Abstract

Sentiment Analysis of Social Media and Financial Markets

Richard Mosse

Stock market prices are somewhat unpredictable. One would expect that they are mainly driven by estimates of future cash flows and profitability as well as being closely tied to the previous day's price, but there are many other factors. This dissertation aims to investigate the relationship between sentiment derived from social media, specifically twitter, and prices in financial markets.

Over the course of this research twitter data was collected for the period December 23rd 2012 to April 8th 2013. Sentiment was extracted from this data using the Rocksteady affect analysis system. Sentiment statistics generated by Rocksteady were aggregated and combined with stock market data. The resultant data was analysed to examine the relationship between sentiment and the returns on financial assets as measured by changes in prices. Returns were examined rather than prices to avoid the effects of multicollinearity among prices as prices are highly correlated with previous prices.

It was found that sentiment was an indicator of returns. However, the impact of sentiment was small and on the majority of days its impact was statistically insignificant. Returns were found to be indicative of sentiment also. Sentiment was found to have a neutral effect on returns in the long run with the overall effect of sentiment on returns being balanced.

The balanced nature of the sentiment implies that the effects of sentiment on returns are temporary which is consistent with existing research in the area for print media.

Sentiment Analysis of Social Media and Financial Markets

Richard Mosse

Master in Computer Science

University of Dublin Trinity College

Supervisor: Professor Khurshid Ahmad

Submitted to the University of Dublin, Trinity College, May 2013

Declaration

I, Richard Mosse, declare that the following dissertation, except where otherwise stated, is entirely my own work; that it has not previously been submitted as an exercise for a degree, either in Trinity College Dublin, or in any other University; and that the library may lend or copy it or any part thereof on request.

Signature　　　　　　　　　　　　　　　　Date

Summary

This dissertation explores the relationship between sentiment derived from the social media platform twitter and financial markets. One of the questions it attempts to answer is if future prices of financial assets can be explained by sentiment derived from social media and vice versa.

The paper explores the theoretical motivations for the research. It is posited that financial traders use the media as a heuristic when making decisions to buy and sell assets. Furthermore, the paper posits that sentiment derived from social media can be used to gauge the effect of this heuristic.

There is some existing research in the field, including that performed by Tetlock (2007) where sentiment derived from the Wall Street Journal column ‘Abreast the market’ was analysed with respect to returns in financial markets which has shown some promising results.

Conducting the present study involved formulating a list of keywords that would match tweets desirable to collect for analysis. Tweets were then logged as they occurred. Tweets were filtered and refined in attempt to reduce the number of off-topic tweets by comparing their contents to a stop list. The language of each tweet was analysed and non-English tweets were discarded. Figures for negative and positive sentiment were generated using the Rocksteady affect analysis system. These sentiment figures were then aggregated and analysed using several statistical measures making use of GRETL, the Gnu Regression, Econometrics and Time-series Library.

It was found that that the role of sentiment as a means of explaining prices was relatively small. It was also found that the impact of sentiment varied from day to day, while the impact of sentiment was statistically significant on certain days it was irrelevant on others. When examining the effect of regression coefficients their overall effect seems to be roughly neutral. This could mean that the effects of sentiment on price are temporary and self-reverting. This is consistent with the work of Tetlock (2007).

With regard to the research aim, it is found that movements in sentiment precede correlated movements in price (indicative sentiment) and vice versa (reactive sentiment). However, this study finds no clear way to distinguish between indicative and reactive sentiment.

Acknowledgements

This dissertation would not have been possible without the help and support of many people. I would like to thank my supervisor Professor Khurshid Ahmad for introducing me to the topic as well as for all his help, support and valuable time throughout the project. I would like to thank all those involved with the Master's programme at Trinity College Dublin. In particular I would like to thank, Dr Mike Brady, Programme Director.

I would like to thank my friends and classmates, In particular Ian Fitzpatrick, Dara Javaherian and Suzanne Ní Bhriain for their support and advice both in conducting the research and during the writing of the dissertation.

I would also like to express my sincere gratitude to my family for their advice and endless encouragement in my academic endeavours, as well as their financial support. Without them this dissertation would not have been possible.

Table of Contents

Declaration i

Summary ii

Acknowledgements iv

1. Introduction 1

    Structure of Dissertation 1

    Research Objective 1

    Motivation 2

        Financial Theories on Price Determination 3

    Sentiment and the Media 5

        Social Media 5

        Symbiosis of Traditional Media and Social Media 6

    Overview of Methodology 7

2. Background 8

    Previous Research 9

    A Review of Existing Sentiment Analysis Systems 11

3. Methodology 13

    Introduction 13

        Goals 13

    Architecture 14

        Overview 14

        The Logging and filtering of tweets 14

        Rocksteady 18

        Aggregation 18

        GRETL 19

        Limitations of approach 21

4. Research Findings 23

    Correlations 23

    Vector Autoregression Analysis (VAR) 25

    Rolling Regressions 26

    Days Influenced 27

Adjusted R-squared 28
Distribution of Regression Coefficients 29
5. Discussion of Results 31
Correlations 31
Vector Autoregression Analysis 32
Adjusted R^2 33
Distribution of Regression Coefficients 33
6. Conclusion 34
Further work 35
References 36

1. Introduction

Structure of Dissertation

Chapter one sets out the research objectives of the dissertation and the motivation for the work. This is followed by a discussion of social and traditional media with respect to sentiment. Chapter two reviews existing work in the field. It provides an account of previous academic research in the area as well as detailing existing systems. Chapter three discusses the approach to the research and the methodology used to conduct the research. The structure of the system and the rationale for the choices made are discussed. The systems used for collecting and analysing the data are described in detail. The chapter concludes with a critical analysis of the limitations of this approach. Chapter four presents the findings of the study. This is then followed by a discussion of the results in chapter five. The paper concludes by describing the implications of the research and outlines recommendations for further research.

Research Objective

The objective of this research is to investigate the relationship between sentiment in social media and financial time series, including commodities. The research aims to determine whether sentiment could have an effect on asset prices and vice versa. The paper discusses price in terms of change in price. Calculations are performed on returns rather than on prices to reduce the effects of multicollinearity among prices. The paper will explore the extent to which sentiment impacts returns. The research does not attempt to predict market prices with sentiment, but merely to assess the role of sentiment in explaining returns.

