| 0.55 | 1.11% |
| 0.75 | 1.00% |

*Table 4: MAPD for successive values of alpha for the KO stock data*

|  |  |
| --- | --- |
| **Alpha Value** | **MAPD** |
| 0.15 | **1.32%** |
| 0.35 | 0.85% |
| 0.55 | 0.72% |
| 0.75 | 0.68% |

From the table above for both the stock data, it is observed that the MAPD for the COST stock data for the alpha value 0.15 is higher as compared to others. Similarly, the MAPD for the KO stock data for alpha value 0.15 is higher and thus for both the stocks the value of alpha for which the most accurate forecast is obtained is 0.15. Since alpha is a percentage of how much importance the model will allocate to the most recent observations compared to the importance of demand history, this value helps in determining the most accurate forecasting for the exponential smoothing.

Now, for the next part, which is the adjusted exponential smoothing, the exponential smoothing forecast of alpha = 0.55 is used for the same with successive values of beta for both the stocks. Here, using the alpha value as 0.55 we compute the MAPE which is the adjusted exponential smoothing absolute error for successive values of beta. The trend parameter is thus computed for the adjusted exponential smoothing after which the forecasting is obtained along with the MAPE value. The MAPE output for the same for both the stocks is as follows.

*Table 5: MAPE for successive values of beta for the COST stock data where alpha is* ***0.55***

|  |  |
| --- | --- |
| **Beta Value** | **MAPE** |
| 0.15 | **1.00%** |
| 0.25 | 0.98% |
| 0.45 | 0.96% |
| 0.85 | 0.95% |

*Table 6: MAPE for successive values of beta for the KO stock data where alpha is* ***0.55***

|  |  |
| --- | --- |
| **Beta Value** | **MAPE** |
| 0.15 | **0.70%** |
| 0.25 | 0.70% |
| 0.45 | 0.69% |
| 0.85 | 0.68% |

As observed, the difference in the MAPE value, which is the Mean Absolute Percentage Error, for values of beta is not much in case of both the stock data but the most accurate results for the forecasting for both the stocks are provided by the beta value of 0.15 which is 1% and 0.70% for COST and KO stocks respectively.

The closing price vs the exponential smoothing and adjusted exponential smoothing graph is represented visually for both the stocks for analysis which is as follows.

*Figure 3: COST closing price vs Exponential Smoothing for alpha = 0.15*

*Figure 4: COST closing price vs Exponential Smoothing for alpha = 0.55*

Similarly, graphs for the COST stock data for other successive values of alpha have been plotted. The graphs for KO stock data for values of beta are as follows.

*Figure 5: KO closing price vs Exponential Smoothing for alpha = 0.15*

*Figure 6: KO closing price vs Exponential Smoothing for alpha = 0.75*

Now, for the successive values of beta for the two stock data, the graphs are represented as below.

*Figure 7: COST closing price vs Adjusted Exponential Smoothing for alpha = 0.55 and beta = 0.15*

*Figure 8: COST closing price vs Adjusted Exponential Smoothing for alpha = 0.55 and beta = 0.85*

**PART 2**

**Long-term Forecasting**

In the part, we perform the long term forecasting which deals with using a 3-period weighted moving average to forecast its value during periods from 1 to 100. The weights given to perform the forecast are, 0.5 (for the most recent period), 0.3 (for the period before the most recent) and 0.2 (for two periods ago). The weighted moving average for the COST and KO stock data is computed for which the MAPE is then calculated.

Since we are given the 3-period weighted moving average, we do not calculate the WMA for the starting three records and began from the 4th period. The weights are thus to be considered while computing the weighted moving average. The WMA is computed for periods from 1 to 100 for which the MAPE is then generated. The Mean Absolute Percentage Error obtained for the COST stock data is 3.79 and that for the KO stock data is 0.37.

Now, we perform the trend based forecasting and analysis i.e., forecast the values using linear trend. The trend and adjusted exponential smoothing would be the same as computed in part 1 for alpha is 0.55 and beta value of 0.15. The trend forecast is computed which takes the values of the adjusted exponential smoothing up to the value of 100 after which for the remaining values we compute the trend value. This gives us the forecasted values using the trend based forecasting. Similar steps were performed for the KO stock data too in order to obtain the forecasted values for the same.

Table

Description automatically generated

*Figure 9: Weighted Moving Average for COST stock data*

Table, Excel

Description automatically generated

*Figure 10: Weighted Moving Average for KO stock data*

The forecasted trend based analysis performed had the following output.

