The impact of tweets on equity trading on the NASDAQ, LSE and JSE

A research report submitted by

Ian Gatley

Student number: 728503

Tel: 083 448 8140

Email: gatley.ig@gmail.com

Supervisor:

Thabang Mokoaleli-Mokoteli

Wits Business School

February 2015

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract

Social media platforms such as Twitter have become embedded in almost every aspect of modern life. Twitter provides a constant and instant stream of news, weather and social updates. The purpose of this report was to ascertain if this stream of data correlated against fluctuations in stock prices can provide valuable insight to stock market traders trading on the JSE. Five months' worth of data was collected for the period between May and September 2014. This data comprised of company specific tweets from 70 companies trading on the JSE, LSE and NASDAQ exchanges; correlating this data against daily price fluctuations and trading volume confirmed that fluctuations twitter activity correlated with days of abnormal trading activity. It can be therefore be suggested that the volume, sentiment and immediacy of information on twitter can provide valuable insight when making trading decisions.

Chapter 1

1.1 Introduction

The purpose of this chapter is to introduce the concepts of and describe how this thesis adds to the existing body of behavioural finance knowledge. The chapter is organised as follows. Section 1.2 presents the context of the study, section 1.3 identifies the research problems, section 1.4 covers the hypothesis used to test the research problems, section 1.5 covers the significance of the study and finally section 1.6 covers the organisation of the remaining chapters of the thesis.

1.2 Context of Study

In recent years, a concept of social trading has emerged as a new way of assisting investors in making trading decisions (Xu, 2012). These social trading platforms provide subscribers with an aggregated view of both market related information and crowd sourced sentiment data. The concept suggests that investors can gain a profit advantage by making trading decisions based on crowd sourced information. This information can be gathered from online communities, microblogging sites such as Twitter and by learning from other online traders. (Xu, 2012). Web sites such as www.seekingAlpha.com and www.stocktwits.com are crowd sourced platforms that aggregate market related information for the purpose of aiding investors in making decisions. One such site www.etoro.com allows subscribers to follow and "copy" trades of other site members. Based on statistics published on their site Tsype (2013) stated that 64% of trades are copied from other traders and that copied trades are on average more profitable than original trades.

A number of related studies (e.g. Xu (2012) and Rao & Srivastava (2013)) have investigated the impact of social media on stock price movement. These studies proved correlation between microblogging activity, news sentiment and price movement. Furthermore, by using data mining algorithms, these studies accurately predicted future price movements.

The key technological and social development that has led to the proliferation of these types of services is Twitter (www.Twitter.com). Twitter provides a microblogging service allowing subscribers to send short messages of 140 characters or less. As of July 2014 Twitter had more than 270 million subscribers sending approximately 500 million tweets per day with 77% of their subscriber base residing outside of the US Twitter.com (2014). This platform can be used for a number of different activities such as; a social and collaboration tool, education, emergency response and Marketing (Kirkpatrick, 2009) (Bruns & Liang, 2012) (Case & King, 2011).

A domain more closely related to this study that has been impacted by Twitter is journalism (Ahmad, 2010). Twitter has created a platform for crowd sourced or citizen journalism (Bruns & Highfield, 2012). The platform has also created a new medium for mainstream journalists, thus reducing the time taken for journalists, including financial and business journalists to break or publish a news story (Newman, 2009).

In order to add additional context, tweets can be augmented with social sentiment (Go, Huang, & Bhayani, 2009). This process involves assigning a polarity score to identify the positive, negative or neutral sentiment implied by the tweet author (Kouloumpis, Wilson, & Moore, 2011).

Wang, Kosinski, Stillwell, & Rust (2014) findings state that social media has an impact on the stock market movement and violates the efficient market hypothesis (EMH). This hypothesis states that the current price of a traded entity already includes all information relating to the entity (Fama, 1970). The violation of EMH supports the behavioural finance model which states that emotions affect the behaviour of investors and how they make their decisions (Barberis & Thaler, 2003)

The aim of this research is to perform a comparative analysis of the correlation between the volume and sentiment of Twitter messages and the share price of stocks listed on the South African, US and London exchanges. The idea is to establish whether there is a differential impact of Twitter messages on the stock prices of developed and emerging stock markets such as the JSE.

1.3 Research Problem

Behavioural finance indicates that stock market movement is not influenced by stock market and firm fundamentals alone but by sentiment as well (Shleifer & Summers, 1990). According to Camelia (2012) this sentiment can be influenced by a number of factors, including how different investors analyse the same piece of information and a range prevailing biases potentially affecting the investor (Chen, Kim, Nofsinger, & Rui, 2007). The extant literature confirms that the social media such as Twitter has an impact on the stock movement. For example, Zhang & Skiena (2010) found that it was possible to design a trading strategy using only user generated content. Furthermore, Oh & Sheng (2011) found a predictive influence of microblogging sentiment on market returns.

Various studies including Xu (2012) and Rao & Srivastava (2013) have investigated the impact of social media on stock price movement. These studies proved correlations between microblogging activity, news sentiment and price movement. Twitter provides a good platform to investigate the correlations between microblogging activity, news sentiment and price movement.

Twitter use has proliferated worldwide and it is estimated that the service currently has approximately 270 million users. Of this user base approximately 77% are from outside the USA (Twitter.com, 2014). The problem is that little research, if any, has been done to investigate whether the use of Twitter has a similar impact on the stock price in other markets other than in the USA. Furthermore there is also a dearth of research that compares the impact of Twitter on the JSE, US and other exchanges. Do tweets related to listed companies accurately correlate to or predict the future stock market activity?

1.4 Research Objectives

The research objectives were stated as follows:

- To investigate the relationship between the volume of tweets and stock market trade volume and trade volatility
- To investigate the relationship between the sentiment of tweets and stock price movement.

- To investigate if markets are responding to news on Twitter or if the inverse is occurring where Twitter is only responding to existing market activity.

1.5 Significance of the study

This study forms part of the body of knowledge of behavioural finance and provides guidance to traders to better understand the impact of social sentiment on the prices of traded entities. The outcomes of this study may also provide insight to intraday traders looking for short term methods of outperforming the market. The results could also potentially assist organisations that make use of or write automated trading algorithms, as these algorithms could benefit from the incorporation of social sentiment. Algorithms of this nature already consume news and other sources of information regarding the economy, companies and sentiment (Groß-Klußmann & Hautsch, 2011).

This research fills the gap by establishing whether the same predictive nature of tweets found in developed markets can be applied to emerging markets such as South Africa.

1.6 Organization of the thesis

The thesis is organised as follows. Chapter 2 presents the extant literature related to the research topic. Chapter 3 discusses the research methodology adopted for this research including data, data sources and research design. Chapter 4 presents the research findings and chapter 5 discusses the results and provides a conclusion to the research.

Chapter summary

This chapter has provided an overview of how this thesis is structured, its context and purpose. Chapter 2 comprises of a literature review of the current body of knowledge.

2 LITERATURE REVIEW

2.1 Introduction

This chapter reviews the extant literature related to the efficient market hypothesis, investment psychology, literature related to computer based natural language processing and text analytics. The chapter is organised as follows: Section 2.2 presents the literature on efficient market hypothesis. Section 2.3 discusses the psychology of investing. Section 2.4 explains natural language processing and text analytics. Section 2.5 addresses the impact of news on stock prices and Section 2.6 summarises the comparative studies conducted in this field.

2.2 The Efficient Market Hypothesis

The principle of the efficient market hypothesis (EMH) as proposed by Malkiel & Fama (1970) dictates that the current price of a traded entity includes all available information and that this information is included in the prevailing price both rationally and instantaneously. Therefore, the more efficient the market becomes, the more random future price movements will be and the less likely it is to make above average profits without taking above average risks.

The strongest form of efficient market hypothesis dictates that all information including privileged information not in the public domain is already factored into the trading price. Assuming this model was possible there would be no advantage to be gained by insider trading. Against this hypothesis Jaffe (1974) conducted empirical research of 200 firms over a six year period from 1962-1968. His conclusion was that insiders will outperform the general market. This research was corroborated by Finnerty (1976) who concluded that insiders were able to identify both upward and downward swings within their own companies.

The intermediate form of semi hard efficiency dictates that all market related information and any publicly available information is already included in the price. To this theory and using stock splits as a proxy for new information Fama, Fisher, Jensen, & Roll (1969) were able to determine that markets react almost immediately to news of a split, thus proving the existence of the EMH. As further proof Fama, Fisher, Jensen, & Roll (1969) concluded that it would not be possible to make abnormal profits based on share splits without inside information, this conclusion again proves the hypothesis of semi hard efficiency. Further evidence of this form has been proven to exist in certain circumstances, examples include research conducted by Groenewold & Kang (1993) analysing the Australian stock market as well as research undertaken by Khan & Ikram (2010) analysing the Indian stock market.

The weak form of the efficient market hypothesis states that the current price includes all stock market related information. However Bondt & Thaler (1985) concluded by comparing previous “winning portfolio’s” with “losing portfolio’s” that the “losing stocks” outperformed “winning stocks” and that in line with the theory of framed decision stock markets do have a tendency to overreact to news (Tversky & Kahneman, 1981). Within the African context, Smith & Dyakova (2014) compared the predictability of eight exchanges and determined that the weak form of EMH does apply in South Africa, Egypt and Tunisia.

In contrast to the theory proposed by Malkiel & Fama (1970), Basu (1977) conducted empirical research comparing price to earnings ratios of firms against share performance. He concluded that based on the correlation of these two factors that a degree of predictability of future price was possible thus contradicting the theory of the efficient market hypothesis.

In 2003, over thirty years after the initial research was published Malkiel (2003) revisited the topic and stated that investors may be subject to certain emotional biases such as the “band wagon effect”, whereby investors are drawn into a market after watching a rising price. However, Malkiel (2003) still defended the efficient market hypothesis stating that markets were efficient and not predictable.

Further limitations of the efficient market hypothesis as raised by Ball (2009) include that the efficient market hypothesis does not deal with the availability or source of information, the assumption is that all information is immediately available to all investors and can be immediately consumed and understood equally by all investors. Other issues highlighted by Ball (2009) are that the theory assumes there are no costs associated with investing or administering the market.

2.3 Psychology of investing

In contrast to the hypothesis proposed by Malkiel & Fama (1970) a number of authors have suggested that markets are never truly efficient and investors are subject to a number of cognitive and emotional biases (Nofsinger, 2005). The study of these biases provides a list of factors that may prohibit individuals from making more informed decisions (Tversky & Kahneman, 1981). These biases in turn impact investor sentiment. With sentiment in this context referring to an investor's appetite for risk (Edelen, Marcus, & Tehranian, 2010). This model of behavioural finance contradicts the theory proposed by Friedman (1953) who deducted that any irrational market position will be corrected by rational investors thus maintaining the EMH status quo.

A number of behavioural and cognitive biases have been identified that may impact how investors make decisions. These biases include the fact that Investors are prone to the "Herding effect" whereby prices will continue to move in a certain direction without any significant news (Nofsinger & Sias, 1999). The price movement is perpetuated as investors follow the current direction of trade. This hypothesis was corroborated by Griffin, Harris, & Topaloglu (2003) who concluded that institutional investors are more likely to invest in securities that performed well the previous day.

Overconfidence is a bias documented by Daniel, Hirshleifer, & Subrahmanyam (1998) that may lead investors to overestimate their ability to process new information , this bias in turn has been proven to lead to more frequent trades and lower returns (Odean, 1999). The degree to which investors are prone to overconfidence also varies between nations with Chen, Kim, Nofsinger, & Rui (2007) highlighting how Chinese investors are more prone to overconfidence than their American counterparts. There is also a gender component as Barber & Odean (2001) proved how men are more prone to this effect than woman.

Home bias is the propensity for investors to gravitate towards local equities as opposed to a more international diversification (Lütje & Menkhoff, 2007). This bias has been shown to exist equally in institutional as well as individual investors. The contributing factors towards home bias highlighted by Lütje & Menkhoff (2007) are linked back to behavioural factors such as overconfidence and a greater degree of optimism for local assets. Part of the reason for home bias also includes a greater volume of information on local assets.

The disposition affect is a phenomenon whereby investors will hold onto losing stocks for too long and sell winning stocks too early. This trend was first identified by Schlarbaum, Lewellen, & Lease (1980). However, research that also shows that investor experience helps mitigate these effects (Da Costa Jr, Goulart, Cupertino, Macedo Jr, & Da Silva, 2013)

Chamorro-Premuzic (2007) concluded that inherit psychological biases such as narcissism have also been proven to impact investor decisions and genetics has also been proven to play a role (Coates, Gurnell, & Rustichini, 2009).