Motivation

According to Ahmad (2013) “There is a realisation that the various stakeholders in financial markets across the world that we do not understand fully how prices of financial instruments change with time” (p. 1).

There are many theories as to what determines the price of an asset traded on financial markets. The efficient market hypothesis would have one believe that the market price of an asset comprises all relevant information about the asset. One would expect that prices are mainly driven by estimates of future cash flows and profitability as well as being closely tied to previous prices. In the field of behavioural finance there is much empirical evidence to contradict the efficient market hypothesis. In 2006 Kumar and Lee noted that “The dynamic interplay between noise traders and rational arbitrageurs establishes prices”

The behavioural finance theories of bounded rationality and herd behaviour suggest that investors will sometimes look to outside sources for guidance. There is evidence to suggest that news may be one source to which traders look to for guidance. Increases in volatility of an asset are observed both around scheduled and unscheduled news announcements of which the asset is the subject. It is clear that news can have an effect on prices. The research on which this paper is based investigates the relationship between news and the stock market. Specifically this paper looks to social media as a source of news. It then examines the relationship between news and returns and attempts to measure the impact of news on price.

Financial Theories on Price Determination

Traditional Finance

Traditional finance theory on price determination is incompatible with the idea that sentiment analysis can be used to predict prices. Kumar and Lee (2006) noted “The traditional view posits that the current price of a stock closely reflects the present value of its future cash flows.”(p. 2451).

Traditional finance suggests that new information arising from news media is incorporated into prices without delay. This is based on the efficient market hypothesis described below.

Efficient Market Hypothesis

The efficient market hypothesis is the traditional view as to how prices are determined. The idea behind this hypothesis is that changes in price are not based on previous prices. When new information becomes available through news, word of mouth or any other medium the price of a given asset adjusts immediately to reflect this information. Given this premise, the idea that a financial trader can consistently make gains by analysis of a company or by statistical analysis of past prices is false. The efficient market hypothesis is compared to a random walk where each change in price is a random movement. The efficient market hypothesis argues that a portfolio of stocks chosen by an expert is **not** more likely to have a greater rate of return than a randomly selected portfolio of stocks. (Malkiel, 2003)

However, there are many criticisms of the efficient market hypothesis. Empirical testing of the random walk model has strongly rejected it as a suitable model for financial markets as described by Lo and MacKinlay (1988).

Behavioural Finance

Behavioural finance provides an alternative approach to explaining market behaviour. Behavioural finance breaks away from the idea that traders behave rationally and explores cognitive biases in traders.

Herbert A. Simon (1979) described his theory of bounded rationality which provides some support for the influence of sentiment on the market. Bounded rationality suggests that people are limited by the information available to them as well as their ability to process this information when making decisions. The theory suggests that people are only rational to a degree in decision making and look to heuristics to aid in decision making. The present study puts forward the idea that sentiment in the media could be used as a heuristic in the decision making process of traders. There is evidence to suggest that media sentiment has an effect on the market. One phenomenon that can be observed is volatility clustering.

Volatility Clustering

Volatility is simply variance in returns. A phenomenon called volatility clustering was first observed by Mandelbrot (1963). Mandelbrot noted that “large changes tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes.” (p. 418).

This clustering of variance in returns was posited to be related to clustering of information arrivals as described by Engle (2004). This lends support to the idea that sentiment stemming from new information arrivals such as news has an effect on the market.

Sentiment Analysis

Sentiment analysis involves programmatically processing natural language and extracting the mood of the writer. Often sentiment analysis systems will associate a degree of negativity or positivity with a piece of natural language. A common approach to sentiment analysis is lexicon based sentiment analysis. This involves breaking natural language up into individual words and associating a weight with each word. For example, the word ‘hate’ might be associated with strong negative sentiment and the word ‘love’ might be associated with strong positive sentiment.

The use of domain specific dictionaries improves results. For example, when analysing text regarding oil, a term that is commonly encountered is ‘Crude Oil’. In a general dictionary the word "crude" would have high negative sentiment. However in the context of oil, "crude" is a neutral word.

Sentiment and the Media

Social Media

This study examines social media. Social media was chosen due to its rapidly evolving nature as a source of information. Platforms such as twitter allow its users to publish new information in minutes. This differs greatly from traditional print media which has daily or less frequent release cycles. An interesting trait of social media is that the information has the potential to be more diverse. Topics are not selected and refined by individuals such as journalists or editors. Topics are created by many different individuals and become widespread through other users by retweeting. Retweeting allows a user to share a status posted by another user and will have the effect of reinforcing the sentiment of the original user. One might initially think that this is desirable assuming an individual would retweet a status due to the fact that they are in agreement with it. However, this may not be the case as it is possible that a user might retweet a status because they are in disagreement with it, in the hopes of sharing the status with like-minded people.

This study differs from past work, such as that conducted by Tetlock (2007), in that it attempts to analyse the relationship between sentiment and the market in a more timely manner. Tetlock conducted his research by analysing newspaper articles which are released on a daily basis. On the other hand twitter content is constantly released as new tweets are posted continuously. The hope is that sentiment derived from tweets will evolve throughout the day.

Symbiosis of Traditional Media and Social Media

A phenomenon observed in the course of this research was the symbiosis of traditional media and social media within twitter. News agencies will often have twitter accounts on which they post links to news articles with a one line summary. For example, the following is a tweet posted by Bloomberg News

> “Texas sues BP, Transocean, others involved in 2010 Gulf of Mexico oil spill
>
> | bloom.bg/19KtDav” (Cronin Fisk, Brubaker Calkins, 2013)

FIGURE I EXAMPLE TWEET

As can be seen above this tweet gives a summary of the major point in the article and includes a link to the full article. Tweets such as these are interesting for this study as they allow sentiment from more traditional media to flow into the dataset.

Overview of Methodology

The study focuses on the behaviour of oil stocks, particularly the behaviour of British Petroleum (BP) and West Texas Intermediate (WTI) which is a grade of oil used as a benchmark in oil pricing. The study involved the logging of twitter data over a period of three and a half months from December 23rd 2012 to April 8th 2013. The data was examined in detail and sentiment figures were generated for the time period using the Rocksteady affect analysis system. The resultant sentiment figures were aggregated with pricing data and analysed using econometric methods to determine the nature of the relationship.

