Candidate number: u1793821

A Study in UK and US Stock Market Indices After the Global Financial Crisis

CS910 Foundations of Data Analytics Project Report

***Abstract*—This paper will analyse how the UK stock market responds compared with the US stock market to the global financial crisis which peaked in 2008 and had giant influence not only on the US but also on the rest of the world. The study is carried out by evaluating the behaviours of two main stock market indices in the UK and US stock markets, FTSE100 and S&P500, focusing on the post-financial crisis period. Data analytics tools are used to identify the pattern of the two stock market indices and also to briefly predict the future trend of the indices in the next calendar year.**

Keywords— FTSE100, S&P500, stock market, financial crisis

I. Introduction

The financial crisis which began in 2007 and peaked in 2008 has become the most severe financial crisis in the recent two centuries after the Great Depression in 1930s [1]. Although it initially became visible as the subprime mortgage crisis in the United States housing market, its later consequences were neither limited in the US market nor in the housing sectors. It led to a tremendous shock in the global stock market and caused a severe downturn in the global economic and financial activities after the breakout in 2008 [1]. It is definite that the United Kingdom, as one of the largest national economies in the world, has also suffered significantly through this process. In this paper, we would like to compare and discuss the behaviours of the UK and US stock markets after the financial crisis. We will take the data from 2007 of two main stock market indices in the UK and US stock markets, the Financial Times Stick Exchange 100 Index and the Standard and Poor's 500 Index. We will then analyse the behaviours of these two indices in the post-financial crisis periods by applying various data analytics tools. Another expected outcome of this study is to provide a rough and preliminary discussion about how these stock market indices will behave in 2018 based on the previous stock data.

II. Background

A. Financial Crisis in 2007-2008

A significant indicator of the global Financial Crisis was the US housing bubble arising in the early 2000s due to the progressive decline in the target federal-funds rate of the US. As the Monetary Policy Report to the Congress in 2002 submitted by the Federal Reserve Board mentioned, the Board underlined the demand of speeding up the economic growth and decided to cut the federal-funds rate to a historically low interest rate below 2%[2]. This encouraged the investment banks to lend money to subprime borrowers and drove up the demand of the American housing market.

What's more, the bankers not only lent their own capitals out. As a result of the deregulations of the Securities and Exchange Commission (SEC) in 2004, the broker-dealers were allowed to operate with a leverage of scores of their initial investments [3]. This boosted up the subprime mortgage market and also formed the secondary market to distribute subprime loans.

We know from the results that those subprime borrowers were not able to pay their money back and the high leverage cost the investment bankers hugely after the bubble disappeared. There were a massive number of financial institutions having suffered from this giant disaster: Bear Sterns, the fifth largest investment bank of the US at that time, was sold to JP Morgan Chase in 2008 after the Fed bailout; Lehman Brothers, which used to be the fourth largest investment bank and was operated for over 150 years, later filed for bankruptcy due to the failure in the subprime mortgage funds; Merrill Lynch, the third largest investment bank back to then, subsequently agreed to be acquired by the Bank of America [4][5]. It was not just the US market which suffered from this systemic crisis. After it burst out, there was a turmoil in the global financial markets. The Northern Rock, for instance, formerly a private-owned British bank, experienced a bank run due to people's lack of confidence after the financial crisis and turned into public ownership in 2008 resulted from insolvency [6].

The scale of the consequences is much larger than just the financial sectors. The United Stated lost 15 million jobs right after the crisis [7]. The UK GDP Growth exceeded a negative 4% in 2009 as Figure 1 shows [8]. It caused a huge shrink of the global trade and also ushered the European sovereign debt crisis since the end of 2009 [7]. Although the United Kingdom is not in the Eurozone, it is also affected by the debt crisis till nowadays.

