

Bond University  
Research Repository



## Predicting FTSE 100 returns and volatility using sentiment analysis

Johnman, Mark; Vanstone, Bruce J; Gepp, Adrian

*Published in:*  
Accounting and Finance

*DOI:*  
[10.1111/acfi.12373](https://doi.org/10.1111/acfi.12373)  
[10.1111/acfi.12373](https://doi.org/10.1111/acfi.12373)

Published: 01/11/2018

*Document Version:*  
Peer reviewed version

[Link to publication in Bond University research repository.](#)

*Recommended citation(APA):*

Johnman, M., Vanstone, B. J., & Gepp, A. (2018). Predicting FTSE 100 returns and volatility using sentiment analysis. *Accounting and Finance*, 58(S1), 253-274. <https://doi.org/10.1111/acfi.12373>, <https://doi.org/10.1111/acfi.12373>

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

For more information, or if you believe that this document breaches copyright, please contact the Bond University research repository coordinator.

# Predicting FTSE 100 Returns and Volatility Using Sentiment Analysis

---

## Abstract

We investigate the statistical and economic effect of positive and negative sentiment on daily excess returns and volatility in the FTSE 100 index, using business news articles published by the Guardian Media Group between 01/01/2000 and 01/06/2016. The analysis indicates that while business news sentiment derived from articles aimed at retail traders doesn't influence excess returns in the FTSE 100 index, it does affect volatility, with negative sentiment increasing volatility and positive sentiment reducing it. Further, an ETF-based trading strategy based on these findings is found to outperform the naïve buy-and-hold approach.

## Keywords

Sentiment analysis, trading strategy, text mining, news, FTSE 100

## JEL Classification

G11

## 1 Introduction

Intuitively, it is reasonable to consider investor sentiment as influencing stock market movements. This is supported by the findings of Blasco, Corredor, and Ferreruela (2012), who showed that sentiment is a key factor in the intensity of investors' herding behaviour. Theoretical models of the effect of investor sentiment usually posit the existence of two types of traders: 'noise traders' (retail investors), who hold random beliefs about future dividends, and 'rational arbitrageurs' (sophisticated investors), who hold Bayesian beliefs (De Long et al. 1990). It has been shown that changes in noise traders' sentiment can result in increased market volatility and temporary deviations in stock prices away from fundamental values (De Long et al. 1990).

The measurement of investor sentiment using media-based data sources (e.g. newspaper articles) is an application of sentiment analysis, which can be broadly defined as analysing textual data using natural language processing techniques to capture people's attitude towards a topic. Methodologies for sentiment analysis vary from matching text against positive and negative word lists to using statistical learning algorithms to identify key words and topics. In terms of finance-specific applications, Klein and Prestbo (1974) can be credited with making one of the first attempts to study the effect of textual information on financial markets. This study pioneered research into how text can be used to derive investor sentiment, predict market movements and exploit them through trading strategies.

A notable portion of the literature refers to the work of Tetlock (2007), who reported that high media pessimism predicts downward pressure on market prices, which is followed by a reversion to fundamentals. In a subsequent paper, Tetlock, Saar-Tsechansky, and Macskassy (2008) confirmed this finding, with return predictability being strongest for news articles focusing on stocks' fundamentals (i.e. containing at least one word with the stem "earn"). In

each of these papers, news-based trading strategies were developed to assess the economic significance of the findings, with the strategies having statistically significant daily excess returns of 4.4 basis points (Tetlock 2007) and 11.8 basis points (Tetlock, Saar-Tsechansky, and Macskassy 2008). Building on this work, Ferguson et al. (2015) applied a similar method of analysing sentiment to stocks in the FTSE 100 between 1981 and 2010, finding that positive (or negative) words predict higher (or lower) next period excess returns. Additionally, the trading strategy used by Ferguson et al. (2015) produced statistically significant risk-adjusted excess returns of 14.2 – 19 basis points per day from 2003 - 2010.

This paper extends the work started by Tetlock (2007) and continued by Ferguson et al. (2015) by focusing on unsophisticated investors. Using all news articles published in the business section by the Guardian Media Group between 01/01/2000 and 01/06/2016, this paper investigates whether measurements of positive and negative sentiment predict daily excess returns and volatility in the FTSE 100 index. In addition, this paper evaluates the economic significance of these sentiment metrics through a trading strategy.

The primary contribution of this paper is that it provides insight into the statistical and economic effect of business news, which is aimed at retail investors, on market returns and volatility. Previous literature analyses sentiment using textual data sources that are aimed towards sophisticated investors, often specifically mentioning the asset being investigated. Analysing business news articles from a less financially focused data source allows us to measure sentiment that will be more likely used by retail investors to make investment choices, whereas a sophisticated investor will likely base trading decisions on financially focused news sources such as Bloomberg or The Wall Street Journal. The effect of this sentiment is assessed using index returns and volatility, as the news articles in the Guardian Media Group generally focus on economy wide issues rather than company specific issues. Our expectation is that the business news sentiment used by retail investors will not have a

statistically significant effect on index returns as noise traders don't add information to the market. However, we expect their behaviour will influence volatility, with negative sentiment increasing volatility and positive sentiment reducing it.

Another contribution of this paper is that it provides insight into how to exploit the temporary deviation in market prices using an ETF-based trading strategy. Since indexes can't be directly traded, strategies that are based on them must purchase all the underlying stocks or use an ETF. ETFs have become extremely popular among retail investors as they reduce strategy complexity by means of only one instrument being traded, instead of all of the assets that comprise the index.

## **2 Literature Review**

Using sentiment analysis to predict financial markets usually involves three major components: data sources, sentiment analysis and a trading strategy. The data sources are the textual data used to measure sentiment and the financial markets data this measurement is evaluated against. Sentiment analysis involves extracting sentiment information from the textual data and representing it in a numerical format, called a sentiment analysis metric (SAM). The trading strategy is the rules used to make trading decisions based on the SAM. Not all research uses a trading strategy, with it being common to focus on evaluating the statistical significance of the SAM, as opposed to its economic significance (see Appendix 1 for references). This literature review discusses the common choices made, as well as outlining some of the key findings that have emerged regarding investor sentiment.