2. Background

There are many claims that investor sentiment can be used to predict stock market prices and can be used in electronic trading strategies. There has been some success in using sentiment as an indicator of price according to an article by King & Peterson (2011). As the latter is simply a news article rather than a research paper it lacks detail and rigour so one might be sceptical with respect to the validity of the conclusions drawn from the article. For example, the processes that claim to use sentiment successfully may merely take it into account as one of many factors in the decision making process.

Harnessing social media to derive sentiment is a concept which much previous work predates. Social media sites offer analysis which has the potential to be more up to date than news articles.

Stock brokers are not journalists by trade and therefore would not necessarily write news articles. The emergence and mass adoption of social media has provided a highly accessible platform which makes it easy for anyone to publish information. The phenomenon of microblogging may provide new insights. When looking at twitter, the length of tweets is quite short. There is room here for short snippets of information that would not necessarily justify a news article but might be quite relevant in terms of sentiment. There is potential here for a new insight which was previously unavailable for analysis.

This chapter discusses previous research in the area as well as some of the existing implementations of sentiment analysis systems. Systems that analyse traditional news articles as well as more recent systems such as the Rocksteady affect analysis system developed at Trinity College Dublin which can be used to analyse social media sites such as twitter as well as blogs are discussed.

Previous Research

There have been many studies into the relationship between investor sentiment and returns in financial markets. The results of some seem to contradict each other. Some early studies found little or no evidence to support the claim that sentiment influences financial markets. A study by Tumarkin & Whitelaw (2001), which analysed online financial bulletin boards found that positive returns influence sentiment but not the inverse. They noted that strong positive returns precede days with abnormal message board activity. Perhaps it is of relevance to their findings that internet usage has grown hugely since the time of their research, not to mention participation in social media and bulletin board sites.

Several years later an Equity Market Sentiment Index (EMSI) was developed as part of a study into the relationship between investor sentiment and returns carried out at the University of Massachusetts in Boston by Bandopadhyaya and Jones, (2006). The EMSI was based on articles in the Boston Globe newspaper. The study examined the relationship between the EMSI and The Massachusetts Bloomberg Index (MBI). The results were quite positive and found that lagged values of the EMSI better explain changes in the market index value than lagged values of the market index itself. Furthermore the study also claimed that. “This has important implications since it appears that short-run changes in the market index value are driven primarily by investor sentiment rather than by the indices own price momentum”. (p. 10).

Later a similar study was carried out to investigate the link between traditional media pessimism and negative returns in stock markets. According to Tetlock (2007) in his study, based around a popular column named “Abreast of the Market” in the Wall Street Journal *“High values of media pessimism induce downward pressure on market prices; unusually high or low values of pessimism lead to temporarily high market trading volume.”* (p. 1166).

Although the study found that the effects were temporary it also noted that the impact of pessimism was particularly high and slow to reverse on small stocks.

The two previous studies are interesting as they were performed using a very narrow dataset. One might wonder whether this means that one article can have a huge influence over the stock market or whether the article simply reflects existing investor sentiment, or some combination of the two. It will also be interesting to see how analysis of traditional news articles fares compared to the narrow range of sources examined in these studies.

Probably the most relevant study I have encountered during the course of my research is entitled “Twitter Mood Predicts the Stock Market” (Bollen, Mao, & Zeng, 2011). The study analysed twitter data for sentiment and found that they could predict the closing value of the Dow Jones Industrial Average Index to a degree of accuracy of 87.6%. This has huge implications for the use of sentiment analysis in forecasting variables in financial markets and shows that there may be huge potential for generating revenue here.

A Review of Existing Sentiment Analysis Systems

There are many sentiment analysis systems available to traders at the moment and companies are gradually beginning to realise the potential of exploiting news. King & Peterson, (2011) in their article “Trading on a World of Sentiment” discussed a now closed investment fund called MarketPsy. The fund took a different approach to making use of sentiment analysis. The Fund looked for negative sentiment in news articles and attempted to spot overreactions in the market from which a stock would later recover. The fund had great success in doing this and went on to outperform the Standard & Poor’s 500-stock index. There are several companies which now conduct sentiment analysis of financial news. A brief discussion of some of the more well-known sentiment analysis systems is given below.

RavenPack is a company which analyses Dow Jones newswires and the Wall Street Journal for sentiment and ties their results back to individual stock market instruments. The Dow Jones alone provides over 19,000 newswires per day. The company takes into account only news articles and does not analyse social media.

Bloomberg L.P. takes a similar approach. They offer an integrated trading product called the “Bloomberg Professional Service”. This contains many different applications which provide different forms of analysis and rankings of financial instruments. The company produces over 6,000 news articles a day itself but also integrates many other news sources into this core product. Bloomberg has a news sentiment figure which it attaches to each financial instrument that is supported by its service. The figure is calculated from all the news articles available through the service. Bloomberg integrates their sentiment figure into their service by displaying it alongside different and unrelated forms of analysis, rather than providing it as a standalone statistic to advise traders. However, the company also provides data through an API which allows the news sentiment to be retrieved programmatically.

Thomson Reuters takes the idea a little further. As well as analysing news articles from various sources they also take into account data from social media sites. Reuters takes articles from 50,000 news sites and four million social media sites. The results from this service are particularly interesting as they combine both forms of media.

The aforementioned are all services available to the public but do not necessarily provide any evidence of success in providing positive returns.

Derwent Capital, a London based hedge fund analyses data from twitter for sentiment and invests accordingly. This firm claims that it can predict the closing prices of the Dow Jones Industrial Average stock market index with 87.6% accuracy. This figure should seem familiar. The work of the hedge fund leverages the work described in the paper “Twitter mood predicts the stock market” mentioned previously, which the fund credits much of their success to.

Sentiment analysis for this dissertation was conducted using the Rocksteady affect analysis system. In the past the system has been used to analyse news and blog articles. Rocksteady was used to analyse news articles preceding the Irish general election in February 2011 and successfully distinguished between the winning coalition and the losing parties (Ahmad, Daly, & Liston, 2011).