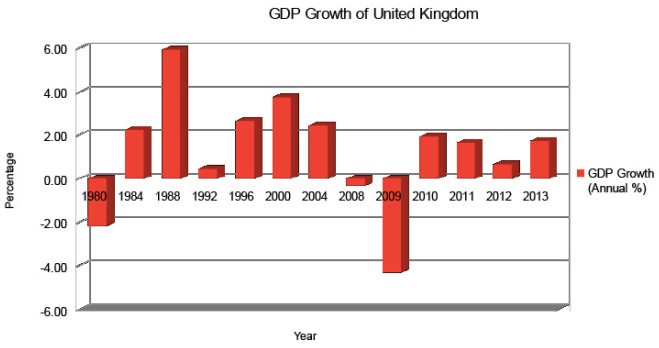

Figure 1: The UK GDP growth from 1980 to 2013

In order to minimise the aftermath of this severe crisis, the government of each national economy released different targeted monetary policies hoping to gather the pace of economic recovery. Figure 2 shows the national target interest rates and the government bond yields of the UK and the US. The base rate announced by the Bank of England after the global financial crisis is always below the US Federal-funds target rate, while Britain has a higher yield of the ten-year government bond [9].

Figure 2: The UK and US monetary policies after 2008

B. *The Financial Times Stock Exchange 100 Index*

The Financial Times Stock Exchange 100 Index (FTSE100) is a market-capitalisation weighted stock index of 101 largest blue-chip companies listed on the London Stock Exchange [10][11]. It was launched in 1984 and is maintained by the FTSE Group headquartered in London [11].

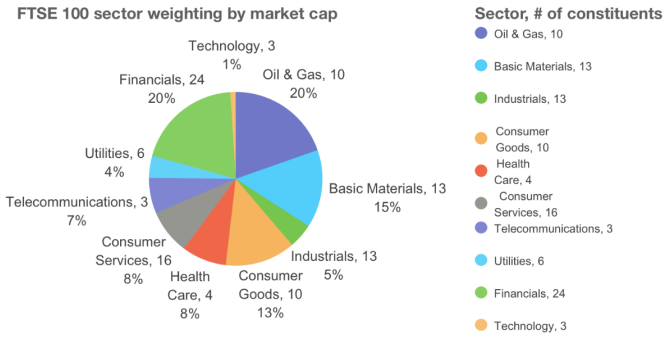

Figure 3: The Sector Breakdown of FTSE100 in 2017

The index is part of the FTSE UK Series and its current Top 5 constituents are HSBC Holdings, British American Tobacco, Royal Dutch Shell A, BP and Royal Dutch Shell B [11]. It can be told from Figure 3 that the Oil and Gas, Financials and Basic Materials are the Top 3 industrial sectors of FTSE100 [12].

C. *The Standard and Poor's 500 Index*

There are a multiple of stock indices in the US stock market. The one we decided to use for this study is the Standard and Poor's 500 Index and there are various reasons why it is selected.

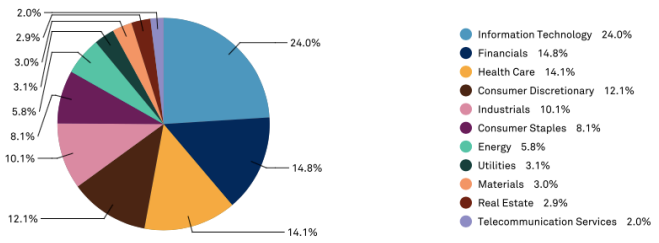

Figure 4: The Sector Breakdown of S&P500 in 2017

The Standard & Poor's 500, often abbreviated as the S&P 500 is an American stock market index created in 1957 based on the market capitalisation of 500 largest American companies having common stocks listed on the New York Stock Exchange (NYSE) or Nasdaq Stock Market [13][14].

The index is operated and its component weightings are determined by S&P Dow Jones Indices [15]. There are some other frequently-used U.S. stock market indices produced by the same company, such as the Dow Jones Industrial Average (the Dow) [15]. In this study, we decide to use S&P500 index, because of its reasonable weighting methodology and good representation of the US stock market. Unlike the Dow which is price-weighted, S&P500 is float-adjusted market-capitalisation weighted, i.e. it is more focusing on the total market share of a specific company than just the individual stock price [14]. This weighting methodology enables the 505 constituents covered by this index to represent as many industries as possible [14]. It is also more comparable with the FTSE100 as both of them are market-capitalisation weighted. The Top 5 Constituents by index weight in 2017 are Apple Inc., Microsoft Corp, Amazon.com Inc., Facebook Inc. and Johnson & Johnson [14]. As Figure 4 shows, the information technology, financials and health care sectors are the Top 3 industrial sectors in the S&P500 constituents [14]. This pattern is slightly different from the FTSE100 sector breakdown shown in Figure 3, as the percentage of energy companies weighs much less in S&P500. On the other hand, the health care and information technology sectors weigh much more than those in FTSE100.