Table

Description automatically generated

*Figure 11: Forecast the values using linear trend for COST stock data*

Table

Description automatically generated

*Figure 12: Forecast the values using linear trend for KO stock data*

The visual representation of the COST closing price and KO closing price with the forecasted trend value is as below.

*Figure 13: COST STOCK DATA*

*Figure 14: KO STOCK DATA*

**PART 3**

**Time Series using R**

In this part, the primary focus is on performing the time series analysis using R for a 5-year worth of data for the KO and COST stock data. To perform the time series analysis for these stock data, the data was loaded using R to perform the further analysis. By performing the time series analysis in R, an overview of the analytical as well as the numerical evaluation would be obtained. The time series analysis performed in excel was the numerical evaluation of the two stock data whereas the analysis performed in R is the analytical evaluation of the stock data. With respect to these evaluations, comparison of the analytical and numerical evaluation can be done in order to check for the results. Arima model would be used in R to perform the analysis to fit the time series model.

After the data is loaded in R, a moving average and exponential smoothing is created for both the plots. A moving average is a calculation which is used to analyze the data points by creating a series of averages of different subsets of the data set. (Fernando, 2021). It is therefore a statistical method which is used for the forecasting long-term trends. Exponential smoothing on the other hand is used for smoothing time series data using the exponential window function which is often used for analytics of time series data.

In order to perform these tasks, required packages and libraries such as ‘quantmod’ and ‘tidyquant’ were installed and loaded in R. The first step here was to load the stock data and display the data and so, following code was written for the same for which the output is as shown below.

**Output: Loading the COST and KO stock data**

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Once the stock data is loaded, the data values are thus displayed and the output of which is as follows.

**Output: Description about the COST stock data** *(head,class,frequency,summary,cycle,plot)*

Text

Description automatically generated

Text

Description automatically generated

Chart, histogram

Description automatically generated

**Output: Description about the KO stock data** *(head,class,frequency,summary,cycle,plot)*

A picture containing text

Description automatically generated

Text

Description automatically generated

Chart, line chart, histogram

Description automatically generated

The chart series is therefore plotted for the COST and KO stock data which is as follows.

**Output: Chart Series for COST stock data**

Chart, line chart

Description automatically generated

Chart

Description automatically generated

**Output: Chart Series for KO stock data**

Chart

Description automatically generated

A picture containing text, sky

Description automatically generated

Now we calculate and compute the moving average and the exponential smoothing for the two stock data. To calculate the moving average for the COST stock data and the KO stock data, rollmean function is used to compute the same and the output of the same is as follows along with the moving average plot.

**Output: Moving average for COST stock data**

Table

Description automatically generated

Chart, histogram

Description automatically generated

**Output: Moving average for KO stock data**

Table

Description automatically generated

Chart

Description automatically generated

After the moving average is computed, we then compute the exponential smoothing for both the stock data and plot the graph for the same. The output for the exponential smoothing for the COST and KO stock data is as shown below.

**Output: Exponential smoothing for COST stock data**

Table

Description automatically generated

**Forecasts from Simple exponential smoothing for COST stock data**

Chart, line chart

Description automatically generated

**Output: Exponential smoothing for KO stock data**

Table

Description automatically generated

**Forecasts from Simple exponential smoothing for KO stock data**

Chart

Description automatically generated

**Output: Actual Closing Price vs Moving Average for COST stock data**

Chart, scatter chart

Description automatically generated

**Output: Actual Closing Price vs Moving Average for KO stock data**

Chart, scatter chart

Description automatically generated

Now, the next part is to fit an AR time series model to KO and COST for which a variable named ‘together’ is used which stores the KO and COST data using the cbind function of R. To fit an AR model, lag function is used to create closing values for the stock data and then we view together in order to check the data stored in the variable.

**Output: Displaying the variable ‘together’ for COST and KO stock data**

Table

Description automatically generated

**Output: Summary of the data ‘together’**

Text

Description automatically generated

From the output above, it is observed that the p-value is less than the significance value of 0.05, and thus we reject the null hypothesis. Similarly, the summary of the COST and KO data for the variable together is displayed which helps in the further analysis. Further, we try to plot the COST and KO stock data for the AR model for which the output is as below.

**Output: COST and KO stock data for the AR model**

Graphical user interface, chart, line chart, scatter chart

Description automatically generated

**Output: Actual Prices vs AR model for COST stock data**

Chart, histogram

Description automatically generated

**Output: Actual Prices vs AR model for KO stock data**

Chart, scatter chart

Description automatically generated

Lastly, we build an Arima model to fit the COST and KO stock data to predict the future value and plot that through a visual representation.

