SCHOOL OF SCIENCE AND ENGINEERING

DOW JONES MARKET VOLATILITY FORECASTING
- New York Stock Exchange Market -

Capstone Design
Spring 2019

Kenza Benomar
Supervised by Dr. Lahcen Laayouni
DOW JONES MARKET VOLATILITY FORECASTING

Capstone Report

Student Statement

“I, Kenza Benomar, pledge that this project is a reflection of my own research and work. I assert that I applied the ethics of design, and rules of safety of the public, and none of them was violated.”

Kenza Benomar

Approved by the Supervisor

Dr. Lahcen Laayouni
Acknowledgements

I would like to sincerely express my gratefulness to Dr. Lahcen Laayouni who supervised this capstone project in every step till the end. Thanks to his permanent assistance and evaluation I have been able to achieve a project in a subject I was still maturing in knowledge and skills wise. I acknowledge the usefulness of this project given the amount of information and skills that I acquired, and that will help me in my future study plans.

I could never thank enough my parents who were my first school in life. My capstone project which is the milestone of my curriculum journey at Al Akhawayn University is the fruit of their love, caring, and strengthening that shaped the person I am today. In addition to my parents, I was lucky to have a sister who was always there to support me. Thank you Sarah.

Last but not least, I would like to express my highest gratitude to my friends who surrounded me with their support and love and were always ready to answer my questions throughout all the process of my capstone. Thank you Yasmine, Leila and Farid. Thank you Fatima Zahra, Omar, Slimane, Marouane, Maâine and others.

All these people were significant to the achievement of my curriculum.
Abstract

The object of this capstone project is to forecast the Dow jones Industrial Average 30. The DJIA is a New York Stock Exchange volatility index which comprises the 30 most significant stocks in the U.S financial market. The model chosen for this study is the Auto Regressive Integrated Moving Average which is a method that consists of taking the historical data of a time series and predict a future path for this series. Throughout the project, different adjustments were performed on the model in order to come up with the optimal variables that will lead to the minimum number of errors. Afterwards, an evaluation of the model is done in order to leave room for innovation in the same direction of this study. This project requires a thorough and in-depth research on the volatility of the financial market and the different factors that affect it. Different studies were made on the prediction of stock returns and option prices, this study suggests a prediction of the financial market as a whole by predicting its volatility. Since the financial market is based on risk and speculations, this project will serve financial analysts and investors as a model to follow in their analysis or investment decisions by decreasing the uncertainty.

**Key words:** Volatility, Dow Jones, Forecasting, ARIMA model.
Résumé

L’objet de ce projet est de prévoir le Dow Jones Industrial Average 30. Le DJIA est un indice de volatilité de la Bourse de New York qui comprend les 30 actions les plus importantes du marché financier américain. Le modèle choisi pour cette étude est la moyenne mobile auto-régressive, méthode qui consiste à prendre les données historiques d'une série chronologique et à prédire l'évolution future de cette série. Tout au long du projet, différents ajustements ont été apportés au modèle afin de déterminer les variables optimales qui conduiraient au nombre minimal d’erreurs. Ensuite, une évaluation du modèle est effectuée afin de permettre de l’innovation dans le même sens que cette étude. Ce projet nécessite une recherche approfondie et approfondie sur la volatilité du marché financier et les différents facteurs qui l’affectent. Différentes études ont été réalisées sur la prévision du rendement des actions et du prix des options. Cette étude suggère une prévision du marché financier dans son ensemble en prévoyant sa volatilité. Étant donné que le marché financier est basé sur le risque et les spéculations, ce projet servira de modèle aux analystes financiers et aux investisseurs en tant que modèle à suivre dans leurs analyses ou décisions d'investissement en réduisant l'incertitude.

Mots clés: Volatilité, Dow Jones, Prévision, Modèle ARIMA
Table of Contents

Acknowledgements ............................................................................................................ iii
Abstract ............................................................................................................................... iv
Résumé ............................................................................................................................... v
Table of Contents ............................................................................................................... vi
List of Figures .................................................................................................................. viii
List of Tables ................................................................................................................... ix
Chapter 1: Literature Review ........................................................................................ 1
  1.1 Introduction to Volatility in the Financial Market .................................................. 1
  1.2 The Dow Jones Industrial Average ....................................................................... 2
  1.3 Factors affecting the Dow Jones .......................................................................... 4
    1.3.1 Economic Factors ......................................................................................... 4
    1.3.2 Political Factors .......................................................................................... 5
  1.4 Feasibility Study ..................................................................................................... 6
  1.5 STEEPLE Analysis ............................................................................................... 7
    1.5.1 Social Impact ............................................................................................... 7
    1.5.2 Technological Impact .................................................................................. 7
    1.5.3 Economic Impact ......................................................................................... 8
    1.5.4 Environmental Impact ................................................................................. 8
    1.5.5 Political Impact ............................................................................................ 8
    1.5.6 Legal Impact ............................................................................................... 8
    1.5.7 Ethical Impact .............................................................................................. 8
Chapter 2: Data and Methodology ............................................................................... 10
  2.1 Descriptive Statistics ......................................................................................... 10
  2.2 Historical Data Method ...................................................................................... 12
    2.2.1 Time Series ............................................................................................... 12
    2.2.2 ARIMA model ........................................................................................... 14
  2.3 Implied Volatility ............................................................................................... 14
    2.3.1 Black Scholes Model .................................................................................. 15
    2.3.2 Limitations of the Black Scholes ................................................................. 16
    2.3.3 Limitations of Implied Volatility ................................................................. 16
Chapter 3: Results and Interpretation ........................................................................ 18
3.1 Data and Variables ........................................................................................................... 18
3.2 Results of ARIMA forecasting .......................................................................................... 20
3.3 Comparison and Interpretation ......................................................................................... 24
  3.3.1 Forecasts and Real Dow Jones .................................................................................... 24
  3.3.2 Interpretation and Explanation ..................................................................................... 28
Chapter 4: Future Work .......................................................................................................... 32
Chapter 5: Conclusion ............................................................................................................ 34
Chapter 6: References ............................................................................................................ 35
Appendix A ............................................................................................................................... 36
Appendix B ............................................................................................................................... 37
Appendix C ............................................................................................................................... 39
List of Figures

Figure 2.1. DIJA statistics (8th February 2018 to the 8th of February 2019) .................................................. 10
Figure 3.1. Sample R code function ........................................................................................................... 18
Figure 3.2. Snapshot of the ARIMA data ...................................................................................................... 21
Figure 3.3. Sample R code for Non-Stationary data (KPSS test) ................................................................. 21
Figure 3.4. Sample output of the R code for Non-Stationary data (KPSS test) ............................................. 22
Figure 3.5. Time series loop .......................................................................................................................... 22
Figure 3.6. The output of the ACF and PACF functions ............................................................................... 22
Figure 3.7. ACF Residual output .................................................................................................................. 23
Figure 3.8. R code which checks for residuals ............................................................................................. 23
Figure 3.9. Final model for the ARIMA forecasting ...................................................................................... 24
Figure 3.10. Change in the future value of the Dow based on our prediction .............................................. 25
Figure 3.11. R code used to model the forecasting ...................................................................................... 25
Figure 3.12. Mean of the future Dow value compared to its actual value in 2019 ......................................... 26
FIGURE 3.13. LOW, UPPER AND MEAN OF OUR PREDICTION OF THE BEHAVIOR OF THE DOW 30 ...... 27
FIGURE 3.14. HISTOGRAM OF RESIDUALS ............................................................................................... 30
FIGURE 3.15. DENSITY OF RESIDUALS .................................................................................................... 30
List of Tables

TABLE 1.1. LIST OF COMPANIES DJIA HAS BEEN INCLUDING AS OF FEBRUARY 8, 2019.............................. 3

TABLE 2.1. DESCRIPTIVE STATISTICS OF OPEN AND CLOSE PRICES OF DJIA (FEB. 8, 2018 - FEB. 8, 2019) 10

TABLE 3.1. SNAPSHOT OF DJI DATA ........................................................................................................ 19

TABLE 3.2. EXPLANATION OF MOVING AVERAGE OVER A ONE YEAR PERIOD ................................. 20

TABLE 3.3. RESIDUALS STATISTICAL RESULTS ......................................................................................... 23

TABLE 3.4. RESIDUALS .............................................................................................................................. 29
Chapter 1: Literature Review

1.1 Introduction to Volatility in the Financial Market

Volatility is a tracing of the behavior of the financial stock market. Volatility determines the variability of the different financial derivatives. Determining the variability of financial securities plays a crucial role in several areas such as the pricing of derivatives, hedging decisions and the calculation of Value-at-Risk measures. Its significance has been most concisely and persuasively summarized by Andersen and Bollerslev in 1998 when they simply state:” Volatility permeates finance”. It is imperative to understand how volatility behaves over time. [1]

It has been proven over time that volatility is time-varying. As observed by Mandelbrot (1963) the financial time series exhibit a prolonged periods of variable volatility, he clearly stated: “Large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes”. [1] Given the permanent variability of the volatility and the influence of current data on future data, it is important to use the previous data of a certain volatility index to predict the future or ongoing one. For that purpose, different forecasting methods have been introduced, such as the use the historical data. The latest is a retrospective volatility measure that focuses on the behavior of the market on a given period and the different events or seasons that influence it. After noticing a trend on the behavior of the market a forecasting of the stock market graph is made.