3. Methodology

Introduction

This chapter gives a detailed description of the research carried out. The aims of the research are stated. A detailed description of the system used to collect and analyse the data is given. Each component of the system is discussed in detail and the rationale for the choices made is provided. The chapter concludes with a critical discussion of the limitations of the study.

Goals

The goal of this research is to determine the relationship between sentiment gathered from the social media platform twitter and returns on financial assets. The research attempts to examine the effect of sentiment on returns as well as the effect of returns on sentiment. The research seeks to discover if there is a causal relationship between sentiment and returns. Furthermore, it attempts to ascertain if the effect of sentiment is constant or if its importance varies across time.

Architecture

Overview

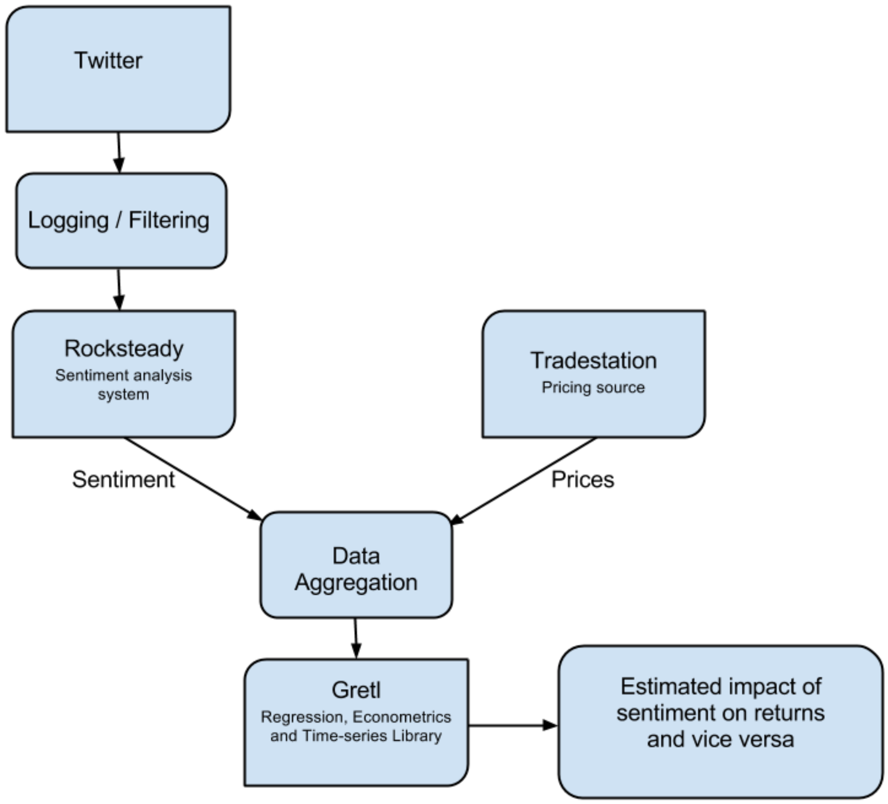

FIGURE II: SYSTEM OVERVIEW

The Logging and filtering of tweets

The first module of the system is involved with logging and filtering of tweets. All tweets matching the elements of a list of predetermined keywords (discussed further on in this chapter) enter the system. Tweets are then filtered and logged. The logged data is then processed in an attempt to reduce the number of irrelevant or "off-topic" tweets.

Twitter Streaming API

Tweet data is collected in real time. Twitter provides a streaming API to provide clients with a low latency method of obtaining tweets. The client provides twitter with a list of keywords in a logical **OR** and logical **AND** format. After this subscription has been made all new tweets which match this query are streamed directly to the client as depicted in Figure III.

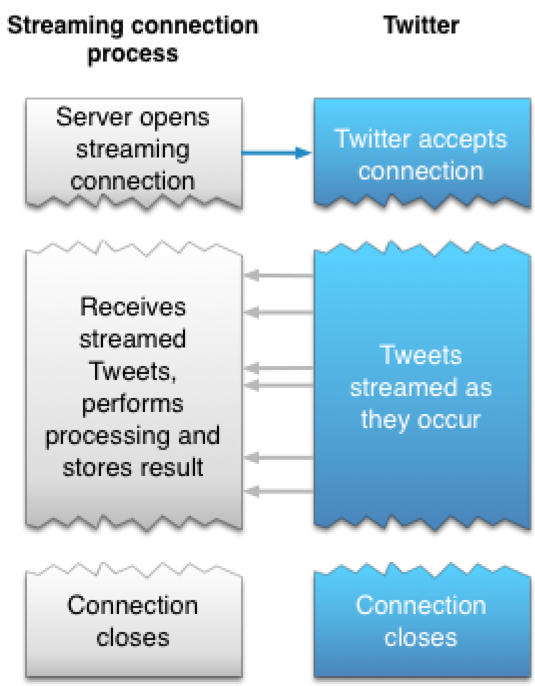

FIGURE III TWITTER STREAMING CONNECTION ("THE STREAMING APIS ", 2012)

The Selection of keywords

A list of 250 terms was selected to determine which tweets would be recorded. Tweets matching any term in the list were streamed to the application. The list consisted mainly of terms refined from a specialist dictionary created for the Rocksteady affect analysis system. The dictionary consisted of terms relating to oil and finance. Ambiguous terms were removed to avoid triggering tweets on unrelated topics. The list consisted of: Stock symbols for supermajor oil companies; names of companies; names of key oil fields, ports, and refineries; as well as oil related terms such as oil sands and tar sands.

Data processing

The data streamed from twitter arrives in JSON format. It includes metadata such as a language tag. For the purposes of this study we are only interested in English language tweets so all non-English tweets are discarded. It was observed that reliance on the language tags provided by twitter was unsatisfactory. Each tweet was analysed before it was logged to determine its language.

Language was determined using the langid library developed at the University of Melbourne by Lui & Baldwin (2012). The system uses a naive Bayes classifier to determine a language. The most probable language is accompanied by a probability score in the range 0 to 1. All tweets determined to be English with a probability score above .9 were accepted. This methodology was found to be highly effective and the number of valid tweets discarded was negligible.

