Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER THESIS

Prediction of Stock Return Volatility Using Internet Data

Author: **Bc. Tomáš Juchelka**

Supervisor: **doc. PhDr. Ladislav Krištoufek, Ph.D.**

Year of defence: **2017**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, July 31, 2017

Signature

Acknowledgment

First of all, I would like to thank my academic supervisor doc. PhDr. Ladislav Krištoufek, Ph.D. from the Institute of Economic Studies, Faculty of Social Sciences, Charles University in Prague, for the guidance, suggestions, comments and knowledge he provided that helped me writing this thesis.

Bibliography Reference

Juchelka, Tomas. *Prediction of Stock Return Volatility Using Internet Data* Prague, 2017. 70 p. Master Thesis (Mgr.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor: doc. PhDr. Ladislav Krištoufek, Ph.D.

Extent of the Thesis

123, 690 (with spaces)

Abstract

The thesis investigates relationship between daily stock return volatility of Dow Jones Industrial Average stocks and data obtained on Twitter, the social media network. The Twitter data set contains a number of tweets, categorized according to their polarity, i.e. positive, negative and neutral sentiment of tweets. We construct two classes of models, GARCH and ARFIMA, where for either of them we research basic model setting and setting with additional Twitter variables. Our goal is to compare, which of them predicts the one day ahead volatility most precisely. Besides, we provide commentary regarding the effects of Twitter volume variables on future stock volatility. The analysis has revealed that the best performing model, given the length and structure of our data set, is the ARFIMA model augmented on Twitter volume residuals. In the context of the thesis, Twitter volume residuals represent unexpected activity on the social media network and are obtained as residuals from Twitter volume auto regression. Plain ARFIMA model was the second best and plain volume augmented ARFIMA was in third place. This means that all three ARFIMA models outperformed all three GARCH models in our research. Regarding the Twitter estimation parameters, we found that higher the activity the higher tomorrow's stock return volatility. This conclusion holds for all Twitter volume variables regardless their polarity.

JEL Classification C22, C52, C55, G12
Keywords GARCH, Volatility, Internet Data

Author's e-mail tom.juchelka@gmail.com
Supervisor's e-mail ladislav.kristoufek@fsv@cuni.cz

Abstrakt

Tato práce zkoumá vztah mezi denní volatilitou akcií z Dow Jones Industrial Average akciového indexu a daty, které byly získány na známé sociální síti Twitter. Twitter data obsahují počet tweetů a jsou kategorizována na základě jejich polarity, a to na pozitivní, negativní a neutrální. V práci jsme využili dva druhy modelů, GARCH a ARFIMA, které jsme zkoumali jak v základním nastavení bez dodatečných proměnných tak s proměnnými, které reprezentují velikost aktivity na této sociální síti. Našim hlavním cílem je zjistit, který z modelů nejpřesněji odhaduje jednodenní predikci realizované volatility. Kromě toho také zkoumáme, jaký effekt mají proměnné získané na Twitteru na budoucí volatilitu. Náš výzkum ukázal, že nejlepším modelem pro predikci volatility je ARFIMA model obohacený o rezidualní objem aktivity na Twitteru. V kontextu naší práce je residuální objem na Twitteru chápán jako proměnná, která reprezentuje neočekávanou aktivitu na této sociální síti. Druhým nejlepším v pořadí byl ARFIMA model bez dodatečných proměnných. ARFIMA s objemem na Twitteru byl potom třetím modelem v pořadí. Pořadí modelů odhalilo, že na našich datech je ARFIMA vhodnějším modelem pro predikci volatility. Co se týče jednotlivých efektů Twitter proměnných, výzkum ukázal, že aktivita na Twitteru pozitivně ovlivňuje budoucí volatilitu. Tento závěr se dá zobecnit bez ohledu na to, jakou má Twitter proměná polaritu.

Klasifikace JEL	C22, C52, C55, G12
Klíčová slova	GARCH, volatilita, Internetová data

Email autora	`tom.juchelka@gmail.com`
E-mail vedoucího práce	`ladislav.kristoufek@fsv@cuni.cz`

Contents

List of Tables

List of Figures

Acronyms

ARCH Autoregressive Conditional Heteroscedastic

GARCH Generalized Autoregressive Conditional Heteroscedastic

AR Autoregressive

ARMA Autoregressive Moving Average

EGARCH Exponencial GARCH

ARFIMA Autoregressive Fractionally Integrated Moving Average

HAR Heterogenous Autoregressive

ACF Auto Correlation Function

AR Autoregressive model

Master Thesis Proposal

Author	Bc. Tomáš Juchelka
Supervisor	doc. PhDr. Ladislav Krištoufek, Ph.D.
Proposed topic	Prediction of Stock Return Volatility Using Internet Data

Topic characteristics Stock return volatility and its features, such as leverage effect or volatility clustering, are very interesting areas of academic research. Due to extensive usage of volatility as a risk measure, entering various pricing models, risk-assessment models or optimal portfolio construction frameworks, it is very important to understand main drivers that allow us to estimate the actual volatility or predict the future volatility. Nowadays, besides the standard data such as trading volume, bid-ask spreads or other variables that are directly related to financial markets, it is possible to obtain a completely new set of relevant variables that are not generated through the pricing or volume mechanism of the financial markets.

The main objective of the thesis will be a construction of two sets of volatility models (GARCH, ARFIMA). First set using standard financial market variables, the second set further augmented on the variables based on internet data that were generated on Twitter in connection to the stocks included in our research. Later we plan to compare the quality as well as prediction capabilities between these two sets of models using various statistical tests to find whether the internet data add any important information to the volatility prediction. We also plan to find the absolute winner to see, which models among the sets performs the best.

Outline

1. Introduction
2. Theoretical Framework

   2.1. Conditional Heteroscedastic Modeling Framework
   2.2. Other Volatility Models
   2.3. Incorporation of Twitter Variables
3. Empirical Research
   3.1. Data and Methodology
   3.2. Results of Research
   3.3. Summary
4. Conclusion

Core bibliography

1. Bollerslev, T (1986): "Generalized autoregressive conditional heteroscedastic ." *Journal of Econometrics* **31**: pp. 307-327.
2. Clements, A.E., Liao, Y., (2013): "5) Modeling and forecasting realized volatility: getting the most out of the jump component." *School of Economics and Finance, Queensland University of Technology,*
3. Nelson, D. (1991): "Conditional heteroscedasticity in asset returns: A new approach." *Econometrica* **59(2)**: pp. 347–370.
4. Granger, C.W.J, Joyeux, R. & W.D. Granger (1980): "An introduction to long-memory time series models and fractional differencing"." *The Journal of Time Series Analysis* **(1)**: pp. 15–30.
5. Tsay, Ruey S., Wiley(2005): *"Analysis of Financial Time Series"* 2nd Edition, John Wiley & Sons, Inc. ISBN: 978-0-471-74618-8
6. Robinson, P.M. (2003): "Time Series With Long Memory." *Oxford University Press.*

Author Supervisor

Chapter 1

Introduction

Stock market features, as well as its prediction, have been widely discussed topic either by academic town and gown or business professionals, working in financial services industry. Such discussions led to birth of various theoretical frameworks that try to explain the rationale behind the "black-box" mechanics of financial markets.

One of the first and also widely accepted theory, so called the Efficient Market Hypothesis, was presented by professor Eugene Fama in his work Efficient Capital Markets, a Review Of Theory And Empirical Work (Fama 1970).The theory states that asset prices fully reflect all available and relevant information. It also directly implies that it is impossible to overperform the market consistently over a long period of time. Based on the theory, the market prices should only react to new information or changes of financial market parameters that are in general unpredictable. Furthermore, this new information, once they are public, are immediately incorporated into asset prices. According to the theory, the asset prices will follow a random walk pattern and cannot be predicted with more than 50 percent accuracy, given the fact that zero drift assumption is satisfied. Such fact would imply, that it is impossible to achieve the abnormal return on a long term basis. Taking a look from the perspective of econometrics, Efficient Market Hypothesis would imply that proxy variables for new information would not stand significantly in econometric models over time. However, a growing amount of research in the area of financial markets and econometrics has disputed the Efficient Market Hypothesis both theoretically and empirically.

In particular, theories from the perspective of Socionomic Theory of Finance written by Robert R. Prechter & Parker (2007) or Behavioral Economics and Behavioral Finance suggests that stock market prices can be, at least to some degree, successfully predicted. These theories say that market participants are not fully rational and their decisions take some time. Furthermore, these decisions are not based solely on facts but also affected by other factors such as mood or emotions and that besides fundamental investors, there are also investors that make decisions based on technical analysis and trends. With the rise of the internet media, there is an overabundance of news so it is hard for the investors to pay attention to them and react accordingly. Such behavior is later transformed into the slower incorporation of news into asset prices, allowing it to be partially predictable. Suggestions of partial predictability were also supported by empirical research, for example in the work that was written by Bo Quian (2007). These researchers investigated predictability of Dow Jones Industrial Average stock market index, showing that not all periods of time are equally random. The research provided empirical evidence that there exist a set of variables, which in combination with the right statistical approach generate stock market predictions with probability more than 65 percent. As such, these findings put into question some of Efficient Market Hypothesis basic assumptions, supporting those of competing theories we mentioned earlier.

In our work, we are not planning to further research the question regarding the predictability of either absolute or relative price movements. We rather focus on predictability of future stock variance or so called, stock return volatility. In general, stock return volatility is a very important variable that enters many asset pricing and risk assessment models that are broadly used by either general public or professionals of financial markets. It possesses many interesting features such as volatility clustering and leverage effect that are sufficiently described in the text written by Ruey S. Tsay (2005). There has already been extensive research on all these features and also methods that are aiming to model the volatility in the most proper way. One of the most prominent method, used either for modeling or predicting future volatility, is the approach suggested by Bollerslev (1986). His General Autocorrelation Conditional Heteroscedastic (GARCH) models are exceptional among competing approaches thanks to their ability to capture both volatility clustering and also non-normality of stock returns. Besides GARCH, there exist other functional models that are broadly used for modeling volatility. These namely are ARIMA and ARFIMA. In our work, we restrict ourselves to using GARCH and ARFIMA models that we compare on the basis of their prediction performance abilities.

Encouraged by fairly solid volatility modeling frameworks, researchers started to put volatility into the context of other explanatory variables. These variables were bid-ask spreads or trading volume, expressed either in financial terms or in a number of stocks traded in a particular period of time. For example Ali F. Darrat (2003) conducted research on all stocks of Dow Jones Industrial Average, using the above-noted GARCH model. He explored that there exist a significant positive relationship between stock return volatility and lagged values of trading volume. Furthermore, Juchelka (2014) examined volume volatility relationship of the top five SP500 companies using GARCH modeling framework. He incorporated trading volume as a proxy for new information arrival and liquidity trading. Namely, he further divided the trading volume into two components, these were expected and unexpected trading volume. Such division showed that there exists asymmetry between liquidity trading, the expected volume and information trading, the unexpected volume. The resulting outcome from his research showed that liquidity, as well as unexpected, event driven, trading volume, lead to volatility increase. Stock trading volume is broadly put into the context of information flows, that is, the volume can be used as a proxy for new information arriving into the market. However, such generalization might be a bit noisy as there can be other reasons, besides liquidity and information trading, that affect trading volume.

In our current research, we are going a level above by using data that are directly connected to the human evaluation of news and sharing opinions related to stocks of Dow Jones Industrial Average. These data were collected from the Twitter, well known social network, which is widely used by the general public as well as financial market professionals who publicly express their opinion related to financial markets. Our data set is composed of a large number of tweets that are directly related to each DJIA's stock and are also sorted based on their mood into three categories. These categories are positive, negative and neutral. On the top of that, we can calculate resulting daily sentiment of tweets based on this three categories.

As the Twitter is becoming widely used and as more people are connected and spending more time on the internet, we believe that these internet data contain very interesting information that can be further incorporated into wide spectrum econometric models. We assume that such employment of these data enhances the quality of results obtained from currently available analytics, either in finance or in other sciences. From the methodology perspective, we include these data as explanatory variables into two different volatility models, GARCH and ARFIMA, to see whether there is any enhancement of the predictive ability compared to models without employment of these variables.

In our thesis, we start with Chapter 2 providing you a comprehensive insight into the current research around the relationship of the internet data and financial market behavior. We also try to put this research in a framework of Behavioral Finance, using twitter data as a proxy for attention. We follow up this topic with explanation of the theory around the volatility modeling frameworks that we use in our analysis. In chapter 3, we provide explanations and description of our data set and methodology as well as giving you the functional forms of models we employ in our research. Chapter 4 is dedicated to the empirical analysis, explaining the results of our research for each of the stocks individually. Moreover, we will try to aggregate the results from the stock level in order to make a general conclusion to this research topic. In the last part of the thesis, we summarize the entire research problem and suggest possible research enhancements that can further improve our research and try to tackle the practical use of our research results.

Chapter 2

Theoretical Framework

In this chapter, we begin with introduction to the investor attraction theory and summarize the main body of the empirical research on the topic of financial market variables and their relation to the internet data. Later, we introduce and expand the theory around volatility modeling frameworks which we utilize in the empirical part of our research by augmenting these basic volatility models by additional explanatory variables originated in Twitter.

Due to the fact that volatility is one of the most important metrics used in the practice of financial markets, let us first hereunder provide the reader with its basic definition as well as with the description of its main features. Volatility, according to Ruey S. Tsay (2005), can be defined as a conditional standard deviation of the underlying asset return. In our case stock returns of DJIA. Volatility is broadly used as a quantification of riskiness of the asset and is not directly observable, hence has to be derived from the realized returns (Ruey S. Tsay 2005).

Such risk measure has also many applications in finance as it enters a large number of financial models. These are, for example, frameworks for construction of the optimal portfolio based on mean-variance optimization or option pricing based on Black-Scholes model that was suggested by Black & Scholes (1973). One very common factor, under which vast majority of financial models operate, is the fact that they broadly use expectations, i.e. the most probable future value of outcome based on certain statistical procedures, commonly represented by the mean value of historical observations. However, using historical average might not be the best choice. Therefore, estimation based on more sophisticated methods, as well as improvement of these methods, by using additional explanatory variables from various data sources can possibly add useful information leading to better results. Therefore, we introduce the reader with more sophisticated volatility modeling approaches later on.

2.1 Literature Review

In this section, we briefly go through the literature that is concerned about research around the topic of social media or internet data in connection to financial markets. As we already stated in the introduction, there has already been extensive research in this area of econometrics and finance.

Let us begin with the work written by Gruhl *et al.* (2005) called The Predictive Power of Online Chatter. In this work, the researchers were examining a phenomenon of blogs. They measured how well the blog chatter activity reflects the comparatively old practice of buying books. The book vertical was chosen in purpose as the book sales are much less likely to be driven by special marketing offers, as opposed to other goods categories.

For the purpose of the research, they employed rather interesting dataset. The first set of data was composed of blog postings and was obtained from IBM's WebFountain project. In total, approximately $200,000$ postings arrived at the database every day and the system contained around three billion webpages and $200,000$ media articles. The second part of the dataset consisted of Amazon sales rank data and it was obtained using Amazon's Web Services. Namely, it contained 120 days of data on all products that at any point during this interval reached a sales rank of less than 300, i.e. 300 best sellers.

Having all the necessary data, they started to detect spikes in a sales rank and blog mentions for each book from the dataset in order to examine the correlation and predict the sales rank around those spikes. The results of the analysis revealed that if there is a spike in the sales rang and there are a lot of blog mentions about the book, then the blog mentions tend to have a spike that exhibits a strong correlation with the sales rank. Furthermore, the research implies that a sudden increase in blog mentions can be a potential predictor of a spike in sales rank.