### **2.1 Data Sources**

The textual data used to analyse investor sentiment vary widely in terms of quantity and type. The main source is news articles from prominent newspapers, such as the Wall Street Journal,

which are utilised because they are read widely by many traders and are considered a stable and trustworthy information source (see Appendix 1 for references). More specifically, most research uses financial news as it is thought to have less noise than general news, being more likely to report market-relevant information, such as economic updates and company news (Khadjeh Nassirtoussi et al. 2014). Within financial news, most papers analyse the full text of articles, while some simply use the headlines of articles, arguing that these are more concise and less noisy than the full text (Huang et al. 2010; Khadjeh Nassirtoussi et al. 2015). Other textual data sources include message and discussion boards, which are usually specific to the assets being investigated (Antweiler and Frank 2004; Das and Chen 2007; Nguyen and Shirai 2015; Sycara and Thomas 2000). Over the last decade, Twitter has emerged as a common alternative source, with tweets being used as they are concise and faster to write than full articles, and thus may precede breaking news (see Appendix 1 for references). Additional sources are more varied, including corporate announcements (Hagenau et al. 2013; Hagenau, Liebmann, and Neumann 2013), Bloomberg macroeconomic news announcements (Chatrath et al. 2014) and company annual reports (Butler and Kešelj 2009). The amount of textual data collected varies dramatically, ranging from data over a few months to nearly thirty years.

Like the textual data, the financial markets data used in investor sentiment analysis are quite varied. Most data used are that of individual stocks from major, predominantly US-based, indexes, such as the S&P 500 (see Appendix 1 for references). In particular, the papers that implement trading strategies almost always do so on stocks. Using stocks from major indexes is often done as they tend to receive more news coverage (Wuthrich et al. 1998). Another choice of instrument is an actual index, with research being done using indexes in several different countries, such as the DJIA, FTSE 100, Hang Seng and Nikkei 225 (see Appendix 1 for references). A relatively small number of papers have also utilised other assets, such as currency exchange rates (Chatrath et al. 2014; Fuehres and Gloor 2012; Khadjeh Nassirtoussi

et al. 2015; Peramunetilleke and Wong 2002) and gold futures (Fuehres and Gloor 2012; Smales 2014).

## **2.2 Sentiment Analysis**

Sentiment analysis can be broken down into three major stages: feature extraction, feature representation and sentiment classification. Feature extraction involves defining the features (variables) of the textual information that will represent investor sentiment and extracting them (e.g. breaking the text up into individual words). Feature representation refers to converting selected features into a numerical format (e.g. counting the number of times a word occurs). The sentiment classification stage involves processing the represented features to classify the textual information (e.g. a news article) as having positive or negative sentiment.

### **2.2.1 Feature Extraction**

The feature extraction stage involves defining and extracting the features of the textual information that are used to represent investor sentiment. A commonly used technique is to represent text content as n-grams, which are collections of phrases consisting of n consecutive words (Khadjeh Nassirtoussi et al. 2014). Another feature extraction technique is Part-Of-Speech (POS) tagging, which involves identifying and extracting certain types of phrases from the text, such as proper nouns (Schumaker and Chen 2009a, 2009b; Schumaker et al. 2012) or noun phrases (Hagenau, Liebmann, and Neumann 2013; Schumaker and Chen 2009b). These methods assume that certain parts of each text document contain more information about investor sentiment than the rest of the document. Once features have been extracted from the text documents, they are often filtered to eliminate features that appear to have less of a material effect on investor sentiment. This is designed to reduce noise and improve the computational efficiency of the sentiment analysis process. Feature selection

techniques include the removal of stop words (e.g. the) (Geva and Zahavi 2014; Gidófalvi 2001; Khadjeh Nassirtoussi et al. 2015; Li, Xie, et al. 2014; Mittermayer 2004; Nguyen and Shirai 2015; Soyland 2015; Zhai, Hsu, and Halgamuge 2007), stemming (reducing words to their base form) (Geva and Zahavi 2014; Hagenau et al. 2013; Hagenau, Liebmann, and Neumann 2013; Mittermayer 2004) and lemmatization (Nguyen and Shirai 2015; Soyland 2015). Other methods include reducing the feature set to features appearing a minimum number of times per document (Butler and Kešelj 2009; Da, Engelberg, and Gao 2015; Schumaker and Chen 2009b; Schumaker et al. 2012) or ranking features according to a feature representation metric and choosing a certain proportion (e.g. 1000 features that occurred the most across all documents) (Antweiler and Frank 2004; Mittermayer 2004; Soyland 2015; Zhai, Hsu, and Halgamuge 2007).

### **2.2.2 Feature Representation**

Feature representation involves assigning a numerical weighting to the selected features. Common feature representation techniques included binary representation (i.e. if a feature is present or not present in a document, it is given a value of 1 or 0) (Antweiler and Frank 2004; Geva and Zahavi 2014; Peramunetilleke and Wong 2002; Schumaker and Chen 2009a, 2009b; Schumaker et al. 2012; Soyland 2015; Tetlock 2007) and frequency (i.e. the number of times a word appears in a document) (Butler and Kešelj 2009; Das and Chen 2007; Li, Xie, et al. 2014; Sycara and Thomas 2000). Another way to represent a feature is to use the ratio of its frequency divided by the total number of features in the document (Ferguson et al. 2015; Fuehres and Gloor 2012; Tetlock, Saar-Tsechansky, and Macskassy 2008; Zhang, Fuehres, and Gloor 2011). Ratios are designed to create more accurate sentiment measurements by ensuring that each measurement is weighted according to the number of features in a document. There are also several variations of ratio metrics, such as the number of positive



features minus the number of negative features divided by the sum of both positive and negative features.

Another well-known feature representation technique is using a term frequency-inverse document frequency (TF-IDF) metric (Hagenau, Liebmann, and Neumann 2013; Li, Wang, et al. 2014; Peramunetilleke and Wong 2002; Zhai, Hsu, and Halgamuge 2007). TF-IDF measures how important a feature is to a document by multiplying a measurement of how often the feature occurs in the document by the inverse of a measurement of how common it is across all documents.