After this process, there remained a large degree of off-topic tweets within the dataset. Further attempts to minimise these tweets were made by filtering them against a stop list. The stop list contained words or phrases commonly found in tweets which were off-topic. Tweets containing any of the terms on the stop list were removed. Furthermore it was noted that the vast majority of tweets containing obscenities or emoticons were off-topic. Hence tweets containing obscenities or emoticons were removed.

It was found that after this process was completed that the majority of the remaining tweets were on the subject of oil and oil markets. It was found that on average 2.88 tweets were logged per minute across the three and half month period.

Fault tolerance

One of the challenges in collecting data in this manner is that the streaming connection with twitter is subject to disconnection. One of the most common reasons the system can be disconnected is if tweets are not parsed quickly enough. This happens when the client is found to be reading tweets too slowly and a backlog of tweets builds up on the server side. When the server side backlog grows too large, twitter breaks the connection with the client.

The connection can also be terminated in the event of a network failure, software failure (either client or server side) or a hardware failure. It is important to minimise data loss in these situations or avoid such situations altogether where possible. The system was designed to attempt to reconnect in the event of a disconnection. The software was monitored and was automatically restarted in the event of a crash. The software was launched on the start-up of the machine to recover in the event of a machine failure.

Rocksteady

About Rocksteady

Rocksteady is an affect analysis system developed at Trinity College Dublin. According to Ahmad, Daly, & Liston, (2011) “The Rocksteady system uses a combination of general purpose affect dictionaries, like Stone’s General Inquirer Dictionary, and an optional domain specific dictionary” (p. 84).

This project uses Rocksteady to calculate positive and negative sentiment figures for the twitter data set using a domain specific dictionary for oil and finance. Rocksteady can compute sentiment figures and aggregate them for any time period. For the purposes of this study data was aggregated on an hourly basis.

Aggregation

Pricing data was obtained through Tradestation which is a technical analysis and trading platform. Using this platform the pricing data can be exported to a text file. Hourly pricing data for British Petroleum (BP) and West Texas Intermediate (WTI) was exported from Tradestation. From this data returns were calculated.

Returns were chosen as opposed to prices as prices are highly correlated. Changes in prices on the other hand have little or no correlation. Specifying the price variables in this form reduces the effects of multicollinearity in the model.

The formula for calculating returns is detailed below in equation I.

$$\mathrm{Return} = \log\left(\frac{\mathrm{P_t}}{\mathrm{P_{t-1}}}\right)$$

$$\mathrm{P\ =\ price, t = today}$$

EQUATION I

The pricing data was then aggregated with the sentiment figures calculated using Rocksteady.

GRETL

About GRETL

GRETL stands for GNU Regression, Econometrics and Time-series Library. Gretl is based on the ESL (Econometric Software Library) developed at the University of California, San Diego. GRETL claims to be the first complete open source econometric software package to be released under the GNU software license. The system has a graphical user interface as well as a command line interface. The command line interface allows the user to create scripts to perform tasks such as rolling regressions discussed in detail in the next section.

The software supports many econometric models for the analysis of financial time series (Baiocchi, Distaso, 2003).

For the purposes of this project GRETL was used for performing Vector Autoregression analysis.

Correlation (Pearson product-moment correlation coefficient)

The first step in the analysis of the dataset was to calculate correlation coefficients. Calculating the correlation coefficients makes evident any direct correlations between sentiment and returns in the dataset. Equation II below gives the formula used to calculate the correlation coefficients.

$$r = \frac{\sum (x-\bar{X})(y-\bar{Y})}{\sqrt{\left[\sum (x-\bar{X})^2\right]\left[\sum (y-\bar{Y})^2\right]}}$$

EQUATION II

The data was analysed for both end of day prices as well as hourly prices. However this approach does not take into account lagged values for returns. This is where regression analysis is used.

Regression Analysis

Regression analysis is a statistical method for estimating the relationship among variables. Specifically, for the purpose of this study Vector Autoregression (VAR) was used to examine the dataset as it allows for analysis of multiple time series.

The formula for calculating VAR estimates is detailed below in Equation III

$$r_t = \alpha_0 + \alpha_1 r_{t-1} + \alpha_2 r_{t-2} + \beta_1 s_t + \beta_2 s_{t-1} + \varepsilon_t$$

r = returns; α, β = coefficients; s = sentiment; ε = error term

EQUATION III

VAR estimates are based on two types of variables namely endogenous variables and exogenous variables. Endogenous variables are those that are internal to the system. For example, lagged values of price are endogenous to price. Similarly exogenous variables are those that are external to the system, i.e. sentiment in this study. Sentiment is an outside factor which is expected to affect price.

VAR allows for the calculation of a value (in this case returns) based on lagged values of itself whilst also taking into account exogenous variables (in this case sentiment). Using VAR we can determine the impact of sentiment on returns and vice versa as well as the statistical significance of this impact.

Adjusted Coefficient of Determination

The basic principle when performing regressions is to create a scatter plot of two variables, in the case of this study returns and sentiment. The next step is to attempt to fit a regression line to the data. The Adjusted Coefficient of Determination often referred to as the Adjusted R^2 describes the goodness of fit of the regression line to the data.

Limitations of approach

The main limitation of this approach is the presence of irrelevant tweets in the data set. Although these have been minimised, their presence still takes away from the purity of the dataset. If for example, the study had looked at oil blogs collected from a domain specific blog site such as “The Oil Drum”; there would be no presence of off-topic data.

Another limitation is that negative and positive sentiment can be interpreted in multiple ways here. The data is on the subject of oil and oil stocks. Tweets may be discussing the price of oil in relation to wars or conflict using strong negative terms. This would generate negative sentiment and would be correlated with a rise in oil prices. The alternative would be people discussing a specific financial asset such as the share price of BP. If tweets here were discussing a fall in oil prices they would be associated with strong negative sentiment. This means that the sentiment figures could be slightly skewed. For the dataset collected the vast majority of the tweets refer specifically to financial assets. However it is possible that the presence of tweets discussing rising oil prices in a negative sense reduce the effect of tweets referring to gains in financial assets in a positive sense and vice versa. This combined with the effect of a certain amount of 'off-topic' tweets within the dataset leaves room for noise within the sentiment calculations. These factors may reduce the accuracy of determining the impact of sentiment in the marketplace.