As an alternative to the use of historical data, we can make use of the implied volatility. It is often referred to as the market’s perception of the future volatility over the remaining life of the option. Implied volatility is different than historical volatility in terms of availability. For instance, historical volatility has a wider availability and could be applied to any variable. On the other hand, implied volatility could be applied only to financial assets on which options are
traded, and is often obtained in conjuncture with a certain option pricing model. [1] The pricing formula we will be working with throughout the implied volatility forecasting is the Black Scholes model. The latest was developed in 1973 by Fisher Black, Myron Scholes and Robert Merton and is still used on a wide range today and recognized as one of the most fair models of determining the prices of options. [2]

1.2 The Dow Jones Industrial Average

The Dow Jones Industrial Average also referred to as the DJIA or the Dow was invented by Charles Dow and his business partner Edward Jones in 1896. It was a price weighted average of only 12 industrial companies. In 1928, it changed its composition to include the 30 most significant companies’ stocks also called “blue-chip” stocks that are traded on the Nasdaq and NYSE (New York Stock Exchange), such as Walt Disney, John Dupoint Morgan Chase, Goldman Sachs etc. The Dow is a market indicator to which investors and traders refer in order to know the state of the market; if the Dow is up the Market is up and vice versa. However, the DJIA did not remain intact and faced a change in its components after the Great Depression; for instance, in 1932, 8 of its stocks were replaced by others such as Coca Cola and Procter & Gamble. The major change the Dow has faced was in the 1997- 1999 when again eight of its stocks were replaced, the major change was when Walgreens Boots replaced general Electric.[2]

The list of companies DJIA has been including as of February 8, 2019 is listed in Table 1.1.
Table 1.1. List of companies DJIA has been including as of February 8, 2019

<table>
<thead>
<tr>
<th>Company Name</th>
<th>Ticker</th>
<th>Exchange</th>
</tr>
</thead>
<tbody>
<tr>
<td>The 3M Company</td>
<td>MMM</td>
<td>NYSE</td>
</tr>
<tr>
<td>The American Express Company</td>
<td>AXP</td>
<td>NYSE</td>
</tr>
<tr>
<td>Apple Inc.</td>
<td>AAPL</td>
<td>NASDAQ</td>
</tr>
<tr>
<td>The Boeing Company</td>
<td>BA</td>
<td>NYSE</td>
</tr>
<tr>
<td>Caterpillar Inc.</td>
<td>CAT</td>
<td>NYSE</td>
</tr>
<tr>
<td>Chevron Corporation</td>
<td>CVX</td>
<td>NYSE</td>
</tr>
<tr>
<td>Cisco Systems, Inc.</td>
<td>CSCO</td>
<td>NASDAQ</td>
</tr>
<tr>
<td>The Coca-Cola Company</td>
<td>KO</td>
<td>NYSE</td>
</tr>
<tr>
<td>DowDuPont Inc.</td>
<td>DWDP</td>
<td>NYSE</td>
</tr>
<tr>
<td>Exxon Mobil Corporation</td>
<td>XOM</td>
<td>NYSE</td>
</tr>
<tr>
<td>The Goldman Sachs Group, Inc.</td>
<td>GS</td>
<td>NYSE</td>
</tr>
<tr>
<td>The Home Depot Inc.</td>
<td>HD</td>
<td>NYSE</td>
</tr>
<tr>
<td>International Business Machines Corp.</td>
<td>IBM</td>
<td>NYSE</td>
</tr>
<tr>
<td>Intel Corporation</td>
<td>INTC</td>
<td>NASDAQ</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>JNJ</td>
<td>NYSE</td>
</tr>
<tr>
<td>JPMorgan Chase &amp; Co.</td>
<td>JPM</td>
<td>NYSE</td>
</tr>
<tr>
<td>McDonald’s Corporation</td>
<td>MCD</td>
<td>NYSE</td>
</tr>
<tr>
<td>Merck &amp; Company, Inc.</td>
<td>MRK</td>
<td>NYSE</td>
</tr>
<tr>
<td>Microsoft Corporation</td>
<td>MSFT</td>
<td>NASDAQ</td>
</tr>
<tr>
<td>Nike, Inc.</td>
<td>NKE</td>
<td>NYSE</td>
</tr>
<tr>
<td>Pfizer Inc.</td>
<td>PFE</td>
<td>NYSE</td>
</tr>
<tr>
<td>Procter &amp; Gamble Co.</td>
<td>PG</td>
<td>NYSE</td>
</tr>
<tr>
<td>The Travelers Companies, Inc.</td>
<td>TRV</td>
<td>NYSE</td>
</tr>
<tr>
<td>UnitedHealth Group, Inc.</td>
<td>UNH</td>
<td>NYSE</td>
</tr>
<tr>
<td>United Technologies Corporation</td>
<td>UTX</td>
<td>NYSE</td>
</tr>
<tr>
<td>Verizon Communications Inc.</td>
<td>VZ</td>
<td>NYSE</td>
</tr>
<tr>
<td>Visa Inc.</td>
<td>V</td>
<td>NYSE</td>
</tr>
<tr>
<td>Walmart Inc.</td>
<td>WMT</td>
<td>NYSE</td>
</tr>
<tr>
<td>Walgreens Boots Alliance, Inc.</td>
<td>WBA</td>
<td>NASDAQ</td>
</tr>
<tr>
<td>The Walt Disney Company</td>
<td>DIS</td>
<td>NYSE</td>
</tr>
</tbody>
</table>

In addition to the DJIA there are different market indicators such as the Nasdaq and the S&P 500. Besides its limitation in number of companies the Dow is the most significant market indicator since it changes permanently if some of its companies faces financial distress, in order not to affect the value of the index. Firms removed from the index face a serious stock price decline in the market since investors would ask for a premium to hedge against holding derivatives that have less available information than others. Historically, the 30 companies on the DJIA had higher trading activity than others and accounted for roughly 25% of the market value of all the NYSE firms. Any changes to Dow are made by the Wall Street Journal without consultation of the NYSE or the firms, or any official agency. [3]

Different political and economic factors affect noticeably the DJIA. Most of the changes that happened to the Dow were from 1929 to 1933, which coincides with the first worldwide
financial crisis and the Great Depression Period. During this period the United states’ economy was unstable which made the DJIA changes infrequent and clustering. [3] The latest had a direct impact on the aggregate stock volatility; for instance; it increased like it never before. The volatility moves and so does the DJIA.

1.3 Factors affecting the Dow Jones

The DJIA is the result of different factors interaction. Every decision or change made on the economic or political level influences the market and thereby the Dow, one of its major indices.

1.3.1 Economic Factors

- **Consumer credit:** One of the greatest influences that can potentially affect the current DJIA is the consumer credit which contributes to over 13% to the Dow.

- **The previous Dow:** The market has a tracking a memory which makes the Dow this month influence by almost 9% the Dow of the next month. This is explained by the fact that the investors will take advantage of the good deals and boost the Dow if the market falls too quickly. On the other hand, if the market rises too quickly investors tend to sell their stocks which makes the Dow drop.

- **Imports and exports:** exports alone account for 7% and imports 4%, which makes it a total 11% influence on the Dow, higher than its previous activity. In fact, exports show how the economy is tied into the world market.

- **Money supply:** The Dow is directly proportional to the money supply, when one increases so does the other, taking into consideration that everything else stays constant. More money supply means more liquidity in the market, here we mean by liquidity cash money circulating, and therefore more demand and more investors, which drives the Dow higher. [4]
• **Federal interest rates**: When the government has a budget deficit, it increases interests for loan and financial institutions such as banks, which drive the latest to also increase their interest rates on their loans. This has a negative impact on the clients which makes them live a shortage of money and have less purchasing power. Therefore, there is less money circulating in the market and its indicators get affected such as the DJIA.