### **2.2.3 Sentiment Classification**

Sentiment classification involves using the represented features to classify the textual information as having positive or negative sentiment. Depending on the analysis, this can be done at several different levels of granularity, such as document-level (e.g. a news article), sentence-level (sentences in a news article) and aspect-level (characteristics about an entity – e.g. a stock) (Medhat, Hassan, and Korashy 2014; Ravi and Ravi 2015). The main approaches to sentiment classification involve using statistical learning techniques, a lexicon or a combination of the two (Medhat, Hassan, and Korashy 2014; Ravi and Ravi 2015).

Statistical learning techniques are essentially algorithms that can learn from and forecast data. The specific statistical learning technique chosen often depends on whether the SAM is being used to predict the future movements of financial markets data. For categorical predictions, such as whether an asset's price will move up or down, support vector machines (SVMs) are predominantly used (see Appendix 1 for references). When actual future prices are being predicted, techniques include support vector regression (SVR), neural networks, naïve bayes, decision trees and random forests (see Appendix 1 for references).

Lexicon-based approaches to sentiment classification involve matching the chosen features to a pre-defined dictionary that contains a collection of words already classified as having positive or negative sentiment (Ravi and Ravi 2015). The Harvard IV-4 dictionary (now part of the General Inquirer dictionary) is often used, along with a variety of others (see Appendix 1 for references). Additionally, Bollen, Mao, and Zeng (2011) have explored how psychological questionnaires assessing people's moods can be translated into dictionaries to assess sentiment. In more recent times, it has been shown that finance-specific dictionaries more accurately capture sentiment than general dictionaries. This approach was pioneered by Loughran and McDonald (2011), who, using 10-Ks from 1994 to 2008, found that almost 75% of the words identified as negative by the Harvard IV-4 dictionary are not considered negative in a financial context. In response to this, Loughran and McDonald (2011) created their own, finance-specific lexicon, which has subsequently been found to measure investor sentiment better than the original Harvard IV-4 dictionary (Ferguson et al. 2015; Li, Wang, et al. 2014; Li, Xie, et al. 2014).

#### **2.2.4 Findings**

There have been relatively few comprehensive comparisons of the effectiveness of different sentiment analysis techniques at predicting asset price movements. While there are no clear conclusions on the most effective approach to sentiment analysis, the literature reviewed strongly suggests that investor sentiment influences asset prices, with it being typically reported that positive (or negative) sentiment induces positive (or negative) pressure on market prices. It is also usually reported that negative sentiment's effect on returns is more pervasive and stronger than that of positive sentiment. Additionally, measures of investor sentiment have been shown to influence volatility in asset prices (Antweiler and Frank 2004; Da, Engelberg, and Gao 2015; Grob-Klubmann and Hautsch 2011; Kumari and Mahakud 2015; Lee, Jiang, and Indro 2002; Olaniyan et al. 2017; Tetlock 2007; Wu, Zheng, and Olson

2014). The literature suggests that positive sentiment decreases volatility (Kumari and Mahakud 2015; Lee, Jiang, and Indro 2002; Olaniyan et al. 2017), while negative sentiment increases volatility (Kumari and Mahakud 2015; Lee, Jiang, and Indro 2002). It has also been reported that negative sentiment has higher predictive power than positive sentiment (Wu, Zheng, and Olson 2014).

While the literature reviewed suggests that investor sentiment can influence asset returns and volatility, there are limitations to the methodologies discussed. For example, classifying entire documents as positive or negative doesn't provide insight into their individual characteristics (e.g. how the positive and negative words are distributed throughout the article) (Wu, Zheng, and Olson 2014; Wu et al. 2010). Additionally, documents may have different effects on asset returns and volatility depending upon the time of their release (e.g. articles written immediately after the market closes may have less of an effect than those written immediately before it opens) (Wu, Zheng, and Olson 2014; Wu et al. 2010).

### **2.3 Trading Strategy**

Trading strategies can often be classified by the purpose of the SAM, such as predicting the actual price of an asset or its direction. For sentiment analysis trading systems predicting actual future asset prices, the main type of trading strategy utilised is to buy or sell the asset if the predicted price  $X$  minutes (e.g. 20) after a news article is released is greater than or equal to a  $Y\%$  movement (e.g. 1%) in the price of the asset at the time of the news article's release (Geva and Zahavi 2014; Li, Wang, et al. 2014; Sycara and Thomas 2000). The exact threshold for the movement is usually found through brute force optimization. For strategies predicting up (or down) movements of the asset, the signals are usually to buy (or sell) the asset if the prediction or SAM is positive (or negative) (Ferguson et al. 2015; Hagenau, Liebmann, and Neumann 2013; Mittermayer 2004; Wuthrich et al. 1998; Zhai, Hsu, and

Halgamuge 2007). Other strategies involve ranking companies by the SAM over a specified timeframe and taking a long (or short) position on the top (or bottom) X% of stocks (e.g. 10%) (Chan 2003; Heston and Sinha 2016; Zhang and Skiena 2010). This is similar to a momentum strategy, where there is a formation period and a holding period. Interestingly, Sinha (2016) found that the short-term reversal effect common to weekly, price-based momentum strategies doesn't appear to occur in weekly, sentiment-based momentum strategies. To apply this strategy to a single instrument (e.g. an index), Tetlock (2007) proposes taking a long (or short) position in the asset if the SAM is in the bottom (or top) X% of the prior calendar year's distribution of a measurement of negative sentiment.

In terms of holding periods, most of the trading strategies maintain their positions over relatively short timeframes, with the most common being daily or intraday (e.g. 20 minutes) (see Appendix 1 for references). Some strategies, particularly the momentum-based ones, have longer holding periods (e.g. weekly) as they use SAMs aggregated over longer timeframes. Rebalancing less regularly can sometimes be more attractive from an implementation perspective as the strategy incurs lower transactions costs (Uhl, Pedersen, and Malitius 2015) and a longer-term SAM is typically less noisy than one compiled over a shorter timeframe (Heston and Sinha 2016).

Overall, the literature reviewed suggests that investor sentiment does influence the returns of financial markets and has the potential to be utilized in a profitable trading strategy. This effect appears to be consistent across several different asset types, such as gold futures and stocks, as well as various textual data sources, such as Twitter tweets and news articles. However, most of the literature reviewed uses textual data sources directed towards more sophisticated investors, thereby not shedding light on the effect of media aimed at retail investors on financial market returns and volatility. Additionally, trading strategies in the

literature reviewed don't tend to explore the use of ETFs, which have the potential to reduce strategy complexity and assist in the implementation of index-based strategies.