Both gender and nationality have been proven to have impact on how investors trade. Beckmann & Menkhoff (2008) concluded that investors follow gender stereotypes whereby women are less likely to be over confident or competitive and are also more risk averse. In analysing cultural differences Ji, Zhang, & Guo (2008) compared investment traits of Canadian investors against their Chinese counterparts and concluded Canadian investors are more likely to believe that the current trend would continue in a linear fashion where Chinese investors are likely to believe in a less linear trend.

Research also shows that in some cases the trading decision is based on a need for thrill seeking (Grinblatt & Keloharju, 2009). The research by Grinblatt & Keloharju (2009) concluded that traders who engage in this personality trait may be prone to other sensation seeking traits such as fast driving or risky sexual behaviour. Dorn & Sengmueller (2009) hypothesised that certain trading decisions are made purely for entertainment reasons. Their research found that traders who gamble turn over their portfolios at twice the rate of their non-gambling compatriots.

In answering the question as to how do investors make decisions, Olsen (1998) highlighted a number of contributing factors, including the fact that decision makers are often searching satisfactory as opposed to optimal outcomes. Furthermore, decision makers adapt to their situation and their final decision may only be made during the final decision making process. These factors may contribute to investors underestimating the levels or risk, or short selling on winning stocks, investing in what seem to be “good” companies as opposed to purely “good” stocks.

Another contributing factor to the decision making process is confirmation bias, whereby investors filter the news based on their chosen position or investors incorporating irrelevant information into their decision making process (Flynn, 2014).

Against this backdrop of behavioural finance Shleifer & Summers (1990) suggested traders can be split into two categories. Firstly rational arbitrage investors who make informed decisions based on analysis of available information and are not subject to sentiment. Secondly, “Noise Traders” whose decisions are not fully rational and are subject to public sentiment. Shleifer & Summers (1990) concluded that certain investment decisions are in response to pseudo signals such as an advice of a “guru”.

The concept that social dynamics can move markets was highlighted by Shiller, Fischer, & Friedman (1984) and the term “noise traders” was coined by Black (1986) in order to differentiate traders who make rational decisions based on information from traders who trade based on noise. Noise was defined as any irrational factor influencing an investment decision (Black, 1986). This research also highlighted that it is the presence of these noise traders that allows information based investors to make a profit in short term trades and that the greater the degree of noise in a market the greater potential to profit from the situation. In contrast, to the hypothesis proposed by Black (1986), DeLong, Shleifer, Summers, & Waldmann (1987) argued that noise traders could impact the price of an entity by investing heavily in more risky stocks. Their conclusion was retested a year later by DeLong, Shleifer, Summers, & Waldmann (1988) who devised a theoretical model to analyse the long term success of “noise traders”. Their results of this model showed that “Noise traders” would survive, but would not make excessive profits. However, a caveat in this model is that they assumed that noise traders would substantially move the market. This model was again retested using empirical evidence of tick data from 1983 – 2001 by Barber, Odean, & Zhu (2006) and concluded that “noise traders” do not substantially move prices.

The impact of noise traders also differs between markets, Lim & Brooks (2010) proposed that emerging markets have a greater degree of price fluctuation than in markets in developed countries. Their hypothesis was that true arbitrage investors may choose markets that have a greater degree of property protection. The absence of arbitrage investors means that noise traders have a greater influence on the market.

Fromlet (2001) described how human emotions can lead to the formation of stock market bubbles; investors are caught in a bandwagon effect and therefore are more likely to follow other investors than make rational decisions. As the price increases investors filter the news they consume to find articles and forecasts that already agree with their chosen position. This psychological sentiment can be further perpetuated by a sense of over confidence that develops as the prices increase.

The impact of social mood on markets was analysed by Nofsinger (2005) who determined that investor activity and risk appetite is impacted by social mood in the same manner as consumers are impacted by the economic sentiment and debt cycles.

Analysis of the impact of social mood using sporting results as the litmus test have had mixed results. Ashton, Gerrard, & Hudson (2003) found a strong correlation between the results of the English national football team and the price movement of the FTSE 100 on the following day. However Vieira (2012) found no evidence of this affect when analysing the results of the 2010 soccer world cup. A wide range of other investment neutral factors have been shown to impact markets, examples include aviation disasters as documented by Kaplanski & Levy (2010) and climate (Kamstra, Kramer, & Levi, 2003).

Lux (2011) investigated the impact emotional sentiment of traders on future price movements. This study analysed the sentiment of traders working in Germany against future price movements of the DAX. The research showed that sentiment has a significant correlation with future price movement. In similar long term study conducted in the US, Dergiades (2012) compared US investor sentiment to monthly time series data collected from 1965 to 2007. This study concluded that sentiment has a predictive power on stock returns.

The results of these studies indicate that a certain amount of market inefficiency may exist based on human psychology and sentiment. Furthermore all investors can be influenced by any number of these biases simultaneously. In order to understand the combined impact of these biases Kourtidis, Šević, & Chatzoglou (2011) categorised a number of Athens based investors into categories of high, medium and low based on their disposition to overconfidence, risk tolerance, self-monitoring and social influence. The results of this study showed that members of the “high” segment outperformed their peers in terms of stock trading.

2.4 Natural language processing and sentiment Analysis

The impact of behavioural factors on investment decisions increases the value of understanding the public mood and the sentiment towards individual securities or the market as a whole. This in turn raises the importance of computerised algorithms that are designed to interpret and understand the intended sentiment of items such as news articles, Facebook posts and tweets.

The term sentiment analysis or opinion mining describes a computer based process of abstracting an author's opinion towards a specific object (Nasukawa & Yi, 2003). This field of study has grown in conjunction with the explosion of user created content on the internet (Liu, 2012). These methods are used in numerous marketing initiatives such as product reviews as well government intelligence (Pang & Lee, 2008).

The process of sentiment analysis as described by Pang & Lee (2008) involves generating a sentiment polarity on a scale of negative, neutral and positive. Most text can include multiple opinions within varying degrees of objective and subjectivity as a result the polarity scores returned from a sentiment analysis engine will be a number in a range and not a binary result.

Liu (2012) identified that analysis can be done at various levels of granularity. For example, at document level, sentence level or entity and aspect level. Entity level analysis is the most granular form and caters for cases where both positive and negative sentiment can be conveyed in a single phrase. This differs from document level which sets out to determine the overall sentiment of an entire article or document.

Go, Huang, & Bhayani (2009) identified that sentiment analysis of microblogs such as Twitter has a number differentiating characteristics, including use of slang and abbreviations and the short message length. However, even with its limitations a large number of studies such as Agarwal, Xie, Vovsha, Rambow, & Passonneau (2011) and Pak & Paroubek (2010) have proven that determining reliable sentiment polarity scores of tweets is possible. Bollen, Mao, & Pepe (2011) showed strong correlation between Twitter sentiment and general public mood around events such as the US elections and public holidays. This type of sentiment analysis has been used extensively in the fields of marketing and brand awareness as well as product reviews (Lee, 2007); (Lee & Bradlow, 2011); (Leung & Chan, 2009).

Specifically when using Twitter as a corpus, sentiment analysis is used as a replacement for polling by political parties (Tumasjan, Sprenger, Sandner, & Welpe, 2010); (Bermingham & Smeaton, 2011). According to research conducted by Ravisankar, Ravi, Raghava Rao, & Bose (2011) sentiment analysis has also been used successfully in financial fraud detection as well as other fraudulent activities such as fraudulent travel reviews (Yoo & Gretzel, 2009). Another use would be government intelligence services who utilise the technology for security purposes (Martineau & Adviser-Finin, 2011); (Chen, Reid, Sinai, Silke, & Ganor, 2008).

2.5 Impact of news on price and volatility

In some of the earliest work, Ross (1989) determined that price volatility is directly impacted by the volume on news related to the security in question. These findings were corroborated by Berry & Howe (1994) who proved a strong correlation between news arrival and volatility but only moderate correlation between news arrival and price. The impact of news on volatility was again proven by Kalev, Liu, Pham, & Jarnecic (2004) who correlated company specific announcements with intraday volatility on the Australian stock exchange. Their findings are in line with the theory of Mixture Distribution Dynamics (MDH). This theory suggests that price volatility is positively correlated with the rate of information arrival.

Further evidence of MDH was found by Darrat, Zhong, & Cheng (2007) who studied the volatility and volume of stocks on NYSE. They categorise stocks into two categories; stocks for which there was public news on that day and stocks that had not publicly available news released on that day. They determined there was a greater degree of volatility for the collection of securities where news was published. Andersen, Bollerslev, Diebold, & Vega (2007) went on to show that not only does the arrival of news impact security prices but that the same news event can move prices in different directions based on the overarching macro-economic conditions facing the country at the time of the announcement. They showed a negative correlation between bad news and price movement in expanding economies. Their proof was achieved by analysing 25 US based macroeconomic announcements against raw tick data from US, German and British stock markets.

Lee (1992) showed individual investors responded differently to news sentiment than their institutional counterparts. Whilst for both classes of investors there was a definite correlation between news releases and trading activity, their research suggests that individual traders are more likely to buy no matter what the sentiment of the news is whereas institutional investors are more likely to trade in the same direction as the sentiment. Geoffrey Booth, Kallunki, & Martikainen (1999) found a similar trend whereby Institutional investors seemed to make more informed decisions than individual investors. This was also corroborated by Barber & Odean (2008) who showed how individual investors are net buyers during periods of greater news activity.

Riordan, Storkenmaier, Wagener, & Sarah Zhang (2013) found that the market did react to news sentiment and that negative news had a greater impact on volatility than positive news. Similar research analysing the impact of news on the Australian market by Smales (2012) showed that unexpected news announcements impacted volatility and spread, again negative news had a greater impact on volatility than positive news.

A causal relationship between media mentions of a security and volumes of trade was proven by Engelberg & Parsons (2011). Their research compared mentions of US based stocks listed on the S&P 500 in local newspapers against the volume of trades that originated from the same local region. A further degree of causality was proven by incorporating extreme weather events into the model. In event of extreme weather the impact of news on volatility was less obvious as the number of readers with access to the news decreased.

Dividend announcements are a specific type of news announcement that theoretically shouldn’t impact price (Merton & Modigliani, 1961). This is due to the fact that the overall value of the investment remains unchanged, however Asquith & Mullins Jr (1983) proposed that the value to the investor may change based on how capital gains are taxed in relation to dividends. Even after taking tax implication into account Statman (1999) hypothesised that there is also a psychological aspect as to how investors react to dividend expectations. Reasons aggregated by Statman (1999) include prospect theory whereby investors place dividend pay-outs and capital growth in two separate mental accounts.

Part of this mental accounting involves allocating capital growth to long term saving whereas dividends pay-outs are seen as “bonus” spending money. The spending of capital gain has also led to investor regret (Kahneman & Tversky, 1979). There is also the fear that investors dipping into their capital gains will be tempted to do it again. As to the actual impact, the empirical results have been mixed. In an analysis of European markets Vieira (2011) found that Portuguese and English markets reacted positively to dividend announcements whereas the French market reacted negatively. This conclusion builds on the work conducted by Michaely, Thaler, & Womack (1995) who determined that prices continue to trend positively (negatively) after an upward (downward) change dividend policy.

Andersen & Bollerslev (1998) confirmed the impact of macroeconomic announcement on both intraday volatility and daily volatility persistence. Similar research was conducted by Jansen & De Haan (2005) who studied the impact of announcements by ECB officials on the Euro-Dollar exchange rate on volatility and mean change in price. In this instance there was correlation between announcements and volatility but there was little evidence of more permanent changes in the exchange rate as a result of ECB announcements. In the same year Bauwens, Ben Omrane, & Giot (2005) studied the impact of both scheduled and unscheduled news announcements on the dollar-Euro exchange rate. They determined that there is an increase in volatility even before a scheduled announcement is made. Evans & Speight (2010) were also able to prove the impact of macroeconomic news announcements of currency fluctuations they conferred that announcements do have an impact, their findings were that US centric announcements have the greatest impact on the value of currencies, however production numbers from the UK and Japanese GDP numbers also impacted the Pound/Euro and Euro/Yen rates.

More recently Dewachter, Erdemlioglu, Gnabo, & Lecourt (2014) corroborated a number of the conclusions highlighted above, concluding that currency markets are impacted by news and that volatility increases prior to scheduled news announcements, however, a differentiating factor was that European markets react with a greater degree of volatility than their US counterparts.