III. Hypothesis

Since the 2008 financial crisis was initiated from the United States, it is a reasonable guess that the US would have suffered more severely in this global economic disaster. However, since the governments published different monetary policies and there are many other factors to consider such as political issues and various impacts on different industrial sectors, it is yet to jump to the conclusion of which national economy is more affected by the global financial crisis. Apart from this supposition, the subsequent hypothesis drawn is that the stock indices in 2008 may have an increasing trend since there are no current disastrous global crisis in the recent years and both countries are under steady economic growth.

In our hypothesis, we postulate that the London Stock Exchange, NYSE and Nasdaq Stock Market operate independently of any political actions. We also assume that the stock price indices are self-contained so that we can draw conclusions from them.

IV. Data Analysis Tools

A. *Weka*

We will use several data analysis tools in this project, first of which is called Weka. Weka, which is in short of Waikato Environment or Knowledge analysis, is developed by the Machine Learning Group at the University of Waikato in New Zealand and is a collection of machine learning algorithms designed for various types of data mining tasks [16]. The group makes Weka open to the public and free for use.

Another advantage of Weka is that it does not require programming skills while using. It is written in Java and has graphical user interfaces as well as command line interfaces

[16]. In this project, we will mainly use Weka to do time series analysis and forecasting, as the stock indices data can be treated as time series.

B. R

R is a more advanced data analysis tool compared with Weka as it requires a considerable amount of programming knowledge. It is especially useful for statistical analysis and data mining and is freely available for the public. R has a wide range of package able to be applied in various contexts. One of the packages we will use in the project is called "quantmod", which is in short of "Quantitative Financial Modelling and Trading Framework" and is developed for quant traders [17]. The "quantmod" package is designed to test and develop statistically based trading models [17]. We can extract and manage stock data, build trading models, plot visualised charts and evaluate period returns using this package [17].

V. THE DATASETS

FTSE100 data is obtained from Yahoo Finance where there are real-time and historical stock data available and free to download [18]. In this project, we extract the FTSE100 index data from the 3rd of January 2007 to the 19th of December 2017, which comprises 2762 trading days. The index is recorded on a daily basis. There are two-day's data missing for every five days since the stock market is closed on weekends. The data is in csv files and is in a size of around 200KB. The attributes are date, open price, close price, the highest price on the day, lowest price on the day, adjusted close price and share volume. The adjusted close price is the close price adjusted after taking into account all applicable splits and dividend distributions [19]. The adjusted close price is more accurate and useful while analysing the historical stock data. We will also use the adjusted close price in the later analysis. As mentioned in the previous section, the stock data can be directly extracted from Yahoo using a specially designed package "quantmod" in R [17]. We also extract the data via this approach in R. Figure 5 shows a plot in R of the adjusted close price in the given period with corresponding volumes at the bottom. The plot is obtained by applying a function called "candleChart" in the "quantmod" package. It can be noticed that there is a significant price drop between 2008 and 2009 and also a general rise afterwards along with a series of small fluctuations. It is also noticeable that the share volume after 2008 has been shrunk to a certain extent. However, the price-against-time plot is still not visualised enough for more detailed discussions and that is why we would like to apply the data analytics tools in the following sections.

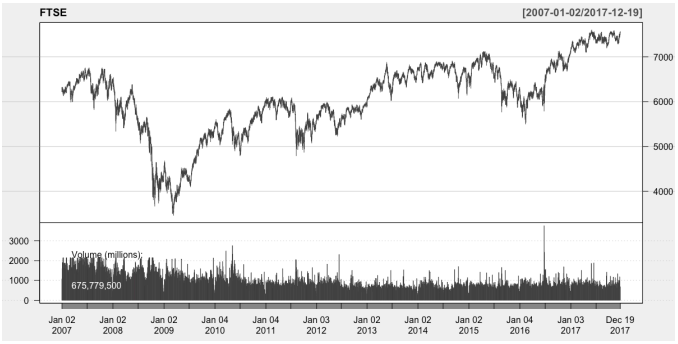

Figure 5: FTSE100 trend and volume in 2007-2017

S&P500 data is also obtained from Yahoo Finance [18]. In this project, we extract the S&P500 index data from the 3rd of January 2007 to the 19th of December 2017, which also comprises 2762 trading days. The index is recorded on a daily basis. There are two-day's data missing for every five days as well. The data is in csv files and is in a size of around 240KB, which is slightly larger than the FTSE100 data since the S&P500 index has larger absolute values. The attributes are the same as FTSE100 data. The trend and volume plot along the given time period produced in R is shown in Figure 6. Apart from the unsurprising drop in 2008-2009, it can also be noticed that the S&P500 index has increased more steadily and faster after the financial crisis compared with the FTSE100 index and the peak value has almost doubled compared with the peak value before the financial crisis. This possibly results from the successful Federal monetary policy of the US in recent years.