### **3 Methodology**

This paper builds on the body of work started by Tetlock (2007) and continued by Ferguson et al. (2015). We investigate whether measurements of positive and negative sentiment predict daily excess returns and volatility in the FTSE 100 index. We also evaluate the economic significance of these metrics by developing a trading strategy that utilises them.

#### **3.1 Data Sources**

The SAMs were derived from news articles from the Guardian Media Group's online application programming interface (API)<sup>1</sup>. The Guardian Media Group is a UK mass media company that publishes widely read UK newspapers such as The Guardian, The Observer and The Guardian Weekly. Their API can be accessed with a developer or commercial key, with the latter opening the entire Guardian content store (developer keys provide access to over 1.9m news articles) and allowing the API to be queried more frequently (developer keys can make up to 12 calls per second and 5000 per day). A developer key was used in this research as there was no need for the additional privileges of a commercial key.

Extracting data between 01/01/2000 and 01/06/2016 ensured that the data set is current and that the analysis was over a timeframe involving both bull and bear markets. All news articles from the business section of the API were used as they are likely to contain less noise (Khadjeh Nassirtoussi et al. 2014). The full text of news articles was chosen to ensure that the entire content of the article was captured. Overall, 79,823 articles were used in the analysis, with the distribution of the daily number of articles being shown by year in Figure 1. While

---

<sup>1</sup> The Guardian's API can be found here: <http://open-platform.theguardian.com/>

the graph shows a decrease in the number of daily articles over the last few years, there are sufficient articles per day to conduct the analysis.

[FIGURE 1 ABOUT HERE]

Since the Guardian Media Group's publications don't have a particular emphasis on finance, readers who use this as a source for investment decisions are more likely to be unsophisticated investors. While more sophisticated investors (i.e. professional traders and fund managers) may read these publications, they are more likely to have access to and use finance-specific news sources, such as Bloomberg, to form views about financial markets.

In terms of financial markets data a UK index was chosen, as the Guardian Media Group is a UK company and has a large UK audience. Specifically, the FTSE 100 was used as it is considered representative of the London Stock Exchange, with its constituents being the most well-known and widely traded stocks on the market. Since the FTSE 100 can't be directly traded, BlackRock's iShares Core FTSE 100 UCITS ETF (ISF) was chosen for the trading strategy. This ETF is a reasonable proxy for the FTSE 100 index, with a three-year tracking error of 0.05 (sample standard deviation of the difference between the return on the ETF and return on the index)<sup>2</sup>. The daily open, high, low, close and volume data for the FTSE 100 index and ETF between 01/01/2001 and 01/06/2016 were sourced from Bloomberg.

### **3.2 Sentiment Analysis**

The sentiment analysis process has been broken into feature extraction, sentiment classification and feature representation stages to mirror the steps outlined in the literature review.

---

<sup>2</sup> More information can be found here: [https://www.blackrock.com/za/literature/fact-sheet/isf-ishares-core-ftse-100-ucits-etf-\(dist\)-fund-fact-sheet-en-za.pdf](https://www.blackrock.com/za/literature/fact-sheet/isf-ishares-core-ftse-100-ucits-etf-(dist)-fund-fact-sheet-en-za.pdf)

### **3.2.1 Feature Extraction**

The news articles were first filtered and grouped so that the articles for trading day  $t$  were those released online between 4:30pm on the previous trading day and 8am on day  $t$ . Using articles only released between the close and open of the market avoids any endogeneity issues in the analysis as no one was able to act on the information in the articles until the market opened on day  $t$ . Using articles released outside of working hours also increases the likelihood that they will have been seen by retail investors, whereas ones released during the trading day may have gone unnoticed as they were at work.

### **3.2.2 Sentiment Classification**

In keeping with Tetlock, Saar-Tsechansky, and Macskassy (2008) and Ferguson et al. (2015), a lexicon-based sentiment classification approach was used. Specifically, words in the grouped articles were classified as positive (negative) if they were in the positive (negative) word lists of the dictionary created by Loughran and McDonald (2011). This dictionary was chosen because finance-specific lexicons have been shown to be better than general ones at capturing sentiment in a finance context (Li, Wang, et al. 2014; Li, Xie, et al. 2014; Loughran and McDonald 2011).

### **3.2.3 Feature Representation**

In order to account for different numbers of news articles being produced each trading day, a ratio feature representation approach was chosen. Specifically, the number of positive (or negative) words for a day was divided by the total number of words in all articles for that day. Following Tetlock, Saar-Tsechansky, and Macskassy (2008), the positive (or negative) figures were further standardised by subtracting the mean and dividing by the standard deviation of the previous calendar year's fraction of positive (or negative) words. The standardisation was done to account for regime changes in these metrics, which may occur if there are changes in the coverage or style of the news articles (Ferguson et al. 2015; Tetlock,

Saar-Tsechansky, and Macskassy 2008). It also helps ensure that the final SAMs are stationary. The formulas for these calculations are as follows:

$$Positive (Negative) = \frac{Number\ of\ Positive\ (Negative)\ Words}{Total\ Number\ of\ Words} \quad (1)$$

$$Std\_Positive (Std\_Negative) = \frac{Positive (Negative) - \mu_{Positive (Negative)\ of\ Previous\ Year}}{\sigma_{Positive (Negative)\ of\ Previous\ Year}} \quad (2)$$

### 3.3 Predicting Excess Returns and Volatility Using Daily Positive & Negative Sentiment

The effects of the measurements of positive and negative sentiment on daily excess returns in the FTSE 100 was investigated using the following OLS regression model.

$$R_i - R_f = \alpha_1 + \beta_1 * Std\_Positive + \beta_2 * Std\_Negative + \varepsilon \quad (3)$$

The  $R_i$  variable represents the daily close-to-close log returns of the FTSE 100. Close-to-close log returns were used as retail investors may take time to act on the news between the close of the previous day and the open of the next (e.g. some may place trades at different times of the day such as during lunch because of work commitments). The  $R_f$  variable was sourced from Gregory et al. (2013), who have constructed daily SMB, HML, UMD, risk-free rate and market return factors for the largest 350 stocks on the UK stock market (i.e. the FTSE 100 and FTSE 250). Since these variables are only available until 30/06/2015, the regression only includes data up to this point.