There are more studies that connected online activity with sales of a particular product through the internet. For example, Mishne *et al.* (2006) used assessments of blog sentiment to predict movie sales and Liu *et al.* (2007) predicted the future product sales using a Probabilistic Latent Semantic Analysis model to extract indicators of sentiment from various internet blogs. Such data might be thought of as granular predictors of financial indicators that are used for company valuation purposes. In such case, there might exist a relationship between those predictors and stock values of the particular companies that sell those products.

Let us now focus on researches that are more closely connected to the employment of internet data in connection to financial markets. For example Schumaker & Chen (2009) examined the connections between breaking financial news and stock price changes. In the research, they conducted analysis employing predictive machine learning approach for financial news articles applied on several textual representations, i.e. bag of words, noun phrases, and named entities. Through this, they examined $9,211$ financial articles and $10,259,042$ stock quotes from SP500 stock market index during the period of five consecutive weeks.

They applied the analysis to estimate a discrete stock price twenty minutes after a particular article was released in the media. Using a Support Vector Machine, they show that the model containing both article terms and stock price at the same time of article release had the best performance in prediction of the actual stock price and the same direction of future price movement as the future price and the highest return using a simulated trading engine.

Similarly, Gidofalvi & Elkan (2001) undertook a research on a very similar topic. They gathered more than $5,000$ financial news articles concerning twelve stocks, and set an interval of time to be twenty minutes before the article was released and twenty minutes after the article was released. Within this period of time, he demonstrated the existence of a weak prediction ability of the direction of security before the market clears itself to equilibrium. He states that a reason for such poor prediction ability is the fact that many of financial news articles are typically reprinted through the various news wire services. Furthermore, he points that a stronger predictive ability may exist in isolating the first release of an article.

Another study written by Bollen *et al.* (2011) deals with an analysis of Twitter social media network in connection to DJIA stock market index. They analyze text content of daily Twitter feeds by mood tracking tools. A Granger analysis and a Self-Organizing Fuzzy Neural Network are then used to investigate the hypothesis that public mood states are predictive of changes in DJIA closing stock prices. For this research, they collected $9,853,498$ of tweets from the period of 10 consecutive months. In addition, they collect a time series of daily DJIA close prices from Yahoo! Finance, and investigate the hypothesis whether public mood can predict future DJIA closing values.

The results of the research indicate that the accuracy of the predictions can be significantly improved if the public mood proxy variables were employed. They found an accuracy of 86.7 percent in predicting the daily up and down changes in the closing values of the stock market index. The results also yielded a reduction of the Mean Average Percentage Error by more than 6 percentage points.

Another research, more closely related to our topic, was undertaken by Ranco *et al.* (2015). The research was dealing with examination of the relationship between micro-blogging platform Twitter and behavior of financial markets. Particularly, they considered the Twitter sentiment and volume about companies that are listed in the Dow Jones Industrial Average index. They collected approximately $1,555,770$ tweets from the period of fifteen consecutive months. These tweets were categorized according to their mood sentiment using supervised learning method. The resulting dataset is in the form of a time series of negative, neutral and positive tweets for each day of the period.

Just to note, Twitter data can be considered as a direct estimation of investor attention, as an investor has to express his opinion on the Twitter. We describe the theory of Imperfect Information and Investor Attention in the section below. Interestingly, the research has revealed statistical relationship between Twitter sentiment and extreme returns during the peaks of Twitter volume, using the event study methodology framework. In general, the main implication of the research was that Twitter sentiment polarity is able to predict the direction of cumulative abnormal returns.

There are plenty of another text that are related to internet data. We have listed here those that were particularly useful for our own research. Another texts regarding the internet data are also presented in the following section, where we put it into the context of Imperfect Information and Investor Attention theory.

2.2 Imperfect Information and Investor Attention

Traditional asset pricing models, (Fama 1970), are based on a set of strong assumptions regarding the behavior of financial markets and their participants, i.e. rational agents, institutions, and framework according to which the markets operate. Indisputably, the central assumption of these theories is the assumption of perfect information. According to this assumption, all investors are able to accommodate and evaluate every information regarding all individual assets. In combination with the assumption of frictionless markets, such evaluation of information leads to immediate incorporation of information into asset prices. However, in practice, it is very unlikely to observe that investors have such strong mental capacity to pay attention to all news regarding their asset holdings. In reality, mental capacity is a valuable cognitive asset (Kahneman 1973) and investors are not able to pay full attention to all information flows.

Knowing this, researchers started to update traditional pricing models to take into account this phenomenon. For example, Merton (1987) has adjusted the capital asset pricing model to accommodate incomplete information of market participants. Namely, he assumed that each investor knows only about a subset of existing securities. Based on this, he examined the impact on equilibrium asset prices affected by such structure of incomplete information. The functional form of his model suggests that we can expect lower market capitalizations of firms compared to market capitalizations of firms under the assumption of perfect information. Moreover, with increasing investor base of the stock, the uninformed market value converges to the perfect information market value. Another implication of the model is that less known stocks with small investor basis have relatively larger expected returns compared to perfect information case. There are many other researches that are easing the assumption of perfect information, not only in connection to financial markets but also in connection to theoretical economics, see for example the work written by Sims (2003).

Up to date, the research on the topic of investor attention has faced a crucial problem of measuring the attention with acceptable precision. As a direct measure of investor attention did not literally exist, researchers had to use indirect proxy variables. For example, Barber & Odean (2008) have used unusual trading volume, news and extreme returns as proxy variables that are connected to attention grabbing events. Using these proxy variables, authors were testing the proposition that retail investors are likely to buy, rather than sell stocks that grab their attention and that retail investors are more affected by attention grabbing stocks compared to the behavior of professional investors. Results of this research indicate that those two propositions might be actually true, based on either of the selected proxy variable.

These proxy variables, however, place the implicit assumption that if the stock was mentioned in a media or experienced significant volume turnover or abnormal return, investors paid attention to it. The fact is that return or turnover might be driven also by other factors that are not related to investor's attention. The same holds for articles in media unless we know that investors actually read those articles. As we know, there are currently thousands of information sources providing continuous and instant information flows via our cell phones, making it harder for people to actually pay attention to that news. The digital consumption, however, helped to establish the entire new set of variables that might be used as a direct estimation of investor attention. As shown in a work of Da *et al.* (2011), one of the possibility to estimate investors attention is to obtain the Google search data in connection to particular stocks. Using the Search Volume Index, the researchers have discovered correlation of the Index with previously mentioned proxies for investor attention. On the top of that, the increase of the Index is able to predict higher stock prices in the next two weeks followed by a reversal in a year time frame. The internet search activity of retail investors also predicts the large first-day return and long run underperformance of IPO stocks.

In our work, we propose to employ the Twitter volumes as well as their sentiment in connection to stock return volatility, rather than stock returns itself. We follow up the previous research that used Twitter data as a proxy variable of investor attention. Moreover, we also have polarity of the Twitter data such that we can discover whether there exist any asymmetry in their relationship. We think that social aspects of financial market participants, estimated by social media data, can be useful to understand and predict financial markets. In fact, financial crises or contagion are often driven by collective investor behavior such as herding or panic (Bouchaud 2009). Such behavior typically leads to higher uncertainty, thus, higher volatility. Therefore, it is very useful to anticipate such behavior in advance. Hence, we base our research on building predictive models that should give us a tool to predict volatility more accurately based on the behavior of investors exhibited on the internet social media. In the next section of the thesis, we provide a theoretical description of volatility modeling frameworks that we utilize in our research.

2.3 Conditional Heteroscedastic Models

2.3.1 The ARCH Process

Let us begin with the very first conditional heteroscedastic model that provides us with the systematic approach of volatility modeling. The Autoregressive Conditional Heteroscedastic (ARCH) model was developed by Engle (1982) and for its comprehensive usefulness underwent a number of generalizations. Before the introduction of the ARCH process in detail, it might be useful to show the modeling of the serially uncorrelated, dependent return series r_t and some of its features such as conditional mean and variance.

In the model, the return series is formed as follows:

$$r_t = \mu + a_t \tag{2.1}$$

Conditional mean of r_t is defined $E(r_t|F_{t-1}) = \mu$ and conditional variance of is given by $Var(r_t|F_{t-1}) = E[(r_t - \mu_t)^2|F_{t-1}] = h_t$. The term F_{t-1} is defined as information set available at time $t - 1$. The term a_t which appears in the equation 2.1 is referred to as shock of return series at time t. It can be expanded into the following form:

$$a_t = \sqrt{h_t}\epsilon_t$$

where $\{\epsilon_t\}$ is an i.i.d. sequence of random variables with zero mean and unity variance.

The main purpose of an ARCH model is to describe the conditional variance h_t.

Let us start with the simple ARCH(1) which we later expand into ARCH(q) model. The ARCH(1) equation is defined as:

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 \tag{2.2}$$

where $a_t = \sqrt{h_t}\epsilon_t$, $\alpha_1 \geq 0$ and $\alpha_0 > 0$ are estimation parameters .

The ARCH(1) model is just a special case of general ARCH(q) model that adds the finite number of lags into the equation and has the following specification:

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \alpha_2 a_{t-2}^2 + \ldots + \alpha_q a_{t-q}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i a_{t-i}^2 \tag{2.3}$$

Where conditions of non-negativity $\alpha_0 > 0$ and $\alpha_i \geq 0$ for all $i > 0$ must hold in order to ensure non-negativity of conditional variance. Coefficients α_i have to also satisfy regularity conditions to obtain the finite unconditional variance of a_t.

From the structure of the equation, we can see that large past squared shocks a_{t-i}^2 cause large conditional variance h_t of a_t. This means that in ARCH framework, large shock tends to be followed by another large shock. Such property describes volatility clustering feature of the asset returns that we have already noted earlier. Therefore, the ARCH(q) model is useful in modeling the financial time series.

Regardless of its useful properties, the ARCH model also has some shortcomings such as its equal respondence to positive and negative shocks. This phenomenon is caused by the conditional variance being dependent on the squares of a_t. ARCH also does not provide any explanation of the source of variations, as it only mathematically describes conditional variance. Another disadvantage is that the model is struggling with is the necessity of high order lag structure due to long memory feature that is often present in financial series data. Eventually, there is also the possibility of violation of non-negativity conditions when it comes to the estimation of the free lag distribution. To handle these specific problems, the ARCH model was later generalized. We introduce these generalizations in the upcoming section of our text.

2.3.2 The GARCH Process

Nevertheless, the ARCH model proved to be effective in volatility modeling, there still was a substantial space for improvements due to the shortcomings that we noted in the previous paragraph. In this part of the text, we aim to introduce the GARCH model that is more general and overcomes many of those problems that ARCH model was unable to deal with.

The GARCH model was introduced by Bollerslev (1986). The extension of the ARCH process to the GARCH is very similar to the extension of the time series of Autoregressive model (AR) process to the general Autoregressive Moving Average (ARMA) process. The model also allows for better description in many situations.

The GARCH(p,q) process is given by these functional forms:

$$r_t = \mu + \epsilon_t = \mu + \sqrt{h_t} z_t \tag{2.4}$$

$$\epsilon_t | \Omega_{t-1} \sim N(0, h_t) \tag{2.5}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j} \tag{2.6}$$

From equation 2.4 we can see that $\epsilon_t = r_t - \mu$ is a real value, discrete time, strictly stationary stochastic process. Term z_t is independent, identically distributed standard normal variable and Ω_{t-1} is the set containing all information available through time $t-1$. Parameters p,q establish the number of lags included in the model and α_0, α_i, β_j are estimation parameters that are subject to the following conditions: $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1 \dots q$ that measure the short term impact of ϵ_t on conditional variance and $\beta_j \geq 0$ for $j = 1 \dots p$ measures the impact in the long-term.

The equation 2.6 shows that if $p = 0$ the GARCH reduces to the ARCH(q) process. Furthermore, if $p = q = 0$ than ϵ_t is simply white noise. As you can see, the specification of the GARCH model suggests that the best predictor of the future variance is a weighted average of the long-term average variance, the variance predicted for the current period and the information in this period captured by the most recent squared residual.

Together with the conditions of non-negativity, the $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$ has to be satisfied in order to ensure the covariance stationarity of GARCH(p,q) process. In order to state it precisely, let us provide you the following theorem that was provided by Bollerslev (1986) on page 310.

Theorem 1. The GARCH(p,q) process defined by 2.5 and 2.6 is wide-sense stationary with $E(\epsilon_t) = 0$ and $Var(\epsilon_t) = \alpha_0(1 - \sum_{i=1}^{q} \alpha_i - \sum_{j=1}^{p} \beta_j)^{-1}$ and $cov(\epsilon_t, \epsilon_s) = 0$ for all $t \neq s$ is and only if $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$

Empirically, the framework offered by GARCH was very successful in modeling as well as predicting financial time series volatility. It takes into account wide range of phenomenon that can be observed in real life practice of financial markets. We must not forget, however, to provide the reader also with the downsides of the GARCH model. As mentioned by Nelson (1991), the non-negativity condition of the estimation parameters, which are imposed in order to ensure the non-negativity of h_t in all time periods, lead to the fact that increasing ϵ_t in any time period also leads to increase of h_{t+m} for all m $\geqslant$ t. Such behavior eliminates the randomness in the fluctuations of h_t. Moreover, simple GARCH model is unable to capture the leverage effect, so the magnitude of h_t does not depend whether $\epsilon_t = r_t - u \leq 0$ or $epsilon_t = r_t - u \geq 0$, though it is observed that $\epsilon_t \leq 0$ is related to higher h_t and $\epsilon_t \geq 0$ is related to lover h_t. The last drawback mentioned by Nelson is difficulty in the evaluation of persistence of shocks to conditional volatility.

As an answer to these problems, the EGARCH model was developed. In our work, however, we aim to use mainly a simple GARCH model, therefore we do not provide the reader with a theoretical description of EGARCH.

2.3.3 Exceptionality of GARCH(1,1)

Though we provided the reader with full-stack theory around the family of Conditional Heteroscedastic Models, we decided to use the simple GARCH(1,1) model in our empirical part. Such decision is based on several reasons. Firstly, the model is very easy to set up, estimate and interpret. Secondly, the model yields precise predictions and good fit on the data. Thirdly, it is able to reproduce some stylized facts of the asset returns, especially volatility clustering and heavy tails. Lastly, simple GARCH is widely researched in the academic sector and is considered to be exceptional among all available options for modeling volatility. Above that, Hansen & Lunde (2005) have evaluated more than 300 different volatility models and found out that any of those models hardly improved prediction precision compared to simple GARCH(1,1) specification.

2.3.4 Augmented GARCH Model

Until now, we provided the theory about classical approach to volatility modeling frameworks using GARCH models. In this section, we explain how we plan to shape these models to incorporate additional information beyond the scope of simple price action information that standard GARCH models usually operate with.

As we already stated in the previous section, we decided to utilize GARCH(1,1) as a base setting for its ease of interpretation and proven empirical performance among other volatility models. The model will be further augmented by additional explanatory variables in relation to Twitter data. For this purpose, we define two sets of additional variables, these are, plain lagged Twitter volumes as defined in section 3.2. Using these variables, we obtain first augmented model for our comparison. The general functional form of variance equation of augmented GARCH model would be as follows:

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + \sum_{j=1}^{5} \gamma_j Tvol_{j,t-1} \tag{2.7}$$

Where α_0, α_1, β_1, γ_j are parameters to be estimated, ϵ_{t-1}^2 is an ARCH term, h_{t-1} is a GARCH term and $Tvol_{j,t-1}$ is a vector that represents Twitter volume variables such that for each integer j we assign a specific twitter variable from the list that we provide in 3.2.