The impact of news sentiment on Foreign exchange markets was analysed by Laakkonen & Lanne (2008) who concluded that negative news has a greater impact than good news during times of a shrinking economy whilst sentiment has less of an impact during times of economic growth. These results are based on fluctuations in the Euro/USD exchange rate for the period between 1999 and 2004

2.6 Impact of social sentiment on security prices

The earliest study in this field undertaken by Niederhoffer (1971) who analysed newspaper headlines from the New York Times against New York Stock Exchange (NYSE) prices and concluded that there was increased price volatility on the days following a world event, he categorised a news item as a world event if it occupied more than five columns on the front page of the newspaper. Niederhoffer (1971) identified a large price change on the first day following an event followed by a corrective period on day's two to five after the event, indicating that markets had a tendency to overreact to bad news.

The digitisation of the media removed the manual limitations faced by Niederhoffer (1971) in analysing the sentiment expressed in large volumes of text. Zhang & Skiena (2009) analysed 10 years of data comparing the price movement of all entities listed on the NYSE against related financial news articles published in over 500 daily newspapers. The purpose of this study was to understand the impact of today's news on tomorrow's stock price. The results indicated that markets took time to react to available news with the strongest correlation between price movements and news mentioning occurring one day after the news was published.

This research was repeated one year later by Zhang & Skiena (2010) and this time incorporating two years of online blogging activity and one year of Twitter data. With Twitter still in its infancy the authors found that on average markets incorporated information from Twitter at a slower rate than information from traditional news and blog sources. The authors concluded that this was due to the time taken before the news contained in the tweet was read by a large volume of people.

Research analysing the interaction between Wall Street Journal articles and price movement Tetlock (2007) found that stocks mentioned in a certain column published in the Wall Street Journal were subject to a greater degree of volatility on the day after publication. The research found correlation between negative sentiment and price movement showing that extreme levels of sentiment also results in a higher degree of volatility. using the same Wall Street Journal columns Dougal, Engelberg, Garcia, & Parsons (2012) proved a causal relationship between Wall Street Journal articles and short term price volatility. Again by using articles from the Wall Street Journal and Dow Jones newswire Smales (2013) was able to show a negative correlation between the VIX index and news announcements and highlighting that negative news has a greater impact on the market than positive news.

An in-depth study of impact of news announcement and news sentiment was undertaken by Groß-Klußmann & Hautsch (2011) who analysed the impact of security specific news for 40 stocks trading on the London Stock Exchange for the period between 2007 and 2008. Their results showed a direct response to news announcements in form of increased trading activity, volatility and spread. They also confirmed a correlation between sentiment and price trend. A similar study conducted by Hoa, Shia, & Zhangb (2014) analysed the impact of twelve years' worth of news announcements and news sentiment for 30 stocks trading on the Tokyo exchange. Their findings were that volatility was impacted to the same extent by both positive and negative news but that news sentiment had a limited impact on volatility persistence.

Antweiler & Frank (2004) analysed 1.5 million comments on internet bulletin boards. The study compared the social sentiment expressed on the Yahoo finance and Raging Bull boards against price movement for 45 companies listed on the Dow Jones Industrial and Dow Jones Internet message boards. The study concluded that statistically significant correlation exists between sentiment and price movement. However, the study also concluded that the prices changes were too small to lead to an economic advantage as the gains for retail investors would be eroded by trading fees. The study also found that a high degree of disagreement in the internet postings correlated with higher trading volumes.

As the use of Twitter became more ubiquitous, further research was undertaken focusing on the correlation between general tweet sentiment and price movement (Zhang, Fuehres, & Gloor (2011). In this instance, by analysing approximately 1% of all tweets over a six month period, the researchers found a negative correlation between user sentiment and price movement of key American indices such as the Dow Jones, S&P 500 and NASDAQ 100. In other words, on days with a high frequency of negative tweets would prelude days of positive price movement. This study however, analysed general market sentiment and index direction and not stock specific sentiment.

A similar study was also undertaken analysing the overall sentiment of Facebook posts to stock market movements. In this instance by using Facebook's "Gross National Happiness index" (GNHI), Siganos, Vagenas-Nanos, & Verwijmeren (2014) were able to prove correlation between positive sentiment and positive price movements. A degree of causality was also implied by the fact that the same correlation exists between general Facebook sentiment on a Sunday and price movement on the Monday. As with Zhang, Fuehres, & Gloor (2011) study, this paper analysed general market sentiment and not stock specific sentiment. The study did however, include 20 international markets. This research corroborated and expanded upon the research done by Karabulut (2011) who conducted a similar study against the American market. Unfortunately the reliability of the GNHI has not been completely validated. Wang, Kosinski, Stillwell, & Rust (2014) tested the GNHI against the Diener's Satisfaction with Life Scale (SWLS) the study included a years' worth of Facebook statuses and no significant correlation could be found.

In analysing the predicative ability of stock specific micro-blogs Oh & Sheng (2011) used a combination of Twitter data and microblogs from StockTwits. Using a sample of over 7200 posts against a combination of NASDAQ and NYSE listed stocks they were able to correlate sentiment with price movement. Their results also showed a higher degree of correlation based on negative sentiment than positive sentiment

In two recent and in-depth studies focusing on Twitter data Rao & Srivastava (2012) and Rao & Srivastava (2013) analysed the sentiment polarity of up to four million stock specific tweets collected between June 2010 and September 2011. The first study focused on 11 technology stocks listed on Dow Jones and NASDAQ and concluded that a strong positive correlation existed between sentiment and price movement. The second study expanded the first study and included spot prices for gold and oil. The second study also included Google search volumes of each company as an additional measure of company interest. Furthermore using this data Rao & Srivastava (2012) were able to build a data mining model to accurately predict future directional price movements with an accuracy of at times in excess of 90%.

Analysing search trends Joseph, Babajide Wintoki, & Zhang (2011) correlated the volume of google searches for a stock code against weekly price movement and volatility. The results found positive correlation for both volatility and price. Mao, Counts, & Bollen (2011) analysed stock specific tweet volumes, tweet sentiment, news sentiment, google search volumes and investor surveys. These sources were compared to movement of the VIX index and they concluded that Twitter sentiment outperformed news sentiment and investors surveys as way of predicting short term movements in stock prices.

Taking a longer term view of publically available information Wang et al. (2014) built a number of hypothetical portfolio's based purely on sentiment analysis off data from online social investment platforms, namely Seeking Alpha and StockTwits. In this instance, a portfolio built using Seeking Alpha would have outperformed the market, where as a StockTwits portfolio would have underperformed in comparison to the market index.

2.7 Conclusion

Based on current research in the long term the price of equities and traded commodities seem to follow the principles of efficient market hypothesis. However this hypothesis is still dependent on the availability of information and the speed at which the market can consume this information.

Furthermore, the current body of knowledge suggests that in the short term prices will be impacted by numerous factors including general social sentiment, investor psychological bias overall public mood and the market sentiment related to the traded entity. Investors are also subject to biases based on their gender and nationality.

In analysing the impact of news arrival on volatility and price a large body of research has proven that news announcements directly impact volatility of individual securities and currencies however the volatility persistence of the fluctuations has varied between studies

Looking at Twitter sentiment specifically Zhang, Fuehres, & Gloor (2011) proved that direction of major American Indices was negatively correlated to average Twitter sentiment. In their two studies Rao et al showed correlation and a predictive influence between stock specific sentiment on Twitter for American technology companies and indices. They were also able to predict the future price movement of these traded entities.

2.8 Research Hypothesis

Based on the objectives specified in section 1.4 and the Literature review above the following hypothesis have been established

- H1: A positive correlation exists between normalised tweet volumes received about a traded entity and the normalised number of trades
- H2: A positive correlation exists between normalised tweet volumes received about a traded entity and normalised trading volumes
- H3: A positive correlation exists between normalised tweet volumes received about a traded entity and normalised trade volatility
- H4: A positive correlation exists between average tweet sentiment and price direction
- H5: Repeating same tests for the hypothesis 1-3 using a time lag between Twitter activity and market activity may also result in a positive correlation. This time lag tests for predictive influence.

The above tests are conducted using the exchange as a level of granularity.

Chapter Summary

This chapter included a review of the extant literature pertinent to this thesis. Chapter 3 covers the data required and the research design used in the study.

3 RESEARCH METHODOLOGY

3.1 Introduction

This chapter explains the research methodology used and how the research design will ensure valid and reliable results. The chapter is organised as follows. Section 3.2 presents the data and data source. Section 3.3 presents the methodology followed to test the hypothesis. Section 3.3.6 describes the analysis techniques. A chapter summary concludes the chapter.

3.2 Data and Data Sources

The data required to conduct this research is attained from two secondary sources; namely *Twitter* and *Bloomberg. Bloomberg* provides all financial data used including trading volumes and price. Appendix B shows the data downloaded per entity. *Twitter* provides the corpus of public opinion - all individual tweets related to the market entities identified in section 3.3.3. The table in Appendix A includes all the fields downloaded using the Twitter API.

3.3 Research Design

The research design process includes six phases, each of these phases is explained in the following sections.

3.3.1 Time Frame Identification

This research is for the period of 2014-05-01 until 2014-09-30. This is in line with comparative studies using Twitter as a corpus, where the time frame ranged from less than three months (Xu, 2012) to eleven months (Rao & Srivastava, 2013). In order to mitigate any potential impact of a comparatively short time frame this research also analyses a larger number of traded entities. The number of stocks from previous research ranged from 10 shares (Oh & Sheng, 2011) to 16 (Xu, 2012).

3.3.2 Market Identification

The second step in the research process is to determine which markets would be used for analysis. Countries with a first language of English was the first filter applied. This selection criteria is required as the process of assigning sentiment scores to non-English text and testing results provides an extra level of complexity.

The NASDAQ exchange was chosen as the baseline as this exchange was used in most of the comparative studies, the technology heavy nature of stocks listed on this index also results in a higher degree of social media attention and as a result provide a greater volume of data for analysis. The lower volumes of tweets from the other exchange may limit the degree of analysis that can be conducted against non-NASDAQ firms or firms that are not subject to a large volume of public comments or tweets.

The LSE was selected as one of the largest exchanges in predominately English speaking countries outside the USA. The JSE was selected as the test case for English speaking emerging markets.

3.3.3 Sample Selection Criteria

From each exchange listed in section 3.3.2, a subset of traded entities were selected. The selection includes the a sample of top 40 shares trading on the JSE, a sample of the top 10 shares trading on the NASDAQ, a sample of 20 of the top 100 shares on the LSE. The sample was also skewed to include a greater number of South African firms. The final sample consisted of 61 trading entities and approximately 311 000 tweets were collected during the period of observation, this sample excludes tweets collected from non-trading days.

3.3.4 Tweet Extraction

Each Individual tweet associated with the selected market entity is downloaded from Twitter using the Twitter search API and predefined search queries; the queries used for each entity are also included in Appendix B. The search queries comprise a combination of hashtags (#) and cashtags ($) tags to identify and filter tweets related to the selected entity. A cashtag is a Twitter standard for identifying companies, specifically in a financial context. A cashtag is constructed using the company’s stock ticker symbol preceded by a dollar sign. i.e. Google would be identified by $Goog. Other variations of this principle include the addition of the market code, i.e., Tweets relating to Truworths trading on the JSE can be found on a Twitter by using any of the following search terms $TRU or $JSE:TRU or $JSETRU.

In order to ensure maximum possible coverage of all tweets posted a combination of all variations will be used. Searches will be automatically conducted against all preselected traded entities using a number of search variations every fifteen minutes.

3.3.5 Sentiment Analysis Scoring

All tweets downloaded are run through a number of comparative cloud based sentiment analysis tools in order to determine which of the cloud services provides the most accurate sentiment analysis of financial tweets. As an example the AlchemyAPI (AlchemyAPI, 2014) engine has been used in previous academic research for classification and sentiment analysis (Quercia, Askham, & Crowcroft, 2012); (Singh, Piryani, Uddin, & Waila, 2013). To illustrate how Sentiment scores are calculated using the Alchemy engine, consider the following example of a tweet related to NASDAQ index:

“$QQQ - Taking Advantage Of Pessimism -> http://t.co/AXx9jr3neU #stock #stocks #stockaction”

Against this tweet the Alchemy sentiment scoring engine concluded the following resulted in a final score of 0.1 out of 1. The engine identified that this tweet consisted of two sentiment related phrases, namely “*Taking Advantage*” and “*Pessimism*”. The first phrase received a positive rating of 0.8 whereas the second phrase received a negative rating of -0.7. The average of these two rating resulted in a final score of 0.1. A score of 0.1 is categorised as a “neutral” score

3.3.6 Statistical Analysis

Based on the data collection process explained above, the process of performing the analysis to meet the research objectives involves linking the individual tweets related to each entity to the Bloomberg trading data of the same entity. Using a similar process to

Zhang & Skiena (2009), Rao & Srivastava (2012) and Rao & Srivastava (2013) a series Pearson correlation coefficients is created between the Twitter and financial measures. A separate model is created per entity and per exchange thus allowing for the statistical comparisons between the individual traded entities as well as the market as a whole.