The effects of the measurements of positive and negative sentiment on daily volatility in the FTSE 100 were investigated using the following OLS regression models.

$$Volatility\ Proxy\ 1: (R_i)^2 = a_2 + \beta_3 * Std\_Positive + \beta_4 * Std\_Negative + \varepsilon \quad (4)$$

$$Volatility\ Proxy\ 2: H_i - L_i = a_3 + \beta_5 * Std\_Positive + \beta_6 * Std\_Negative + \varepsilon \quad (5)$$



The  $H_i$  and  $L_i$  variables represent the daily high and low of the FTSE 100. Two different proxies of volatility, namely squared returns and the high-low range, were used to increase the reliability and robustness of the results. Squared returns were chosen as a proxy as they can be interpreted as a conditionally unbiased estimator of the true unobserved conditional variance of an asset (Patton 2011), while the high-low range has also been shown to be a proxy of volatility (Parkinson 1980).

#### **4 Analysis of Statistical Significance**

The purpose of this paper is to investigate the statistical and economic effect that measures of positive and negative sentiment have on the excess returns and volatility of the FTSE 100 index. A summary of the regression results for the excess returns and volatility models are shown in Table 1 and Table 2.

[TABLE 1 ABOUT HERE]

[TABLE 2 ABOUT HERE]

As per our expectation, neither measurements of sentiment appear to influence daily excess returns. In contrast to the excess returns model, both measurements of sentiment appear to have an effect on daily volatility in the FTSE 100. These results are consistent with the literature, with negative sentiment increasing volatility (Kumari and Mahakud 2015; Lee, Jiang, and Indro 2002) and showing greater statistical significance than positive sentiment (Wu, Zheng, and Olson 2014), which reduces volatility (Kumari and Mahakud 2015; Lee, Jiang, and Indro 2002; Olaniyan et al. 2017). Overall, the results suggest that while the activity of retail investors based on sentiment doesn't influence market returns, their behaviour does add significant noise to the market. The results also indicate that sentiment metrics derived from data sources that are aimed at different types of investors can have

different effects on financial markets, with previous research showing that sentiment metrics derived from more financially focused data sources can influence asset returns (Ferguson et al. 2015; Tetlock, Saar-Tsechansky, and Macskassy 2008). Our findings supports the theoretical model of investor sentiment posited by De Long et al. (1990) and suggests that noise traders and rational arbitrageurs have different effects on financial markets.

#### **4.1 Robustness Checks**

The variance inflation factors (VIF) of the regression coefficients of the models were calculated, revealing no indications of multicollinearity problems. These factors, along with a correlation matrix of the variables in the regressions, are shown in Table 3 and Table 4.

[TABLE 3 ABOUT HERE]

[TABLE 4 ABOUT HERE]

Additionally, several other variations to the sentiment analysis process were examined, such as removing stop words and excluding articles from non-trading days. The conclusions were not materially affected by these variations.

### **5 Economic Significance**

A trading strategy has been designed to exploit the regression results, which suggest that the behaviour of retail investors increases volatility in the FTSE 100 when there is an increase in negative news, but doesn't cause a fundamental change in the index's returns. This suggests that a short-term reversal strategy may be able to exploit the additional noise by entering the market when prices have diverged from their fundamental values and exiting when they revert to them. Consequently, if today's *Std\_Negative* value, calculated using news from the previous trading day's close to today's open, is greater than the 70<sup>th</sup> quantile of the prior calendar year's *Std\_Negative* values, our strategy takes a long position at today's close. The

strategy enters positions at the close to give retail investors time to add additional noise to the market and distort prices from their fundamental values. A threshold of 70% has been used to help ensure the strategy only takes positions when there is a substantially negative signal. Starting with an equity of \$1,000,000 and assuming zero transactions costs, the strategy uses all available equity when entering positions, holding them until the close of the day the *Std\_Negative* value falls back below the 70% threshold (i.e. when noise has reduced and prices have reverted to their fundamental values). The testing period is from 01/01/2002 (when the ETF first started) - 01/06/2016, containing an 11-month out-of-sample period (01/07/2015 – 01/06/2016), which was not included in the regression analysis. The strategy averages 3-4 trades per month, with some example trades illustrated in Figure 2.

[FIGURE 2 ABOUT HERE]

The example trades in Figure 2 provide insight into the typical behaviour of the strategy. The strategy buys on Wednesday 11/06/08 when the negative sentiment metric (0.930) is greater than the threshold (0.694) and holds this position while the metric remains above the threshold on Thursday 12/06/08 and Friday 13/06/08. The position was then closed out at the close of Monday 16/06 when the negative sentiment metric dropped below the threshold. This first trade is a clear example of a successful trade, with the strategy entering into a position after the price has fallen as a result of retail investor activity and then closing out after it has reverted upwards a few days later. The second trade is an example of an unsuccessful trade, with the strategy entering into a position after the *Std\_Negative* metric rose above the threshold and holding it for two days. During this time, the price continued to trend downwards with an increase in negative sentiment instead of reverting after its initial fall. Finally, the third trade is similar to the first, buying at the close of Tuesday 24/06/08 and holding the position for one day before profitably exiting at the close of Wednesday 25/06/08.

As shown in Figure 3 and Table 5, when considered across the entire testing period, the trading strategy strongly outperforms a simple Buy & Hold strategy on both a risk-adjusted and absolute basis. It has a larger Sharpe Ratio, as well as lower drawdown throughout the testing period. Additionally, the active strategy's average daily return for the days in which it is in the market is over 13 times that of the Buy & Hold strategy.