The second set of variables we plan to augment the GARCH with are variables defined in 3.2. Again, we will use first lag to make the predictions possible. The general functional form of the variance model is defined as:

$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} + \sum_{j=1}^{5} \gamma_j RTvol_{j,t-1} \tag{2.8}$$

Where α_0, α_1, β_1, γ_j are parameters to be estimated, ϵ_{t-1}^2 is an ARCH term, h_{t-1} is a GARCH term and $RTvol_{j,t-1}$ is a vector representing residual or unexpected Twitter volume variables obtained from the regression 3.3 such that for each integer j we assign a specific residual Twitter variable from the list that we provide in 3.2. We assume that unexpected Twitter volume variables can introduce interesting information into the variance equation of the GARCH model as they may cluster around interesting news or events that attracted a lot of attention of investors, who shared their opinion publicly on Twitter. In the next section, we introduce ARMA model and its fractional generalizations. Such models will form another family of models for our comparison.

2.4 ARMA And Its Fractional Generalizations

2.4.1 ARMA Model

In this section, we aim to introduce another group of econometric models that we later use in our empirical research. Namely, we provide you with the theory of Autoregressive Moving Average (ARMA) models and their extensions that are widely used in the research practice, see work written by Box *et al.* (1994). These models were developed as a reaction to applications in which simple Autoregressive and Moving Average models were not sufficient, mainly in situations with dynamic structure of the data. Fundamentally, these models combine the ideas of AR and MA models into a compact form that keeps the number of parameters sufficiently small, thus, it is easy to apply them in practice. Regarding their application in finance, the chances to use these models for return series is fairly small, however, it has been discovered that their application in volatility modeling is highly relevant.

Let us start with the general ARMA(p,q) model that has the following form:

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i}, \tag{2.9}$$

where a_t is a white noise series and p, q are non-negative integers. As you can see from the equation 2.9, AR and MA models are just a special case of general ARMA model. The model can be rewritten using the back-shift operator into the following form:

$$(1 - \phi_1 B - ... - \phi_p B^p) r_t = \phi_0 + (1 - \theta_1 B - ... - \theta_q B^q). \tag{2.10}$$

Where the polynomial $(1 - \phi_1 B - ... - \phi_p B^p)$ represents the AR part of the equation and the polynomial $(1 - \theta_1 B - ... - \theta_q B^q)$ represents the MA part accordingly. It is required that there are no common factors between those two polynomials, otherwise, the order (p, q) of the model can be reduced. Same as in the pure AR model, the AR polynomial introduced the characteristic equation of an ARMA model. In a situation where all solutions of the characteristic equation are less than one in absolute value, the process is weakly stationary.

In practice, however, financial time series are rarely stationary as trends and periodicities are often present in the data. Therefore, there is a need to remove these effects before application of econometric models. For this purpose, we introduce ARIMA and ARFIMA models that can overcome such issues.

2.4.2 ARIMA Model

Considering previously introduced ARMA model, ARIMA model is an extension that allows the AR polynomial to have 1 as its characteristic root. In such case, ARMA becomes the Autoregressive Integrated Moving Average model that is said to be unit-root non-stationary. Same as random-walk model, an ARIMA model possesses a feature of long memory as the coefficients of the MA representation do not converge to zero over time. The model, however, deals with the unit-root non-stationarity by employing a method of differencing,that is, using the changes of the variable rather than the actual values.

The general ARIMA(p,d,q) model can be written in terms of the following equation:

$$A(L)(1 - L)^d y_t = \alpha + B(L)\varepsilon_t. \tag{2.11}$$

The first set of p parameters define the autoregressive polynomial in the lag operator L:

$$A(L) = 1 - \rho_1 L - ... - \rho_p L^p \tag{2.12}$$

The second set of q parameters define the moving-average polynomial in the i.i.d. disturbance process:

$$B(L) = 1 + \theta_1 L + ... + \theta_q L^q \tag{2.13}$$

The third parameter, d, expresses the integer order of differencing to be applied to the series before estimation to render it stationary. Thus, we speak of an ARIMA(p,d,q) model, with $p + q$ parameters to be estimated. In order to estimate such model, the d-differenced time series is required to be stationary such that the AR polynomial in the lag operator may be inverted. For a more thorough discussion regarding the properties of the model, see the work written by Box *et al.* (1994).

2.4.3 ARFIMA Model

Estimating an ARIMA model, the researchers have to choose the integer order of differencing d, to ensure that the resulting time series $(1 - L)^d y_t$ is a stationary process. The ARFIMA models, however, fill in the gap between the extreme cases of unit root models and stationary models that impose exponential decrease of the autocorrelations and, therefore, a spectrum bounded in the zero frequency. ARFIMA models are models of autoregressive moving average, where the differentiation is fractional. The differencing parameter d is not an integer but a real number. These models cover the intermediate case that exists between the unitary root, long-range dependence, ARIMA processes and the ARMA processes. To test whether the series exhibits such process, one can use a combination of unit root tests with the opposite null hypotheses. One such approach would be to use an augmented Dickey–Fuller test introduced by Dickey & Fuller (1979) in combination with Kwiatkowski–Phillips–Schmidt–Shin test (KPSS) by Kwiatkowski *et al.* (1992).

The model of an autoregressive fractionally integrated moving average of a time series of order (p,d,q), denoted by ARFIMA(p,d,q), with mean μ, may be written using the following form:

$$\Phi(L)(1-L)^d(y_t-\mu) = \Theta(L)\varepsilon_t, \varepsilon_t \ i.i.d., (0, \sigma_\varepsilon^2) \tag{2.14}$$

where L is a back-shift operator,$\Phi(L) = 1-\phi_1-...-\phi_p L^p$, $\Theta(L) = 1+\upsilon_1 L+...+\upsilon_q L^q$, and $(1-L)^d$ is fractional differencing operator defined by the following:

$$(1-L)^d = \sum_{k=1}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)} \tag{2.15}$$

Where $\Gamma(.)$ denotes a generalized factorial function. The parameter d is allowed to assume any real value. Setting arbitrary restriction for d to integer values, the model becomes simply ARIMA process. For further elaboration on the topic of model's properties, please revisit the book written by Ruey S. Tsay (2005). Having introduced the modeling framework of our research, we can move to data and methodology description.

2.4.4 Augmented ARFIMA Model

In our research, we will use ARFIMA(1,d,0) as our benchmark for plain ARFIMA model. We later augment such model on both Twitter volume variables as well as residual Twitter volume variables that represent unexpected Twitter volume. Same as in the case of GARCH, we use two sets of variables that we add to the ARFIMA equations. You can find these variables and their explanations in 3.2 and 3.2. The general functional forms of our augmented ARFIMA models with mean μ are as follows:

$$\Phi(L)(1-L)^d(y_t-\mu-\sum_{j=1}^{5}\gamma_j Tvol_{j,t-1}) = \Theta(L)\varepsilon_t, \varepsilon_t \ i.i.d., (0, \sigma_\varepsilon^2) \tag{2.16}$$

$$\Phi(L)(1-L)^d(y_t-\mu-\sum_{j=1}^{5}\gamma_j RTvol_{j,t-1}) = \Theta(L)\varepsilon_t, \varepsilon_t \ i.i.d., (0, \sigma_\varepsilon^2) \tag{2.17}$$

Where $Tvol_{j,t-1}$ and $RTvol_{j,t-1}$ are vectors representing Twitter volumes and unexpected Twitter volumes, γ_j are parameters representing Twitter effects. Term L is a back-shift operator,$\Phi(L) = 1 - \phi_1 - ... - \phi_p L^p$, $\Theta(L) = 1 + \upsilon_1 L + ... + \upsilon_q L^q$, and $(1-L)^d$ is fractional differencing operator defined by the following:

$$(1-L)^d = \sum_{k=1}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)} \tag{2.18}$$

In our empirical research, we will specifically use ARFIMA(1,d,0) model setting for our augmented models. Given the fact that we provided sufficient amount of detailed theoretical background, it is time to move to the following part of the thesis, where we comment and describe the data and methodology that we utilize in our research.

Chapter 3

Data and Methodology

As we mentioned earlier, we examine all 30 stocks that constitute well known Dow Jones Industrial Average stock index. Dow Jones was established and published for the first time in the year of 1896. It is an index that shows how 30 large, publicly owned companies based in the United States, have traded during a standard trading session in the stock market. It is widely used as a proxy for indication of stock market performance and also one of the most known stock market indices in the world. Later in the research part of this work, we compose models that aim to predict future stock volatility using data generated by Dow Jones Industrial Average stocks and data collected on Twitter. In the following part of the thesis, we provide a thorough description regarding the data we utilize in our research as well as methodological procedures that we use to adjust our data for the purpose of our research. We list all the companies as well as their tickers and number of tweets in the table below.

3.1 Return Data

We collected daily market data, consisting of open, close, daily high and daily low prices, for all 30 Dow Jones companies. The period of collection was set according to availability of Twitter data, i.e. 15 months starting on 31/5/2013 ending on 18/9/2014. We obtained the data from the paper that was written by Ranco *et al.* (2015).

For each stock, we have a total of 329 observations regarding each type of the price data. This corresponds to the fact that there are some non-trading days during working weeks and there is also no trading during the weekends.

Table 3.1: DJIA components

Ticker	Company Name	Tweets
TRV	Travelers Companies Corp	12184
UNH	UnitedHealth Group Inc	15020
UTX	United Technologies Corp	16123
MMM	3M Co	17001
DD	E I du Pont de Nemours and Co	17340
AXP	American Express CO	21941
PG	Procter & Gamble Co	25751
NKE	Nike Inc 29,220	29220
CVX	Chevron Corp	29477
HD	Home Depot Inc	30923
CAT	Caterpillar Inc	38739
JNJ	Johnson & Johnson	40503
V	Visa Inc	43375
VZ	Verizon Communications Inc	45177
KO	Coca-Cola Co	45339
MCD	McDonald's Corp	45971
XOM	Exxon Mobil Corp	46286
DIS	Walt Disney Co	46439
BA	Boeing Co	51799
MRK	Merck & Co Inc	54986
CSCO	Cisco Systems Inc	57427
GE	General Electric	61836
WMT	Wal-Mart Stores Inc	63405
INTC	Intel Corp	68079
PFE	PFE Pfizer Inc	71415
T	AT&T Inc	75886
GS	Goldman Sachs Group Inc	91057
IBM	International Business Machines Co	101077
JPM	JPMorgan Chase and Co	108810
MSFT	Microsoft Corp	183184
Total		**1,555,770**

Source: author's computations.

For our return series, we use the traditional logarithmic returns series that is defined by the following equation:

$$r_t = ln\left(\frac{P_t}{P_{t-1}}\right) = ln(P_t) - ln(P_{t-1}) \tag{3.1}$$

Where P_t is the close price at day t, P_{t-1} is the close price of the previous day, ln is a natural logarithm and r_t is daily log return of day t. It is also good to notice that log returns are preferable when working with financial series due to several reasons. Firstly, log returns are closer to symmetrical distribution compared to classical return series. Secondly, they stack up when you sum them up over any period of time. The stock price data were collected from Bloomberg terminal. Summary statistics for returns of all 30 stocks are provided in the table B.2.

From the table, we can see that Microsoft was the stock with the most negative daily return that accounted for -11.5 percentage points. On the other hand, the highest daily return was generated by Intel Corp. and accounted for 9.2. The average daily return of all Dow Jones stocks was approximately 0.05 percent.

At this point, we should notice a possible problem of discontinuity of our return series in relation to non-trading days and weekends. As the equation 3.1 suggests the series is calculated using today's and yesterday's close prices. Thus, when non-trading day occurs the return is instantly longer than daily return. Given the continuous flow of information, the non-trading day returns may contain more information and thus be statistically different in their size compared to daily returns. To control for this possible effect, we propose a solution to create a dummy variable that marks each non-trading day return in our series, later adding it into our regression and discovering whether there is an effect whatsoever. Such mean model would have a form defined by the following equation:

$$r_t = \mu + \beta Ntdr_t + \epsilon_t \tag{3.2}$$

where μ is a constant term and $Ntdr$ is a variable that stands for the "non-trading day return" with effect parameter β. If we discover such variable significant, we can simply use residuals ϵ_t of the regression 3.2 as our new return series. Such practice is similar in methodology to controlling for trends and seasonality in standard time series regression analysis. Now, let us move to another section where we discuss the data collected on Twitter.

3.2 Twitter Data

Another data source for our research is Twitter. Namely, we will operate with data representing tweets about Dow Jones stocks including their frequency and sentiment polarity. Originally, the data were collected using Twitter Search API, where the search query consisted of the stock cash-tag, such as $JPM for J.P. Morgan Chase, for all Dow Jones stocks. The data cover the period of 15 consecutive months starting on 1/6/2013 and ending on 18/9/2014, again, the data were obtained from the paper that was written by Ranco *et al.* (2015).

The Twitter data sample consists of 1.5 million data points collected over the period we mentioned above. The data were furthermore classified according to their sentiment, thus we also have a qualitative information regarding the tweets. The resulting data sample consists of the following variables:

- Twitter volume, tw_t, the total number of tweets at day t
- Positive tweets, ptw_t, the number of positive tweets at day t
- Negative tweets, ntw_t, the number of negative tweets at day t
- Neutral tweets, ztw_t, the number of neutral tweets at day t
- Sentiment polarity, p_t, the overall polarity of daily tweets calculated as $p_t = \dfrac{ptw_t - ntw_t}{ptw_t + ntw_t}$

Although we have not classified the tweets on our own, for the sake of completeness, we will briefly introduce the methods that were used for sentiment classification. It is not an easy task to determine the sentiment polarity of any information, as financial experts often disagree whether the tweet represents positive or negative information. Furthermore, the experts are not always consistent with themselves as well. In order to set up an efficient automatic classification method, a significantly large set of tweets has to be manually classified by human experts. In our case, such sample consists of $100,000$ tweets.

Above that, in order to maximize the agreement regarding the polarity between financial experts, over $6,000$ tweets were annotated twice by two different experts. The classification approach applied on our data was based on supervised machine learning method. Such method basically consists of four steps, which are the following:

- Manual annotation of sentiment of sample of tweets
- Training and tuning the classifier on labeled set of tweets
- Evaluation of the classifier by cross-validation and comparison to inter-annotator agreement
- Application of the classifier to the whole set of collected data

The classifier applied to our data set was based on Support Vector Machine, a widely used, supervised learning algorithm that is well suited for large scale text categorization tasks. The researchers have applied wrapper approach that was described by Frank & Hall (2001), which utilizes two linear-kernel Support Vector Machine classifiers as described by Vapnik (2013). Given the fact that the classes are ordered, two classifiers are sufficient to partition the space of tweets into sentiment categories that we introduced earlier. The two classifiers were trained to distinguish between positive and negative-or-neutral and between negative and positive-or-neutral. If the target class could not be determined due to the disagreement between the classifiers, the tweet was tagged as neutral. The more detailed description regarding exactly the same methodology, employed on the sentiment classification, can be found in a paper written by Ranco *et al.* (2015).

Having the Twitter series efficiently categorized, we can treat it as a time series and further research its properties. If the autocorrelation was discovered in the series, we model the Twitter series using a well-known Autoregression model. There might be an interesting rationale behind the autocorrelation presence in the Twitter series that evolves from the logic behind Twitter social network and behavior of its users. Imagine that there is a news regarding any of the stocks that we use in our analysis. You can expect a Twitter user to share his opinion regarding the news on the social network, which possibly can lead into series of re-tweets attracting more people involved in the discussion over time.