• **Building permits**: that are well known to scholars and forecasters to be a forward-looking indicator of a future economic activity.

### 1.3.2 Political Factors

Stock volatility in general has always been triggered by different political events, some of which shacked the worldwide stock market badly.

• **Historical turning points**: One of the history recalls is the Great Depression that refers to the greatest economic depression that came after the 1929 crisis and lasted until 1939. Stock volatility during this period was three times higher than any other period in the history of the American financial market. Scholars before that period suggested that the largest stock volatility spike can be predicted when the buildings permits increases; however, during the Great Depression the predictions were labeled “volatility puzzle” because no logical reasoning could explain the severe rise in stock volatility. One of the major factors of this event is the irrational behavior of the investors as they were going with overrated loans, which resulted in many borrowers defaulting and banks being bankrupt. [5]

• **International Trades**: we could recall the latest China/US trade, that caused an undecisive reaction of the market, since there are no terms agreed on yet.
This cold war has caused the stock market to be on hold and indices live a stagnation.

- **Governmental decisions:** recently president Donald Trump said that U.S and China have a good chance of reaching a deal. However, if both parties did not reach a satisfactory end, Trump has threatened to increase the tariffs on Chinese goods. This hold on the trade with china affected U.S stocks; for instance, they all closed higher, the Dow Jones closed with +0.23% rose by 60.14 points. Global Equities climbed after Donald Trump’s tweet pf extending the deadline of boosting the tariffs on Chinese till March first.

- **Government Shutdown:** The shutdown of the government drove the stock market on severe fall by mid- December till mid- January, however, a fast recovery has been noticed afterwards. Further details about this period’s stock behavior will be explained in detail in the next chapter.

### 1.4 Feasibility Study

- **Technology and Knowledge Feasibility**
Technology considerations made with regards to this project are mainly about developing a technique of knowing which method is better for forecasting the Dow volatility index. With choosing the best method, in addition to time saving, we avoid errors and assure certainty. Software used will be SPSS and excel and other software could be used depending on the technical need. Knowledge needed to conduct this project is mainly about the financial market. Before starting the analysis and implementation of methods, enough information should be gathered about what factors influence the market and volatility index.

- **Organization and Schedule Feasibility**
To make sure that the project is time efficient, an organization of schedule and tasks is mandatory:

- Agreement on the specific subject and set up of initial specifications of the project with the supervisor (by the 25th of January)

- Guidance instructions about the literature review: How the Dow volatility index works? What factors influence it? First diary: An estimated outline of the different parts of the project including the computation to be made and the time scale to be taken (By 31st of January)

- Late February: Extracting the data from Yahoo finance of the DJIA

- Beginning March: Doing a descriptive statistic of the Data in Excel

- Mid March: Performing an ARIMA model and comparing the predictions with the actual data

- Beginning April: exploring other prediction methods and comparing them with ARIMA model.

1.5 STEEPLE Analysis

1.5.1 Social Impact

The project has no direct social implication. However, on the long term, making a good use of one of the major tools in finance (volatility) will not only decrease the risk of loss for companies but also stabilize the employees social and financial situation.

1.5.2 Technological Impact

One direct technological impact of this project is on technology is to develop a new way of approaching index of stock market. ARIMA model already existed, however, fitting it to the DJIA time series is new. It was first applied to stock prices and returns. Having this model ready to use and to be adapted to other indexes such as the S&P 500 and Nasdaq is the technological impact of our study.
1.5.3 Economic Impact

While conducting their investigations on volatility index, companies would look up for the best method to forecast the future of their stocks. The volatility index of DJIA is one of the major indicators of the NYSE (New York Stock Exchange). Therefore, choosing the method with the least possible errors would assure the accuracy of their forecasting, and thereby save them time and money.

1.5.4 Environmental Impact

Our study has no direct impact on the environment. However, it is environment friendly and does not include any natural resources use, the whole study is done on a virtual level.

1.5.5 Political Impact

Our study aims to develop a technique to forecast to the DJIA, which is the index of the most significant companies in the US. Being able to predict and estimate its value through an adequate model can help predicting the stock market as well. Based on the historical data of the index, we could detect a pattern for the DJIA, this will lower the unexpected crashes in the economy and thus impact the political decisions.

1.5.6 Legal Impact

This project legal impact could potentially be on the financial laws issued concerning the financial market and its derivatives, since the DJIA could be considered as the leading index.

1.5.7 Ethical Impact

Ethics are very crucial in the financial market. The latest through different tools and departments makes sure that every individual involved in the financial market is ethical. One of the aims of this study, is to be able to hedge against the future surprises of the market by a simple modeling of its history, and thus have no need to use illegal and unethical methods to know the future of a stock price or a leading index before investing.
Chapter 2 : Data and Methodology

2.1 Descriptive Statistics

The data set was chosen for the DJIA from the of 8\textsuperscript{th} February 2018 to the 8\textsuperscript{th} of February 2019 from Yahoo Finance. [6] The descriptive statistics of the open and close prices for that period is shown in Table 2.1 and in Figure 2.1.

Table 2.1. Descriptive statistics of open and close prices of DJIA (Feb. 8, 2018 - Feb. 8, 2019)

<table>
<thead>
<tr>
<th>Close</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>24925.76324</td>
</tr>
<tr>
<td>Standard Error</td>
<td>52.29521311</td>
</tr>
<tr>
<td>Median</td>
<td>24909.535</td>
</tr>
<tr>
<td>Mode</td>
<td>#N/A</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>826.8599207</td>
</tr>
<tr>
<td>Sample Variance</td>
<td>683697.3284</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>0.74672447</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.250380462</td>
</tr>
<tr>
<td>Range</td>
<td>5036.19</td>
</tr>
<tr>
<td>Minimum</td>
<td>21792.2</td>
</tr>
<tr>
<td>Maximum</td>
<td>26828.39</td>
</tr>
<tr>
<td>Sum</td>
<td>6231440.81</td>
</tr>
<tr>
<td>Count</td>
<td>250</td>
</tr>
<tr>
<td>Confidence Level(95.0%)</td>
<td>102.9973502</td>
</tr>
</tbody>
</table>

Figure 2.1. DIJA statistics (8th February 2018 to the 8th of February 2019)
Observations:

- Standard Deviation: we notice a standard deviation of 826 from the mean. The values cannot deviate from the mean more than 830 points on a scale of 20000. The Dow Jones has known a monotone trend until the end of the year due to the government shutdown.

- Skewness: we notice a skewness of -0.25, that is between -1/2 and +1/2. We can say in that case that the distribution is approximately symmetric. It could be again inferred that the Dow is stable because the values are close to the mean; however, they are more likely to be inferior than the mean.

- Kurtosis = 0.74, which is approximately a value of $1 < 3$. This means that the distribution of the sample is not sharp. We have a platykurtic distribution, where the tails are shorter and thinner and central peak lower and broader.

Interpretation of the large drop down:

It is noticeable in late December and beginning of January a sharp drop in stock prices including the Dow Jones; however, right after we see a rise that was described as a recovery. The immediate rise in stock price followed the 4 weeks government shutdown, which is similar to what happened after October 2013 government shutdown that was followed by four months of huge gains. A similar scenario is happening here, the decline in the leading indexes was a result of the inability of the government to provide economic reports during the shutdown, which pushed the housing market and housing market to apply the brakes. Afterwards, the Dow index has known an increase of approximately 11% in February, that happened to be the period of the end of the government shutdown and progress on the U.S-China trade war. It could be also explained by the indications of likely extension for the 1st of March tariff deadline that Donald Trump set up.
2.2 Historical Data Method

The historical data method uses the previous volatility index data and detect a pattern in other to be able to predict to future data. Financial derivatives and assets have known large fluctuations during the past years. Investigating these changes is crucial to the study of the financial market, since it has a large impact on the real economy, especially influencing financial crashes such as the Great Depression. Estimating the risk comes through getting used to the behavioral pattern of the stock. According to Catherine Lubochinsky: “Historical volatility is used to analyze past or present prices and expected volatility is used to predict future price changes. Under this approach, the only real problem for the various agents in the financial economy stems from unexpected volatility” [7]. Questions about which time period to use and the one that will be more representative of the pattern of the index, are still unanswered. Choosing 30 trading days or a whole financial year can lead eventually to different results, however, not always one is better than the other. Different models have been used to represent the dynamics of the volatility, and it is usually an auto regressive model such as GARCH (general autoregressive conditional heteroskedasticity) or ARIMA( autoregressive integrated moving average). [7]