Having established a Pearson Correlation Coefficient, Granger analysis (Granger, 1969) is conducted at multiple time lag intervals in order to test for predictive influence. This type of analysis is used to determine the predictive nature of one event against another. Granger analysis is used extensively within the fields of economic and finance. This technique was also used in comparative studies by Xu (2012), Bianconi, Hua, & Tan (2013) and Rao & Srivastava (2013).

3.4 Statistical Measures

The following table describes the various measures used as part of this research

3.4.1 Twitter Activity Measures

Measure Name	Definition
Tweet Count	The total number of tweets received
Variance tweet count	The difference between the average number of tweets received for this entity for this period, and the actual number recorded

Table 1: Twitter activity measures

3.4.2 Market Activity Measures

Measure Name	Definition
Tick Count	Number of individual trades
Variance tick count	The variance from the mean of the number of individual trades recorded for this entity for this period
Volume	The total volume of shares traded
Variance volume	The variance from the mean of the total volume of shares traded for this entity for this period
Value	The total value of shares traded recorded in the local currency
Variance Value	he variance from the mean of the total value of shares traded

Table 2: Market activity measures

3.4.3 Market Direction Measures

Measure Name	Definition
Price Range	Price Close – Price Open
Volatility	Price High – Price Low

Table 3: Market direction measures

3.4.4 Tweet sentiment Measures

Measure Name	Definition
Positive tweet count	Count of positive tweets. A positive tweet was defined as having a sentiment score> 0.1
Negative tweet count	Count of negative tweets. A negative tweet was defined as having a sentiment score < -0.1
Polarity	(p - n) / (p + n)

Table 4: Tweet sentiment measures

The formulae for polarity was the same formula used by Zhang & Skiena (2009)

3.5 Validity and reliability

3.5.1 External validity

External Validity is the process of ensuring that the results of the study are representative of the broader population (Roe & Just, 2009). In order to maximise the external validity this research will include the analysis of a variety companies and indices from three exchanges namely, NASDAQ, LSE and JSE.

The NASDAQ was selected as entities on this exchange are responsible for the most number of tweets generated. Furthermore this exchange has been used in a number of comparative studies such as (Oh & Sheng, 2011), (Xu, 2012) and (Rao & Srivastava, 2013).

This study will attempt to prove that assuming similiar results are attained again that the same trend can be seen in South African and English markets.

3.5.2 Internal validity

Internal Validity is primarily focused with the principle of causality, previous studies such as Schram (2005) have highlighted the inherit trade-off between internal and external validity specifically in economic experiments.

In conducting this research, it is impractical to draw a causal link between events on Twitter and stock market fluctuations. The purpose of the study is instead to show the predictive nature of external events such as social sentiment on markets.

3.6 Delimitations of the study

This research includes a stock level analysis of a subset of shares traded on the JSE top 40, FTSE top 100 and NASDAQ Exchanges.

A second delimitation is that only Twitter data will be used as the corpus, no additional data sources such as news articles or blogs will be included in the research.

Thirdly this research relies on natural language processing algorithms to successfully determine the sentiment and context of tweets. Whilst sentiment analysis of tweets is possible (Pak & Paroubek, 2010), the lexicon of microblogging which includes emoticons, abbreviations and different grammar styles limits the accuracy of sentiment analysis.

The study will focus only on English speaking countries and tweets. As the sentiment analysis of foreign languages adds an additional layer of complexity.

A further delimitation of this study relates to the restrictions imposed by Twitter on using the Twitter search API. Twitter (2014b) restricts search results to a maximum of 100 results per call with a total limit of 1500 results in a fifteen minute window. Twitter does not guarantee that a search query will return all results. Due to these restrictions the delimitation is that only a subset of tweets per company were analysed and that the study will focus on a limited time period.

Chapter Summary

This chapter presented the data required and research techniques that used to conduct the research. Chapter 4 will present the results of the statistical analysis

4 EMPERICAL RESULTS

4.1 Introduction

The purpose of this chapter is to present the empirical statistical findings that form part of the research. This chapter starts by presenting the overall descriptive statistics as well as the descriptive statistics split per exchange. The remainder of the chapter analyses each of the individual data sets constructed encompassing of tweet volumes as well as tweet sentiment. The chapter also illustrates using time lagged intervals the speed at which news is consumed by the market.

4.2 Univariate analysis of different exchanges aggregated per day

Table 5 below presents the univariate statistics of the tweets volumes collected per exchange per day over a period between May to September 2014. The results show that, as expected, the number of tweets received for NASDAQ completely outweighs the number of tweets collected for the other markets. Over 511 000 tweets were collected against NASDAQ listed firms, this is 89% of the total sample. Comparatively the JSE accounted 7% of the sample. This translated into an average of 577 tweets per day per firm for NASDAQ listed firms. In comparison an average of 20 and 22 tweets per day were collected against JSE and LSE listed firms respectively. There is also a recognisable pattern which shows that majority of tweets are collected during week days. Appendix C contains the breakdown of the number of tweets collected against each firm per month.

Exchange Name	**JSE**	**LSE**	**NASDAQ**
Company count	35	17	9
Tweet count	40,301	26,032	511,764
Minimum daily tweet count	1	1	14
Maximum daily tweet count	1 649	624	8 203
Average daily tweet count	20	23	578
Median daily tweet count	5	7	249

Table 5: Univariate Analysis per exchange

Figure 1 illustrates the weekday/ weekend cycle and the comparative difference between the tweets collected against the NASDAQ and other two exchanges.

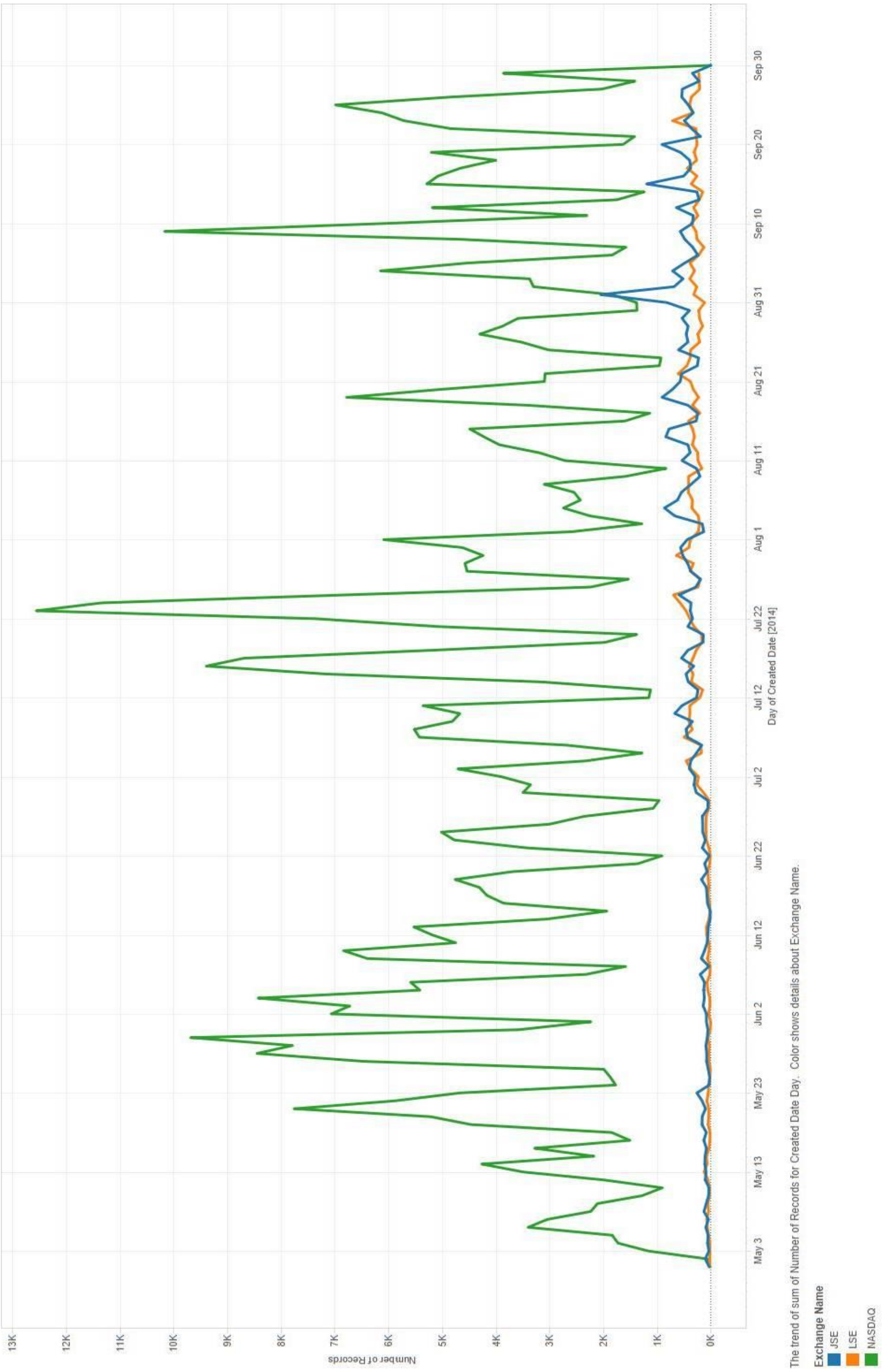

Figure 1: Tweet volumes per day

The graph also highlights that the week of the 21st of July; the week that both Apple and Facebook reported financial earnings also resulted in the highest number of tweets. The time period of analysis also included the date of Apple's launch event on the 9th of September. This date produced the single largest spike in Twitter activity with over 8200 tweets collected during the period

The data was further analysed to calculate the total number of tweets posted per hour over the total period of analysis. Following the US centric pattern as seen above the daily dispersion of tweets aligns with the trading hours of the US markets, for the both the JSE and the LSE more trades occur in the afternoon as the time aligns more closely to the US trading hours. All tweets were time stamped to South African Standard Time (SAST). In this time zone the trading hours of the NASDAQ exchange is from 14:30 to 22:00.

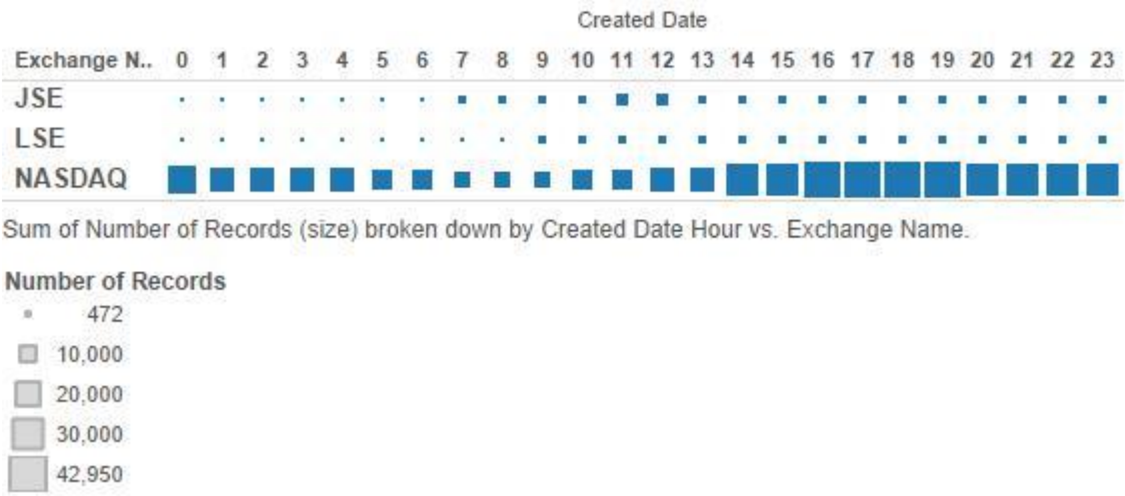

Figure 2: Total tweets per hour

4.3 Univariate Analysis of different measures per exchange

The following table displays the statistical analysis of the variables used in this chapter as presented in chapter 3. All results for the remainder of this chapter are aggregated at a daily level and grouped per exchange. One key metric is the average number of tweets received per day per entity, on the JSE this number was 20, compared to 578 on the NASDAQ. N in this instance refers to the total number of trading days for all companies that was included in the analysis.