The more people react over time, the more attraction it generates just like the snow-ball effect. Therefore, it might be reasonable to expect autocorrelation in the series. To put this behavior into some reasonable framework, we predict and save the residuals from the AR regression, generating new series that we interpret as unexpected Twitter volumes. We expect this series to contain interesting information about abnormal twitter volumes, possibly capturing extraordinary events. The functional form of Twitter volume models can be generalized as follows:

$$Tvol_t = \mu + \sum_{i=1}^{j} \beta_i Tvol_{t-i} + \epsilon_i \tag{3.3}$$

Where j is a positive integer that determines the lag length of the AR equation, μ is a vector of constants, $Tvol_t$ is a vector representing the set of our Twitter variables tw_t ,ntw_t, ptw_t, ztw_t and p_t. Residual terms ϵ_i that we predict from our equation are interpreted as unexpected Twitter volumes. These AR regression residuals produce the entire new set of variables related to their Twitter volume counterparts, these variables are:

- Unexpected Twitter volume, Rtw_t, the total number of unexpected tweets at day t
- Unexpected positive tweets, $Rptw_t$, the number unexpected of positive tweets at day t
- Unexpected negative tweets, $Rntw_t$, the number of unexpected negative tweets at day t
- Unexpected neutral tweets, $Rztw_t$, the number of unexpected neutral tweets at day t
- Unexpected sentiment polarity, Rp_t, the overall unexpected sentiment polarity of daily tweets

As we stated before, such variables arise as residuals from the equation 3.3 based on their autocorrelation and partial autocorrelation functions. We will employ these variables in separate regression models, comparing them to those without any additional variables as well as to those where we plug simple twitter volume series. Let us move now to another section that is related to modeling realized volatility that we use as a comparison benchmark for all sets of our models.

3.3 Proxy of Realized Volatility

The financial volatility process is central to financial asset return modeling. However, volatility itself is unobservable. Hence, one often has to rely on proxies when specifying, estimating and evaluating volatility models. For our purposes, we need to choose such proxy of realized volatility that will allow us to compare prediction performance results of our GARCH and ARFIMA model families. The most common proxy for realized volatility is just simple square return series. Such estimator is however very imprecise and not very efficient. Better estimation results can be obtained by the employment of different proxy estimators. For example, a family of price range estimators employs more information than just a daily close prices. A significant practical advantage of the price range estimators is that for many assets, daily opening, highest, lowest, and closing prices are readily available, as most data suppliers provide these data as summaries of daily price activity. Thus, range-based volatility proxies can be easily calculated. When using this record, the additional information yields a great improvement when used in financial applications. Roughly speaking, knowing these records allows us to get closer to the real underlying process, even if we do not know the whole path of asset prices. Let us now define the underlying variables that are needed to come up with our price range estimators:

- Daily close price, C_t, is the price at which the market closed at day t
- Daily open price, O_t, is the price at which the marked opened at day t
- Daily high price, H_t, the highest price of an asset at day t
- Daily low price, L_t, the lowest price of an asset at day t

Instead of using just daily close prices and calculating squared returns, four data points, as the opening, closing, the highest and lowest prices, might give us extra information. Garman & Klass (1980) propose several volatility estimators based on the knowledge of the opening, closing, highest and lowest prices. In our research, we decided to employ volatility estimator that is defined by the following equation:

$$\sigma^2_{GK} = 0.5[ln\left(\frac{H_t}{L_t}\right)]^2 - [2ln2 - 1][ln\left(\frac{C_t}{O_t}\right)]^2 \tag{3.4}$$

Where ln is natural logarithm, and H_t, L_t, O_t and C_t are variables as defined in 3.3. In our research, we employ the Garman Klass estimator in two ways. First, we use it as an input variable for our ARFIMA model family. Second, we use it as a common denominator for cross-family model comparison. After rescaling predicted variance from the GARCH models by a certain scaling factor, we are able to directly compare GARCH and ARFIMA models. The scaling factor is defined as follows:

$$c = \frac{\phi\sigma^2_{GK,insample}}{\phi\sigma^2_{GARCH,insample}} \tag{3.5}$$

Where $\phi\sigma^2_{GK,insample}$ is average Garman Klass estimator estimated within our training sample interval, $\phi\sigma^2_{GARCH,insample}$ is average GARCH estimator estimated within our training sample interval for every model of comparison, i.e. model with Twitter volumes, without Twitter volumes and with unexpected Twitter volume variables as defined in 3.2. Scaling factors c are then used to rescale GARCH volatility predictions, making them comparable to Garman Klass predictions and also ARFIMA predictions.

Chapter 4

Results of the Research

In this section, we describe the actual application of the methodology on the real data. We begin with thorough elaboration on the stock of "Walt Disney Co.". We go through all the tests and regression we performed step by step, building and describing the entire course of action. The modeling procedure remains similar for the remaining stocks. We elaborate and comment only on the five selected stocks that have interesting results. For the rest, we also provide aggregated results, aiming to discover whether there is any possibility to generalize our overall research and possibly draw conclusions regarding the model performance and the estimated polarity of the Twitter effects. In the end, we summarize the entire research in the summary part of the work.

4.1 Results of the Selected Stocks

4.1.1 Walt Disney Co.

For explanatory purposes, we decided to present analysis conducted on a stock of the famous Walt Disney Co., which is traded on New York Stock Exchange under the "DIS" ticker. The company operates in the entertainment sector, operating a vast array of broadcast, cable, radio, publishing and digital businesses.

Before we start with implementing our volatility modeling frameworks, we should check the underlying assumptions beneath it.
Let us start with the GARCH model. In general, conditional heteroscedasticity models operate under the assumptions of stationarity of a time series or at

least under no unit root assumption. Furthermore, suitability of these model is indicated by the presence of serial correlation of mean model residuals and presence of arch terms in return series.

We begin with a stationarity assumption. For this purpose, we combine two tests with opposite null hypotheses. These tests are the augmented Dickey–Fuller test and the Kwiatkowski–Phillips–Schmidt–Shin test applied on our log return series. The ADF test yields a test statistic of -18.5540, which at a critical value of -2.877 at 5 percent significance level means strong rejection of the null hypothesis, that is, rejection of unit root presence in the series. The KPSS test works the other way around. The null hypothesis states that the series is stationary. Critical value at 5 percent significance level is equal to 0.1460. The test statistics obtained through lags 1 - 16 ranged between 0.0266 - 0.0403 which effectively means the failure of rejection, hence we can not rule out the stationarity of the time series of log returns.

Yet we know that we most likely deal with stationary time series, so we can move on to examine the possible form of the mean equation of our GARCH model. For such purpose, it is useful to examine autocorrelation function of the log return series. Figure A.1 shows that there is likely no autocorrelation in a return series for DIS stock, hence the form of a mean equation will probably be a just simple mean regression with a dummy variable for non trading day return as we explained in equation 3.2. To check for possible ARCH effects in the return series, we will estimate the mean equation right ahead and test its residuals with the Lagrange-Multiplier test. Such test is simply equivalent to the usual F statistic for testing the joint hypothesis in a linear regression of squared residuals on its lagged values. The estimation results of the mean equation are provided in table 4.1. As you can see, the variable representing non trading day stands insignificant in the equation. Such results lead to the simplest form of DIS mean equation, consisting simply of the constant term.

Table 4.1: DIS mean equation

Variable	Coeff.	Std. Err.	t	p
Ntdr	-0.0007	0.0015	-0.4600	0.6460
cons	0.0012	0.0007	1.7700	0.0770

Source: author's computations.

Having checked the *Ntdr* variable, we can proceed to the Lagrange-Multiplier test applied on mean equation's residuals. The outcome of the test can be found in the table B.1. As you can see from the p values in the table, we can strongly reject the null hypothesis of no ARCH effects over higher lag order values. Therefore, it is possibly reasonable to employ GARCH model for our analysis.

As we mentioned in section 3.2, we need to model the Twitter volume series as well. Employing 3.3 equations to volume and unexpected volume variables, we show the regression results for the completeness. But first, let us elaborate on the general modeling procedure applied on the variable representing a number of tweets with negative sentiment. Again, we first started with the check of stationarity of the ntw_t time series. Performing ADF and KPSS test, the results show that there is no unit root problem and according to KPSS, the series is likely stationary. Proceeding to actual examination of lag-length of the regression model, we perform the auto correlation function. The resulting figure A.2 suggests that there is strong first order auto correlation followed by insignificant higher order lags. Therefore, the *ntw* AR model will contain only first lag of the variable. The results of such regression are presented in table 4.2, where variable *L.ntw* represents lagged values of negative Twitter volumes. From such regression, one can conclude that yesterday's Twitter activity with negative polarity positively affects today's twitter activity with negative sentiment. We conducted the same analysis also for the rest of Twitter volume variables that we provided in 3.2. In aggregate, all the Twitter volume variables exhibited an auto correlation to some degree. From each regression, we also predicted and saved residuals that we later utilize as a proxy of unexpected Twitter volume variables.

Table 4.2: AR regression of ntw

Variable	Coeff.	Std. Err.	t	p
L.ntw	0.3268	0.0524	6.2400	0.0000
cons	3.3448	0.4470	7.4800	0.0000

Source: author's computations.

At this point, we have all the variables needed for specifying our GARCH and

ARFIMA models. Let us begin with the simple GARCH(1,1) specification and then continue to elaborate on the augmented GARCH models.

The entire analysis is performed on the data of 230 observations that corresponds to approximately 70 percent of the entire sample. As we already presented, we will use the simple mean equation, that is, the equation where only intercept is present.

Table 4.3: GARCH Regression of DIS

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0015	0.0009	1.6900	0.0920
Variance Equation				
const	0.0003	0.0000	26.5700	0.0000
L.arch	0.0214	0.0134	1.5900	0.1120
L.garch	0.7823	0.0309	-34.0700	0.0000

Source: author's computations.

Interpretation of the simple GARCH model is fairly straightforward. Beginning with the mean equation, we can see that intercept term is not significant at 5 percent significance level. Let us focus, however, on the variance equation that yields more interesting information for our research topic. As you can see from the table 4.3, arch effect captured by the *L.arch* is insignificant at 5% significance level. Such variable represents the response of volatility to previous period shocks in return series. The variable representing the persistence of volatility, in the table under the name of *L.garch* is statistically significant. As for the magnitude, yesterday's volatility leads to another period of high today's volatility. We use our simple GARCH(1,1) model just as a benchmark model. After augmenting GARCH on Twitter volume variables, we obtain results that we present in the table below.

Interpretation of such regression is more interesting as there are new additional variables. Let us start again with the arch and garch effects that are represented by variables *L.arch* and *L.garch* respectively. The coefficients and significance remain very similar to the case of simple GARCH setting with no additional variables. It is good to note that the negativity of the arch effect estimate is offset by its insignificance, thus it should comply with standard arch

Table 4.4: Augmented GARCH Regression of DIS

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0013	0.0008	1.6600	0.0980
Variance Equation				
const	-8.2400	0.0980	-84.0800	0.0000
L.ntw	0.0065	0.0028	2.3600	0.0190
L.ptw	0.0033	0.0006	5.6000	0.0000
L.arch	-0.0297	0.0330	-0.9000	0.3690
L.garch	0.7747	0.0708	-12.3500	0.0000

Source: author's computations.

assumptions and should not cause any problems in our regression. Focusing on Twitter variables, we can say that lagged Twitter activity represented by variables *L.ntw* and *L.ptw* positively affects future volatility. Starting with variable *L.ntw*, representing lagged Twitter volume with negative sentiment, is statistically significant with the magnitude of estimation parameter equal to 0.0065. Positive Twitter volume variable has the magnitude of 0.0033. Comparing the variables, Twitter with negative sentiment polarity is in its magnitude 1.93 times higher than *L.ptw*. Compared to the main explanatory variable, *L.garch*, the magnitudes of *L.ntw* and *L.ptw* are marginal.

Another GARCH model differs by employing residuals from Twitter volume AR equations, that is unexpected Twitter volumes.

Table 4.5: Volume Residual Augmented GARCH Regression of DIS

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0012	0.0008	1.5500	0.1220
Variance Equation				
const	-8.1504	0.0943	-86.4500	0.0000
L.Rntw	0.0086	0.0014	6.0800	0.0000
L.Rptw	0.0025	0.0005	5.4700	0.0000
L.arch	-0.0448	0.0362	-1.2400	0.2160
L.garch	0.7123	0.0472	-17.2200	0.0000

Source: author's computations.

Interpretation of such regression is interesting, as the unexpected Twitter volume variables capture social buzz around events that generate an abundance of attention on social networks. Let us start again with the arch and garch effects that are represented by variables *L.arch* and *L.garch* respectively. The coefficients and significance remain very similar to the case of simple GARCH setting with no additional variables. Focusing on unexpected Twitter variables, we can say that lagged unexpected Twitter activity represented by variables *L.Rntw* and *L.Rptw* positively affect future volatility. Starting with variable *L.Rntw*, representing lagged unexpected Twitter volume with negative sentiment, is statistically significant with the magnitude of estimation parameter equal to 0.00862. Positive unexpected Twitter volume variable has the magnitude of 0.002541 and is significant as well. Comparing the variables, unexpected Twitter volume with negative sentiment polarity is in its magnitude 3.39 times higher than *L.Rptw*. Compared to the main explanatory variable, *L.garch*, the magnitudes of *L.ntw* and *L.ptw* are again marginal. If we were to compare effects between plain and residual Twitter volumes, we would discover that effect of *L.Rntw* is approximately 1.32 times larger than of *L.ntw*. Similarly, *L.Rptw* accounts approximately for 76% of the magnitude of *L.ptw*.

To be able to compare prediction capability of the models, it is necessary to predict fitted values of variance equations. To remind the reader, we have estimated the models on a sub sample of our data set. Our benchmark for prediction precision is Garman Klass volatility estimator that we introduced in section 3.3. To be able to compare GARCH outcomes with Garman Klass, we have to rescale the variance fitted values by scaling parameter presented in 3.5. The scaling parameters for our GARCH, augmented GARCH and residual augmented GARCH are equal to 0.67750, .06597 and 0.66640 respectively. Proceeding further, we can compare the prediction precision among family of GARCH models using test introduced by Diebold & Mariano (2002). The test is performed on mutual pairs of predictors to be able to determine their order. Among GARCH models, the test results and their outcome is presented in the table 4.6.

Table 4.6: DM test for GARCH of DIS

GARCH	AGARCH	RGARCH
3	1	2

Source: author's computations.

As you can see, the ordering of the models works clearly in favor of models with additional Twitter variables. AGARCH representing GARCH model with plain Twitter volumes rank in the first position, RGARCH representing a model with residual Twitter volumes ends up to be the second and plain GARCH model is on the third position. Given the ordering of the models, we also provide you with the p value from the testing the first and the third model, i.e. the residual GARCH and the plain GARCH. The p value of the test is equal to 0.1749, thus it seems the performance difference is not statistically significant within the GARCH family. To make the research even more comprehensive, we now move on to ARFIMA modeling to be able to make comparison across these two families of models.

ARFIMA models are useful in cases where time series exhibit behavior between unit root and stationary process. For our modeling purposes, we use the Garman Klass series as a time series in our ARFIMA regression. Performing ADF and KPSS show that Garman Klass series does not exhibit unit root problem and we can not reject that the series is trend stationary. We can therefore easily continue to set up our base model, that is, ARFIMA(1,d,0). The estimation results of the model are provided in table 4.7.

Table 4.7: ARFIMA(1,d,0) of DIS

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	17.3600	0.0000
L.ar	0.2010	0.1581	1.2700	0.2040
d	-0.0401	0.1321	-0.3000	0.7610

Source: author's computations.