Twitter allows users to "retweet". This process involves a user repeating verbatim a tweet posted by another user. On first impressions one might interpret this as signifying a user in agreement with the original tweet. In the case of this system retweets have the effect of strengthening sentiment in the direction of the original tweet. However this may not be the ideal behaviour. It is possible that a user would retweet a post because he or she strongly disagrees with it and the sentiment attached to it.

4. Research Findings

In this chapter the results and findings of the research are presented. Data was collected and analysed as described in chapter three. The aim of the analysis was to investigate correlations in the data as described in section 1. The next step was to perform vector autoregression analysis. The analysis then went a step further as to fragment the data into individual trading weeks to determine if the relationship was consistent from week to week. Similarly the data was broken into single days to determine on how many days there was a statistically significant impact on returns from sentiment and vice versa.

The coefficient of determination or (Adjusted R-squared) was examined to understand the overall "goodness of fit" of the regression line. Finally, the effect of the sentiment regression coefficients on returns as a whole was examined.

Correlations

Correlations were performed on both daily and hourly data. Table I depicts correlation coefficients calculated for a period of 61 trading days. Sentiment figures for tweets were aggregated on a daily basis. Weekends and days in which trading did not occur were excluded. The correlation coefficient is scaled between 0 and 1 with 1 indicating a perfect correlation and 0 indicating no correlation. The sign of the correlation coefficient indicates whether the relationship was positive or negative.

	Articles	Negative	Positive	WTI_Return	BP_Return
Articles	1.00				
Negative	0.41	1.00			
Positive	-0.39	-0.23	1.00		
WTI_Return	-0.22	0.05	-0.03	1.00	
BP_Return	-0.04	-0.11	-0.11	0.35	1.00

TABLE I DAILY CORRELATION COEFFICIENTS

Similarly Table II depicts correlation coefficients on an hourly basis. It may be noted that the coefficients for daily data are stronger than the coefficients for hourly data in every case.

	Articles	Negative	Positive	WTI_Return	BP_Return
Articles	1.00				
Negative	0.20	1.00			
Positive	-0.12	-0.15	1.00		
WTI_Return	0.02	-0.01	-0.01	1.00	
BP_Return	0.00	-0.01	-0.03	0.05	1.00

TABLE II HOURLY CORRELATION COEFFICIENTS

Vector Autoregression Analysis (VAR)

Vector Autoregression analysis was performed on hourly data for the entire data set. Detailed in table III are the P-values for comparisons of all variables. In all cases where the result was found to be statistically significant this was the case for a single term in the equation only. The lag column in table III indicates the term of the regression equation in which the result was significant where each number represents the number of hours lag the term represented.

Endogenous variable	Endogenous variable	Lag Term	P-value (Significance level)
WTI Return	Negative	5	90+%
WTI Return	Positive	1	99+%
WTI Return	BP Return	-	-
BP Return	Negative	5	90+%
BP Return	Positive	5	90+%
BP Return	WTI Return	5	90+%
Negative	Positive	1	99+%
Negative	BP Return	1	99+%
Negative	WTI Return	1	99+%
Positive	Negative	1	99+%
Positive	BP Return	1	99+%
Positive	WTI Return	1	99+%

TABLE III P-VALUES FOR VECTOR AUTOREGRESSION ANALYSIS

Rolling Regressions

Further from the Vector Autoregression analysis detailed in the previous section the data was broken into units each unit representing a single trading week. There were 15 trading weeks over all. Table III details the results with the fourth column indicating the number of weeks in which the results were statistically significant. It was found that in cases where there was a statistically significant result for the dataset as a whole, this was not the case for each fragment when the dataset was fragmented. This implies that the dataset is heteroskedastic.

Endogenous Variable	Endogenous variable	Lag Term	Significant weeks
WTI Return	Negative	5	5
WTI Return	Positive	1	4
WTI Return	BP Return	-	-
BP Return	Negative	5	1
BP Return	Positive	5	1
BP Return	WTI Return	5	2
Negative	Positive	1	3
Negative	BP Return	1	4
Negative	WTI Return	1	4
Positive	Negative	1	2
Positive	BP Return	1	2
Positive	WTI Return	1	2

TABLE IVROLLING REGRESSIONS

Days Influenced

Further from the Rolling regressions on weekly data the same analysis was performed on daily data, the results are detailed in Table V. In total 62 days were examined. The number of days in which there was a statistically significant impact is detailed in each case in the third column. Similarly to Table IV where the results from the analysis of the data on a per week basis are specified, the dataset was found to be heteroskedastic.

Endogenous variable	Endogenous variable	Days Influenced
WTI Return	Negative	16
WTI Return	Positive	12
WTI Return	BP Return	-
BP Return	Negative	5
BP Return	Positive	4
BP Return	WTI Return	2
Negative	Positive	10
Negative	BP Return	6
Negative	WTI Return	12
Positive	Negative	13
Positive	BP Return	5
Positive	WTI Return	10

TABLE V STATISTICAL SIGNIFICANCE ON A PER DAY BASIS

Adjusted R-squared

The data as a whole was examined to determine the values for adjusted R^2. It was found that the overall impact of sentiment on returns was quite small. The impact of returns on sentiment was higher although still relatively small. The values for the Adjusted R^2 range from 0 to 1 and are detailed below in table VI.

Endogenous Variable	Endogenous variable	Adjusted R-Squared
WTI Return	Negative	.001
WTI Return	Positive	.0009
WTI Return	BP Return	.006
BP Return	Negative	.001
BP Return	Positive	.0007
BP Return	WTI Return	.0006
Negative	Positive	.29
Negative	BP Return	.31
Negative	WTI Return	.29
Positive	Negative	.19
Positive	BP Return	14
Positive	WTI Return	.19

TABLE VI ADJUSTED R^2

Distribution of Regression Coefficients

Regression coefficients describe the unit change effect of one variable as a function of the other variable. All statistically significant regression coefficients where negative and positive sentiment were the exogenous variables and returns was the endogenous variable were aggregated. The purpose of this experiment was to determine the overall effect of sentiment on returns. Figure IV shows the results.