2.2.1 Time Series

During the latest years more interest had been given to modeling of the volatility of stocks. From a financial perspective, understanding stock volatility is crucial in investment decisions. It is the simplest but most accurate risk measure and plays an indispensable role in option pricing. In financial literature, scholars used various linear time series models to investigate volatility of stocks, such as “Poterba and Summers” and “Stambaugh”. [8] Using time series data in volatility analysis has been a method used long time ago through history. Time series analysis main goal in most cases is forecasting future data. In finance and business in general, time series is used to investigate the correlation between different variables and to develop
regression analysis for volatile data. [9] A standardized model for time series could be written as follows:

\[ z_t = f(\Psi_{t-1}) + \alpha_t \]  

In between parentheses is the data set available at that time and alpha is a sequence of random variables. [9] Despite the given information, throughout the history of time series analysis, the latest was most of the time divided into two approaches: Time Domain and Frequency Domain. Obviously, there was a controversy debate between the two approaches’ proponents. In time domain approach parametric models and autocorrelation functions are used, an example of that is the ARIMA model developed by Box and Jenkins. On the other hand, the frequency model approach mainly studies the applications of time series analysis by using the distribution of power or spectral analysis. [9]

Today, there is no more debate between the two approaches. Now the analyst is more concerned about which factors makes one approach better than the other. Also, the development of different software and models made it easier to perform time series analysis, one can just create a model using R language (which we will be using) or SPSS and to fit the model to different data and change the parameters and variables accordingly.

Future research could be done in this subject. In the Journal of the American Statistical Association, Ruey Tsay said: ” An important driving force of future research in time series analysis is the advance in high-volume data acquisition. Consider, for instance, transaction-by-transaction data common in financial markets, or communications networks, and in e-commerce on the internet” [9]. Ruey imagines a special dimension of the use of the time series features, one that could be adapted to different data sizes and discrete variables. For instance, according to his viewpoint, different subjects in the financial market will generate more need for time series models. He clearly says: “In my personal opinion, the use of multivariate models
either in a vector ARMA frame-work or a state-space form will increase, partly because of the need to study the dynamic relationships between variables and partly because of the advances in computing facilities” [9]. He also expects to flourishment of nonlinear and non-Gaussian models that are not within the ARIMA assumption, which we will discuss in detail in the next section.

### 2.2.2 ARIMA model

An important reference for time series analysis is the publication of Box and Jenkins named “Time Series Analysis: Forecasting and Control”, it made the ARIMA( autoregressive integrated moving average) model known throughout the world by using a method consisting of iterating the model based on identification, estimation, and model checking. [9] When built and found to be accurate, expectations of the ARIMA model are used to forecast future values of the time series. ARIMA(p,d,q) adopts the following equation:

\[
 f(\Psi_{t-1}) = c + \sum_{i=1}^{p} \phi_i w_{t-i} - \sum_{j=1}^{q} \theta_j a_{t-j},
\]

The above figure is a generalized function of the ARIMA model, and p, d, and q are positive integers and wt is the differenced series. One of the key assumptions made in ARIMA model is that the ‘at’ function is Gaussian and ‘wt’ contains stationary parameters. However, the Auto Regressive Moving Average has known many developments since its creation, its results and analysis became no longer limited to linear and Gaussian time series models. [9]

### 2.3 Implied Volatility

All trader knows volatility (known as history), the more the subjacent one fluctuates, the more one says that it is volatile. Let’s take a real-life case, an action as ArcelorMittal is more volatile than an action as Total because Arcelor has price changes more important than Total. An active trader will prefer trader Arcelor because there will be much more opportunities. Other volatility
is the Implicit Volatility, which is a concept that the trader uses only on options. Indeed, in this case the mathematical model used for pricing the bonus of the options, the implied volatility is one of the fundamental data. Implied volatility is another way to volatility information which is mostly called the perception of future over the life of any given option.

2.3.1 Black Sholes Model

This method use the well-known mathematical model that was developed by Black and Scholes in 1973, known as the Black Scholes option pricing model which is based on an equilibrium solution. [1] Black Scholes equation value a dividend that pays a call option on any given asset as follows:

\[
\begin{align*}
    c &= S N(d_1) - X e^{-r(T-t)} N(d_2) \\
    d_1 &= \frac{\ln(S/X) + (r + \frac{\sigma^2}{2})(T-t)}{\sigma \sqrt{T-t}} \\
    d_2 &= d_1 - \sigma \sqrt{T-t}
\end{align*}
\]  

[1]

S: the asset price

X: Exercise Price

R: Risk Free rate of return

\( \sigma \): the volatility

(T-t): Time remaining to maturity

The N function used in the first equation is the accumulated profitability.

These types of values are recognized at the time t, except \( \pi \), and since the marketplace price of the decision choice can also be observed, the implied volatility over the option's remaining life can be recurrently calculated. Unfortunately, there is no specific formulation for implied
volatility since the components of Black-Scholes cannot be inverted with the volatility parameter $\pi$ appreciated. However, it is common to use the Newton-Raphson procedure in which case in a relatively small number of iterations a quite accurate estimate of $\pi$ can be obtained. [1]

### 2.3.2 Limitations of the Black Scholes

The Black-Scholes formula has some restrictive hypotheses, some of which we mentioned earlier. One of these is that during the remaining life of the option, the underlying asset does not pay dividends. On the other hand, the surge of profit installments can be thought to be consistent over the rest of the life of the option, as proposed by Merton in 1973. In that case $S$ operating at a profit Scholes model can be substituted with $Se^{-\delta(T-t)}$ where $\delta$ indicates the persistent profit yield. One of the other limitations of the Black Scholes is the possibility of its application only on European options which cannot be exercised before maturity unlike the American ones. [1]

### 2.3.3 Limitations of Implied Volatility

Implied volatility forecasts, given a number of empirical studies and tests is supposedly more accurate than the historical volatility. Historical volatility, as we will see later in this study provided larger scale predictions and close to the real data. As found by Canina and Figlewski:”implied volatility is virtually uncorrelated with subsequent realized volatility” [1], she also claimed that a simple historical volatility measure is more accurate. In their examination they analyzed the data estimation of every day Standard and Poor's 100 stock record bring alternatives over the period May 1973 to May 1987. Both the in this manner got suggested volatilities and a straightforward 60-day authentic unpredictability measure were relapsed on acknowledged instability for assessment purposes. Although the relapse coefficients were
statically immaterial for the suggested volatilities, recorded unpredictability seemed to have
some illustrative power which was affirmed in their encompassing regression. [1]
Chapter 3: Results and Interpretation

3.1 Data and Variables

The data chosen is the historical data of the Dow Jones Industrial Average 30 of the year 2018. The data was extracted from Yahoo Finance using R language function: GetSymbols. GetSymbols is an R function used to wrap and load any kind of data from different sources, in our case we had the choice to retrieve the data from different sources either Bloomberg or Yahoo finance or any other NYSE internet platform. After finding the data, the latest is fetched through one of the getSymbols methods and then, it is either saved or returned to the function caller. The function in question is assigned directly to the variable as it is spelled and named in the given environment, which is in this case ^DJI in Yahoo Finance.

The R code for this specific function is shown in Figure 3.1

```r
getSymbols("^DJI", src="yahoo", from="2018-1-1", to="2018-12-31")
tsdisplay(DJI$DJI.Close)
```

Figure 3.1. Sample R code function

The value DJI components are high, low, open, close and adjusted close. Since the close value of the DJIA in the market is almost 100% of the time the same as the adjusted close value, the value we are interested in is the close value and is the column on which we will base our forecasting.