Variable *L.ar* represents the auto regression term in the equation, according to the results, the coefficient is not significantly different from zero, that is, yesterday's volatility does not affect today's volatility in the series. Parameter d is not significant as well, we can, therefore, assume that its value is equal to zero. Such value of d is often interpreted such that the time series exhibits short memory and is stationary.

When we add Twitter volume variables to the model, we obtain regression results presented in table 4.8.

Table 4.8: Twitter Volume Augmented ARFIMA(1,d,0) of DIS

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	17.3600	0.0000
L.ntw	0.0000	0.0000	2.0700	0.0390
L.ptw	0.0000	0.0000	2.0800	0.0370
L.ar	0.1321	0.1618	0.8200	0.4140
d	-0.0387	0.1349	-0.2900	0.7740

Source: author's computations.

In general, variables representing Twitter volumes, *L.ntw* and *L.ptw* are statistically significant at 5% significance level. Lagged positive Twitter volume, *L.ptw*, positively affects one step ahead values of the Garman Klass volatility estimator. Lagged negative Twitter volume parameter values are also positive. As for their magnitude, we can clearly see that *L.ntw* has a stronger effect in our model. Classical ARFIMA variables, *L.ar* and *d* are again insignificant in the regression. If we should connect the results with a real behavior on social networks and financial markets, we would draw the conclusion that activity on social networks implies higher future volatility in financial markets. Such results will also be interesting in the case of unexpected Twitter volumes model which we present in table 4.9.

Table 4.9: Residual Twitter Volume Augmented ARFIMA(1,d,0) of DIS

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	15.6200	0.0000
L.Rntw	0.0000	0.0000	2.5100	0.0120
L.Rptw	0.0000	0.0000	3.4400	0.0010
L.ar	0.0278	0.1434	0.1900	0.8470
d	0.0290	0.1175	0.2500	0.8050

Source: author's computations.

Again, unexpected Twitter volume variables *L.Rntw* and *L.Rptw* are statistically significant in the model. As for their effects, they both positively contribute to future volatility represented by Garman Klass series. Comparing the magnitudes of residual volumes to those of simple Twitter volumes, we can state that both residual volume parameters are larger in their magnitudes. Also, the p values of residual volumes are lower compared to those of simple Twitter volumes.

Advantage of using Garman Klass as our benchmark estimator for ARFIMA model is that we do not need to rescale any predictions of our models. Again, we have estimated both ARFIMA models on a subsample that consists of 230 data points. The rest of data points were used to predict out of sample predictions. Applying Diebold Mariano test on our three ARFIMA models, we obtain the following comparison of models.

Table 4.10: DM test for ARFIMA of DIS

ARFIMA	AARFIMA	RARFIMA
3	2	1

Source: author's computations.

The comparison results are very similar to those of GARCH case. The most efficient model, according to our testing methodology, is model with residual Twitter volumes. The second scores model with simple Twitter volumes and the plain ARFIMA model ends up in a third place. Such ordering suggests that Twitter variables improve the predictive power of the ARFIMA model, regardless of whether it is simple or residual Twitter volumes. Furthermore, the test statistic from the comparison of the first and the third model seems to be significant with p value equal to 0.0391.

The last and maybe the most important step is to figure out ordering between GARCH and ARFIMA models. Again, we apply the Diebold Mariano test on all six models and sort them accordingly. The test statistics yield the ordering that we present in the table 4.11

Table 4.11: DM test for GARCH and ARFIMA of DIS

GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
6	4	5	3	2	1

Source: author's computations.

The test results reveal that ARFIMA models overperform the entire family of GARCH models. The augmented models, however, overperform the plain benchmark GARCH and ARFIMA. Thus, it seems that in the case of Walt Disney Co., the Twitter variables can be used to predict the one day ahead volatility forecast more precisely. The first option for selecting a model, given our available data frame, would be the ARFIMA with residual Twitter volume variables, *L.Rntw* and *L.Rptw*, representing unexpected positive Twitter volume and unexpected negative Twitter volume respectively. The p value from testing the first and the last model indicates statistical significance, p values of the entire set of the stocks is provided in the appendix. Both of these variables exhibited positive effects on future volatility. In the next section, we present results based on the other stocks.

4.1.2 American Express Co.

Another analysis that we present is the analysis performed on the American Express company. American Express is financial services company with its core business in credit cards, charge cards, and traveler's cheques. The company is listed on the New York Stock Exchange and is a component of various stock indices such as SP100, SP500 and finally DJIA. In this subsection and following sections, we will present only shortened version of the analysis, as the general procedures are the same as in the case of analysis performed on DIS in the section 4.1.1.

For the beginning, let us start with estimating the plain GARCH model with no additional explanatory variables. The estimation results are presented in the table 4.12. As you can see, the arch term representing a reaction to recent shocks is marginally significant with p value of 0.057. Garch term is, however, statistically significant with p value equal to zero.

Obtaining such results can suggest that volatility features are present in the data and their effects can be properly captured by the GARCH model setting.

Another task is to augment the GARCH by the Twitter volume variables. In the AXP case, we found that there is a dependence between the volatility and

Table 4.12: GARCH Regression of AXP

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0008	0.0008	1.0600	0.2900
Variance Equation				
const	0.0000	0.0000	1.4000	0.1630
L.arch	0.1303	0.0684	1.9100	0.0570
L.garch	0.6908	0.1716	4.0300	0.0000

Source: author's computations.

the number of positive tweets. The regression results are presented in the table 4.13.

Table 4.13: Augmented GARCH Regression of AXP

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0007	0.0008	0.8700	0.3860
Variance Equation				
const	0.0000	0.0000	1.4000	0.1630
L.ptw	0.0143	0.0054	2.6500	0.0080
L.arch	0.0715	0.0647	1.1100	0.2690
L.garch	-0.1025	0.2101	-0.4900	0.6260

Source: author's computations.

From the regression results, you can see that both the traditional variance equation variables and their parameters turned statistically insignificant, thus we can state that the arch and garch terms are both equal to zero, not affecting the regression. Again, the negativeness of the garch effect is offset by its insignificance. Additional variable *L.ptw* representing lagged values of positive Twitter volumes are statistically significant and exhibit positive effect on tomorrow's volatility prediction. Such effect in practice means, the more Twitter activity with positive sentiment, the more abnormal return can we expect in the upcoming day.

Now we move on to estimate and comment on the results of our final GARCH model, to which we add unexpected Twitter volume variables. The regression results are presented in the table 4.14 below.

Table 4.14: Residual Augmented GARCH Regression of AXP

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0007	0.0008	0.9200	0.3560
Variance Equation				
const	-8.9111	0.2072	-43.0100	0.0000
L.Rptw	0.0211	0.0065	3.2600	0.00100
L.arch	0.0819	0.0673	1.2200	0.2230
L.garch	-0.0112	0.1558	-0.0700	0.9430

Source: author's computations.

Again, the results show that after adding residual Twitter volume variable to the variance equation of the model, *L.arch* and *L.garch* parameters become statistically insignificant. The variable *L.Rptw* represents unexpected Twitter volume with positive sentiment polarity. As for the effect polarity and magnitude, *L.Rptw* positively affects future volatility and its effect is roughly 1.47 times the effect of plain positive Twitter volume.

Nevertheless, primary task of our thesis is not to show and explain the Twitter effects but to elaborate on the prediction accuracy of the Twitter and plain models. For such reason, we employ the test statistics that can help us with sorting the models accordingly. The test results are presented in the table 4.15. As you can see, despite significance of Twitter and residual Twitter variables, the plain GARCH model scores the first place in the comparison. Comparing the results of Twitter and residual Twitter variables, a model with plain Twitter positive sentiment volume overperforms unexpected Twitter volume model. The p value from testing plain GARCH and residual GARCH is equal to 0.7780. Let us move on now to the ARFIMA model results.

Table 4.15: DM test for GARCH of AXP

GARCH	AGARCH	RGARCH
1	2	3

Source: author's computations.

We start with estimating benchmark ARFIMA model with no additional variables, later we estimate the model with additional variables and elaborate on estimation results. The estimation results of the plain model are presented in the table 4.16. To briefly comment on the results, the *L.ar* variable representing autoregressive term in the equation turns insignificant. Variable d representing fractional parameter is, however, statistically significant. Given the magnitude of the parameter, the time series seem to exhibit long memory or long-range positive dependence.

Table 4.16: ARFIMA(1,d,0) of AXP

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	6.7900	0.0000
L.ar	-0.0343	0.1018	-0.3400	0.7360
d	0.1825	0.0785	2.3200	0.0200

Source: author's computations.

Adding Twitter volume variable to the ARFIMA equation yields results that we present in the table 4.17. Again, we see the insignificance of parameter *L.ar*, meaning that there is no evidence for first order auto correlation in the series. Parameter d still suggests the same properties as in the plain ARFIMA model case. Variable *L.ptw*, representing positive Twitter volume, shows that positive Tweets tend to increase the future volatility of the American Express stocks.

Table 4.17: Twitter Volume Augmented ARFIMA(1,d,0) of AXP

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	5.6500	0.0000
L.ptw	0.0000	0.0000	4.8900	0.0000
L.ar	-0.0839	0.1007	-0.8300	0.4050
d	0.1941	0.0788	2.4600	0.0140

Source: author's computations.

The model that includes residual Twitter volumes is presented in the table 4.18. Despite the p value of *L.ar* variable has decreased, the variable is still statistically insignificant. Parameter d is still in the same value interval that suggests long memory property of the time series. The variable *L.Rptw* that stands for unexpected Twitter volume with positive polarity is positive and statistically significant. Compared to *L.ptw*, *L.Rptw* is larger in its magnitude. Thus, it can be interpreted that the effect of unexpected Twitter volume is stronger compared to plain Twitter volume parameter.

Table 4.18: Residual Twitter Volume Augmented ARFIMA(1,d,0) of AXP

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	5.9700	0.0000
L.Rptw	0.0000	0.0000	5.5300	0.0000
L.ar	-0.1425	0.0995	-1.4300	0.1520
d	0.2312	0.0781	2.9600	0.0030

Source: author's computations.

Having all the three ARFIMA models estimated, we can perform the DM test to compare their prediction accuracy results. The results of the test are presented in the table 4.19. Unlike the case of GARCH, both Twitter volume ARFIMA models perform better compared to the plain ARFIMA model. The p value from testing the first and the third model is equal to 0.3053.

Table 4.19: DM test for ARFIMA of AXP

ARFIMA	AARFIMA	RARFIMA
3	1	2

Source: author's computations.

Finally, we can compare GARCH and ARFIMA models among themselves. We present the comparison results in the table 4.20.

Table 4.20: DM test for GARCH and ARFIMA of AXP

GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
3	5	6	4	1	2

Source: author's computations.

The comparison results are a little bit more interesting compared to the previous case of DIS. As you can see, the test reveals mixed results. The first position is captured by AARFIMA model that represents augmented ARFIMA model, second is captured by the RARFIMA model representing residual augmented

ARFIMA model and the third scores the simple GARCH model. From such results, we can draw a conclusion that augmented ARFIMA models overperform the entire family of GARCH models. The p value from testing the first and the last model indicates marginal significance, i.e. 0.0565.

4.1.3 General Electric

General Electric is an American multinational corporation that operates in the fields of aviation, current, digital, energy connections, financial services and other industries. From such broad spectrum of activities, GE is regarded as a widely diversified company. The stocks of the company are traded on the New York Stock Exchange and the company is a constituent of the DJIA stock index.

Let us begin again with the estimation of plain GARCH model later followed by models with Twitter volume variables. We again checked all the necessary assumptions that allow us to utilize our modeling frameworks. The results of the plain GARCH model are presented in the table 4.21.

Table 4.21: GARCH Regression of GE

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0006	0.0008	0.8200	0.4110
Variance Equation				
const	0.0000	0.0000	0.5400	0.5920
L.arch	0.0315	0.0390	0.8100	0.4190
L.garch	0.8103	0.3132	2.5900	0.0100

Source: author's computations.

The estimation results show that *L.arch* variable, representing shock response, is insignificant. Variable *L.garch* is statistically significant with the estimation parameter equal to 0.810339. The estimation condition that the sum of arch and garch parameters must be less than one is satisfied.

Following the same methodology as in previous cases, we augment the model by Twitter variables. In the General Electric's case, the significant variable was *L.ptw*. We present the estimation results in the table 4.22.

Table 4.22: Augmented GARCH Regression of GE

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0006	0.0007	0.8400	0.4030
Variance Equation				
const	-10.6816	0.8610	-12.4100	0.0000
L.ptw	-0.0484	0.0259	-1.8700	0.0620
L.arch	0.0626	0.0435	1.4400	0.1500
L.garch	0.8560	0.1029	8.3200	0.0000

Source: author's computations.

The regression results reveal interesting information. Arch and garch parameters remain the same regarding their significance. Positive Twitter volume, however, shows to be significant at the level of 6.2%. Interestingly, the parameter is negative which is a bit surprising knowing the results from previous stocks. Given the fact that we operate with five percent significance level, we can interpret such results as *L.ptw* is not statistically different from zero.

Conducting the analysis with residual Twitter volumes, we obtain interesting results as well. Unlike previously, we obtained different significant residual Twitter variable, that is, the unexpected polarity of Tweets represented by *L.Rpol*. The results are presented in the table 4.23.

Table 4.23: Residual Augmented GARCH Regression of GE

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0006	0.0008	0.7500	0.4530
Variance Equation				
const	-11.5797	1.3422	-8.6300	0.0000
L.Rpol	2.3722	1.2422	1.9100	0.0560
L.arch	0.0139	0.0240	0.5800	0.5640
L.garch	0.8779	0.1363	6.4400	0.0000

Source: author's computations.

As you can see, the arch parameter remains insignificant and garch parameter remains significant in the equation. Variable *L.Rpol* has the strongest effect among Twitter variables so far. The p value of the *L.Rpol* is 0.564 which is close to being significant at 5% significance level. As we stated before, the variable represents unexpected daily polarity of tweets regarding the GE stock. Real world interpretation of this variable is not easy at all but one can say that such variable might be interpreted as a surprise to investors in relation to some news or events.

Given the fact that we have estimated all three models, we can move to a comparison of their prediction capability. The test statistic yielded the results that we present in the table 4.24. Interestingly, the results of the test show that augmented GARCH model is better for out of sample prediction of volatility in the case of General Electric. The best model of our comparison is GARCH wit *L.Rpol* followed by a model with *L.ptw*. Plain GARCH model with no additional variables is in the last place in our comparison. The p value from comparing the best and the worst model is equal to 0.1236, thus is statistically insignificant.

Table 4.24: DM test for GARCH of GE

GARCH	AGARCH	RGARCH
3	2	1

Source: author's computations.

Proceeding with estimation of ARFIMA models, we provide the results of the plain regression in the following table 4.25. As you can see, both *L.ar* and *d* are statistically insignificant.

Table 4.25: ARFIMA(1,d,0) of GE

Variable	Coeff.	Std. Err.	t	p
const	0.0000	0.0000	18.6600	0.0000
L.ar	0.1047	0.1159	0.9000	0.3660
d	-0.0045	0.0935	-0.0500	0.9620

Source: author's computations.

When we add Twitter volume variables to the model, we find out that *L.ntw* representing negative Twitter volume is significant. We present the regression outcomes in the table 4.26. From the table you can see that again the Twitter volume variable positively affects tomorrow's volatility. Such outcome is in line with our general expectation that the more volume the more attention of investors translates into bigger return magnitude. Traditional ARFIMA variables, *L.ar* and *d* are not significant even after adding an additional explanatory Twitter variable.