Exchange	N	Variable	Mean	Minimum	Maximum	Median
JSE	1975	Tweet Count	20	1	1,649	5
		Variance Tweet Count	10	(99)	2,056	(29)
		Variance Tick Count	0	(90)	486	(9)
		Variance Volume	0	(95)	879	(12)
		Variance Value	0	(95)	842	(12)
		Tick Count	2,880	104	21,078	2,409
		Volume	1,644,998	28,583	21,082,827	990,956
		Value	31,447,491,999	375,681,991	572,549,858,509	18,469,764,395
		Polarity	0	(1)	1	0
		Price Day Range	(35)	(4,900)	11,437	(14)
		Volatility	586	14	14,144	355
LSE	1138	Tweet Count	23	1	624	7
		Variance Tweet Count	6	(88)	1,980	(20)
		Variance Tick Count	0	(85)	730	(7)
		Variance Volume	0	(90)	1,036	(10)
		Variance Value	0	(90)	947	(10)
		Tick Count	3,605	139	26,091	2,652
		Volume	3,741,735	14,439	87,654,977	1,635,620
		Value	2,573,327,977	21,784,354	29,563,974,773	1,404,803,690
		Polarity	0	(1)	1	0
		Price Day Range	(2)	(164)	94	(1)
		Volatility	24	2	185	19
NASDAQ	886	Tweet Count	578	14	8,203	249
		Variance Tweet Count	0	(97)	479	(17)
		Variance Tick Count	0	(84)	335	(2)
		Variance Volume	0	(73)	380	(4)
		Variance Value	0	(71)	423	(3)
		Tick Count	70,962	0	689,506	57,944
		Volume	20,674,748	0	176,320,406	18,171,909
		Value	1,365,310,887	0	17,588,987,245	788,150,272
		Polarity	0	(1)	1	0
		Price Day Range	0	(72)	55	0
		Volatility	6	0	80	1

Table 6: Univariate analysis per exchange

4.4 Relationship between tweet volumes and trade volumes

The following tables illustrates the strength of correlation between the measures pertaining to Twitter volume and trading activity. The two key metrics in this series of tables are the correlation between tweet count and value and secondly the correlation between the variance in tweet count and the variance in trading value.

			Trading Measures					
			Tick Count	Volume	Value	Variance Tick Count	Variance Volume	Variance Value
Twitter	JSE	Tweet Count	0.13	0.076	0.056	0.062	0.054	0.054
			(<.0001)	-0.0007	-0.0134	-0.0058	-0.0166	-0.0164
		Variance Tweet Count	0.144	0.092	0.12	0.204	0.179	0.183
			(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)
	LSE	Tweet Count	0.335	0.735	0.288	0.109	0.09	0.096
			(<.0001)	(<.0001)	(<.0001)	-0.0002	-0.0024	-0.0012
		Variance Tweet Count	0.155	0.045	0.114	0.298	0.255	0.268
			(<.0001)	-0.1284	-0.0001	(<.0001)	(<.0001)	(<.0001)
	NASDAQ	Tweet Count	0.626	0.63	0.838	0.257	0.269	0.265
			(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)
		Variance Tweet Count	0.289	0.308	0.228	0.529	0.524	0.515
			(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)	(<.0001)

Table 7: Twitter activity

Across all three markets statistically significant positive correlations were calculated for both the raw values of tweet count correlated to value, and the variance in tweet counts compared to the variance in value. In comparing tick count to tweet count the Rho for the JSE was 0.13 increasing to 0.26 to the NASDAQ, for the same comparison the LSE was 0.335. In analysing the correlation scores relating to variance calculations the variance in Tweet count for the JSE was 0.06 compared to 0.529 for the NASDAQ. In this case the Rho for the LSE was 0.298. In every instance the strongest correlation scores were calculated for the NASDAQ and the weakest scores were calculated for the JSE.

4.5 Relationship between Twitter volatility and Price volatility

The following table illustrates the strength of correlation between the variance in tweet volumes and the maximum price range.

		Volatility
Variance Tweet Count	JSE	0.142
		(<.0001)
	LSE	0.12293
		(<.0001)
	NASDAQ	0.081
		(0.0237)

Table 8: Correlation between variance in tweet volumes and price volatility

Table 8 shows that the JSE, which received the lowest number of tweets per day per entity had the highest degree of correlation between Variance tweet count and volatility. This was the only test where the Rho scores were highest for the JSE and lowest for the NASDAQ, inverting the trend of the other tests undertaken. This result is indicative of the amount of "noise" associated with the stocks trading on the NASDAQ, whereby a larger number of tweets are not related to the share. The tweets related to the NASDAQ listed firms also include a large degree of spam. An example of this type of tweet is "*Check out how I'm making HUNDREDS of $$$ every month with the profit.ly affiliate program! http://t.co/5GZewWhwc3 $zpin $hurc $FB $live*"

4.6 The relationship between tweets' sentiment polarity and price direction

This section highlights the impact of social sentiment on price direction. The sentiment analysis engine used for the research is http://sentiment.vivekn.com. The comparative tests done between this engine and the Alchemy sentiment engine produced similar results on a subset of the data, therefore all future mentions of sentiment refer to the sentiment score generated using this API. It is important to note that the accuracy of sentiment analysis engines can be greatly enhanced by training the model against a specific content type (Nann, Krauss, & Schoder, 2013). The models used in this report were untrained, thus the strength of correlation shown is potentially weaker than if a trained model had been used.

		Day Range Percent
Polarity	JSE	0.012
		(0.5865)
	LSE	-0.009
		(0.7712)
	NASDAQ	0.081
		(0.0227)

Table 9: Sentiment correlation per exchange

Table 9 shows that NASDAQ has a positive and significant (at 5% significant level) between polarity and the direction of stock price movement while the relationship is insignificant in JSE and LSE. The strength of correlation for the NASDAQ is in line Siganos, Vagenas-Nanos, & Verwijmeren (2014) who calculated an average correlation between stock market returns and the Facebook's Gross National Happiness Index of 0.031. It is also in line with the strength of correlation between general polarity of newspaper articles and daily stock market returns Zhang & Skiena (2009) calculated a correlation of 0.0487.

4.7 Twitter activity and the time lag of share price movement

In order to ascertain if the tweets are a prelude to market activity or if the noise on Twitter is only in response to past market activity Granger analysis was used. In this instance, the variance in trading value was correlated with variance in the number tweets. The tests were then rerun correlating the variance in the number of tweets from yesterday (Lag 1 day) against today's variance in trading value. This was repeated for five days, i.e. the final test was the variance in the number of tweets from five days ago correlated against today's variance in trading volume. Figure 3 illustrates how rapidly the strength of correlation decreases as the number of days lagged increases. When analysis with a lag of zero days the correlation scores are the same scores recorded in Table 7. I.e. the score for the NASDAQ exchange was 0.515 and the score for the JSE was 0.183. With a lag of one day the strength of correlation for the NASDAQ dropped to 0.25 and the score for the JSE dropped to 0.038.

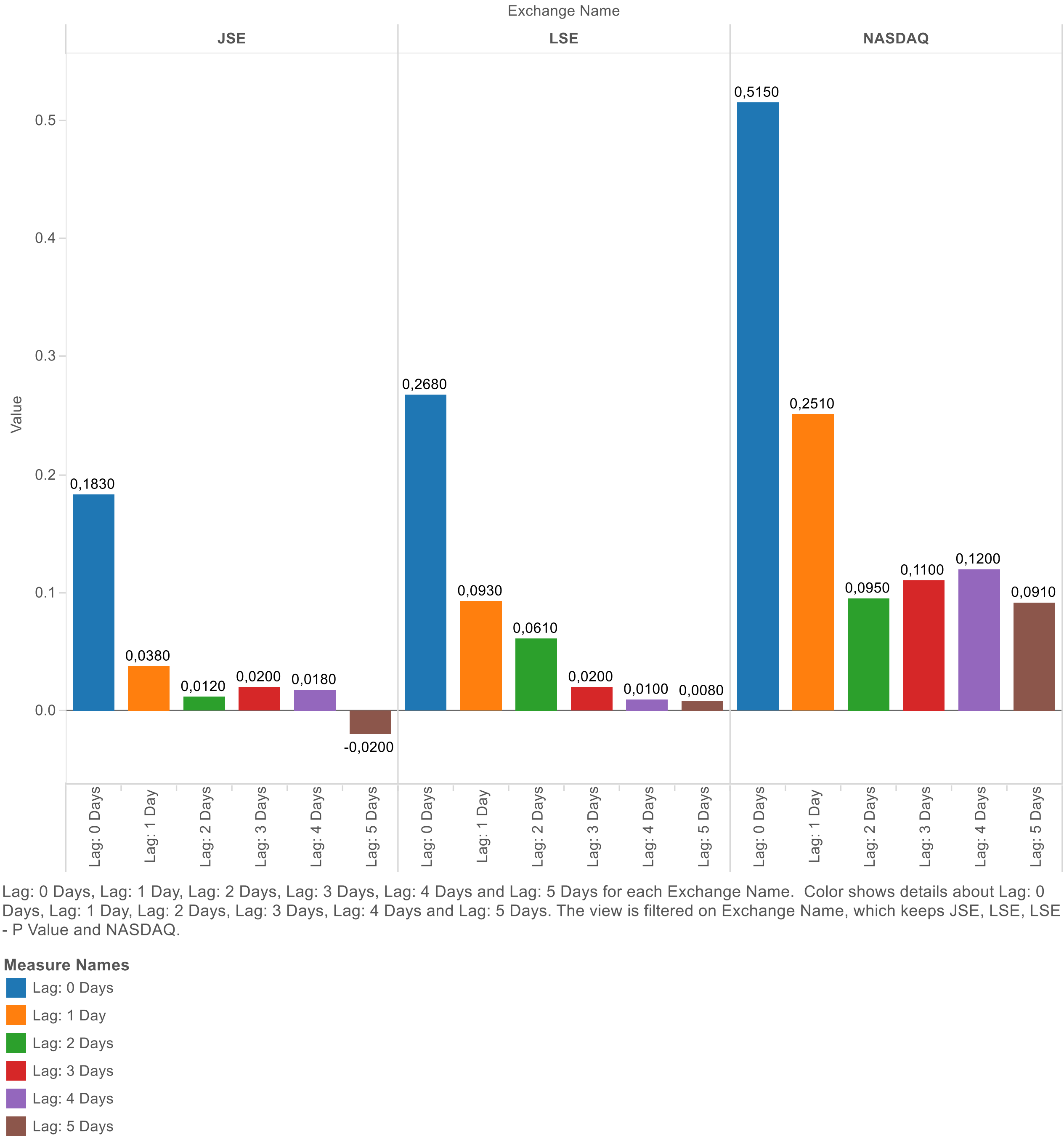

Figure 3: Granger analysis

Table 10 further illustrates how as degree of significance also decreases as the number of days lagged increases. Across all three exchanges the strength of correlation is less than half on a lag of one day than it was on a lag of 0 days. By a lag of three days for all exchanges except the NASDAQ the results are no longer significant. This result indicates the immediacy at which information is both shared on Twitter and how quickly the market adjusts to new information.

		Variance Value					
		Lag: 0 Days	**Lag: 1 Day**	**Lag: 2 Days**	**Lag: 3 Days**	**Lag: 4 Days**	**Lag: 5 Days**
Variance Tweet Count	JSE	0.183	0.038	0.012	0.020	0.018	-0.020
		(<.0001)	(0.0908)	(0.5874)	(0.3615)	(0.4107)	(0.3521)
	LSE	0.268	0.093	0.061	0.020	0.010	0.008
		(<.0001)	(0.0016)	(0.0397)	(0.4997)	(0.7437)	(0.7836)
	NAS DAQ	0.515	0.251	0.095	0.110	0.120	0.091
		(<.0001)	(<.0001)	(0.0042)	(0.0009)	(0.0003)	(0.0058)

Table 10: Granger analysis

Chapter Summary

The purpose of this chapter was to present the empirical findings of the research. Chapter 5 will provide context to the number that were presented in this chapter.