[FIGURE 3 ABOUT HERE]

[TABLE 5 ABOUT HERE]

## 6 Conclusions

This paper provides insight into the statistical and economic effect of the sentiment of general business news, which is read by retail investors, on market returns and volatility. Using news articles from the Guardian Media Group between 01/01/2000 and 01/06/2016, positive and negative measurements of sentiment were calculated and regressed against daily excess returns and proxies for volatility in the FTSE 100 index. The results indicate that the sentiment metrics don't influence excess returns, but do influence volatility, with negative sentiment increasing volatility and positive sentiment reducing it. These findings suggest that measurements of sentiment from textual data sources directed towards retail investors have a different effect on asset prices to sources that are more likely to be read by sophisticated investors. This is because noise traders don't add information to the market, but can increase volatility and cause prices to temporarily deviate from their fundamental values (De Long et al. 1990).

The empirical results suggest that the findings have economic significance, with the active strategy outperforming the Buy & Hold strategy on a risk-adjusted basis. This shows that an ETF-based trading strategy can be used to exploit the effects of sentiment on an index.

Overall, the findings of this paper suggest that while sentiment metrics derived from data sources that are aimed at retail investors don't influence excess returns, they do affect market volatility, with the additional noise providing opportunities that have the potential to be exploited.

## **7 Future Research**

This research has shown that sentiment measurements from a data source targeted at retail investors has a statistically significant and economically exploitable effect on volatility in the FTSE 100. Extensions to the trading strategy, such as accounting for transaction costs and considering the market effect of trades, would provide further insight into how a sentiment trading strategy could be implemented in practice. Additionally, there is evidence that the effect of investor sentiment on asset returns is stronger for smaller stocks with lower market capitalization, with their reduced visibility potentially slowing investors' reactions to news information (Chan 2003; Ferguson et al. 2015; Tetlock, Saar-Tsechansky, and Macskassy 2008). Further, Coulton, Dinh, and Jackson (2016) have shown that firms vary in their sensitivity to investor sentiment and that this sensitivity influences the timeliness of price discovery. Repeating the process followed in this paper on an index consisting of smaller stocks (e.g. FTSE 250) could provide insight into how the effect of general business news sentiment changes depending on the size of the index's constituents.

## Appendix 1

Characteristics of studies that measure investor sentiment using media-based data sources are shown in the table below. The fields shown were chosen to highlight key characteristics referred to in the Literature Review section.

Research Study	Sentiment Data Source	Financial Security Type	Sentiment Classification Method	Trading Strategy
Antweiler and Frank (2004)	Stock message boards	Stock index	Support Vector Machine (SVM), Naïve Bayes	No strategy
Bollen, Counts, and Mao (2015)	Twitter tweets	Stock index	Dictionary	No strategy
Bollen, Mao, and Zeng (2011)	Twitter tweets	Stock index	Neural Network	No strategy
Butler and Kešelj (2009)	Company annual reports	Individual stocks	SVM	No strategy
Chan (2003)	Newspaper articles	Individual stocks	None	Monthly
Chatrath et al. (2014)	Bloomberg macroeconomic news	Currency exchange rates	None	No strategy
Da, Engelberg, and Gao (2015)	Internet search data (volume)	Stock index	Harvard IV-4 dictionary	No strategy
Das and Chen (2007)	Stock message boards	Individual stocks	Naïve Bayes, Dictionary	No strategy
Ferguson et al. (2015)	Newspaper articles	Individual stocks	Dictionary	Daily
Fuehres and Gloor (2012)	Twitter tweets	Stock index, currency exchange rates, gold futures	Dictionary	No strategy
Geva and Zahavi (2014)	Newspaper articles	Individual stocks	Neural Network, decision tree	Intraday
Gidófalvi (2001)	Newspaper articles	Individual stocks	Naïve Bayes	No strategy
Grob-Klubmann and Hautsch (2011)	Newspaper articles	Individual stocks	None	No strategy
Hagenau et al. (2013)	Corporate announcements	Stock index	Tonality (statistical frequency measure)	Monthly
Hagenau, Liebmann, and Neumann (2013)	Corporate announcements	Individual stocks	SVM, Harvard IV-4 dictionary	Daily
Hahsler and Nagar (2012)	Newspaper articles	Individual stocks	Dictionary	No strategy
Heston and Sinha (2016)	Newspaper articles	Individual stocks	Reuters NewsScope Sentiment Metric	Daily
Huang et al. (2010)	Newspaper articles	Individual stocks	Weighted association rules	No strategy
Khadjeh Nassirtoussi et al. (2015)	Newspaper articles	Currency exchange rates	SVM	No strategy

<b>Research Study</b>	<b>Sentiment Data Source</b>	<b>Financial Security Type</b>	<b>Sentiment Classification Method</b>	<b>Trading Strategy</b>
Li, Wang, et al. (2014)	Newspaper articles	Individual stocks	Support Vector Regression (SVR), Dictionary	Intraday
Li, Xie, et al. (2014)	Newspaper articles	Individual stocks	SVM, Harvard IV-4 dictionary	No strategy
Mittermayer (2004)	Newspaper articles	Individual stocks	SVM	Intraday
Nguyen and Shirai (2015)	Stock message boards	Individual stocks	SVM	No strategy
Peramunetilleke and Wong (2002)	Newspaper articles	Currency exchange rates	Dictionary	No strategy
Schumaker and Chen (2009a)	Newspaper articles	Individual stocks	SVR	Intraday
Schumaker and Chen (2009b)	Newspaper articles	Individual stocks	SVR	Intraday
Schumaker et al. (2012)	Newspaper articles	Individual stocks	SVR	Intraday
Sinha (2016)	Newspaper articles	Individual stocks	Reuters NewsScope Sentiment Metric	Weekly
Smales (2014)	Newspaper articles	Gold futures	Reuters NewsScope Sentiment Metric	No strategy
Soyland (2015)	Newspaper articles, Twitter tweets	Individual stocks	Random forest	No strategy
Sycara and Thomas (2000)	Stock message boards	Individual stocks	Genetic algorithm	Daily
Tetlock (2007)	Newspaper articles	Stock index	Harvard IV-4 dictionary	Daily
Tetlock, Saar-Tsechansky, and Macskassy (2008)	Newspaper articles	Individual stocks	Harvard IV-4 dictionary	Daily
Uhl, Pedersen, and Malitius (2015)	Newspaper articles	Stock index	Reuters NewsScope Sentiment Metric	Monthly
Vu et al. (2012)	Twitter tweets	Individual stocks	Decision tree, Dictionary	No strategy
Wuthrich et al. (1998)	Newspaper articles	Stock index	Neural Network, Naïve Bayes, Dictionary	Daily
Zhai, Hsu, and Halgamuge (2007)	Newspaper articles	Individual stocks	SVM	Daily
Zhang and Skiena (2010)	Newspaper articles, Twitter tweets	Individual stocks	Dictionary	Daily
Zhang, Fuehres, and Gloor (2011)	Twitter tweets	Stock index	Dictionary	No strategy