Table 4.26: Twitter Volume Augmented ARFIMA(1,d,0) of GE

Variable	Coeff.	Std. Err.	t	p
const	0.0000	0.0000	14.3600	0.0000
L.ntw	0.0000	0.0000	1.9600	0.0500
L.ar	0.0457	0.1143	0.4000	0.6890
d	0.0111	0.0881	0.1300	0.9000

Source: author's computations.

The last model that we estimate for the General Electric is ARFIMA augmented on residual Twitter variable. In this case, *L.Rztw*, variable that represent unexpected neutral Twitter volume turns significant in the regression. As you can see from the table 4.36, *L.Rztw* variable is statistically significant with positively affecting future volatility. As for the auto regressive term and difference parameter, these variables remain insignificant in the regression.

Table 4.27: Residual Twitter Volume Augmented ARFIMA(1,d,0) of GE

Variable	Coeff.	Std. Err.	t	p
const	0.0000	0.0000	15.7200	0.0000
L.Rztw	0.0000	0.0000	2.7800	0.0060
L.ar	-0.0036	0.1170	-0.0300	0.9760
d	0.0504	0.0893	0.5600	0.5720

Source: author's computations.

Comparison of prediction capabilities of the model is presented in the table 4.28. The best among competing model is the model with residual neutral Twitter volume. The second is plain ARFIMA with no additional variables. The third place in our comparison is captured by the model that includes *L.ntw*, that is, the model with negative Twitter volume variable. The p value from the comparison of the third and the first model indicates statistical significance, i.e. 0.0173.

Table 4.28: DM test for ARFIMA of GE

ARFIMA	AARFIMA	RARFIMA
2	3	1

Source: author's computations.

Having all models estimated, we can move on to compare all six models to discover which one of them has the best volatility prediction performance. Such comparison is presented in the table 4.29.

Table 4.29: DM test for GARCH and ARFIMA of GE

GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
4	2	1	5	6	3

Source: author's computations.

When we look at the results, we can draw a conclusion that the augmented models are in the first three positions. The best performing model is the residual Twitter GARCH followed by Twitter volume GARCH and residual Twitter ARFIMA. Finally, the p value from the best and the worst model indicates statistical significance and is equal to 0.0012.

4.1.4 3M co.

Minnesota Mining and Manufacturing Company is an American multinational conglomerate corporation that produces a wide spectrum of products in the field of electronic materials, medical supplies, car-care products, and various other categories. The company's stocks are traded on the New York Stock Exchange. The company is a component of various stock indices such as SP500, SP100, and DJIA.

Let us begin again with estimation of plain GARCH model later followed by models with Twitter volume variables. Same as in the case of previous stocks, we checked all the necessary assumptions that allow us to utilize our modeling frameworks. The results of the plain GARCH model are presented in the table 4.30.

Table 4.30: GARCH Regression of 3M

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0009	0.0007	1.3400	0.1810
Variance Equation				
const	0.0001	0.0000	1.5300	0.1270
L.arch	0.0860	0.0501	1.7200	0.0860
L.garch	0.2047	0.4707	0.4300	0.6640

Source: author's computations.

The estimation results show that *L.arch* and *L.garch* variables, representing shock response and long memory, are statistically insignificant. Given the insignificance of arch and garch terms, estimation condition that the sum of arch and garch parameters must be less than one is satisfied.

Moving further, we augment the model by Twitter volume variables. In the 3M's case, the statistically significant variable is again *L.ptw*. We present the estimation results in the table 4.31.

Table 4.31: Augmented GARCH Regression of 3M

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0010	0.0007	1.4800	0.1390
Variance Equation				
const	-9.6637	0.3737	-25.8600	0.0000
L.ptw	0.0218	0.0092	2.3700	0.0180
L.arch	0.0469	0.0588	0.8000	0.4250
L.garch	0.0776	0.2492	0.3100	0.7560

Source: author's computations.

The estimation results show interesting information. Arch and garch parameters remain the same regarding their insignificance. Positive Twitter volume, however, shows to be significant at the level of 1.8%. The parameter sign is in line with our general expectation and most of the previous regression results.

Running the regression with residual Twitter volumes, we obtain interesting results as well. According to our analysis, *L.Rptw* is the significant variable in this case. The regression results are presented in the table 4.32.

Table 4.32: Residual Augmented GARCH Regression of 3M

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0010	0.0007	1.4900	0.1360
Variance Equation				
const	-9.6193	0.3392	-28.3600	0.0000
L.Rptw	0.0219	0.0106	2.0600	0.0390
L.arch	0.0485	0.0602	0.8100	0.4200
L.garch	0.1949	0.2319	0.8400	0.4010

Source: author's computations.

As you can see, the arch and garch parameters remain statistically insignificant again. Variable *L.Rptw* positively affects the future volatility. The p value of the residual Twitter variable is 0.039 which means being significant at an even lower level than five percent. Such results are again in line with our general expectation. As we stated before, the variable represents unexpected activity with positive sentiment on the Twitter social network.

Having all the three models estimated, we can move on to comparison of their prediction capability. According to our Diebold-Mariano test statistics, we present the results in the table 4.33. Again, the results of the test show that augmented GARCH models are better for out of sample prediction of volatility in the case of 3M stocks. The best model of our comparison is GARCH with positive Twitter volume variable followed by a model with *L.Rptw*. Plain GARCH model with no additional variables is in the last place in our comparison again. However, the test between the best and the worst model has a p value of 0.8217, i.e. the prediction precision between the models seems not to be statistically different.

The estimation results of the plain ARFIMA regression are presented in the table 4.34. As you can see, *L.ar* representing auto regression term is insignif-

Table 4.33: DM test for GARCH of 3M

GARCH	AGARCH	RGARCH
3	1	2

Source: author's computations.

icant. Difference parameter d is statistically significant and indicates that the series exhibits long memory feature.

Table 4.34: ARFIMA(1,d,0) of 3M

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	4.2700	0.0000
L.ar	0.0723	0.1516	0.4800	0.6330
d	0.2599	0.1238	2.1000	0.0360

Source: author's computations.

When we add Twitter volume variables to the model, we discover that none of our additional variables is statistically significant. Therefore, we choose the one with the lowest p value. Such variable is in our case *L.ptw* representing positive Twitter volume. We present the regression outcomes in the table 4.35. From the table, you can see that our Twitter volume variable is insignificant with p value equal to 0.3070. Regression outcome can be interpreted such that in the case of 3M, the Twitter variable does not affect future volatility. Traditional ARFIMA variable, $L.ar$, is not significant again and d still suggests the long memory of the volatility series.

Table 4.35: Twitter Volume Augmented ARFIMA(1,d,0) of 3M

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0001	4.0600	0.0000
L.ptw	0.0000	0.0000	1.0200	0.3070
L.ar	0.0633	0.1499	0.4200	0.6730
d	0.2589	0.1224	2.1100	0.0340

Source: author's computations.

The last model that we estimate for the General Electric is ARFIMA augmented on the residual Twitter variable. In this case, $L.Rztw$, variable that represent unexpected neutral Twitter volume turns to be least insignificant among tested variables. Again, the insignificance of the residual Twitter variable can be interpreted as the coefficient of the variable was equal to zero. As for the auto regressive term and difference parameter, these variables remain yielding the same practical results as in the previous case.

Table 4.36: Residual Twitter Volume Augmented ARFIMA(1,d,0) of GE

Variable	Coeff.	Std. Err.	t	p
const	0.0000	0.0000	4.9300	0.0000
L.Rztw	0.0000	0.0000	-1.2800	0.1990
L.ar	0.1587	0.1783	0.8900	0.3740
d	0.2181	0.1416	1.5400	0.1240

Source: author's computations.

Comparison of prediction capabilities of the model is presented in table 4.37. The best among competing model is the plain model with no additional Twitter variables. Such results are in line with our expectation, as Twitter volumes turned to be insignificant in our ARFIMA regressions. The second is ARFIMA with plain Twitter volume and the third place in our comparison is captured by the model that includes residual Twitter volume variable *L.Rztw*. In this case, the p value of the best versus the worst model equals to 0.0471.

Table 4.37: DM test for ARFIMA of GE

ARFIMA	AARFIMA	RARFIMA
1	2	3

Source: author's computations.

Having all models estimated, we can move on to compare all six models to discover which one of them has the best volatility prediction performance. Such comparison is presented in the table 4.38.

Table 4.38: DM test for GARCH and ARFIMA of 3M

GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
6	4	5	1	2	3

Source: author's computations.

When we look at the results, we can draw a conclusion that ARFIMA model is more suitable for modeling volatility. In this case, the best among all models is the plain ARFIMA with no additional variables. The test between the best and the worst model yields the p value of 0.0006, i.e. indicates significant difference.

4.1.5 UnitedHealth Group Inc.

UnitedHealth Group Inc. is an American health care company headquartered in Minnesota. The company offers various products and services in the health care and managed care sector. Every year, the company serves more than 70 million individuals. UnitedHealth is listed on the New York Stock Exchange and is a constituent of SP500, SP100 and DJIA stock market indices.

Starting with the simple setting of the GARCH model, we follow the same methodology as previously. Again, we checked all the necessary assumptions that allow us to utilize our modeling frameworks. The results of the plain GARCH model are presented in the table 4.39.

Table 4.39: GARCH Regression of UNH

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0014	0.0006	2.4300	0.0150
Variance Equation				
const	0.0001	0.0000	4.8900	0.0000
L.arch	0.4689	0.1051	4.4600	0.0000
L.garch	-0.0600	0.1228	-0.4900	0.6250

Source: author's computations.

The regression results show that arch parameter is statistically significant with positive effect. Garch effect, however, has a negative sign before the estimation parameter. Nevertheless, the p value of garch effect shows that the estimation parameter is not different from zero. Given the sign of arch and insignificance of garch term, estimation condition that the sum of arch and garch parameters must be and less than one is satisfied.

Another step of our research is to augment the model by Twitter volume variables. In the case of UNH, the statistically significant variable is *L.pol*. We present the estimation results in the table 4.40.

Table 4.40: Augmented GARCH Regression of UNH

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0006	0.0008	0.8000	0.4230
Variance Equation				
const	-8.9260	0.2023	-44.1200	0.0000
L.pol	-1.2123	0.2763	-4.3900	0.0000
L.arch	0.0346	0.0603	0.5800	0.5640
L.garch	0.3817	0.1007	3.7900	0.0000

Source: author's computations.

The regression results show interesting information. Arch and garch parameters switched positions regarding their insignificance. Garch effect, as for the sign, turned positive and significant. Arch effect is insignificant after adding polarity variable. Twitter polarity seems to negatively affect the future volatility. In other words, if the daily polarity of tweets is negative, tomorrow's volatility of returns tends to be higher.

Running the regression with residual Twitter volumes, we obtain interesting results as well. According to our analysis, *L.Rpol* is the significant variable in this case. The regression results are presented in the table 4.41.

Table 4.41: Residual Augmented GARCH Regression of UNH

Variable	Coeff.	Std. Err.	t	p
Mean Equation				
const	0.0006	0.0008	0.6800	0.4940
Variance Equation				
const	-9.6146	0.2536	-37.9100	0.0000
L.Rpol	-1.35902	0.249987	-5.44	0.0000
L.arch	0.0166	0.0517	0.3200	0.7470
L.garch	0.4470	0.0942	4.7500	0.0000

Source: author's computations.

As you can see, the arch and garch parameters remain the same regarding their significance levels. Variable *L.Rpol* negatively affects the future volatility. The p value of the residual Twitter variable is zero which means being significant at even lower level than five percent. Such results are again in line with our general expectation. As we stated before, the variable represents unexpected polarity of daily tweets on the Twitter social network.

Having all the three models estimated, we can move on to the comparison of their prediction capability. According to our Diebold-Mariano test statistics, we present the results in the table 4.42. Again, the results of the test show that the augmented GARCH model is better for out of sample prediction of volatility in the case of UNH stocks. The best model of our comparison is GARCH with no additional explanatory variable. The second position is captured by the model with residual Twitter daily polarity. The model with standard Twitter variable, *L.pol*, is in the last place in our comparison regarding the GARCH family models. In case of the first and the third model, the DM test has a p value of 0.1828.

Table 4.42: DM test for GARCH of UNH

GARCH	AGARCH	RGARCH
1	3	2

Source: author's computations.

The estimation results of the plain ARFIMA regression are presented in the table 4.43. As you can see, *L.ar* representing auto regression term is insignificant. Difference parameter *d* is statistically insignificant as well.

Table 4.43: ARFIMA(1,d,0) of UNH

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	17.4100	0.0000
L.ar	0.3534	0.2733	1.2900	0.1960
d	-0.1186	0.2391	-0.5000	0.6200

Source: author's computations.

Adding the Twitter volume variables to the model, we discover that the model works the best with the variable representing neutral Twitter volume. Such variable is in our case *L.ztw* and has a positive effect on future volatility. We present the regression outcomes in the table 4.44. Traditional ARFIMA variable, *L.ar*, is not significant and model parameter *d* is not different from zero as well.

Table 4.44: Twitter Volume Augmented ARFIMA(1,d,0) of UNH

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	6.9800	0.0000
L.ztw	0.0000	0.0000	0.0000	0.0020
L.ar	0.1191	0.1683	0.7100	0.4790
d	0.0296	0.1345	0.2200	0.8260

Source: author's computations.

The last model that we estimate for the UNH is the ARFIMA augmented on the residual Twitter variable. In this case, *L.Rztw*, a variable that represents unexpected neutral Twitter volume turns to be significant in the analysis. Same as in the previous case, the variable has a positive effect on one day ahead volatility. As for the auto regressive term and difference parameter, these variables remain yielding the same practical results as in the previous case.

Table 4.45: Residual Twitter Volume Augmented ARFIMA(1,d,0) of UNH

Variable	Coeff.	Std. Err.	t	p
const	0.0001	0.0000	10.4300	0.0000
L.Rztw	0.0000	0.0000	0.0000	0.0020
L.ar	0.0486	0.1792	0.2700	0.7860
d	0.0760	0.1385	0.5500	0.5830

Source: author's computations.

Comparison of prediction capabilities of the model is presented in the table 4.46. The best among competing model is the model with additional Twitter variables. The second is the ARFIMA with residual Twitter volume and the third place in our comparison is captured by the model that does not include any additional Twitter volume variables. However, there seems to be no statistical difference between the first and the third model, as the p value of the test is equal to 0.4040.

Table 4.46: DM test for ARFIMA of UNH

ARFIMA	AARFIMA	RARFIMA
3	1	2

Source: author's computations.

Having all models estimated, we can move on to compare all six models to discover which one of them has the best volatility prediction performance. Such comparison is presented in the table 4.47.

Table 4.47: DM test for GARCH and ARFIMA of UNH

GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
4	6	5	3	1	2

Source: author's computations.

From the results, we can draw a conclusion that the ARFIMA model is more suitable for modeling volatility in the case of UNH stocks. In this case, the best among all models is the ARFIMA model that contains additional information in the form of Twitter volume variable. Again, the p value from testing the best and the worst model indicates marginal significance, i.e. 0.0812.

UNH was the last stock that we provided the reader with a thorough description of analysis and results. In the next section, we will take more distant perspective, trying to draw a more general conclusion from the results of the research.

4.1.6 All DJIA stocks

In this section, we aim to provide results on all thirty DJIA stocks. We will comment namely on the signs of Twitter variable effects, trying to conclude what effects they have on the future volatility. The second task will be to draw a conclusion regarding ranking of our four different models as for their prediction performance. Let us begin with analyzing Twitter effects one by one for each of our two models.

The first effect that we aim to describe is the effect represented by variable *L.Rntw*. This variable captures the effect of unexpected Twitter volume with negative sentiment polarity. The results for the stocks, where this variable was significant, are presented in the table 4.48.