5 DISCUSSION AND CONCLUSION

5.1 Introduction

The previous chapter presented the findings of the research and this chapter discusses the findings and concludes the thesis. The chapter is organised as follows: Section 5.2 discusses the research findings. Section 5.3 provides a conclusion to the research and section 5.4 provides discussion about possible further research on the current research topic.

5.2 Discussion

A series of tests were conducted to prove the relationship between Twitter activity and market activity. These tests proved a degree of correlation for all three markets. These results are in line with other news related studies such as Groß-Klußmann & Hautsch (2011) that showed markets do respond to news events in the form of increased trading activity and volatility. Looking specifically at Twitter activity the results of this study were in line with studies undertaken by Rao & Srivastava (2012) and Oh & Sheng (2011). This research was also able to prove that a similar although less pronounced trend was found for the JSE. This finding proves that Twitter is an accurate indicator for changes in trading activity for both emerging and developed markets.

The first hypothesis tested if shares which are consistently associated with high trading volumes also experience high volumes of Twitter activity. This hypothesis would not necessarily consider variations in trading activity as the market responds to abnormal events.

To test for the second hypothesis, another series of tests was conducted to compare variations in Twitter activity to daily trading activity and price volatility. In this instance the strongest correlation index was again found for the US market, the values for tick count, volume and value were 0.53 for tick count, 0.54 and 0.54 (significant to < 0.0001). Comparatively the JSE's correlation scores were 0.20, 0.18 and 0.18 for the same three measures (significant to < 0.0001). The JSE scores were similar to the LSE scores of 0.30, 0.25 and 0.27. These findings highlight the existence of the theory of Mixture Distribution

Dynamics (Kalev, Liu, Pham, & Jarnecic, 2004). These results also show this theory can be applied to emerging markets and that Twitter accurately reports on events that may result in abnormal trading activity.

In analysing variances in Twitter activity correlated to volatility, the strongest correlation index was found for the JSE, followed by the LSE. In this instance the comparative strength of correlation was inverted in comparison to all the previous tests, as the JSE was the strongest and the NASDAQ the weakest. This finding could be due to higher volume of non-trading related tweets created for NASDAQ listed stocks.

The results of the granger analysis show that the strongest correlation exists when analysing activity within the same time periods. This result indicates the immediacy at which information is both shared on Twitter and how quickly the market adjusts to new information. This immediacy is an indicator that all three markets analysed in this study follow the efficient market hypothesis (Fama, 1970). Introducing a lag of one day still produces a degree of correlation in both the US and London markets however this tapers off substantially by day 3. This tapering proves that markets are responding to news that is communicated on Twitter and that Twitter is not only reporting on past events.

Including sentiment scores in the model, produced mixed results; the correlation scores for the NASDAQ exchange was similar to previous studies, however no significant correlation could be found for the JSE or LSE exchanges. Two potential reasons for this are that an untrained sentiment model was used for this research and secondly that the number of tweets including sentiment for non US listed companies was insufficient, thus accurate sentiment analysis was not possible.

5.3 Conclusion

In this paper we analysed the correlation between company specific Twitter activity and trading activity. The activity was tracked against a number of companies listed on the JSE as well as two other world markets namely the LSE and NASDAQ. The results proved a strong positive correlation between Twitter activity and trading activity for the US market. This finding is in line with the previous studies undertaken. Furthermore a moderate degree of correlation was proven for the JSE and the LSE market. A similar pattern was observed in tracking the variance of Twitter activity to market activity, this trend indicates that periods with abnormal news activity and noise on Twitter are correlated with periods of abnormal trading activity.

By analysing the lagged impact of tweets on future trading activity indicated the immediacy at which Twitter data is consumed by the markets.

Finally including sentiment into the model produced less reliable results with only the US showing a slight positive correlation between Twitter sentiment and market sentiment.

5.4 Future Research

Traders are continuously searching for ways of gaining a short term advantage when making decisions, as a result future research topics could include comparisons, more granular time based analysis between the arrival of new tweets and market reaction. As an example, in highly traded stocks correlations at an hourly granularity may exist.

Secondly this research used comparatively simple sentiment analysis tools, future research should include more advanced models or potentially machine learning algorithms to identify if different types of tweets of phrases have more impact on the market.

References

Agarwal, Apoorv, Xie, Boyi, Vovsha, Ilia, Rambow, Owen, & Passonneau, Rebecca. (2011). *Sentiment analysis of twitter data.* Paper presented at the Proceedings of the Workshop on Languages in Social Media.

Ahmad, Ali Nobil. (2010). Is Twitter a useful tool for journalists? *Journal of Media Practice, 11*(2), 145-155.

AlchemyAPI. (2014). Retrieved 2014-07, 2014, from http://www.alchemyapi.com/company/about-us/

Andersen, Torben G, & Bollerslev, Tim. (1998). Deutsche mark–dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *the Journal of Finance, 53*(1), 219-265.

Andersen, Torben G, Bollerslev, Tim, Diebold, Francis X, & Vega, Clara. (2007). Real-time price discovery in global stock, bond and foreign exchange markets. *Journal of International Economics, 73*(2), 251-277.

Antweiler, Werner, & Frank, Murray Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance, 59*(3), 1259-1294.

Ashton, JK, Gerrard, Bill, & Hudson, Robert. (2003). Economic impact of national sporting success: evidence from the London stock exchange. *Applied Economics Letters, 10*(12), 783-785.

Asquith, Paul, & Mullins Jr, David W. (1983). The impact of initiating dividend payments on shareholders' wealth. *Journal of business*, 77-96.

Ball, Ray. (2009). The global financial crisis and the efficient market hypothesis: what have we learned? *Journal of Applied Corporate Finance, 21*(4), 8-16.

Barber, Brad M, & Odean, Terrance. (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), 785-818.

Barber, Brad M, Odean, Terrance, & Zhu, Ning. (2006). *Do noise traders move markets?* Paper presented at the EFA 2006 Zurich meetings paper.

Barber, Brad M., & Odean, Terrance. (2001). Boys will be Boys: Gender, Overconfidence, and Common Stock Investment. *The Quarterly Journal of Economics, 116*(1), 261-292. doi: 10.1162/003355301556400

Barberis, Nicholas, & Thaler, Richard. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance, 1*, 1053-1128.

Basu, Sanjoy. (1977). Investment performance of common stocks in relation to their price-earnings ratios: A test of the efficient market hypothesis. *The Journal of Finance, 32*(3), 663-682.

Bauwens, Luc, Ben Omrane, Walid, & Giot, Pierre. (2005). News announcements, market activity and volatility in the euro/dollar foreign exchange market. *Journal of International Money and Finance, 24*(7), 1108-1125.

Beckmann, Daniela, & Menkhoff, Lukas. (2008). Will women be women? Analyzing the gender difference among financial experts. *Kyklos, 61*(3), 364-384.

Bermingham, Adam, & Smeaton, Alan F. (2011). On using Twitter to monitor political sentiment and predict election results.

Berry, Thomas D, & Howe, Keith M. (1994). Public information arrival. *The Journal of Finance, 49*(4), 1331-1346.

Bianconi, Marcelo, Hua, Xiaxin, & Tan, Chih Ming. (2013). Determinants of Systemic Risk and Information Dissemination: Department of Economics, Tufts University.

Black, Fischer. (1986). Noise. *The journal of finance, 41*(3), 529-543.

Bollen, Johan, Mao, Huina, & Pepe, Alberto. (2011). *Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena.* Paper presented at the ICWSM.

Bondt, Werner FM, & Thaler, Richard. (1985). Does the stock market overreact? *The Journal of finance, 40*(3), 793-805.

Bruns, Axel, & Highfield, Tim. (2012). Blogs, Twitter, and breaking news: the produsage of citizen journalism. *Produsing Theory in a Digital World: The Intersection of Audiences and Production in Contemporary Theory, 80*, 15-32.

Bruns, Axel, & Liang, Yuxian Eugene. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday, 17*(4).

Camelia, Oprean. (2012). A Behavioral Finance Perspective Of The Efficient Market Hypothesis. *Annals of University of Craiova - Economic Sciences Series, 3*(40), 159-164.

Case, Carl J, & King, Darwin L. (2011). Twitter Usage in the Fortune 50: A Marketing Opportunity? *Journal of Marketing Development and Competitiveness, 5*(3), 94-103.

Chamorro-Premuzic, Tomas. (2007). *Personality and individual differences*: Blackwell publishing.

Chen, Gongmeng, Kim, Kenneth A, Nofsinger, John R, & Rui, Oliver M. (2007). Trading performance, disposition effect, overconfidence, representativeness bias, and experience of emerging market investors. *Journal of Behavioral Decision Making, 20*(4), 425-451.

Chen, Hsinchun, Reid, Edna, Sinai, Joshua, Silke, Andrew, & Ganor, Boaz. (2008). *Terrorism informatics: Knowledge management and data mining for homeland security* (Vol. 18): Springer.

Coates, John M., Gurnell, Mark, & Rustichini, Aldo. (2009). Second-to-fourth digit ratio predicts success among high-frequency financial traders. *Proceedings of the National Academy of Sciences, 106*(2), 623-628. doi: 10.1073/pnas.0810907106

Da Costa Jr, Newton, Goulart, Marco, Cupertino, Cesar, Macedo Jr, Jurandir, & Da Silva, Sergio. (2013). The disposition effect and investor experience. *Journal of Banking & Finance, 37*(5), 1669-1675. doi: http://dx.doi.org/10.1016/j.jbankfin.2012.12.007

Daniel, Kent, Hirshleifer, David, & Subrahmanyam, Avanidhar. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance, 53*(6), 1839-1885.

Darrat, Ali F, Zhong, Maosen, & Cheng, Louis TW. (2007). Intraday volume and volatility relations with and without public news. *Journal of Banking & Finance, 31*(9), 2711-2729.

DeLong, J Bradford, Shleifer, Andrei, Summers, Lawrence H, & Waldmann, Robert J. (1987). The economic consequences of noise traders: National Bureau of Economic Research Cambridge, Mass., USA.

DeLong, J Bradford, Shleifer, Andrei, Summers, Lawrence H, & Waldmann, Robert J. (1988). The survival of noise traders in financial markets: National Bureau of Economic Research Cambridge, Mass., USA.

Dergiades, Theologos. (2012). Do investors' sentiment dynamics affect stock returns? Evidence from the US economy. *Economics Letters, 116*(3), 404-407.

Dewachter, Hans, Erdemlioglu, Deniz, Gnabo, Jean-Yves, & Lecourt, Christelle. (2014). The intra-day impact of communication on euro-dollar volatility and jumps. *Journal of International Money and Finance*.

Dorn, Daniel, & Sengmueller, Paul. (2009). Trading as entertainment? *Management Science, 55*(4), 591-603.

Dougal, Casey, Engelberg, Joseph, Garcia, Diego, & Parsons, Christopher A. (2012). Journalists and the stock market. *Review of Financial Studies, 25*(3), 639-679.

Edelen, Roger M, Marcus, Alan J, & Tehranian, Hassan. (2010). Relative sentiment and stock returns. *financial analysts journal, 66*(4), 20-32.

Engelberg, Joseph E, & Parsons, Christopher A. (2011). The causal impact of media in financial markets. *The Journal of Finance, 66*(1), 67-97.

Evans, Kevin, & Speight, Alan. (2010). International macroeconomic announcements and intraday euro exchange rate volatility. *Journal of the Japanese and International Economies, 24*(4), 552-568.

Fama, Eugene F. (1970). Efficient capital markets: A review of theory and empirical work*. *The journal of Finance, 25*(2), 383-417.

Fama, Eugene F, Fisher, Lawrence, Jensen, Michael, & Roll, Richard. (1969). The adjustment of stock prices to new information. *International economic review, 10*(1), 1-21.

Finnerty, Joseph E. (1976). Insiders and market efficiency. *The Journal of Finance, 31*(4), 1141-1148.

Flynn, Simone I. (2014). Behavioral Finance (pp. 1-1): Great Neck Publishing.

Friedman, Milton. (1953). *Essays in positive economics*: University of Chicago Press.

Fromlet, Hubert. (2001). Behavioral finance-theory and practical application. *Business Economics, 36*(3), 63-69.

Geoffrey Booth, G, Kallunki, Juha-Pekka, & Martikainen, Teppo. (1999). Earnings news and the behaviour of large and small traders in the Finnish stock market. *Applied Economics Letters, 6*(12), 771-774.

Go, Alec, Huang, Lei, & Bhayani, Richa. (2009). Twitter sentiment analysis. *Entropy, 17*.