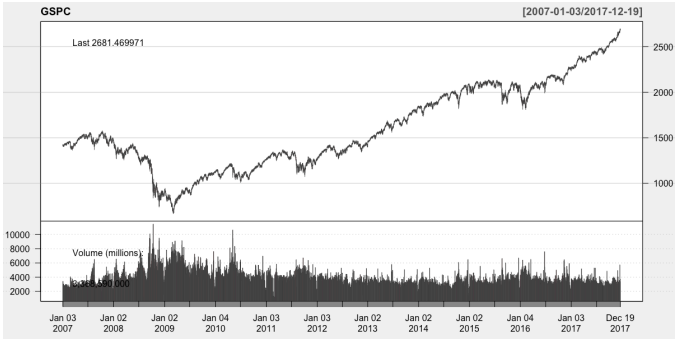

Figure 6: S&P500 trend and volume in 2007-2017

Before further analysing the data using, the preprocessing we need to do is about the "date" attribute. It initially appears as a nominal attribute in Weka, but we would like to treat it as a date attribute so that we can apply time series analysis to it. Therefore, we convert the ".csv" file to an ".arff" file in Weka and edit the head of the ".arff" file so that the "date" attribute can be read as a date attribute in Weka.

VI. DATA PROCESSING AND ANALYSIS

A. Log returns

The rate of returns of a stock index over a single period is $P_t - P_{t-1}$, where P_t is the value of the price indices for period t and t is measured in days in our case. The log return of a stock index R_t is the natural logarithm of the rate of returns, which is $R_t = \ln(P_t) - \ln(P_{t-1})$[20]. The log returns are approximately equal to the ordinary rate of returns when the return value is small. Log returns are preferred while evaluating the stock indices, since logarithm changes at a slower rate than linear models so that only significant and meaningful changes are noticed. Moreover, it is symmetric about zero and its zero-value indicates that there is no change based on previous days, which is more visualised than the ordinary return model. In this subsection, we calculate and plot the daily log returns of the two stock indices using R. The value of the price indices against time is chosen to be the adjusted close price.