Table 4.48: *L.Rntw* results in GARCH model

Stock	Coeff.	p
TRV	-0.2289	0.4520
XOM	0.0133	0.0130
DIS	0.0086	0.0000
WMT	0.0008	0.0000
IBM	0.0065	0.0010
MSFT	0.0045	0.0000

Source: author's computations.

As you can see in the majority of the stocks, the effects of the unexpected negative volume are positive. This means that with an increase of variable *L.Rntw*, you can expect the future volatility to increase as well. Only in the case of TRV stock, you can see the coefficient to be negative. In this case, however, the p value indicates statistical insignificance and thus we can say that the coefficient is equal to zero. The average effect of the variable that we calculated is equal to 0.005621.

The results of ARFIMA model are presented in table 4.49. The effects of coefficients are as for their signs very similar to the GARCH case. The average effect is in this case equal to 0.0000007136 which is basically equal to zero.

Table 4.49: *L.Rntw* results in ARFIMA model

Stock	Coeff.	p
DIS	0.00000212	0.0120
BA	0.000000940	0.3980
PFE	0.000000933	0.0020
IBM	0.000000292	0.0000
MSFT	0.000000223	0.0460

Source: author's computations.

To conclude, the negative unexpected Twitter volume is estimated to positively influence the tomorrow's prediction of volatility. To follow up, we present the results of plain negative Twitter volume variable in the following paragraphs and tables.

For both GARCH and ARFIMA models, the estimated effects are positive. Such results are in line with our general expectation as well as with the results we presented earlier.

Table 4.50: *L.ntw* results in GARCH model

Stock	Coeff.	p
TRV	0.1336	0.0040
XOM	0.0117	0.0260
DIS	0.0065	0.0190
WMT	0.0012	0.0000
GS	-0.0026	0.0570
IBM	0.0059	0.0000
MSFT	0.0044	0.0000

Source: author's computations.

Table 4.51: *L.ntw* results in ARFIMA model

Stock	Coeff.	p
DIS	0.000001700	0.0390
BA	0.000000910	0.2780
GE	0.000000822	0.0500
IBM	0.000000196	0.0000
MSFT	0.000000223	0.0460

Source: author's computations.

Comparing the coefficients between those two models, we can see that in the GARCH model the magnitude of effects is higher. The average effect in GARCH case is equal to 0.0233 whereas in the case of ARFIMA only 0.0000005982.

Following with variable that captures unexpected positive Twitter volume, we discover very similar behavior as previously. We present the results in tables 4.52 and 4.53. Again, this variable exhibits similar behavior in both models, positively affecting future volatility. Moreover, the magnitudes of coefficients in GARCH model are higher compared to ARFIMA model.

Table 4.52: *L.Rptw* results in GARCH model

Stock	Coeff.	p
UTX	-0.0142	0.1660
MMM	0.0219	0.0390
DD	-0.0632	0.1250
AXP	0.0211	0.0010
XOM	0.0056	0.0420
DIS	0.0025	0.0000
BA	-0.0013	0.1080
JPM	0.0010	0.3710

Source: author's computations.

Table 4.53: *L.Rptw* results in ARFIMA model

Stock	Coeff.	p
TRV	0.00000034	0.5320
UTX	-0.000001270	0.0270
AXP	0.000001900	0.0000
DIS	0.000000616	0.0010
MRK	0.000000120	0.0360
T	0.000001240	0.0000

Source: author's computations.

Another interesting fact that can be drawn from the tables is that whenever the effect is estimated to be negative, its p value indicates insignificance. In such cases, we replace the negative coefficient value with zero and calculate the average of the effect. In the GARCH case, the average effect is equal to 0.00638 and in the ARFIMA case, it is 0.0000004343.

Similarly, we present the results of plain positive Twitter volume in the table 4.54 for the GARCH model and in table 4.55 for the ARFIMA model.

Table 4.54: *L.ptw* results in GARCH model

Stock	Coeff.	p
UTX	-0.0127	0.0560
MMM	0.0218	0.0180
DD	-0.0621	0.1460
AXP	0.0143	0.0080
XOM	0.0054	0.0400
DIS	0.0033	0.0000
BA	-0.0016	0.1070
GE	-0.0485	0.0620

Source: author's computations.

Table 4.55: *L.ptw* results in ARFIMA model

Stock	Coeff.	p
TRV	0.000000124	0.8140
UTX	-0.000001260	0.0350
MMM	0.000000320	0.3070
AXP	0.000001620	0.0000
DIS	0.000000347	0.0370
T	0.000001210	0.0000

Source: author's computations.

In general, the effects are almost in most cases positive for both the GARCH and ARFIMA models. However, the UTX stock in ARFIMA model and the GE stock in GARCH model are estimated to have a negative effect while remaining statistically significant. On average, the effect of positive volume in GARCH model is equal to 0.0056 and 0.0000003195 in ARFIMA model. Again, the average magnitude is higher in the GARCH model.

Our next variable to present is the unexpected neutral Twitter volume. The regression results are presented for both the GARCH and ARFIMA models in table 4.56 and 4.57 respectively.

Table 4.56: *L.Rztw* results in GARCH model

Stock	Coeff.	p
CAT	0.0065	0.0000
VZ	0.0149	0.0040
INTC	0.0066	0.0000
GS	-0.0077	0.0830

Source: author's computations.

Table 4.57: *L.Rztw* results in ARFIMA model

Stock	Coeff.	p
UNH	0.000001360	0.0020
MMM	-0.000000223	0.1990
DD	0.000000249	0.3540
CVX	0.000000128	0.1490
VZ	0.000000388	0.1910
GE	0.000000173	0.0060
WMT	-0.00000003	0.1760
INTC	0.000000173	0.0170
JPM	0.000000086	0.0030

Source: author's computations.

In the GARCH case, we can observe that the effects are positive in all stocks except GS. In the ARFIMA case, we can see quite a lot insignificant stocks with p values higher than our selected benchmark of five percent. The average effect in the case of GARCH is equal to 0.00699 and 0.0000001992 in the case of ARFIMA. In both cases, we included insignificant variables as zero for calculation of the average.

Another variable we present is the unexpected Twitter volume without any sentiment. Simply said, the total number of unexpected tweets regardless of their polarity. The results can be found in the tables 4.58 and 4.59. The results are in line with previous cases where positive effect dominated among presented stocks. In few cases of the GARCH model, the effects are negative and significant. The resulting average effect is, however, still positive and equal to 0.00078. Taking a closer look at the ARFIMA results we can observe basically the same. The average effect is positive but its magnitude is 0.00000007082 which is smaller compared to the GARCH case.

Table 4.58: *L.Rtw* results in GARCH model

Stock	Coeff.	p
TRV	-0.0249	0.4400
UTX	0.0044	0.2240
PG	0.0041	0.1540
NKE	0.0037	0.0000
CVX	0.0055	0.0150
HD	-0.0005	0.0380
V	0.0015	0.0490
VZ	-0.0101	0.0200
KO	0.0015	0.0620
MCD	0.0034	0.0000
CSCO	0.0037	0.0000
PFE	0.0011	0.0000
T	0.0023	0.0040
MSFT	0.0005	0.0320

Source: author's computations.

Table 4.59: *L.Rtw* results in ARFIMA model

Stock	Coeff.	p
UTX	0.000000437	0.0630
PG	0.000000160	0.0320
VZ	-0.000000269	0.1760
MCD	0.000000054	0.0080
CSCO	0.000000140	0.0000

Source: author's computations.

The results of plain Twitter volume variable are presented in the same manner and can be found in the tables 4.58 and 4.61. The plain Twitter volume exhibits some divergence from the results we obtained in the variables we described earlier. It is not obvious for the first sight, but in the GARCH case, the magnitude of TRV variable coefficient is larger than the sum of all remaining significant coefficients. Given the fact that the TRV coefficient is negative, the average effect is equal to -0.000445 for the plain Twitter volume variable. The ARFIMA model exhibits similar behavior to previous cases with the average effect equal to 0.00000009194.

Table 4.60: *L.tw* results in GARCH model

Stock	Coeff.	p
TRV	-0.0242	0.0010
UTX	0.0048	0.1460
PG	0.0030	0.1180
NKE	0.0033	0.0000
CVX	0.0043	0.0200
HD	0.0005	0.6350
V	0.0006	0.0020
VZ	-0.0082	0.2200
KO	0.0005	0.0000
MCD	0.0021	0.0020
CSCO	0.0029	0.0010
PFE	0.0011	0.0000
T	0.0020	0.0090
JPM	0.0001	0.6110
MSFT	0.0007	0.0060

Source: author's computations.

Table 4.61: *L.tw* results in ARFIMA model

Stock	Coeff.	p
UTX	0.000000402	0.0700
PG	0.000000166	0.0120
HD	0.000000116	0.0080
MCD	0.000000048	0.0210
CSCO	0.000000130	0.0000

Source: author's computations.

The last variable set from our list are the variables that capture the overall daily polarity of tweets. These variables are represented by *L.Rp* for unexpected polarity and *L.p* for the simple daily polarity of tweets. Beginning with the unexpected polarity, we present the results in the tables 4.62 and 4.65. As you can see, the residual polarity in the GARCH model does not exhibit uniform behavior. We obtain positive, negative and neutral coefficients across various stocks. Average of these coefficients is equal to -0.2095. If we take a look at the ARFIMA model, we see the effect to be insignificant for both stocks, the average is thus estimated to be zero.

Table 4.62: *L.Rp* results in GARCH model

Stock	Coeff.	p
UNH	-1.3590	0.0000
NKE	-0.5703	0.0420
JNJ	-0.2345	0.2100
MRK	0.6718	0.0020
GE	2.3722	0.0560
GS	0.0000	0.0650

Source: author's computations.

Table 4.63: *L.pt* results in ARFIMA model

Stock	Coeff.	p
JNJ	0.0000	0.3190
V	0.0000	0.5510

Source: author's computations.

Continuing with plain Twitter polarity, we present the results in the tables 4.64 and 4.63. Compared to the previous case, we obtain the Twitter polarity to be significant for all stocks in the GARCH model. As for the effect sign, we predominantly obtain negative coefficients. Interpretation of such effect is quite intuitive. If there is the negative sentiment on the social media, the volatility tends to increase as multiplication of the variable with its coefficient yields positive number and thus higher expected volatility. The average of the coefficients is equal to -0.2930. In the ARFIMA models, we obtain significance for the JNJ stock. Same as previously, the effect is negative thus having the same interpretation.

Having provided the reader with results sorted by Twitter variable type, we can conclude that in most cases both expected and unexpected volumes tend to positively affect the future volatility. In case of polarity, the effect is in line with our expectation as well, connecting negative sentiment with an increase in volatility. The last thing to present in our research are the overall results of the model comparison across all stocks on which we performed our analysis.

Table 4.64: *L.p* results in GARCH model

Stock	Coeff.	p
UNH	-1.21203	0.0000
NKE	-0.46910	0.0370
JNJ	-0.37060	0.0320
MRK	0.58660	0.0010
GS	-0.00003	0.0460

Source: author's computations.

Table 4.65: *L.Rpt* results in ARFIMA model

Stock	Coeff.	p
JNJ	-0.00002	0.0270
V	-0.00007	0.3110

Source: author's computations.

The table 4.66 reveals a number of interesting facts. If we compare those two model families, we can state that the ARFIMA models perform better more frequently compared to the GARCH models. For example, the ARFIMA model captured all the first three places in a particular stock in total 19 times, whereas the GARCH model only once in the case of the BA stock. To compare the prediction ability based on whether the model uses Twitter variables or not, we calculated the average score for each model. The average score is presented in the table 4.67. The average scores also suggest better performance of the ARFIMA family compared to the GARCH family. Within the ARFIMA family, we can clearly see that the model with residual Twitter volumes has the best prediction capabilities. The second place is captured by the ARFIMA model with no additional variables followed by the model containing plain Twitter volumes. Within the family of the GARCH model, plain GARCH setting seems to be more suitable than settings with additional Twitter variables. The p values from testing the best versus the worst models are provided in the table in the appendix. From the the table, it is obvious that in total of 17 stocks, there is a significant difference between the worst and the best model.

Table 4.66: Overall Model Comparison

Stock	GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
TRV	4	6	5	2	3	1
UNH	4	6	5	3	1	2
UTX	4	6	5	1	3	2
MMM	6	4	5	1	2	3
DD	5	4	6	3	2	1
AXP	3	5	6	4	1	2
PG	3	6	2	5	4	1
NKE	2	6	5	1	4	3
CVX	4	5	6	3	2	1
HD	4	6	5	2	3	1
CAT	4	5	6	2	1	3
JNJ	4	5	6	1	3	2
V	1	3	5	2	6	4
VZ	5	6	4	1	3	2
KO	5	6	4	3	1	2
MCD	6	4	5	3	2	1
XOM	6	5	4	3	1	2
DIS	6	4	5	3	2	1
BA	2	3	1	5	6	4
MRK	4	5	6	3	2	1
CSCO	4	5	6	2	3	1
GE	4	2	1	5	6	3
WMT	6	4	5	1	2	3
INTC	4	5	6	3	1	2
PFE	2	6	5	1	4	3
T	2	6	5	1	4	3
GS	1	3	2	2	1	3
IBM	6	1	2	5	4	3
JPM	6	4	5	1	2	3
MSFT	6	4	1	5	3	2

Source: author's computations.

Table 4.67: Average Score of GARCH and ARFIMA Models

GARCH	AGARCH	RGARCH	ARFIMA	AARFIMA	RARFIMA
4.1000	4.6670	4.4667	2.5670	2.7330	2.1670

Source: author's computations.

4.2 Summary

In our research, we have conducted analysis on all thirty components of the DJIA stock market index. Selection of the subject of analysis was based mainly on Twitter data availability. We worked with a rather short time frame in our analysis, starting on 31/5/2013 and ending on 18/9/2014. For this period of time, we collected price data for every particular component of the stock index and also data generated on the Twitter social network.

The main focus of our analysis was to make volatility predictions using models that utilize Twitter components to predict the future volatility. The family of models consisted of two different model types, these are, GARCH and ARFIMA models. For each group, we further created three different models that differ by employing Twitter volume variables. Thus, we have created six different model types that we conducted the research on.

The research has revealed a number of interesting outcomes regarding prediction abilities as well as effects of Twitter variables on the future volatility. As for the suitability of modeling framework, ARFIMA model has outperformed GARCH on our sample of stock during the time frame of our analysis. ARFIMA models were better for prediction compared to the entire group of GARCH models. ARFIMA outperformed GARCH in total 19 times whereas GARCH outperformed ARFIMA only once. For the remaining 10 stocks, we obtained mixed results of prediction performance. To quantify the performance, we have calculated the average score for each model. The scores suggested that ARFIMA models are on average better than GARCH for our sample. From the score, we can also suggest that the models that employ residual Twitter variables are well suited for prediction of future volatility using ARFIMA model.

Regarding the effects of additional Twitter variables, we can generally say that the higher the Twitter volumes, the higher the stock return volatility. If we focus on daily Twitter polarity, the results suggest a reverse effect on the future volatility. Such effect simply means that if there is negative sentiment, tomorrow's volatility tends to be higher, whereas if the sentiment is positive, the future volatility shrinks regardless of the volatility modeling framework.

Chapter 5

Conclusion

Although stock return volatility is quite extensively researched topic, the interaction of users on the Internet and social media generates data of unprecedented scale and content of information. The rise of usage of mobile devices and unlimited access to the Internet will further continue to supply us with data suitable for utilization for various topics including analysis of financial markets.