A snapshot of how the data was generated, shown in Table 3.1.
Table 3.1. Snapshot of DJI data

<table>
<thead>
<tr>
<th>Date</th>
<th>DJI Open</th>
<th>DJI High</th>
<th>DJI Low</th>
<th>DJI Close</th>
<th>DJI Volume</th>
<th>DJI Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-01-02</td>
<td>23058.61</td>
<td>23413.47</td>
<td>22928.59</td>
<td>23346.24</td>
<td>321570000</td>
<td>23346.24</td>
</tr>
<tr>
<td>2019-01-03</td>
<td>23176.39</td>
<td>23176.39</td>
<td>22638.41</td>
<td>22666.22</td>
<td>262440000</td>
<td>22686.22</td>
</tr>
<tr>
<td>2019-01-04</td>
<td>22894.92</td>
<td>22518.64</td>
<td>22894.92</td>
<td>23433.16</td>
<td>396020000</td>
<td>23433.16</td>
</tr>
<tr>
<td>2019-01-07</td>
<td>23474.26</td>
<td>23687.74</td>
<td>23301.59</td>
<td>23531.35</td>
<td>334000000</td>
<td>23531.35</td>
</tr>
<tr>
<td>2019-01-08</td>
<td>23608.32</td>
<td>23664.65</td>
<td>23601.45</td>
<td>23707.45</td>
<td>217400000</td>
<td>23707.45</td>
</tr>
<tr>
<td>2019-01-09</td>
<td>22944.27</td>
<td>23965.45</td>
<td>22776.36</td>
<td>23879.12</td>
<td>325700000</td>
<td>23879.12</td>
</tr>
<tr>
<td>2019-01-10</td>
<td>23111.11</td>
<td>24014.78</td>
<td>23703.25</td>
<td>24001.92</td>
<td>338150000</td>
<td>24001.92</td>
</tr>
<tr>
<td>2019-01-11</td>
<td>23140.01</td>
<td>23996.32</td>
<td>23798.16</td>
<td>23995.95</td>
<td>262650000</td>
<td>23995.95</td>
</tr>
<tr>
<td>2019-01-14</td>
<td>23800.53</td>
<td>23964.90</td>
<td>23765.24</td>
<td>23909.84</td>
<td>277560000</td>
<td>23909.84</td>
</tr>
<tr>
<td>2019-01-15</td>
<td>23914.11</td>
<td>24099.14</td>
<td>23887.93</td>
<td>24085.59</td>
<td>291570000</td>
<td>24085.59</td>
</tr>
</tbody>
</table>

The reason we chose a recent year is to be up to date to the financial changes of the New York stock exchange market. For instance, changes in the market happen on a long term and close period of time tend to have approximately the same behavior over the seasons. We have been watching the behavior of the DJIA throughout many years, and how historical events such as the Great Depression or crashes in general, affected the level of volatility of the Dow 30.

In the following section, we will be working on the previous data and fitting it to our model which is the ARIMA (Auto-Regressive Integrated Moving Average) model to forecast the data of 2019. Before starting the forecast of 2019 Dow Jones, we have been following its behavior according to the different political events affecting New York stock exchange daily. As an example of that we could notice the first unusual change in the behavior of the Dow by the end of 2018 and beginning of 2019 which coincides with the government’s shutdown. Right after this crash, the DJIA has known an immediate recovery and jumped up to its highest value of the year 2018. We notice a similar behavior between all NYSE indexes in 2013 government shutdown and the latest shutdown. In 2013, after a dramatical crash, the DJIA along with other indexes such as the S&P 500 recovered the same way as it did in 2018. This tells us how the historical pattern of any given index could help us predict a seasonal trend of its behavior according to different circumstances.
3.2 Results of ARIMA forecasting

The Auto Regressive Moving Average is one of the most used methods of forecasting that is majorly used for time series. ARIMA model has two major components:

- **The Auto-Regressive component**: or also called the Lag component, which consists of the correlation between a certain value in a time series with the previous one. It is the difference between two values in a time series that is repeating itself, which usually hints towards a potential redundant behavior of the time series and thus enable the forecasting. For our case the Auto-Regressive component is DJI close.

- **Moving Average components**: the moving average is basically the average of consecutive values of the time series in different periods of time. Its order can vary from a lag to another. If used via excel, when we drag the moving average formula over the whole time series column it gives us the forecasts of the future time period of the series. Table 3.2 shows an example of a time series moving average:

```
Table 3.2. Explanation of moving average over a one-year period

<table>
<thead>
<tr>
<th>Month</th>
<th>Sales</th>
<th>Movave2(Sales)</th>
<th>Movave3(Sales)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-15</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Feb-15</td>
<td>101</td>
<td>101.5</td>
<td>101.5</td>
</tr>
<tr>
<td>Mar-15</td>
<td>98</td>
<td>100.5</td>
<td>100.33</td>
</tr>
<tr>
<td>Apr-15</td>
<td>116</td>
<td>107</td>
<td>105.67</td>
</tr>
<tr>
<td>May-15</td>
<td>120</td>
<td>118</td>
<td>111.33</td>
</tr>
<tr>
<td>Jun-15</td>
<td>100</td>
<td>110</td>
<td>112.00</td>
</tr>
<tr>
<td>Jul-15</td>
<td>130</td>
<td>115</td>
<td>116.67</td>
</tr>
<tr>
<td>Aug-15</td>
<td>133</td>
<td>131.5</td>
<td>121.00</td>
</tr>
<tr>
<td>Sep-15</td>
<td>104</td>
<td>118.5</td>
<td>122.33</td>
</tr>
<tr>
<td>Oct-15</td>
<td>137</td>
<td>120.5</td>
<td>124.67</td>
</tr>
<tr>
<td>Nov-15</td>
<td>143</td>
<td>140</td>
<td>128.00</td>
</tr>
<tr>
<td>Dec-15</td>
<td>105</td>
<td>124</td>
<td>128.33</td>
</tr>
</tbody>
</table>
```

ARIMA modeling goes through five major steps:
**Reading the data and plotting it:** for the retrieval of the data, we used `getSymbols` as stated in the previous section. A second display of the data is provided in the ARIMA model as follows, a snapshot of the graph command in R is shown in Figure 3.2. The time period units are in years with increments of 0.2 years.

![Figure 3.2. Snapshot of the ARIMA data](image)

1- **Checking seasonality and volatility:** after checking one can go for two options which are either to differentiate or to fit ARIMA model to the seasonality of the data. For our case we fit the ARIMA model, using the following function:

```r
fit_TS = ts(DJI$DJI.Close, start=c(2018,01,02),frequency = 250) # This makes TS
```

2- **Non-Stationary data (KPSS test):** In this step we check if the data is stationary, otherwise we adjust for the non-stationary data by taking the differences of the data. A stationary data returns a p-value bigger than 0.05. In our case the data is stationary and returned a p-value of p= 0.075. The code used in R is shown in Figure 3.3:

```r
adf = adf.test(fit)
kpss = kpss.test(fit)
adf
kpss
```

![Figure 3.3. Sample R code for Non-Stationary data (KPSS test)](image)

The output values after running the KPSS test are shown in Figure 3.4.
> kpss

KPSS Test for Level Stationarity
data:  fit
KPSS Level = 0.40368, Truncation lag parameter = 5, p-value = 0.07557

Figure 3.4. Sample output of the R code for Non-Stationary data (KPSS test)

3- ACF and PACF functions: to identify orders of p, d and q the arguments of the ARIMA model, we use the auto correlation coefficient, to analyze the correlation between values of one variable in different times, and we also use the partial auto correlation function which is the conditional correlation among the values analyzed in the previous function. the loop we did for all the values of the time series is shown in Figure 3.5:

```r
> TSgraph=function(series, nlag=30){
17      layout(1:3)
18      plot(series)
19      acf(series, nlag)
20      pacf(series, nlag)
21      layout(1)
22 }
```

Figure 3.5. Time series loop

The output of the ACF and PACF functions are shown in Figure 3.6.

Figure 3.6. The output of the ACF and PACF functions

Since the order of our moving average is 1 , we are noticing only one significant correlation coefficient in the first lag of the ACF function. ACF1 = 0.061, means the lack of correlation between our data points in the following lags.
4- **Residuals Checking:** check if residuals are uncorrelated or normally distributed or have a constant variance. The output of our residuals is shown in Figure 3.7 and Table 3.3.

Figure 3.7. ACF Residual output

![ACF Residual output](image1)

Table 3.3. Residuals statistical results

<table>
<thead>
<tr>
<th>Test set</th>
<th>ME</th>
<th>RMSE</th>
<th>MAE</th>
<th>MPE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>789.8518</td>
<td>888.923</td>
<td>807.6327</td>
<td>3.087859</td>
<td>3.166236</td>
</tr>
</tbody>
</table>

5- **Complete the forecast and Fit the model:** after performing the previous tests and checking for residuals.

```r
fitarima <- arima(fit_TS, order = c(1, 0, 0))
summary(fitarima)
tsdig(fitarima)
## this is a test to see how your model fits, if all the dots are above the line
future2 = forecast(fitarima, h = 63)
plot(future2)
```

Figure 3.8. R code which checks for residuals

⇒ **Output final model for the ARIMA forecasting:**
Figure 3.9. Final model for the ARIMA forecasting

\[ P = 1 : \text{number of auto regressive terms.} \]

\[ D = 0: \text{number of seasonal differences needed to achieve stationarity. It is null because our data is already stationary.} \]

\[ Q = 0: \text{number of lagged forecast errors in prediction equation. It is also null since our model gives a mean of the predictions and is accurately fitted to the real data; therefore, errors are almost inexistent.} \]

The results of our ARIMA forecasting is an immediate recovery in 2019 after the crash due to the government shutdown by the end of the year 2018. Our forecasting lead to an envelope of an interval of +/- 800 on a scale of 20000 which shows the accuracy of our model predictions for 2019.