## Bibliography

- Antweiler, W., and M. Z. Frank, 2004, Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards, *The Journal of Finance* 59, 1259-1294.
- Blasco, N., P. Corredor, and S. Ferreruella, 2012, Market sentiment: a key factor of investors' imitative behaviour, *Accounting & Finance* 52, 663-689.
- Bollen, J., S. Counts, and H. Mao, 2015, Quantifying the effects of online bullishness on international financial markets, Working paper (ECB Statistics Paper Series, European Central Bank).
- Bollen, J., H. Mao, and X. Zeng, 2011, Twitter mood predicts the stock market, *Journal of Computational Science* 2, 1-8.
- Butler, M., and V. Kešelj, 2009, Financial Forecasting Using Character N-Gram Analysis and Readability Scores of Annual Reports, Paper presented at the 22nd Canadian Conference on Artificial Intelligence, Kelowna.
- Chan, W. S., 2003, Stock price reaction to news and no-news: drift and reversal after headlines, *Journal of Financial Economics* 70, 223-260.
- Chatrath, A., H. Miao, S. Ramchander, and S. Villupuram, 2014, Currency Jumps, Cojumps and the Role of Macro News, *Journal of International Money and Finance* 40, 42-62.
- Coulton, J. J., T. Dinh, and A. B. Jackson, 2016, The impact of sentiment on price discovery, *Accounting & Finance* 56, 669-694.
- Da, Z., J. Engelberg, and P. Gao, 2015, The Sum of All FEARS Investor Sentiment and Asset Prices, *Review of Financial Studies* 28, 1-32.
- Das, S. R., and M. Y. Chen, 2007, Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web, *Management Science* 53, 1375-1388.
- De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann, 1990, Noise Trader Risk in Financial Markets, *Journal of Political Economy* 98, 703-738.
- Ferguson, N. J., D. Philip, H. Y. T. Lam, and J. M. Guo, 2015, Media Content and Stock Returns: The Predictive Power of Press, *Multinational Finance Journal* 19, 1-31.
- Fuehres, H., and P. Gloor, 2012, Predicting Asset Value through Twitter Buzz, in: J. Altmann, U. Baumöl and B. Krämer eds, *Advances in Collective Intelligence 2011*, Vol. 113 (Springer, Berlin, Heidelberg), 23-34.
- Geva, T., and J. Zahavi, 2014, Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news, *Decision Support Systems* 57, 212.
- Gidófalvi, G., 2001, *Using News Articles to Predict Stock Price Movements* (University of California, San Diego, CA).
- Gregory, A., R. Tharyan, and A. Christidis, 2013, Constructing and Testing Alternative Versions of the Fama–French and Carhart Models in the UK, *Journal of Business Finance & Accounting* 40, 172-214.
- Grob-Klubmann, A., and N. Hautsch, 2011, When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions, *Journal of Empirical Finance* 18, 321-340.



- Hagenau, M., M. Hauser, M. Liebmann, and D. Neumann, 2013, Reading All the News at the Same Time: Predicting Mid-term Stock Price Developments Based on News Momentum, Paper presented at the 46th Hawaii International Conference on System Sciences, Maui.
- Hagenau, M., M. Liebmann, and D. Neumann, 2013, Automated news reading: Stock price prediction based on financial news using context-capturing features, *Decision Support Systems* 55, 685-697.
- Hahsler, M., and A. Nagar, 2012, Using Text and Data Mining Techniques to extract Stock Market Sentiment from Live News Streams, Paper presented at the 2012 International Conference on Computer Technology and Science, New Delhi.
- Heston, S. L., and N. R. Sinha, 2016, News versus Sentiment: Predicting Stock Returns from News Stories, Working paper (Finance and Economics Discussion Series, Federal Reserve Board).
- Huang, C.-J., J.-J. Liao, D.-X. Yang, T.-Y. Chang, and Y.-C. Luo, 2010, Realization of a news dissemination agent based on weighted association rules and text mining techniques, *Expert Systems With Applications* 37, 6409-6413.
- Khadjeh Nassirtoussi, A., S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, 2014, Text mining for market prediction: A systematic review, *Expert Systems With Applications* 41, 7653-7670.
- Khadjeh Nassirtoussi, A., S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, 2015, Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment, *Expert Systems With Applications* 42, 306-324.
- Klein, F. C., and J. A. Prestbo, 1974, *News and the Market* (Regnery, Chicago, Ill).
- Kumari, J., and J. Mahakud, 2015, Does investor sentiment predict the asset volatility? Evidence from emerging stock market India, *Journal of Behavioral and Experimental Finance* 8, 25-39.
- Lee, W. Y., C. X. Jiang, and D. C. Indro, 2002, Stock market volatility, excess returns, and the role of investor sentiment, *Journal of Banking & Finance* 26, 2277-2299.
- Li, Q., T. Wang, P. Li, L. Liu, Q. Gong, and Y. Chen, 2014, The effect of news and public mood on stock movements, *Information Sciences* 278, 826-840.
- Li, X., H. Xie, L. Chen, J. Wang, and X. Deng, 2014, News impact on stock price return via sentiment analysis, *Knowledge-Based Systems* 69, 14-23.
- Loughran, T., and B. McDonald, 2011, When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks, *The Journal of Finance* 66, 35-65.
- Medhat, W., A. Hassan, and H. Korashy, 2014, Sentiment analysis algorithms and applications: A survey, *Ain Shams Engineering Journal* 5, 1093-1113.
- Mittermayer, M. A., 2004, Forecasting Intraday Stock Price Trends with Text Mining Techniques, Paper presented at the 37th Annual Hawaii International Conference on System Sciences, Big Island.
- Nguyen, T. H., and K. Shirai, 2015, Topic Modelling based Sentiment Analysis on Social Media for Stock Market Prediction, Paper presented at the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing, Beijing.
- Olaniyan, R., D. Stamate, L. Ouarbya, and D. Logofatu, 2017, Sentiment and stock market volatility predictive modelling : a hybrid approach, *DSAA (2015), Paris*, Vol. (IEEE, Piscataway, NJ).