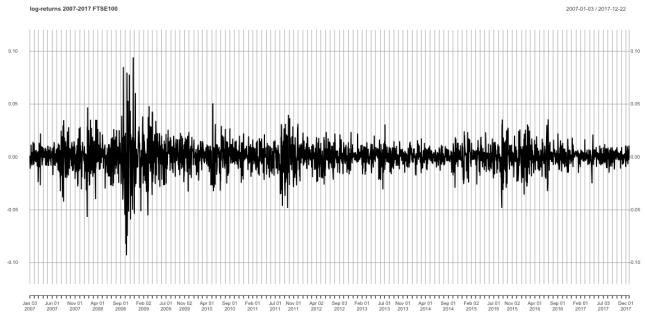

Figure 7: Daily log-returns of FTSE100 in 2007-2017

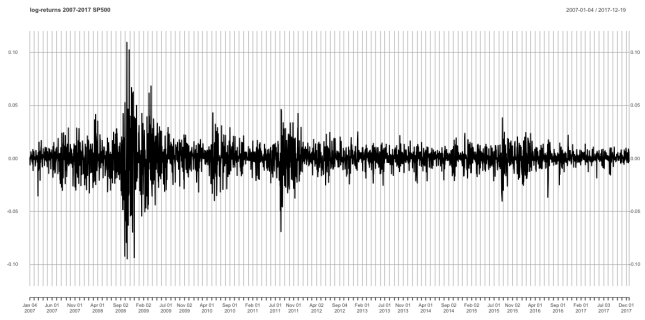

Figure 8: Daily log-returns of S&P500 in 2007-2017

Figure 7 shows the log-returns of FTSE100 index over the whole period that we are interested in and Figure 8 shows the daily log-returns of S&P500 index in 2007-2017 period. Their axis ranges are set to be the same for the convenience of comparison. The most remarkable point is that the S&P500 index has stronger fluctuations, especially over the financial crisis period between 2007 and 2009, which agrees with our previous hypothesis to a certain extent. There are a series of fluctuations after the financial crisis period, possibly due to the aftermath of the financial crisis. It is noticeable that there is a significant change in log returns in 2011. This is mainly caused by the world-known stock market crash in August 2011, led by people's fear of the European sovereign debt crisis, which is one of the famous aftermaths. The change in log returns becomes much milder after that period and the S&P500 index seems to behave more stable than the FTSE100 index in the recent two years according to the figures.

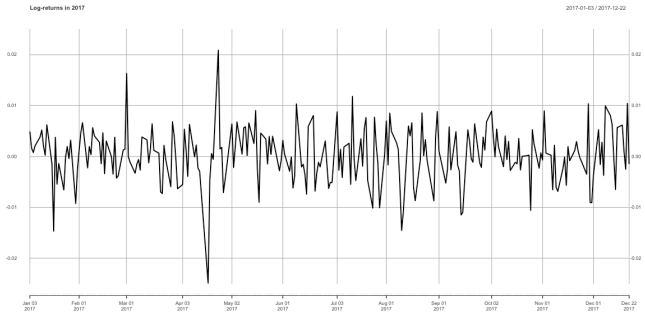

Figure 9: Daily log-returns of FTSE100 in 2017

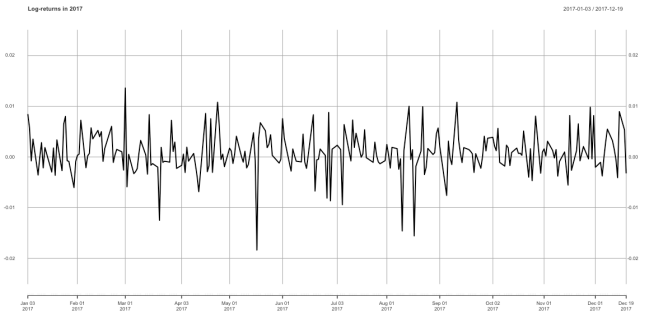

Figure 10: Daily log-returns of S&P500 in 2017

After analysing the whole dataset, we are interested in the log returns in the current year as well, which is shown in Figure 9 and Figure 10. The previous log return plots (Figure 7 and 8) range from -0.12 to 0.12, while these two figures have y-axis bounded by -0.025 and 0.025. Therefore, it can be told from the scale that the stock market becomes much more involatile nowadays. However, the FTSE100 index fluctuated relatively markedly in April and May 2017, while the S&P500 index did not respond as significantly as the FTSE100 did. This was probably due to the Brexit in 2017. As we know, Prime Minister Teresa May signed and invoked the Article 50 to the European Council on the 29th of March 2017 [21]. This could result in a vibration of the stock market and the US market was less affected as the most companies listed in the London Stock Exchange and included in the FTSE100 are Europe- or UK-based companies. Other than this observation, the rest pattern of this year's log returns is stable and satisfying.

B. Probability of volatility

The log returns are always calculated based on previous data. Since it is a lag approach, we are unable to predict and get an idea of what will happen in the next time period by simply observing the log return plots. Therefore, in this section, another approach is introduced, the probability of volatility. In finance, the term volatility represents "a generic measure of the magnitude of market fluctuations" [22]. In this context, we consider the volatility as the standard deviation of the daily logarithmic rate of returns, i.e. $\sigma_t = \sqrt{R_t}$, where R_t is the log return as introduced in the previous subsection [20]. The probability of volatility here is a conditional probability based on the previous variances [23]. Since the log returns are continuous, we use their Gaussian emission probability density function to calculate the probability.