3.3 Comparison and Interpretation

3.3.1 Forecasts and Real Dow Jones

For our auto-regressive moving average model, we performed a forecasting based on different assumptions. Our first assumption is our choice of the time period on which to perform the forecasting which is a one-year time period. We choose the year 2018 because
it is the latest time period before the first quarter of 2019, that is, the behavior of the Dow Jones industrial average in that period would be a kind of reflection of what could happen in 2019, since the market and stock prices would have approximately the same tendency for in two or three years. If we supposedly chose a period of 3 years, the market and the DJIA would be more volatile given the bigger scope of time chosen, and this will make our predictions less accurate. It is very common in financial analysis in general, to use the most recent historical data of the company or its derivatives in order to forecast its future values. The one-year time period we choose gave the most accurate and close results to the real close value of the Dow Jones. Figure 3.10 shows the change of the future value of the Dow based on our prediction, as stated previously the forecasting starts by the first quarter of 2019:

![Figure 3.10. Change in the future value of the Dow based on our prediction](image)

The code used for this model is shown in Figure 3.11.

```r
40 pred <- predict(fitarima, n.ahead = 22*3) # predict 3 months
41 predl <- as.vector(as.matrix(pred$pred))
42 transformeddata <- as.data.frame(DJI)
43 y <- seq(as.Date('2019-01-01'), as.Date('2019-04-02'), by = 1)
44 y <- y[!weekdays(y) %in% c('Saturday', 'Sunday')]
45 y <- c(as.Date(rownames(transformeddata)), y)
46 x <- c(as.vector(transformeddata[, 4]), predl)
```

![Figure 3.11. R code used to model the forecasting](image)
Comparing this result with the ARIMA model graph shown in the previous section, shows the accuracy of our prediction. In the figure above, we see a linear recovery of the DJIA that is within the envelope of the ARIMA result in the previous section.

To test and visualize the accuracy of our forecasting using the ARIMA model, we plotted the future value with the actual value of the Dow Jones in 2019. Figure 3.12 shows the mean of the future value of the Dow compared to its actual value in 2019.

![Figure 3.12. Mean of the future Dow value compared to its actual value in 2019](image)

The x-axis is in days, showing 3 months of 2019. As we can notice in the graph above, the mean of our future DJIA tend almost all the time to be less than that the actual value. This result shows that we are more likely to underestimate the DJIA value rather than overestimating it. However, we can still claim the accuracy of our prediction since the fluctuations of the actual Dow 30 follow the same line trend we got as a result. Additionally, we can notice that the highest error occurred in the beginning of the year 2019, which is of a value of -1500.

In the figure below, we are showing the result of our forecasts, given multiple data. The graph shows the lowest and highest expectations of the Dow Jones in year 2019’s first quarter. When performing ARIMA model of forecasting, we got an envelope which represents the interval of
our prediction of the behavior of the Dow 30, a plot of this interval is shown in Figure 3.13.

Below, is the graph result that was generated when performing the following function on R: “
```
matplot(cbind(future2$mean,future2$lower,future2$upper,DJI$DJI.Close),type='l')
```

![Graph of Asset 30 Prediction](image)

**Figure 3.13. Low, upper and mean of our prediction of the behavior of the Dow 30**

The Y-axis represents the values of the Dow jones in 2019 first quarter, the X-axis represents the days of the first three months. The blue lines represent the upper future of the Dow Jones, and the red and green line represent the lower future of the Dow Jones, the black line is the mean of our predictions and the purple fluctuated line represent the real data of the Dow. We notice the real Dow fluctuates on the upper envelope of the ARIMA forecasting. However, in the beginning of the year2019, we notice an inaccuracy between the prediction and the real Dow, in fact, the purple line fluctuates in the lower envelope, going in an opposite direction of our prediction.

Before getting those results, we have conducted a residual study of the Dow, and tried to choose the optimal variable of the ARIMA that would lead to the least number of residuals. We found out that this is the best fit for our forecasting, which is the first order ARIMA with non-zero mean.
In the following section, we will be interpreting the results, and try to explain the reason behind the seemingly inaccurate prediction in the beginning of 2019.

**3.3.2 Interpretation and Explanation**

- **Test Results:**

Given the results we obtained from the KPSS test, it returned a p-value = 0.07 > 0.05, which showed that our data is stationary. This could be interpreted as the following: our time series is around a linear trend and following a systematic pattern. However, we cannot be decisive on that since occasional events happen to affect the Dow Jones, as an example of that the U.S government shutdown that drove the stock prices down and the U.S China trade that caused a stagnation in the stock market. Further in our interpretation of residuals we will explore how these affairs affected the accuracy of our prediction. For our ACF results we can notice only one significant correlation between the residuals which is in the first lag, which shows again the highest error or seemingly inaccurate result in the first 10 days of 2019.

ACF1: Auto Correlation function at lag 1, gave a result of $\text{ACF} = 0.06$ which is almost null. This means that our data points are uncorrelated to the previous data points, meaning that the value of the Dow Jones on a market day is less likely to affect the one of the following day.

- **Residuals:**

When treating our residuals, we used in our forecasting the optimal variables. For instance, choosing zero seasonal component was the most optimal solution given the kind of time series we are dealing with. The Dow Jones represents 30 companies, some are from the same industries some are not. This makes our model a prediction for multiple companies, and since the Dow is an average of these companies, any change in seasons means a change in the entire average of these companies at the same time. This cannot be true or logical, since the economy is permanently changing, and each industry has a different change according to seasons. This
means that a seasonal component in this case would be meaningless and could not be effective unless we choose a large set of data extending to almost 100 years. The residuals data is presented in Table 3.4

<table>
<thead>
<tr>
<th>Test set</th>
<th>ME</th>
<th>RMSE</th>
<th>MAE</th>
<th>MPE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>789.8518</td>
<td>888.923</td>
<td>807.6327</td>
<td>3.087859</td>
<td>3.166236</td>
</tr>
</tbody>
</table>

We could interpret our residuals as the following:

ME: which is the mean error that represents the mean difference between modelled and observed values. ME= 789.85 which represents 3% of the range of our data, this means that our prediction has very low number of errors.

RMSE: Root Mean Squared Error, which takes the square root of each difference between the observed and the model. Using RMSE gives us a value of 888.

MAE: Mean Absolute Error. It is like ME, however the errors in MAE are positive which means that the positive values do not cancel negative ones, meaning that it measures the average magnitude of a forecasting without taking into consideration the direction. It gave us a value of 807 that is smaller than RMSE, which means that there is an increase in the variance of the frequency distribution of errors.

MPE: Mean Percentage Error, where each error is an expression of the percentage of forecast estimate. It is not relevant in our case since it is used to compare the accuracy of forecasts between different time series, and we only have one time series.

MAPE: Mean Absolute Percentage Error, it is the same as MPE, but all errors inversed into positive values, therefore no cancelation happens between positive and negative values. MAPE= 3.16 meaning that our forecast is off by 3.16%.
To evaluate our residuals we plotted a density graph of residuals to represent the distribution of our errors, and obtained the results shown in Figure 3.14 and Figure 3.15.

Figure 3.14. Histogram of residuals

Figure 3.15. Density of residuals
We notice that our residuals are normally distributed, meaning that most of the error values are around zero, and that the mean and median of our residuals are the same. This result asserts the accuracy of our data.
Chapter 4 : Future Work

Financial analysts use various methods in order to forecast future values of indexes or stock prices or derivatives in general. One of the widely used methods is the Generalized Autoregressive Conditional Heteroskedastic model, known as GARCH. To explore further in our research, we looked up this model as an alternative to ARIMA model of forecasting. As a future work in the direction of our study, one could explore GARCH model method and evaluate its result for the futures of the Dow Jones.