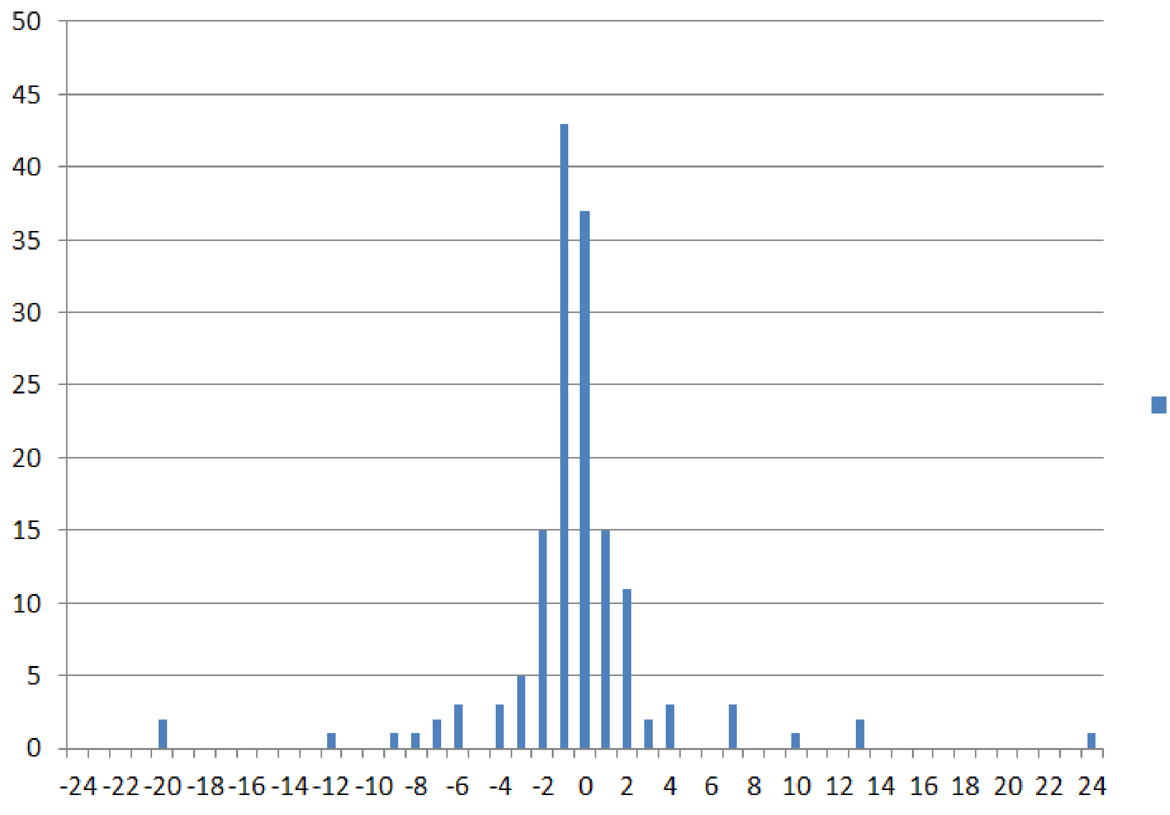

FIGURE IV REGRESSION COEFFICIENTS SENTIMENT VS. RETURNS

As can be seen from figure IV the results are somewhat evenly distributed with most results grouped around 0.

5. Discussion of Results

Correlations

When examining the correlation coefficients, one of the most striking features of the results is that the daily figures are stronger than the hourly figures without exception. A possible explanation is detailed below.

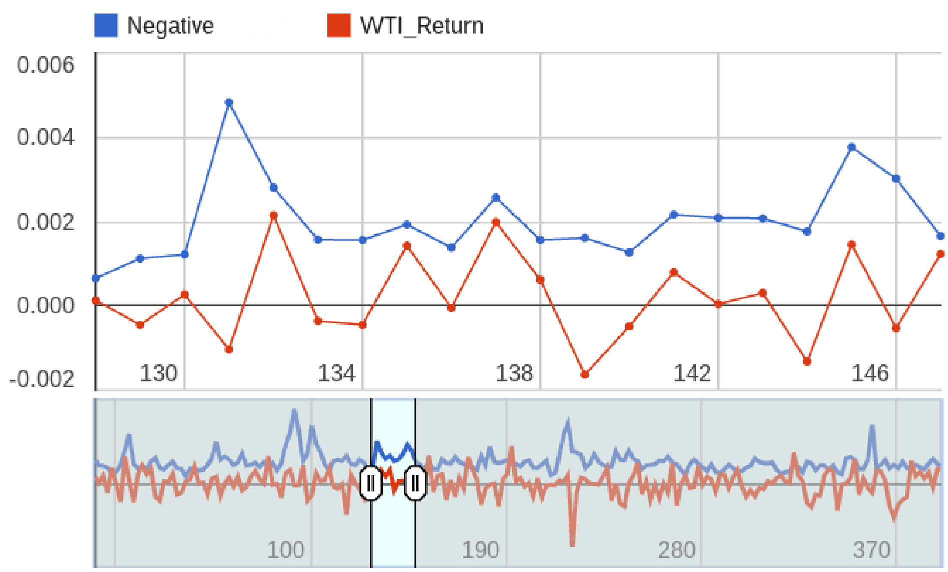

FIGURE V NEGATIVE SENTIMENT VS. WTI (**HOURLY DATA**)

Illustrated above in figure V is a graph detailing returns on WTI and negative sentiment for a cross section of the dataset. It is clear from the trend graph above that sentiment and returns follow a similar pattern. However, after closer inspection, it is noticeable that at data point 130 there is a spike in sentiment which is followed directly by a spike in returns at data point 131. In this case the change in sentiment precedes the change in returns. Conversely at data point 139 a spike in returns precedes a spike in sentiment. Similarly there are many instances in the dataset where a rise or a fall in sentiment coincides with a rise or fall in returns. On a daily basis however, prices are end of day prices and sentiment is accumulated throughout the day. This means that the phenomenon of changes in sentiment occurring before, occurring simultaneously with, or occurring after changes in returns is masked. This is due to the fact that the sentiment figures for the day are aggregated and variations on an hourly scale are hidden.