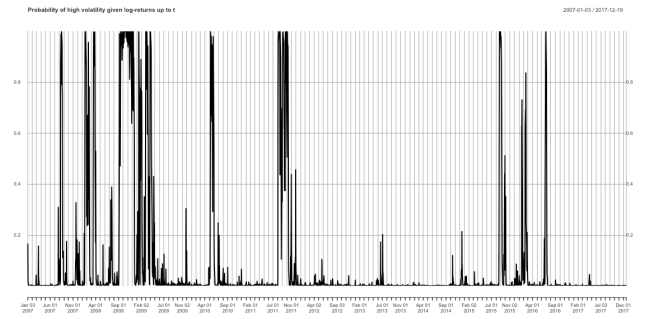

Figure 11: Probability of volatility of FTSE100 in 2007-2017

By applying the conditional probability recursively over the whole dataset, we get Figure 11 which shows the probability of volatility of the FTSE100 index in the given period. From the comparison with the log returns in Figure 7, we can see that the probability of high volatility peaks roughly when there are lots of fluctuations in the log-returns which suggests our calculations are a good fit with the data.

High volatility corresponds to a rapidly changing trading prices of FTSE 100, which normally corresponds to highly fluctuated trading prices. This agrees with our previous conclusion on the significant economic and financial events. There is very small volatility in the end of 2017, which indicates that the stock index in 2018 is unlikely to be fluctuating at least at the beginning of next year.

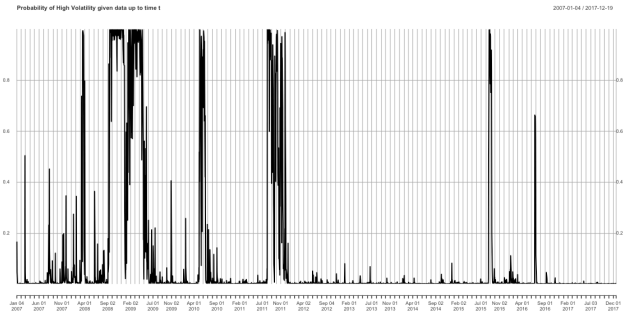

Figure 12: Probability of volatility of S&P500 in 2007-2017

Figure 12 shows the probability of volatility of S&P500 index in 2007-2017. It can be learned from the figure that the S&P500 index had larger probability of volatility in 2011 based on previous data and smaller probability of volatility in 2016 compared with Figure 11 of the FTSE100 index. This shows that the US stock market back to then was more unsteady. However, it is also noticed that the probability of vitality in 2017 is extremely small, which is a good sign of a stable market in next year.

C. Time series forecasting

As we conclude from the previous subsection, the probability of volatility at the end of 2017 is very low and thus the change in stock price indices should be assumed to be steady. We use the time series forecasting tool in Weka to further analyse the potential behaviour of the two indices in 2018. The time series forecasting is a built-in feature in Weka, which can be installed via the package manager. One of the restrictions is that the package is only available in Weka 3.7.3 or later [24]. Once installed, it will appear in the Explorer's main menu. As mentioned in the datasets section, the "date" attribute in the original data is converted to date attributes in Weka in order to do time series analysis.

We plot the adjusted close price against the time and compare different prediction functions built in Weka's package, which can be selected in the advanced configuration menu. In this study, we apply the K-nearest neighbour classifiers with different k values for k=1 and k=10 and the linear regression classifier for prediction.

We do the prediction on a quarter-year basis and thus make four predictions for the next year stock indices. The forecasting results of FTSE100 are shown in Figure 13, with K-nearest neighbour k=1, K-nearest neighbour k=10 and linear regression from the top to the bottom respectively.

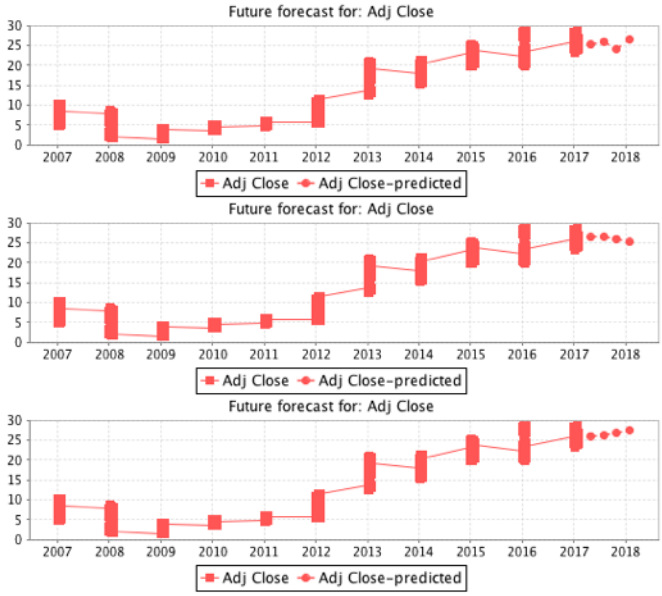

Figure 13: Time series forecasting of FTSE100

As Figure 13 shows, the K-nearest neighbour classifier with k=1 provides a fluctuating prediction of the next year's stock index, while the same classifier with k=10 suggests the FTSE100 stock index to be declined by a small amount in the next year. The linear regression model, on the other hand, predicts the stock index to increase steadily in 2018. However, it is worth remarking that the stock time series model is apparently non-linear based on previous datasets.