Most finance and banking related applications are multivariate issues with unpredictability figures as one of the inputs. Arrangements of fluctuation is generally effectively done using the Generalized Autoregressive Regressive Heteroskedastic (GARCH) model, for example the restrictive fluctuation network is demonstrated as an element of past returns. The original of models, for instance the vector mistake remedy (VEC) model of Bollerslev and the BEKK model of Engle and Kroner, are immediate expansions of the univariate GARCH model of Bollerslev. These models are general and take into consideration adaptable elements for the contingent change grid. They have been broadly used to show unpredictability overflows and in applications, for example, restrictive capital resource evaluating model (CAPM) and hedges of futures. [10]

Lately, there is more focus on bigger scale issues, for example, elements of relationships among value and bond returns, portfolio choice and incentive in danger. In these applications, the numerical assessment of original models progresses toward becoming unfeasible. Both the quantity of parameters and the multifaceted nature of the probability work tend to detonate quickly with the quantity of arrangement. Elective methodologies for accomplishing increasingly reasonable results. However, one of the limitations GARCH model could be facing is in the case of a multivariate time series, the model cannot predict the change in the
variance over time. Like ARIMA model of forecasting time series, GARCH model uses lags as parameters, for instance its second parameter $q$ that represents the squared residual errors could also be set to 0 or 1. Generally speaking this approach expects the series to be stationary, if not a future analyst would want adjust for the differences. [10]

In our case, GARCH model could be used to analyze the volatility of the market. The latest is in fact represented by the Dow 30 that groups the 30 most significant stocks of the New York Stock Exchange. Furthermore, this model could be extended to option valuation of financial derivatives. GARCH model results would be a good estimation for the index returns as used in many research papers previously.
Chapter 5 : Conclusion

In our study, we explored different financial and mathematical techniques in order to come up with a method that will enable us to forecast the Dow Jones Industrial average, which is the most significant index of New York Stock Exchange. The project started by studying the different advancement achieved in this field through a detailed literature review. The latest was the focus of the study, since the subject was new and comprising advanced mathematical models. Various models were discovered such as the GARCH model, Black-Scholes model, that were used in previous capstone projects. We performed and evaluated the results of ARIMA model of forecasting on other finance related subjects, such as stock pricing series and sales series of a given company. We came up to the conclusion that the model could also be used to forecast a market index, which would potentially lead to forecasting instead of a stock, a whole market.

With the adjustment for residuals and the optimal choice of variables we came choose the Auto Regressive Moving Average of first order without a seasonal component. Results of our forecasting for the first quarter of 2019, had the same trend as the actual Dow Jones. More importantly, we could notice a logic in our predictions given the actual changes in the U.S economy. As a future recommendations, financial analyst should be more in the field of practice when evaluating the financial situation of a certain company, in order to come up with a more realistic evaluation that would help them better predict the futures of its stock price, and decrease the degree of uncertainty in the market.
Chapter 6: References


Appendix A

```r
library(quantmod)
library(ggplot2)
library(Forecast)
library(tseries)
library(fpp)
library(lambda.r)

getSymbols("^ADJI", src="yahoo", from="2018-1-1", to="2018-12-31")
tsdDisplay(DJISDJIClose)

fit_TS = ts(DJISDJIClose, start=c(2018,01,02),frequency = 250)

# plotting function
Tgraph_function(series, nlag=30){
  layout(1:3)
  plot(series)
  acf(series, nlag)
  pacf(series, nlag)
  layout(1)
}

## Firstly, we should think if the TS is stationary or not, it looks more or less stationary, so let's continue

Tgraph(fit_TS,40) # PACF has first lag out of bounds. It suggests AR 1

# It also as many ACFs out of bounds, but we prefer AR 1 over MA 24, because MA 24 will have too much parameters
# also, there is no concrete seasonality, so we stick with AR 1 without seasonality.

fitarima <- arima(fit_TS, order = c(1, 0, 0))
summary(fitarima)
tsdia(fitarima)

## this is a test to see how your model fits, if all the dots are above the line
future2 = forecast(fitarima, h = 63)
plot(future2)

plot(future2mean)

getSymbols("^ADJI", src="yahoo", from="2019-1-1", to="2019-04-03")
matplot(cbind(future2mean, future2slower, future2supper, DJISDJIClose), type='l')
nrow(DJII)
length(future2mean)
accuracy(future2mean, DJISDJIClose)

model <- lm (future2mean ~ DJISDJIClose)
resid(model) # List of residuals
plot(density(resid(model))) # A density plot
hist(residuals(model))
```
### Appendix B