Vector Autoregression Analysis

One of the problems encountered when examining the correlation coefficients alone is that this process does not take into account lagged values of the variables. This problem is overcome using vector autoregression analysis. It is clear from the results described in chapter 4 that the effect of returns on sentiment is small and the effect of sentiment on returns is smaller. However, we have determined that sentiment appears to play some small role in determining price.

Further from the findings in Vector Autoregression analysis of the entire dataset, when the dataset was fragmented to examine if the results were the same from week to week, this was found not to be the case. It was found that while sentiment was influential some weeks it was statistically insignificant other weeks. These were the first signs in the results that the importance of sentiment's influence varied over time.

When breaking the data up further and exploring the influence of sentiment on returns for individual trading days, the results showed that its influence was only significant on 8-25% of days. Again this shows that sentiment does not always play a significant role in determining price.

Adjusted R^2

The results for the adjusted R^2 again show us that for the dataset as a whole the effect of sentiment on returns is quite minimal. These figures lend support to the argument that the influence of sentiment over time varies.

Distribution of Regression Coefficients

The distribution of regression coefficients as depicted in figure IV are quite interesting. Most of the coefficients are grouped around 0 with the rest dispersed somewhat symmetrically on both the positive and negative sides of the graph. Figure IV shows that the average effect of sentiment is nearly neutral. This behaviour suggests that the effect of sentiment on returns displays the characteristics of mean reversion.

Mean reverting behaviour implies the effects of sentiment on returns are temporary. This is consistent with the findings of Tetlock (2007). There is also some evidence for mean reversion of prices in financial markets as noted by Balvers, Wu and Gilliland (2000). Assuming that sentiment performs in this way in other areas of the market, the mean reverting behaviour of sentiment seen in this study may have the potential to reinforce mean reversion in stock prices.

6. Conclusion

The aim of this research was to investigate the relationship between sentiment and returns in financial markets. Chapter one provided an introduction to the theory which motivated this investigation. In Chapter two existing research in the area was examined and an overview of existing implementations of sentiment analysis systems was given. Chapter three went on to describe the approach to the research and gave a detailed description of how the research was carried out. This was followed by a discussion of some of the limitations of this approach. In Chapter four the findings of the research were presented followed by a discussion of the results in Chapter 5.

This research has found that sentiment derived from social media, specifically twitter, plays only a small role in explaining prices. It was found that the explanatory power of sentiment varies across time and is not statistically significant the majority of the time. It was found that the regression coefficients for the impact of sentiment on returns were distributed in such a way that they imply mean reversion. This suggests that the effect of sentiment on returns is balanced over time.

With regard to the aim of the research it appears sentiment does have a small impact on returns and vice versa. There is evidence to suggest that both phenomena are happening at the same time. Hence the ability to use sentiment as a predictor for asset prices may be significantly reduced. There is no clear method of distinguishing between sentiment which may lead to future price movements and sentiment that is generated by past price movements.

Further work

Arising from this research some significant questions have been raised. A compelling question is to whether it is possible to determine the degree to which sentiment can explain future prices. If sentiment was to be used as part of a strategy to predict prices, would it be possible to determine when to take it into account and how much influence it should have?

For the purposes of this study, oil stocks were focused on as the underlying financial asset. It would be interesting to see if the results differ greatly for other financial instruments.

More research into methods to better refine the input data would also be useful to improve the accuracy of the findings. While the presence of off-topic tweets was minimised in this experiment, some irrelevant data was still present which had the potential to skew results.

One of the issues raised by this research is whether it is possible to distinguish between sentiment that may lead to price movements and sentiment that is caused by price movements.

References

Ahmad, K. (2013, February 15). *Utility Theory and Beyond*. Retrieved April 21, 2013, from https://www.cs.tcd.ie/khurshid.ahmad/Teaching/Behavioural_Finance/Lecture 3_Utility_Theory.pdf

Ahmad, K., Daly, N., & Liston, V. (2011). What is new? News media, General Elections, Sentiment, and named entities. *Sentiment Analysis where AI meets Psychology (SAAIP)*, 80.

Baiocchi, G., & Distaso, W. (2003). GRETL: Econometric software for the GNU generation. *Journal of Applied Econometrics*, *18*(1), 105-110.

Balvers, R., Wu, Y., & Gilliland, E. (2000). Mean reversion across national stock markets and parametric contrarian investment strategies. *The Journal of Finance*, *55*(2), 745-772.

Bandopadhyaya, A., & Jones, A. L. (2006). Measuring investor sentiment in equity markets. *Journal of Asset Management*, *7*(3), 208-215.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1-8.

Cronin Fisk , M., & Brubaker Calkins, L. (n.d.). *BP, Transocean Are Sued by Texas Over 2010 Gulf Oil Spill - Bloomberg*. Retrieved from http://www.bloomberg.com/news/2013-05-18/bp-transocean-are-sued-by-texas-over-2010-gulf-oil-spill.html

Engle, R. (2004). Risk and volatility: Econometric models and financial practice. *The American Economic Review*, *94*(3), 405-420.

King, R., & Peterson, R. (2011). Trading on a World of Sentiment. *Bloomberg Businessweek.*

Kumar, A., & Lee, C. (2006). Retail investor sentiment and return comovements. *The Journal of Finance*, *61*(5), 2451-2486.

Lo, A. W., & MacKinlay, A. C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of financial studies*, *1*(1), 41-66.

Lui, M., & Baldwin, T. (2012, July). langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations* (pp. 25-30). Association for Computational Linguistics.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, *17*(1), 59-82.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business*, *36*(4), 394-419.

Simon, H. A. (1979). Rational decision making in business organizations. *The American economic review*, *69*(4), 493-513.

The Oil Drum | Discussions about Energy and Our Future. Retrieved May 21, 2013, from http://www.theoildrum.com/.

The Streaming APIs [Diagram]. (2012, September 24). Retrieved from https://dev.twitter.com/sites/default/files/images_documentation/streaming-intro-2_1.png

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

Tumarkin, R., & Whitelaw, R. F. (2001). News or noise? Internet postings and stock prices. *Financial Analysts Journal*, 41-51.