Granger, Clive WJ. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.

Griffin, John M, Harris, Jeffrey H, & Topaloglu, Selim. (2003). The Dynamics of Institutional and Individual Trading. *The Journal Of Finance, 58*(6).

Grinblatt, Mark, & Keloharju, Matti. (2009). Sensation seeking, overconfidence, and trading activity. *The Journal of Finance, 64*(2), 549-578.

Groenewold, Nicolaas, & Kang, Kuay Chin. (1993). The Semi-Strong Efficiency of the Australian Share Market*. *Economic Record, 69*(4), 405-410.

Groß-Klußmann, Axel, & Hautsch, Nikolaus. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance, 18*(2), 321-340.

Hoa, Kin-Yip, Shia, Yanlin, & Zhangb, Zhaoyong. (2014). News Sentiment and High-Frequency Volatility Dynamics in the Japanese Stock Market. *Handbook of Asian Finance: REITs, Trading, and Fund Performance*, 285.

Honey, Courtenay, & Herring, Susan C. (2009). *Beyond microblogging: Conversation and collaboration via Twitter.* Paper presented at the System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on.

Jaffe, Jeffrey F. (1974). Special Information and Insider Trading. *The Journal of Business, 47*(3), 410-428. doi: 10.2307/2352458

Jansen, David-Jan, & De Haan, Jakob. (2005). Talking heads: the effects of ECB statements on the euro–dollar exchange rate. *Journal of International Money and Finance, 24*(2), 343-361.

Ji, Li-Jun, Zhang, Zhiyong, & Guo, Tieyuan. (2008). To buy or to sell: Cultural differences in stock market decisions based on price trends. *Journal of Behavioral Decision Making, 21*(4), 399-413.

Joseph, Kissan, Babajide Wintoki, M, & Zhang, Zelin. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting, 27*(4), 1116-1127.

Kahneman, Daniel, & Tversky, Amos. (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, 263-291.

Kalev, Petko S, Liu, Wai-Man, Pham, Peter K, & Jarnecic, Elvis. (2004). Public information arrival and volatility of intraday stock returns. *Journal of Banking & Finance, 28*(6), 1441-1467.

Kamstra, Mark J, Kramer, Lisa A, & Levi, Maurice D. (2003). Winter blues: A SAD stock market cycle. *American Economic Review*, 324-343.

Kaplanski, Guy, & Levy, Haim. (2010). Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics, 95*(2), 174-201.

Karabulut, Yigitcan. (2011). Can Facebook predict stock market activity? *SSRN eLibrary*.

Khan, AQ, & Ikram, Sana. (2010). Testing Semi-Strong Form of Efficient Market Hypothesis in Relation to the Impact of Foreign Institutional Investors'(FII's) Investments on Indian Capital Market. *International Journal of Trade, Economics & Finance, 1*(4).

Kirkpatrick, Marshall. (2009). How one teacher uses twitter in the classroom. *ReadWriteWeb. Retrieved May, 20*, 2011.

Kouloumpis, Efthymios, Wilson, Theresa, & Moore, Johanna. (2011). Twitter sentiment analysis: The good the bad and the omg! *ICWSM, 11*, 538-541.

Kourtidis, Dimitrios, Šević, Željko, & Chatzoglou, Prodromos. (2011). Investors' trading activity: A behavioural perspective and empirical results. *The Journal of Socio-Economics, 40*(5), 548-557.

Laakkonen, Helinä, & Lanne, Markku. (2008). Asymmetric news effects on volatility: Good vs. bad news in good vs. bad times.

Lee, Charles. (1992). Earnings news and small traders: An intraday analysis. *Journal of Accounting and Economics, 15*(2), 265-302.

Lee, Thomas Y. (2007). *Needs-based analysis of online customer reviews.* Paper presented at the Proceedings of the ninth international conference on Electronic commerce.

Lee, Thomas Y, & Bradlow, Eric T. (2011). Automated marketing research using online customer reviews. *Journal of Marketing Research, 48*(5), 881-894.

Leung, Cane, & Chan, Stephen CF. (2009). Sentiment Analysis of Product Reviews.

Lim, Kian-Ping, & Brooks, Robert D. (2010). Why do emerging stock markets experience more persistent price deviations from a random walk over time? A country-level analysis. *Macroeconomic Dynamics, 14*(S1), 3-41.

Liu, Bing. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies, 5*(1), 1-167.

Lütje, Torben, & Menkhoff, Lukas. (2007). What drives home bias? Evidence from fund managers' views. *International Journal of Finance & Economics, 12*(1), 21-35.

Lux, Thomas. (2011). Sentiment dynamics and stock returns: the case of the German stock market. *Empirical economics, 41*(3), 663-679.

Malkiel, Burton G. (2003). The efficient market hypothesis and its critics. *Journal of economic perspectives*, 59-82.

Malkiel, Burton G, & Fama, Eugene F. (1970). Efficient capital markets: A review of theory and empirical work*. *The journal of Finance, 25*(2), 383-417.

Mao, Huina, Counts, Scott, & Bollen, Johan. (2011). Predicting financial markets: Comparing survey, news, twitter and search engine data. *arXiv preprint arXiv:1112.1051*.

Martineau, Justin, & Adviser-Finin, Tim. (2011). *Identifying and isolating text classification signals from domain and genre noise for sentiment analysis*: University of Maryland at Baltimore County.

Merton, H. Miller, & Modigliani, Franco. (1961). Dividend Policy, Growth, and the Valuation of Shares. *The Journal of Business, 34*(4), 411-433. doi: 10.2307/2351143

Michaely, Roni, Thaler, Richard H, & Womack, Kent L. (1995). Price reactions to dividend initiations and omissions: overreaction or drift? *The Journal of Finance, 50*(2), 573-608.

Nann, Stefan, Krauss, Jonas, & Schoder, Detlef. (2013). Predictive Analytics On Public Data-The Case Of Stock Markets.

Nasukawa, Tetsuya, & Yi, Jeonghee. (2003). *Sentiment analysis: Capturing favorability using natural language processing.* Paper presented at the Proceedings of the 2nd international conference on Knowledge capture.

Newman, Nic. (2009). The rise of social media and its impact on mainstream journalism. *Reuters Institute for the Study of Journalism, 8*(2), 1-5.

Niederhoffer, Victor. (1971). The Analysis of World Events and Stock Prices. *The Journal of Business, 44*(2), 193-219. doi: 10.2307/2351663

Nofsinger, John R. (2005). Social mood and financial economics. *The Journal of Behavioral Finance, 6*(3), 144-160.

Nofsinger, John R, & Sias, Richard W. (1999). Herding and feedback trading by institutional and individual investors. *The Journal of Finance, 54*(6), 2263-2295.

Odean, Terrance. (1999). Do Investors Trade Too Much? *The American Economic Review*.

Oh, Chong, & Sheng, Olivia. (2011). Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement.

Olsen, Robert A. (1998). Behavioral finance and its implications for stock-price volatility. *Financial analysts journal*, 10-18.

Pak, Alexander, & Paroubek, Patrick. (2010). *Twitter as a Corpus for Sentiment Analysis and Opinion Mining.* Paper presented at the LREC.

Pang, Bo, & Lee, Lillian. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval, 2*(1-2), 1-135.

Quercia, Daniele, Askham, Harry, & Crowcroft, Jon. (2012). *TweetLDA: supervised topic classification and link prediction in Twitter.* Paper presented at the Proceedings of the 3rd Annual ACM Web Science Conference.

Rao, Tushar, & Srivastava, Saket. (2012). Twitter Sentiment Analysis: How To Hedge Your Bets In The Stock Markets. *arXiv preprint arXiv:1212.1107*.

Rao, Tushar, & Srivastava, Saket. (2013). *Modeling movements in oil, gold, forex and market indices using search volume index and Twitter sentiments*. Paper presented at the Proceedings of the 5th Annual ACM Web Science Conference, Paris, France.

Ravisankar, Pediredla, Ravi, V, Raghava Rao, G, & Bose, Indranil. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems, 50*(2), 491-500.

Riordan, Ryan, Storkenmaier, Andreas, Wagener, Martin, & Sarah Zhang, S. (2013). Public information arrival: Price discovery and liquidity in electronic limit order markets. *Journal of Banking & Finance, 37*(4), 1148-1159.

Roe, Brian E, & Just, David R. (2009). Internal and external validity in economics research: Tradeoffs between experiments, field experiments, natural experiments, and field data. *American Journal of Agricultural Economics, 91*(5), 1266-1271.

Ross, Stephen A. (1989). Information and volatility: The no-arbitrage martingale approach to timing and resolution irrelevancy. *The Journal of Finance, 44*(1), 1-17.

Schlarbaum, Gary G, Lewellen, Wilbur G, & Lease, Ronald C. (1980). Realized returns on common stock investments: The experience of individual investors. *NBER Working Paper*(R0021).

Schram, Arthur. (2005). Artificiality: The tension between internal and external validity in economic experiments. *Journal of Economic Methodology, 12*(2), 225-237.

Shiller, Robert J, Fischer, Stanley, & Friedman, Benjamin M. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 457-510.

Shleifer, Andrei, & Summers, Lawrence H. (1990). The noise trader approach to finance. *Journal of Economic perspectives, 4*(2), 19-33.

Siganos, Antonios, Vagenas-Nanos, Evangelos, & Verwijmeren, Patrick. (2014). Facebook's daily sentiment and international stock markets. *Journal of Economic Behavior & Organization*(0). doi: http://dx.doi.org/10.1016/j.jebo.2014.06.004

Singh, VK, Piryani, R, Uddin, A, & Waila, P. (2013). *Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification.* Paper presented at the Automation, Computing, Communication, Control and Compressed Sensing (iMac4s), 2013 International Multi-Conference on.

Smales, Lee A. (2012). *Non-Scheduled News Arrival and High-Frequency Stock Market Dynamics: Evidence from the Australian Securities Exchange.* Paper presented at the 25th Australasian Finance and Banking Conference.

Smales, Lee A. (2013). News sentiment and the investor fear gauge. *Finance Research Letters*.

Smith, Graham, & Dyakova, Aneta. (2014). African Stock Markets: Efficiency and Relative Predictability. *South African Journal of Economics, 82*(2), 258-275. doi: 10.1111/saje.12009

Statman, Meir. (1999). Behaviorial finance: Past battles and future engagements. *Financial Analysts Journal*, 18-27.

Tetlock, Paul C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance, 62*(3), 1139-1168.

Tsype, Abby. (2013, 2013-05-21). Our Social Trading Network Celebrates 50 million Positions. Retrieved 2014-08, 2014, from http://www.etoro.com/blog/markets/21052013/our-social-trading-network-celebrates-50-million-positions/?dl=30000716/&utm_medium=Media&utm_source=46599&utm_content=4731&utm_serial=infograph5.aspx&utm_campaign=infograph5.aspx&utm_term=http://paid.outbrain.com/network/redir#

Tumasjan, Andranik, Sprenger, Timm Oliver, Sandner, Philipp G, & Welpe, Isabell M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM, 10*, 178-185.

Tversky, Amos, & Kahneman, Daniel. (1981). The framing of decisions and the psychology of choice. *Science, 211*(4481), 453-458.

Twitter.com. (2014). About Twitter, Inc Retrieved 2014-07, 2014, from https://about.twitter.com/company

Vieira, Elisabete F Simões. (2012). Investor sentiment and market reaction: evidence on 2010 FIFA World Cup. *International Journal of Economics and Accounting, 3*(1), 51-76.

Vieira, Elisabete Simões. (2011). Investor sentiment and the market reaction to dividend news: European evidence. *Managerial Finance, 37*(12), 1213-1245.

Wang, Gang, Wang, Tianyi, Wang, Bolun, Sambasivan, Divya, Zhang, Zengbin, Zheng, Haitao, & Zhao, Ben Y. (2014). Crowds on Wall Street: Extracting Value from Social Investing Platforms. *arXiv preprint arXiv:1406.1137*.

Wang, N, Kosinski, M, Stillwell, DJ, & Rust, J. (2014). Can well-being be measured using Facebook status updates? Validation of Facebook's Gross National Happiness Index. *Social Indicators Research, 115*(1), 483-491.

Xu, Feifei. (2012). Data Mining in Social Media for Stock Market Prediction.

Yoo, Kyung Hyan, & Gretzel, Ulrike. (2009). *Comparison of Deceptive and Truthful Travel Reviews.* Paper presented at the ENTER.

Zhang, Wenbin, & Skiena, Steven. (2009). Trading strategies to exploit news sentiment. *Submitted for publication*.