In our work, we analyzed DJIA stock in relation to data generated on the Twitter social media network. Our main task was to discover the relationship between Twitter interaction and stock return volatility. The analysis was performed using different modeling frameworks that were later compared performance wise to provide us with the suggestion of the most suitable framework.

Given the fact that we use data that are generated by many users on the Twitter, we can put our analysis into theoretical framework related to investor attention and imperfect information. Up to date, it was very hard to measure investor's attention as no direct measure of such variable was available. Researchers tried to use various proxy variables such as unusual trading volume, news or extreme returns. All these variables are somehow connected to attention grabbing events, but none of them is precise. The main drawback of these proxy variables is the assumption that investors pay attention and behave according to these variables. Imagine articles in the press, if you proxy attention by a number of articles, you implicitly assume that the articles are read by investors. Such assumption lacks scale and precision or even does not have to be true at all. On the other hand, social media give us more precise proxy for investor's attention, as the data are generated by user interaction. Given the quality of the internet data, their econometric usage evolves rapidly. For example, Ranco *et al.* (2015) performed analysis regarding the relationship between Twitter volume variables and behavior of financial markets. The research has revealed the statistical relationship between Twitter sentiment and extreme returns during the peaks of Twitter volume. In general, the main implication of the research was that Twitter sentiment polarity is able to predict the direction of cumulative abnormal returns of the DJIA stocks.

In our analysis, we were rather interested in relating Twitter to stock return volatility. In short, we generated two classes of Twitter variables. First, the plain Twitter volume variables representing a number of tweets with the perspective sentiment. Second, unexpected Twitter volumes representing abnormal Twitter volumes generated as residuals from Twitter auto regression models. The entire analysis was performed on all thirty Dow Jones Industrial Average stocks during the period of 15 months starting in 2013 and ending in 2014. As our task was to compare the models and find the right one with the best prediction performance, we had to select a comparison benchmark to proxy the realized volatility. Taking into account our volatility modeling frameworks, we decided to use Garman Klass proxy for volatility.

As we stated before, we were mainly interested in the effects of Twitter variables and prediction performance of the models. Let us start with a description of the Twitter effects on future volatility. In general, we can say that almost all the variables have positive effects on tomorrow's volatility. The interpretation is straightforward as well, the higher the number of tweets regardless their polarity, the higher the volatility on the other trading day. The only exception was the variables representing the overall daily polarity as defined in section 3.2. The effect of the polarity variables was negative. Such results mean that if there is the negative daily polarity of tweets, the tomorrow's volatility increases. If there is positive daily polarity, the tomorrow's effect is on the other hand negative, thus, decreases the volatility. Such effects might have a connection to the so-called leverage effect in the case that negative tweets would also cause negative price motion on the markets. The second task was to discover which model is the best among the models we compared. For such reason, we used various testing procedures described in the earlier sections. The results of the comparison are quite easy to interpret. For each stock, we just ordered the models according to results of the test performed on mutual pairs. From this procedure, we obtained the final order of the models. To generalize the results, we calculated average order for each model across all stocks. The results are provided in the table 4.67, where you can see that the ARFIMA model augmented on residual Twitter variables was on average the best model. In general, the model comparison reveals that the ARFIMA model performs much better than the GARCH in the time frame of our analysis. To be honest, we should mention that the time series we performed our analysis on was rather a short one with numbers of non trading days. Such data structure was caused by the relatively difficult availability of Twitter data.

Finally, we can suggest ways for possible improvement and development of the analysis regarding Twitter data and volatility. Firstly, we would suggest using longer periods of time to perform the analysis on. There is a possibility that GARCH models would become more efficient on the longer time periods. Secondly, it would be very interesting to employ the Twitter data on high-frequency financial markets data. Such structure would allow us to discover immediate reactions of financial markets to Twitter user interactions. Thirdly, analysis of volatility around periods with extreme Twitter activity. Such analysis would be most likely connected to surprising or unexpected events regarding particular stocks. Lastly, the analysis performed during the period of the economic downturn would be interesting as well.

Bibliography

Ali F. Darrat, Shafiqur Rahman, M. Z. (2003): "Intraday trading volume and return volatility of the DJIA stocks: A note." *Journal of Banking and Finance* **(27)**: pp. 235–243.

Barber, B. M. & T. Odean (2008): "All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors." *Review of Financial Studies* **21(2)**: pp. 785–818.

Black, F. & M. Scholes (1973): "The pricing of options and corporate liabilities." *Journal of political economy* **81(3)**: pp. 637–654.

Bo Quian, K. R. (2007): "Stock market prediction with multiple classifiers." *Applied Inteligence* **26**: pp. 25–33.

Bollen, J., H. Mao, & X. Zeng (2011): "Twitter mood predicts the stock market." *Journal of computational science* **2(1)**: pp. 1–8.

Bollerslev, T. (1986): "Generalizet autoregresive conditional heteroscedastic." *Journal of Econometrics* **31**: pp. 307–327.

Bouchaud, J.-P. (2009): "The (unfortunate) complexity of the economy." *Physics World* **22(04)**: p. 28.

Box, G., G. Jenkins, & G. Reinsel (1994): *Time Series Analysis: Forecasting and Control*. Forecasting and Control Series. Prentice Hall.

Da, Z., J. Engelberg, & P. Gao (2011): "In search of attention." *The Journal of Finance* **66(5)**: pp. 1461–1499.

Dickey, D. A. & W. A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American statistical association* **74(366a)**: pp. 427–431.

DIEBOLD, F. X. & R. S. MARIANO (2002): "Comparing predictive accuracy." *Journal of Business & economic statistics* **20(1)**: pp. 134–144.

ENGLE, R. (1982): "Autoregresive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation." *Econometrica* **50(4)**: pp. 987–1007.

FAMA, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* **25**: pp. 383–417.

FRANK, E. & M. HALL (2001): "A simple approach to ordinal classification." In "European Conference on Machine Learning," pp. 145–156. Springer.

GARMAN, M. B. & M. J. KLASS (1980): "On the estimation of security price volatilities from historical data." *Journal of business* pp. 67–78.

GIDOFALVI, G. & C. ELKAN (2001): "Using news articles to predict stock price movements." *Department of Computer Science and Engineering, University of California, San Diego* .

GRUHL, D., R. GUHA, R. KUMAR, J. NOVAK, & A. TOMKINS (2005): "The predictive power of online chatter." In "Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining," pp. 78–87. ACM.

HANSEN, P. R. & A. LUNDE (2005): "A forecast comparison of volatility models: does anything beat a garch (1, 1)?" *Journal of applied econometrics* **20(7)**: pp. 873–889.

JUCHELKA, T. (2014): "Trading volume and volatility in the us stock markets." *Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Supervisor: PhDr. Boril Sopov, Msc., LL.M.* p. 42.

KAHNEMAN, D. (1973): *Attention and effort.* Prentice-Hall series in experimental psychology. Prentice-Hall.

KWIATKOWSKI, D., P. C. PHILLIPS, P. SCHMIDT, & Y. SHIN (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of econometrics* **54(1-3)**: pp. 159–178.

Liu, Y., X. Huang, A. An, & X. Yu (2007): "Arsa: a sentiment-aware model for predicting sales performance using blogs." In "Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval," pp. 607–614. ACM.

Merton, R. C. (1987): "A simple model of capital market equilibrium with incomplete information." *The journal of finance* **42(3)**: pp. 483–510.

Mishne, G., M. De Rijke *et al.* (2006): "Capturing global mood levels using blog posts." In "AAAI spring symposium: computational approaches to analyzing weblogs," pp. 145–152.

Nelson, D. (1991): "Conditional heteroscedasticity in asset returns: A new approach." *Econometrica* **59(2)**: pp. 347–370.

Ranco, G., D. Aleksovski, G. Caldarelli, M. Grčar, & I. Mozetič (2015): "The effects of twitter sentiment on stock price returns." *PloS one* **10(9)**: p. e0138441.

Robert R. Prechter, J. & W. D. Parker (2007): "The Financial/Economic Dichotomy in Social Behavioral Dynamics: The Socionomic Perspective." *The Journal Behavioral of Finance* **8**: pp. 84–108.

Ruey S. Tsay, W. (2005): *Analysis of Financial Series.* Chicago: John Wiley & Sons, Inc.

Schumaker, R. P. & H. Chen (2009): "Textual analysis of stock market prediction using breaking financial news: The azfin text system." *ACM Transactions on Information Systems (TOIS)* **27(2)**: p. 12.

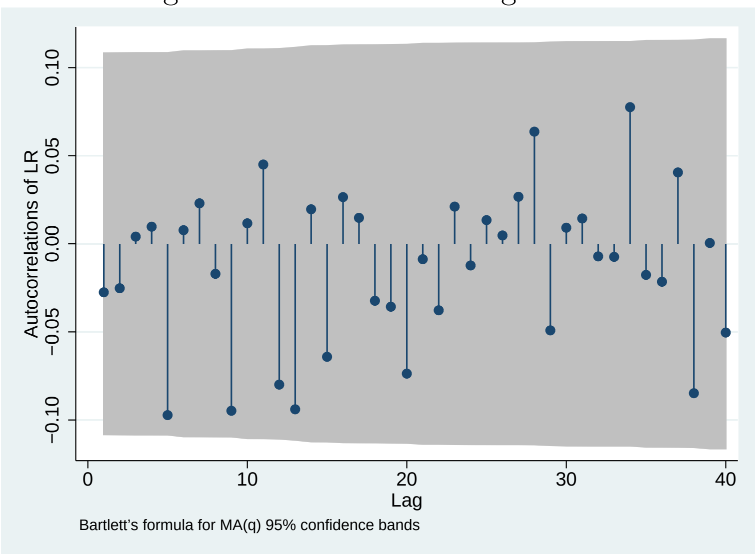
Sims, C. A. (2003): "Implications of rational inattention." *Journal of monetary Economics* **50(3)**: pp. 665–690.

Vapnik, V. (2013): *The nature of statistical learning theory.* Springer science & business media.

Appendix A

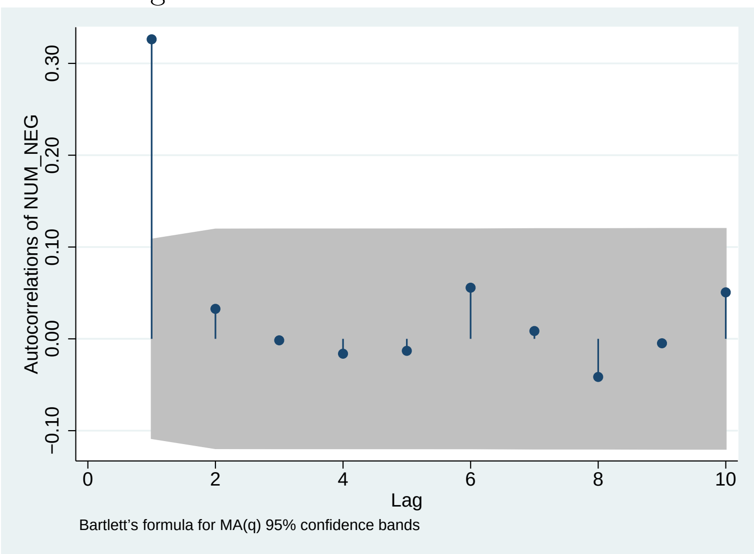
Figures

Figure A.1: ACF of DIS log returns



Source: author's computations.

Figure A.2: ACF of DIS ntw Variable



Source: author's computations.

Appendix B

Tables

Table B.1: Lagrange-Multiplier test DIS

Lag	chi2	df	p
1	0.0320	1	0.8576
2	0.5110	2	0.7746
3	13.3540	3	0.0039
4	13.5870	4	0.0087
5	14.2290	5	0.0142
6	15.5300	6	0.0165
7	16.1350	7	0.0239
8	16.0190	8	0.0421
9	17.1390	9	0.0466
10	17.6590	10	0.0610
11	18.1000	11	0.0793
12	19.9200	12	0.0686
13	20.2350	13	0.0895
14	19.3900	14	0.1506
15	19.5580	15	0.1895
16	22.0280	16	0.1423
17	37.2120	17	0.0031
18	37.4070	18	0.0046
19	37.6860	19	0.0065
20	38.4580	20	0.0078
21	38.6550	21	0.0108
22	38.6720	22	0.0154
23	39.0840	23	0.0194
24	40.7380	24	0.0178
25	40.8080	25	0.0240
26	40.5080	26	0.0348
27	40.3590	27	0.0474
28	44.2280	28	0.0264
29	45.4640	29	0.0265
30	44.9550	30	0.0390
31	46.5600	31	0.0360
32	46.5100	32	0.0469
33	46.6430	33	0.0581
34	47.4970	34	0.0620
35	49.2380	35	0.0558
36	50.3050	36	0.0571
37	50.3850	37	0.0700
38	51.1230	38	0.0757
39	52.6510	39	0.0710
40	52.6560	40	0.0867

Source: author's computations.

Table B.2: Return Summary Statistics

Ticker	Mean	Minimum	Maximum	St. Dev.
TRV	0.000414044	-0.03841	0.022228	0.008703
UNH	0.001248148	-0.0507	0.066515	0.011815
UTX	0.000448069	-0.0332	0.029769	0.009429
MMM	0.000968739	-0.0452	0.033564	0.008670
DD	0.000879947	-0.0332	0.052583	0.010074
AXP	0.000593695	-0.0377	0.052190	0.011235
PG	0.000412021	-0.0302	0.029868	0.007873
NKE	0.000972766	-0.0512	0.048467	0.011168
CVX	0.000111872	-0.0409	0.024731	0.008457
HD	0.000544174	-0.0306	0.054785	0.010525
CAT	0.000658525	-0.0606	0.059417	0.010821
JNJ	0.000713538	-0.0259	0.022897	0.008151
V	0.000796722	-0.0704	0.044594	0.012588
VZ	7.50575E-05	-0.0428	0.034921	0.010376
KO	0.000188573	-0.0319	0.037707	0.008169
MCD	-0.00008563	-0.0263	0.037605	0.007098
XOM	0.000281205	-0.0421	0.031418	0.008788
DIS	0.001259170	-0.0368	0.052954	0.011265
BA	0.000903585	-0.0515	0.054253	0.012819
MRK	0.000818016	-0.0262	0.064755	0.010252
CSCO	0.000354578	-0.1087	0.060526	0.012248
GE	0.000444596	-0.0336	0.046108	0.009736
WMT	0.000050065	-0.0271	0.023963	0.007076
INTC	0.001234882	-0.0543	0.092026	0.012789
PFE	0.000394627	-0.0299	0.041612	0.009712
T	0.000044055	-0.0408	0.034577	0.009171
GS	0.000457930	-0.0377	0.040205	0.011308
IBM	-0.00009755	-0.0616	0.036228	0.009969
JPM	0.000344317	-0.0364	0.044522	0.011155
MSFT	0.001022211	-0.1142	0.073479	0.014511

Source: author's computations.

Table B.3: P Values for DM Test

Ticker	p value
TRV	0.0494
UNH	0.0812
UTX	0.0851
MMM	0.0006
DD	0.0233
AXP	0.0565
PG	0.0000
NKE	0.2270
CVX	0.1752
HD	0.1591
CAT	0.2487
JNJ	0.0024
V	0.0000
VZ	0.0000
KO	0.0000
MCD	0.1380
XOM	0.0000
DIS	0.0398
BA	0.0492
MRK	0.1504
CSCO	0.1121
GE	0.0120
WMT	0.0000
INTC	0.2015
PFE	0.1546
T	0.2446
GS	0.0019
IBM	0.0000
JPM	0.0000
MSFT	0.0000

Source: author's computations.