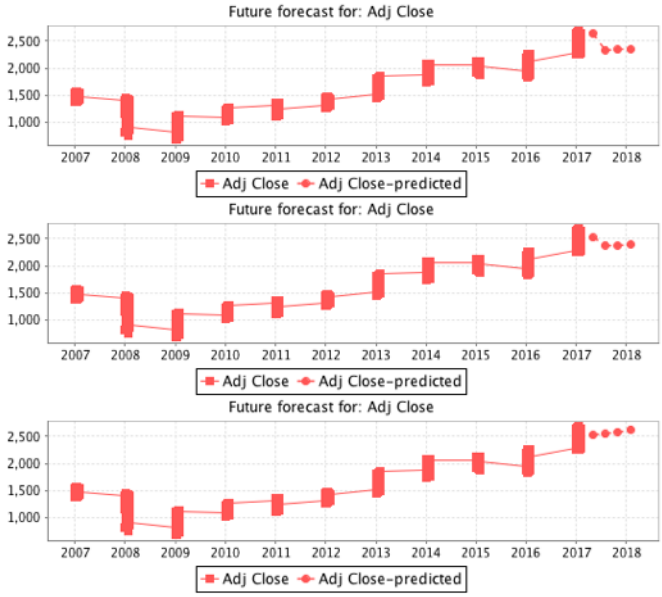

Figure 14: Time series forecasting of S&P500

Figure 14 presents the forecasting results of S&P500, with K-nearest neighbour k=1, K-nearest neighbour k=10 and linear regression from the top to the bottom respectively. Similar to what Figure 13 shows, the k=1 nearest neighbour classifier

predicts the index trend with a dramatic change, since it only takes the nearest value of each point and that may affect its judgement of the whole trend. The k=10 classifier still predicts the trend to be downwards, but the pattern is more concave compared with the FTSE100 pattern. Therefore, there might be a turn over afterwards. The linear regression method performs exactly the same as in FTSE100, indicating that there will be a steady increase in the stock index in 2018.

VII. Conclusion

A. Conclusions

After applying log returns, probability of volatility and time series forecasting methods to the two stock indices data, we could conclude that the US stock market suffered more severely during and after the financial crisis in 2008 than the UK stock market in terms of the log returns. It was also fluctuating in the following periods when the aftermaths occurred such as the European debt crisis. Meanwhile, the UK stock market became more volatile in 2016 after the world-shocking Brexit announcement. This led to a less steadiness of the UK stock market in the recent two years.

We used time series forecasting in Weka to briefly predict and forecast the picture of the two stock indices in 2018. Although different models generate different outcomes, we can learn from the patterns that the stock market in the next year should not be too volatile based on previous data. It also agrees with some expertise from the Wall Street, though they all predicted an increase in the S&P500 price indices by 5-12% [25].

B. Limitations and future recommendations

1) Data sources.

It is such a brief and preliminary report that there are many limitations inside it and plenty of aspects able to be developed. In terms of the data sources, the FTSE100 and S&P500 indices only cover large-cap companies in the stock markets. However, median- and small-cap companies tend to suffer more significantly in the financial crisis and will draw more fluctuations in the return plots. Since the FTSE100 and S&P500 data are easily extracted both from the Yahoo Finance and via the R "quantmod" package, this project is limited to these two indices. More locally-indicated indices could be evaluated in the future such as the FTSE250 index.

It is also worth mentioning that the stock price indices themselves do not contain all the information required for this evaluation. There are various other factors such as the interest rates of the national economy, the up-to-date exchange rate and derivatives and commodity markets. No matter if we want to identify the trend after the global financial crisis, or want to roughly draw a picture of the stock market in the coming year, it is not enough just to consider the stock market indices. Another remarkable note is that the history does not say everything and the future pattern can be completely different and unrelated compared with the historical ones. The stock indices prediction should only be treated as a brief guideline.