**Output: Arima model for COST stock data**

Table

Description automatically generated

Text

Description automatically generated

Chart, line chart

Description automatically generated

**Output: Arima model for KO stock data**

Table

Description automatically generated

Text

Description automatically generated

Chart

Description automatically generated

We finally build a model to predict the future values of the COST and KO stock data for which the output is as below.

**Output:**

Table

Description automatically generated

The process of model building using the Arima model is repeated to perform the time series forecasting for the dry wine prices. The dry\_wine.csv contains 2 parameters, namely, the dates and the dry wine prices. In order to forecast the prices for the time series analysis, Arima model is therefore used for the same. The output of the time series forecasting is as follows.

**Output: Arima model with visual representation for dry wine data**

Table

Description automatically generated with medium confidence

Text

Description automatically generated

A picture containing shape

Description automatically generated

**ANALYSIS**

The forecasting financial time series deals with short term forecasting, long term forecasting and time series analysis in R for the two stock data and for the dry wine data. In the short term forecasting, which is the part 1 of the analysis, exponential smoothing and adjusted exponential smoothing is performed which helps in computing the MAPD (Mean Absolute Percentage Deviation) and MAPE (Mean Absolute Percentage Error) for the further analysis. To perform the exponential and adjusted exponential smoothing in order to forecast the time series for both the stock data, successive values of alpha and beta were given respectively.

As mentioned above, analysis of the alpha value which yielded the most accurate forecast for each stock has been done. The alpha value for which the most accurate forecast was obtained is 0.15 and this is because for smaller values of alpha the fitted lines are smoother as they give more weight to past observations as compared to the higher values of alpha. With respect to the adjusted exponential smoothing with alpha value of 0.55, and successive values of beta for both the stock data, we computed the MAPE. For the beta values, there was not much of a difference between the MAPE values, but again we can observe that for beta value of 0.15, we obtained the most accurate forecast.

In the long term forecasting, we use the 3-period weighted moving averages to forecast the values during the mentioned periods using the linear trend forecasting methods. The MAPE obtained for the 3-period weighted moving averages for COST and KO stock data are 3.79 and 0.37 respectively. For the trend based forecasting, the MAPE obtained is 0.44% and 0.67% for COST and KO stock data respectively. If compared this to the above analysis, we observe that the MAPE in the exponential smoothing forecasting was high as compared to the trend based forecasting for both the stock data. And since, MAPE value determines the average difference between the forecasted value and actual value, we can conclude that the trend based forecasting was more accurate as compared to the other as the difference between the values there is less. Therefore, trend based forecasting method has yielded the most accurate forecast for COST and KO stocks.

The time series analysis using R deals with performing moving average and exponential smoothing and fitting an AR model to KO and COST. The numerical evaluation of the stock data would be obtained in R using the arima function whereas the analytical evaluation was obtained using the Excel. The moving average and exponential smoothing were performed on both the stock data and visual representation of both were plotted as mentioned earlier which gave an understanding about the data and the forecasting time series analysis. The AR time-series model will help in choosing the best model and predicts the future value of the forecasted time series for the data. Similar analysis was performed for the dry wine data in order to predict the future and fit an AR model.

**REFERENCES**

Anish, A. (2020, November 24). Time Series Analysis. Medium. Retrieved January 29, 2022, from <https://medium.com/swlh/time-series-analysis-7006ea1c3326>

DD. (2018, April 2). How to download stock prices in R. Coding Finance. Retrieved January 31, 2022, from <https://www.codingfinance.com/post/2018-03-27-download-price/>

Using R for Time Series Analysis. (2010). Avril Coghlan. <https://a-little-book-of-r-for-time-series.readthedocs.io/en/latest/src/timeseries.html#arima-models>

Fernando, J. (2021, October 5). Moving Average (MA). Investopedia. Retrieved January 31, 2022, from <https://www.investopedia.com/terms/m/movingaverage.asp>

Exponential Smoothing in R Programming. (2022, January 25). GeeksforGeeks. Retrieved January 31, 2022, from <https://www.geeksforgeeks.org/exponential-smoothing-in-r-programming/#:~:text=Exponential%20Smoothing%20in%20R%20Programming.%20Last%20Updated%20%3A,time%20the%20exponential%20functions%20assign%20exponentially%20decreasing%20weights>.

Vandeput, N. (2019, November 12). Simple Exponential Smoothing for Time Series Forecasting. Towards Data Science. Retrieved January 31, 2022, from <https://towardsdatascience.com/simple-exponential-smoothing-749fc5631bed#:~:text=alpha%20is%20a%20ratio%20%28or%20a%20percentage%29%20of,certain%20weight%20%28alpha%29%20to%20the%20last%20demand%20occurrence>.