<table>
<thead>
<tr>
<th>Date</th>
<th>DJI.Open</th>
<th>DJI.High</th>
<th>DJI.Low</th>
<th>DJI.Close</th>
<th>DJI.Volume</th>
<th>DJI.Adjusted</th>
</tr>
</thead>
<tbody>
<tr>
<td>2019-01-02</td>
<td>23058.61</td>
<td>23413.47</td>
<td>22928.59</td>
<td>23346.24</td>
<td>321570000</td>
<td>23346.24</td>
</tr>
<tr>
<td>2019-01-03</td>
<td>23176.39</td>
<td>23176.39</td>
<td>22638.41</td>
<td>22586.22</td>
<td>424240000</td>
<td>22666.22</td>
</tr>
<tr>
<td>2019-01-04</td>
<td>22894.92</td>
<td>23518.64</td>
<td>22894.92</td>
<td>23433.16</td>
<td>395020000</td>
<td>23433.16</td>
</tr>
<tr>
<td>2019-01-07</td>
<td>23474.26</td>
<td>23687.74</td>
<td>23301.59</td>
<td>23531.35</td>
<td>334200000</td>
<td>23531.35</td>
</tr>
<tr>
<td>2019-01-08</td>
<td>23680.32</td>
<td>23864.65</td>
<td>23581.45</td>
<td>23787.45</td>
<td>317420000</td>
<td>23787.45</td>
</tr>
<tr>
<td>2019-01-09</td>
<td>23844.27</td>
<td>23985.45</td>
<td>23776.56</td>
<td>23879.12</td>
<td>324570000</td>
<td>23679.12</td>
</tr>
<tr>
<td>2019-01-10</td>
<td>23811.11</td>
<td>24014.76</td>
<td>23703.25</td>
<td>24001.92</td>
<td>338150000</td>
<td>24001.92</td>
</tr>
<tr>
<td>2019-01-11</td>
<td>23940.01</td>
<td>23996.32</td>
<td>23798.16</td>
<td>23995.95</td>
<td>262550000</td>
<td>23995.95</td>
</tr>
<tr>
<td>2019-01-14</td>
<td>23880.53</td>
<td>23964.90</td>
<td>23765.24</td>
<td>23909.84</td>
<td>277560000</td>
<td>23909.84</td>
</tr>
<tr>
<td>2019-01-15</td>
<td>23914.11</td>
<td>24099.14</td>
<td>23887.93</td>
<td>24065.59</td>
<td>291570000</td>
<td>24065.59</td>
</tr>
<tr>
<td>2019-01-16</td>
<td>24139.91</td>
<td>24288.61</td>
<td>24119.72</td>
<td>24207.16</td>
<td>302330000</td>
<td>24207.16</td>
</tr>
<tr>
<td>2019-01-17</td>
<td>24147.09</td>
<td>24474.46</td>
<td>24088.90</td>
<td>24370.10</td>
<td>288590000</td>
<td>24370.10</td>
</tr>
<tr>
<td>2019-01-18</td>
<td>24534.19</td>
<td>24750.22</td>
<td>24459.03</td>
<td>24706.35</td>
<td>372970000</td>
<td>24706.35</td>
</tr>
<tr>
<td>2019-01-22</td>
<td>24607.76</td>
<td>24607.76</td>
<td>24244.31</td>
<td>24404.48</td>
<td>338400000</td>
<td>24404.48</td>
</tr>
<tr>
<td>2019-01-23</td>
<td>24577.25</td>
<td>24700.98</td>
<td>24307.17</td>
<td>24575.62</td>
<td>318500000</td>
<td>24575.62</td>
</tr>
<tr>
<td>2019-01-24</td>
<td>24579.96</td>
<td>24626.30</td>
<td>24422.73</td>
<td>24553.24</td>
<td>320170000</td>
<td>24553.24</td>
</tr>
<tr>
<td>2019-01-25</td>
<td>24687.21</td>
<td>24860.15</td>
<td>24676.75</td>
<td>24737.20</td>
<td>375890000</td>
<td>24737.20</td>
</tr>
<tr>
<td>2019-01-28</td>
<td>24596.08</td>
<td>24596.08</td>
<td>24323.94</td>
<td>24528.22</td>
<td>347170000</td>
<td>24528.22</td>
</tr>
<tr>
<td>2019-01-29</td>
<td>24519.62</td>
<td>24674.87</td>
<td>24504.04</td>
<td>24579.96</td>
<td>330870000</td>
<td>24579.96</td>
</tr>
<tr>
<td>2019-01-30</td>
<td>24825.52</td>
<td>25109.62</td>
<td>24790.90</td>
<td>25014.86</td>
<td>397330000</td>
<td>25014.86</td>
</tr>
<tr>
<td>2019-01-31</td>
<td>24954.48</td>
<td>25049.62</td>
<td>24842.09</td>
<td>24999.67</td>
<td>507350000</td>
<td>24999.67</td>
</tr>
<tr>
<td>2019-02-01</td>
<td>25025.31</td>
<td>25193.15</td>
<td>24982.49</td>
<td>25063.89</td>
<td>372000000</td>
<td>25063.89</td>
</tr>
<tr>
<td>2019-02-04</td>
<td>25062.12</td>
<td>25239.91</td>
<td>24977.67</td>
<td>25239.37</td>
<td>293540000</td>
<td>25239.37</td>
</tr>
<tr>
<td>2019-02-05</td>
<td>25287.93</td>
<td>25427.32</td>
<td>25287.65</td>
<td>25411.52</td>
<td>310550000</td>
<td>25411.52</td>
</tr>
<tr>
<td>2019-02-06</td>
<td>25371.57</td>
<td>25439.04</td>
<td>25312.06</td>
<td>25390.30</td>
<td>247590000</td>
<td>25390.30</td>
</tr>
<tr>
<td>2019-02-07</td>
<td>25265.81</td>
<td>25314.26</td>
<td>25000.52</td>
<td>25169.53</td>
<td>295460000</td>
<td>25169.53</td>
</tr>
<tr>
<td>2019-02-08</td>
<td>25042.36</td>
<td>25106.39</td>
<td>24803.04</td>
<td>25106.33</td>
<td>248210000</td>
<td>25106.33</td>
</tr>
<tr>
<td>Date</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
<td>Value 5</td>
<td>Value 6</td>
</tr>
<tr>
<td>------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>2019-02-25</td>
<td>26126.15</td>
<td>26241.42</td>
<td>26080.66</td>
<td>26091.95</td>
<td>281600000</td>
<td>26091.95</td>
</tr>
<tr>
<td>2019-02-26</td>
<td>26051.61</td>
<td>26155.29</td>
<td>25966.01</td>
<td>26057.98</td>
<td>257050000</td>
<td>26057.98</td>
</tr>
<tr>
<td>2019-02-27</td>
<td>25995.00</td>
<td>26039.68</td>
<td>25877.24</td>
<td>25985.16</td>
<td>253380000</td>
<td>25985.16</td>
</tr>
<tr>
<td>2019-02-28</td>
<td>25984.28</td>
<td>26029.21</td>
<td>25896.56</td>
<td>25916.00</td>
<td>338520000</td>
<td>25916.00</td>
</tr>
<tr>
<td>2019-03-01</td>
<td>26019.67</td>
<td>26143.92</td>
<td>25914.37</td>
<td>26026.32</td>
<td>280800000</td>
<td>26026.32</td>
</tr>
<tr>
<td>2019-03-04</td>
<td>26122.19</td>
<td>26155.09</td>
<td>25611.55</td>
<td>25819.65</td>
<td>295510000</td>
<td>25819.65</td>
</tr>
<tr>
<td>2019-03-05</td>
<td>25829.07</td>
<td>25877.15</td>
<td>25725.63</td>
<td>25806.63</td>
<td>252260000</td>
<td>25806.63</td>
</tr>
<tr>
<td>2019-03-06</td>
<td>25818.76</td>
<td>25837.61</td>
<td>25633.71</td>
<td>25673.46</td>
<td>255250000</td>
<td>25673.46</td>
</tr>
<tr>
<td>2019-03-07</td>
<td>25645.45</td>
<td>25645.45</td>
<td>25352.55</td>
<td>25473.23</td>
<td>239140000</td>
<td>25473.23</td>
</tr>
<tr>
<td>2019-03-08</td>
<td>25347.38</td>
<td>25466.14</td>
<td>25252.46</td>
<td>25450.24</td>
<td>270500000</td>
<td>25450.24</td>
</tr>
<tr>
<td>2019-03-11</td>
<td>25208.00</td>
<td>25661.63</td>
<td>25208.00</td>
<td>25650.88</td>
<td>309790000</td>
<td>25650.88</td>
</tr>
<tr>
<td>2019-03-12</td>
<td>25600.30</td>
<td>25675.44</td>
<td>25522.17</td>
<td>25554.66</td>
<td>317730000</td>
<td>25554.66</td>
</tr>
<tr>
<td>2019-03-13</td>
<td>25637.23</td>
<td>25776.49</td>
<td>25571.31</td>
<td>25702.09</td>
<td>331410000</td>
<td>25702.09</td>
</tr>
<tr>
<td>2019-03-14</td>
<td>25592.31</td>
<td>25752.84</td>
<td>25621.31</td>
<td>25709.94</td>
<td>302790000</td>
<td>25709.94</td>
</tr>
<tr>
<td>2019-03-15</td>
<td>25720.96</td>
<td>25927.91</td>
<td>25649.70</td>
<td>25848.87</td>
<td>615910000</td>
<td>25848.87</td>
</tr>
<tr>
<td>2019-03-18</td>
<td>25801.89</td>
<td>25924.77</td>
<td>25785.66</td>
<td>25914.10</td>
<td>399530000</td>
<td>25914.10</td>
</tr>
<tr>
<td>2019-03-19</td>
<td>25987.87</td>
<td>26109.68</td>
<td>25814.92</td>
<td>25887.38</td>
<td>342250000</td>
<td>25887.38</td>
</tr>
<tr>
<td>2019-03-20</td>
<td>25867.79</td>
<td>25929.52</td>
<td>25670.63</td>
<td>25745.67</td>
<td>321980000</td>
<td>25745.67</td>
</tr>
<tr>
<td>2019-03-21</td>
<td>25688.44</td>
<td>26009.90</td>
<td>25657.78</td>
<td>25962.51</td>
<td>335250000</td>
<td>25962.51</td>
</tr>
<tr>
<td>2019-03-22</td>
<td>25644.65</td>
<td>25677.01</td>
<td>25501.45</td>
<td>25502.32</td>
<td>386510000</td>
<td>25502.32</td>
</tr>
<tr>
<td>2019-03-23</td>
<td>25490.72</td>
<td>25603.27</td>
<td>25372.26</td>
<td>25516.83</td>
<td>303120000</td>
<td>25516.83</td>
</tr>
<tr>
<td>2019-03-24</td>
<td>25649.56</td>
<td>25796.29</td>
<td>25544.78</td>
<td>25657.73</td>
<td>305110000</td>
<td>25657.73</td>
</tr>
<tr>
<td>2019-03-27</td>
<td>25676.34</td>
<td>25758.17</td>
<td>25425.27</td>
<td>25625.59</td>
<td>270340000</td>
<td>25625.59</td>
</tr>
<tr>
<td>2019-03-28</td>
<td>25693.32</td>
<td>25743.41</td>
<td>25576.69</td>
<td>25717.46</td>
<td>237170000</td>
<td>25717.46</td>
</tr>
<tr>
<td>2019-03-29</td>
<td>25827.31</td>
<td>25949.32</td>
<td>25771.67</td>
<td>25928.68</td>
<td>300620000</td>
<td>25928.68</td>
</tr>
<tr>
<td>2019-04-01</td>
<td>26075.10</td>
<td>26280.90</td>
<td>26071.69</td>
<td>26258.42</td>
<td>295570000</td>
<td>26258.42</td>
</tr>
<tr>
<td>2019-04-02</td>
<td>26213.55</td>
<td>26221.24</td>
<td>26122.31</td>
<td>26179.13</td>
<td>250940000</td>
<td>26179.13</td>
</tr>
</tbody>
</table>
Appendix C

Initial Specifications:

The aim of this capstone project is to compare different indicators of forecasting the market volatility index of the US; DJIA (Dow Jones Industrial Average). The latest is a price-weighted average of the 30 most significant stocks in the New York Stock Exchange Market.

The DJIA forecast will consist of two forecasting indicators: Historical data using the ARIMA model of forecasting. The analysis of those two methods will go through three main stages. First, literature review of the main years during which the market index was affected by the current affairs. Second the collection of the data and choice of the time period that will best fit our model. The last step is to apply the ARIMA model and adjust for variables.

Using the results and accounting for the errors in every indicator, we will be able to evaluate the accuracy of our model and compare our prediction with the real data.

Throughout the process, we will make use of two software: R Studio and Microsoft Excel. The sample chosen will be on a daily basis, taking the data of 2018 and predicting the data of the first quarter of 2019.