2) Methodology.

In this project, R and Weka are the main data analytics tools used as they are open and free for use. Since the stock data has really complex properties, the tools are selected based on our current capacity, but are far too simply compared with the actual tools used by the Wall Street and Canary Wharf analysts. What's more, due to the restriction of page limits, we are unable to split the data into smaller periods and do more detailed analysis of a certain period (for example, April and May 2017 during Brexit). It would be helpful and should be carried out in the future analysis.

There are also more functions available in Weka for time series prediction apart from linear regression and k-nearest neighbour. Various functions should be attempted in future analysis. We should also compare the prediction made by the tools with existing data to evaluate the accuracy of the prediction.

REFERENCE

[1]: A. Murphy. (2008) An Analysis of the Financial Crisis of 2008: Causes and Solutions.
https://ssrn.com/abstract=1295344 or http://dx.doi.org/10.2139/ssrn.1295344
[2]: 2002 Monetary Policy Report to the Congress
https://www.federalreserve.gov/boarddocs/hh/2002/February/FullReport.htm
[3]: Securities and Exchange commission rules in 2004
https://www.sec.gov/rules/final/34-49830.pdf
[4]: A.N. Berger. (2014). The Oxford Handbook of Banking, 2nd Edn. Oxford University Press, Oxford, 2014, pp. 1104. Econ Rec, 92: 674–676. doi:10.1111/1475-4932.12305
[5]: M. Dubofsky. (2013). The Oxford Encyclopedia of American Business, Labor, and Economic History.
[6]: S. H. Shin. (2009). Reflections on Northern Rock: The Bank Run that Heralded the Global Financial Crisis. The Journal of economic perspectives, 23(1), 101-119.
[7]: A. Malliaris, & L. Shaw, & H. Shefrin. (2016). The Global Financial Crisis and Its Aftermath: Hidden Factors in the Meltdown. 3-22. 10.1093/acprof:oso/9780199386222.003.0001.
[8]: Sourced from world bank
http://www.eiiff.com/economy/united-kingdom
[9]: Controlling interests - monetary policy after the crash. The Economist. Sept 2013.
https://www.economist.com/news/schools-brief/21586527-third-our-series-articles-financial-crisis-looks-unconventiona
[10]: http://www.ftserussell.com
[11]:www.ftse.com/Analytics/FactSheets/Home/DownloadSingleIssue?issueName=UKX
[12]: Sourced from FTSE webpage
http://frankinvestments.company/commentary/2014/1/20/how-to-manage-the-balancing-act
[13]: https://www.intercontinentalexchange.com/about
[14]: http://us.spindices.com/indices/equity/sp-500
[15]: https://www.spindices.com
[16]: https://www.cs.waikato.ac.nz/ml/weka/
[17]: https://cran.r-project.org/web/packages/quantmod/quantmod.pdf
[18]: Yahoo Finance home page https://uk.finance.yahoo.com
[19]: https://help.yahoo.com/kb/SLN28256.html
[20]: S.H. Poon, S. J. Taylor. (1992). Stock returns and volatility: An empirical study of the UK stock market. In Journal of Banking & Finance, Volume 16, Issue 1, 1992, Pages 37-59. ISSN 0378-4266, https://doi.org/10.1016/0378-4266(92)90077-D.
[21]: http://www.lisbon-treaty.org/wcm/the-lisbon-treaty/treaty-on-european-union-and-comments/title-6-final-provisions/137-article-50.html
[22]: T. Kaizoji. (2005) Statistical Properties of Absolute Log-Returns and a Stochastic Model of Stock Markets with Heterogeneous Agents. In: Lux T., Samanidou E., Reitz S. (eds) Nonlinear Dynamics and Heterogeneous Interacting Agents. Lecture Notes in Economics and Mathematical Systems, vol 550. Springer, Berlin, Heidelberg
[23]: A. R. Pagan, G.W. Schwert. Alternative models for conditional stock volatility, In Journal of Econometrics, Volume 45, Issues 1–2, 1990, Pages 267-290, ISSN 0304-4076, https://doi.org/10.1016/0304-4076(90)90101-X.
[24]:https://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka
[25]: https://finance.yahoo.com/news/wall-streets-stock-market-forecasters-already-rewriting-2018-outlooks-152401967.html