Zhang, Wenbin, & Skiena, Steven. (2010). *Trading Strategies to Exploit Blog and News Sentiment.* Paper presented at the ICWSM.

Zhang, Xue, Fuehres, Hauke, & Gloor, Peter A. (2011). Predicting stock market indicators through twitter "I hope it is not as bad as I fear". *Procedia-Social and Behavioral Sciences, 26*, 55-62.

APPENDIX A: Tweet Fields

Column Name	Description	Sample Value
TweetID	Twitter's internal unique identifier	487426067695935000
Date Inserted	Date and time the tweet was saved to the local database	11/07/2014 04:45
Created Date	The date and time (in local time) the tweet was posted	11/07/2014 04:40
Text	The tweet text	The #Lew and #Woolworths saga continues - is he hiding derivatives under his belt thus sitting mum? #DavidJones #CountryRoad
is Retweet	Boolean value of if the tweet was a retweet	0
Retweet Count	The number of times this tweet was retweeted	0
Coordinate Longitude	Coordinates from where the tweet was posted	18
Coordinate Latitude	Coordinates from where the tweet was posted	-34
Creator Name	The full name of the person posting the tweet	Akanksha Malik-Nair
Creator Description	The Twitter bio of the person who created the tweet	Full of life, Investment Professional with an insatiable passion to learn & grow. A thinker, a doer, an optimist & an absolutely wonderful wife ;) living in
Creator Follower Count	The number of followers following the creator of the tweet	1221
Creator ID	Twitters internal ID identifying the creator	403973061
Creator ScreenName	The screen name of the person posting the tweet	AkankshaMalikNr
Creator Verified	Boolean value of If the profile has been verified by Twitter	0
InReply To ScreenName	In the case where the tweet is in reply to another tweet, this will be the screen name of the person replied to	NULL
InReply To StatusID		NULL
InReply To UserID		NULL
Hashtag Count	The number of hashtags used in the tweet	4
Hashtag Serialised	A serialised list of all hashtags used in the tweet	Lew, Woolworths, DavidJones, CountryRoad
URL Count	The number of links to external web sites included in the tweet	0
URL Serialised	A Serialised list of all external web sites included in the tweet	NULL
Media Count	The number of links to external media such as images included in the tweet	NULL
Media URL Serialised	A serialised list of links to external media such as images included in the tweet	NULL

APPENDIX B: Bloomberg Data Fields

Field Name	Description	Sample Value
Symbol	The exchange symbol for selected entity	Goog
Date	The date and time in local time zone	01/04/2014 15:35
Price Open	The opening price during the timing window	560.48
Price Close	The closing price of the timing window	560.60
Low	The lowest trade price during timing window	560.48
High	The Highest trade price during timing window	562.28
Volume	Volume of shares traded	100
Value	The Value in local currency of shares traded	56048
Tick Count	The number of individual trades during snapshot window	129

APPENDIX C: Tweet Volumes

Exchange	Company Name	May	June	July	August	September	Total
JSE	Africa Rainbow Minerals		2	12	17	29	60
	Anglo American	87	70	281	161	98	697
	Anglo Platinum		5	11	45	49	110
	AngloGold Ashanti		1	40	123	186	350
	Aspen	2	4				6
	Assore		1			2	3
	Bidvest		4	59	132	112	307
	British American Tobacco	1,106	938	826	780	1,211	4,861
	Discovery		4	17	60	68	149
	Exxaro		17	40	85	46	188
	Glencor	356	289	220	344	185	1,394
	Goldfields		2	5	44	26	77
	GrowthPoint		5	26	87	32	150
	Impala Platinum		5	67	86	44	202
	Imperial		2	4	26	16	48
	Investec Limited		4	4	18	14	40
	Kumba		6	180	88	74	348
	Massmart		1	28	88	43	160
	Mediclinic		5	36	68	54	163
	Mondi Limited			8	12	10	30
	Naspers	154	172	117	129	56	628
	Nedbank		47	237	318	318	920
	Old Mutual		28	187	261	174	650
	Rand Merchant Bank			2	23	21	46
	Remgro		26	104	60	98	288
	Richemont	57	20				77
	SAB Miller	208	167	311	301	757	1,744
	Sanlam		14	93	137	122	366

	Sasol	231	417	441	403	342	1,834
	Shoprite		63	729	563	672	2,027
	Standard Bank	24	121	3,329	5,489	5,152	14,115
	Tiger Brands		8	24	55	183	270
	Truworths		7	39	103	62	211
	Vodacom		27	1,612	911	1,135	3,685
	Woolworths		29	985	1,131	1,088	3,233
LSE	Aberdeen Asset Management	47	33	107	19	48	254
	Admiral Group	310	254	1,005	582	326	2,477
	Aggreko	24	62	73	123	58	340
	Antofagasta	362	424	576	352	456	2,170
	Ashtead Group	247	155	153	114	155	824
	Associated British Foods PLC		4	95	20	150	269
	AstraZeneca PLC		1		1	1	3
	Aviva PLC		40	1,330	797	1,245	3,412
	Babcock International Group		21	34	34	43	132
	Barclays PLC			4,083	3,504	4,083	11,670
	Barratt Developments PLC		35	101	69	136	341
	BG Group PLC		61	627	260	123	1,071
	BHP Billiton PLC		61	189	182	168	600
	British American Tobacco PLC		1	28	11	11	51
	British Land Co PLC		20	109	29	63	221
	British Sky Broadcasting Group		38	637	203	228	1,106
	EasyJet	114	91	177	76	114	572
NASDAQ	Apple	22,654	40,882	47,465	29,047	65,688	205,736
	Cisco	4,113	2,735	3,271	4,305	2,224	16,648
	Facebook	22,977	20,223	31,124	15,458	15,253	105,035

	Gilead Sciences	2,102	3,763	7,026	5,438	4,850	23,179
	Google	19,935	19,820	16,138	9,611	5,552	71,056
	Intel Corporation	2,167	4,869	6,165	1,585		14,786
	Microsoft	12,582	9,974	17,173	7,974	8,113	55,816
	QUALCOMM Incorporated	940	1,286	2,914	1,446	999	7,585

Table 11: Tweet Volumes per company

APPENDIX D: Company Selection

Exchange Name	Company Name	Search Query
FTSE	Aberdeen Asset Management	$ADN OR $LSEADN OR $LSE:ADN
FTSE	Admiral Group	$ADM OR $LSEADM OR LSE:ADM
FTSE	Aggreko	$AGK OR $LSEAGK OR LSE:AGK OR #Aggreko
FTSE	Anglo American	$AAL OR LSE:AAL OR LSE:AAL
FTSE	Antofagasta	$ANTO OR #Antofagasta
FTSE	ARM Holdings	$ARM
FTSE	Ashtead Group	$AHT OR LSE:AHT OR LSEAHT
FTSE	Associated British Foods PLC	$ABF OR #associatedBritishFood OR $LSEABF OR $LSE:ABF
FTSE	AstraZeneca PLC	$AZN OR #astraZeneca $LSEAZN OR $LSE:AZN
FTSE	Aviva PLC	$avivaplc OR #AVIVAPLC OR #AVIVA OR $AV OR $LSE:AV OR $LSEAV
FTSE	Babcock International Group	$BAB OR $LSEBAB OR $LSE:BAB
FTSE	BAE Systems PLC	$BAE OR $LSEBAE OR $LSE:BAE OR #BAESystems OR #BAESystem
FTSE	Barclays PLC	$BARC OR $LSEBARC OR $LSE:BARC OR #barclays
FTSE	Barratt Developments PLC	$BDEV OR #BDEV OR $LSE:BDEV OR LSE:BDEV OR #barratt
FTSE	BG Group PLC	$BG OR $LSEBG OR LSE:BG OR #BGGroup
FTSE	BHP Billiton PLC	$BLT OR $LSEBLT OR $LSE:BLT
FTSE	British American Tobacco PLC	$BATS OR $LSEBATS OR $LSE:BATS
FTSE	British Land Co PLC	$BLND OR $LSE:BLND OR $LSE:BLND OR #BritishLand
FTSE	British Sky Broadcasting Group	$BSY OR #BskyB OR $LSEBSY OR $LSE:BSY
FTSE	EasyJet	$EZJ OR LSE:EZJ OR LSEEZJ
JSE	ABSA	$JSE:ASA OR #ABSA OR $ASA OR $JSEASA
JSE	Africa Rainbow Minerals	$JSEARI OR $JSE:ARI
JSE	Anglo American	$AGL OR $JSEAGL OR $JSE:AGL OR #AngloAmerican
JSE	Anglo Platinum	$JSE:AMS OR #Angloplat OR $JSEAMS
JSE	AngloGold Ashanti	$JSEANG OR #Anglogold
JSE	Aspen	$JSE:APN OR #AspenPharma OR #aspen or $JSEAPN
JSE	Assore	$JSEASR OR $JSE:ASR #Assore
JSE	BHP Billiton	$JSE:BIL OR #bhpbilliton or $JSEBIL
JSE	Bidvest	$BVT OR $JSE:BVT OR $JSEBVT OR #Bidvest
JSE	British American Tobacco	$BTI OR $JSEBTI OR $JSE:BTI OR "British American Tobacco"
JSE	Discovery	$JSEDSY OR $JSE:DSY
JSE	Exxaro	$JSE:EXX OR $JSEEXX OR #Exxaro

JSE	First Rand Group	$JSE:FSR or #FirstRand OR $FSR OR $JSEFSR
JSE	Glencor	$GLN OR JSEGLN OR $JSE:GLN OR #Glencore
JSE	Goldfields	$JSE:GFI OR $JSEGFI OR #GoldFields
JSE	GrowthPoint	$JSE:GRT OR JSEGRT OR #GrowthPoint
JSE	Impala Platinum	$JSEIMP OR $JSE:IMP OR #Implats
JSE	Imperial	$JSEIPL OR $JSE:IPL
JSE	Investec	$JSEINP OR $JSE:INP OR #Investec
JSE	Investec Limited	$JSEINL OR $JSE:INL
JSE	Kumba	$JSE:KIO OR #Kumba OR $KIO OR $JSEKIO
JSE	Massmart	$JSE:MSM OR $JSEMSM OR #Massmart
JSE	Mediclinic	$JSEMDC OR $JSE:MDC OR #MediClinic
JSE	Mondi	$JSEMNP OR $JSE:MNP
JSE	Mondi Limited	$JSEMND OR $JSE:MND
JSE	MTN	$JSE:MTN OR JSEMTN OR $MTN -vail
JSE	Naspers	$NPN OR $JSENPN OR $JSE:NPN OR #Naspers
JSE	Nedbank	$NED OR $JSE:NED OR #Nedbank
JSE	Old Mutual	$JSE:OML OR #OldMutual OR $OML OR $JSEOML
JSE	Rand Merchant Bank	$JSE:RMH OR $JSERMH
JSE	Remgro	$REM OR $JSE:REM OR #Remgro
JSE	Richemont	$JSE:CFR OR $JSECFR #Richemont -cullen
JSE	SAB Miller	$SAB OR $JSESAB OR $JSE:SAB OR #SABMiller
JSE	Sanlam	$JSESJ OR $JSE:SJ OR #Sanlam
JSE	Sasol	$JSE:SOL OR #SASOL OR $JSESOL
JSE	Satrix top 40	$STX40
JSE	Shoprite	$SHP OR JSE:SHP OR #Shoprite
JSE	Standard Bank	$SBK OR $JSESBK OR $JSE:SBK OR "Standard Bank"
JSE	Steinhoff	$SHF OR $JSE:SHF or $JSESHF
JSE	Tiger Brands	$JSETBS OR $JSE:TBS OR #TigerBrands
JSE	Truworths	$JSE:TRU OR $JSETRU OR #Truworths
JSE	Vodacom	$JSE:VOD OR #Vodacom OR $VOD OR $JSEVOD
JSE	Woolworths	$WHL OR $JSEWHL OR JSE:WHL OR #Woolworths
NASDAQ	Apple	$AAPL
NASDAQ	Cisco	$CSCO OR #CSCO
NASDAQ	Facebook	$FB
NASDAQ	Gilead Sciences	$GILD OR #GILD
NASDAQ	Google	$Goog
NASDAQ	Intel Corporation	$INTC OR # INTC
NASDAQ	Microsoft	$MSFT OR #MSFT

NYSE	Oracle	$ORCL OR #ORCL
NASDAQ	QUALCOMM Incorporated	$QCOM OR #QCOM