- Parkinson, M., 1980, The Extreme Value Method for Estimating the Variance of the Rate of Return, *The Journal of Business* 53, 61-65.
- Patton, A. J., 2011, Volatility forecast comparison using imperfect volatility proxies, *Journal of Econometrics* 160, 246-256.
- Peramunetilleke, D., and R. K. Wong, 2002, Currency Exchange Rate Forecasting from News Headlines, Paper presented at the Thirteenth Australasian Database Conference, Melbourne.
- Ravi, K., and V. Ravi, 2015, A survey on opinion mining and sentiment analysis: Tasks, approaches and applications, *Knowledge-Based Systems* 89, 14-46.
- Schumaker, R. P., and H. Chen, 2009a, A quantitative stock prediction system based on financial news, *Information Processing and Management* 45, 571-583.
- Schumaker, R. P., and H. Chen, 2009b, Textual analysis of stock market prediction using breaking financial news: the AZFinText system, *ACM Transactions on Information Systems* 27, 12.
- Schumaker, R. P., Y. Zhang, C.-N. Huang, and H. Chen, 2012, Evaluating Sentiment in Financial News Articles, *Decision Support Systems* 53, 458.
- Sinha, N. R., 2016, Underreaction to News in the US Stock Market, *Quarterly Journal of Finance* 6, 1-46.
- Smales, L. A., 2014, News Sentiment in the Gold Futures Market, *Journal of Banking and Finance* 49, 275-286.
- Soyland, C., 2015, *Interday News-Based Prediction of Stock Prices and Trading Volume*, (Chalmers University of Technology, Goteborg, Sweden).
- Sycara, K., and J. D. Thomas, 2000, Integrating Genetic Algorithms and Text Learning for Financial Prediction, Paper presented at the Genetic and Evolutionary Computing Conference, Las Vegas.
- Tetlock, P. C., 2007, Giving Content to Investor Sentiment: The Role of Media in the Stock Market, *The Journal of Finance* 62, 1139-1168.
- Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008, More than Words: Quantifying Language to Measure Firms' Fundamentals, *The Journal of Finance* 63, 1437-1467.
- Uhl, M. W., M. Pedersen, and O. Malitius, 2015, What's in the News? Using News Sentiment Momentum for Tactical Asset Allocation, *Journal of Portfolio Management* 41, 100-112.
- Vu, T. T., S. Change, Q. T. Ha, and N. Collier, 2012, An Experiment in Integrating Sentiment Features for Tech Stock Prediction in Twitter, Paper presented at the 24th International Conference on Computational Linguistics, Bombay.
- Wu, D. D., L. Zheng, and D. L. Olson, 2014, A Decision Support Approach for Online Stock Forum Sentiment Analysis, *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 44, 1077-1087.
- Wu, Y., F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu, 2010, OpinionSeer: Interactive Visualization of Hotel Customer Feedback, *IEEE Transactions on Visualization and Computer Graphics* 16, 1109-1118.
- Wuthrich, B., D. Permunetilleke, S. Leung, V. Cho, J. Zhang, and W. Lam, 1998, Daily Prediction of Major Stock Indices from Textual www Data, Paper presented at the 4th International Conference on Knowledge Discovery and Data Mining, New York.

- Zhai, Y., A. Hsu, and S. Halgamuge, 2007, Combining News and Technical Indicators in Daily Stock Price Trends Prediction, Paper presented at the 4th International Symposium on Neural Networks, Nanjing.
- Zhang, W., and S. Skiena, 2010, Trading Strategies to Exploit Blog and News Sentiment, Paper presented at the Fourth International AAAI Conference on Weblogs and Social Media, Washington DC.
- Zhang, X., H. Fuehres, and P. A. Gloor, 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.

**Table 1 – Excess Returns Regression Results**

<b>Variable</b>	<b>Coefficient</b>	<b>T-Statistic</b>	<b>P-Value</b>
Intercept	-8.927e-05	-0.437	0.662
Std_Positive	6.195e-05	0.309	0.757
Std_Negative	1.652e-05	0.088	0.930
<b>F-Statistic</b>	0.048		0.953

\* (\*\*) significant at 95% (99%) level of confidence

**Table 2 - Volatility Regression Results**

Volatility Proxy	$(R_i)^2$			$H_i - L_i$		
	Coefficient	T-Stat	P-Value	Coefficient	T-Stat	P-Value
Intercept	1.467e-04	20.144	0.000 **	76.6704	101.375	0.000 **
Std_Positive	-1.746e-05	-2.442	0.015 *	-2.4163	-3.255	0.001 **
Std_Negative	5.482e-05	8.206	0.000 **	11.3370	16.338	0.000 **
<b>F-Statistic</b>	46.770		0.000 **	169.600		0.000 **

\* (\*\*) significant at 95% (99%) level of confidence

**Table 3 - Excess Returns Regression Correlation Matrix & Variance Inflation Factors**

	<b>R<sub>i</sub> - R<sub>f</sub></b>	<b>Std_Positive</b>	<b>Std_Negative</b>
R <sub>i</sub> - R <sub>f</sub>	1.0000	0.0049	0.0000
Std_Positive	0.0049	1.0000	-0.2979
Std_Negative	0.0008	-0.2979	1.0000
VIF	N/A	1.0974	1.0974

**Table 4 - Volatility Regression Correlation Matrix & Variance Inflation Factors**

	$(R_i)^2$	$H_i - L_i$	Std_Positive	Std_Negative
$(R_i)^2$	1.0000	0.6878	-0.0836	0.1528
$H_i - L_i$	0.6878	1.0000	-0.1346	0.2866
Std_Positive	-0.0836	-0.1346	1.0000	-0.2979
Std_Negative	0.1528	0.2866	-0.2979	1.0000
VIF (Eq. 4)	N/A	N/A	1.0974	1.0974
VIF (Eq. 5)	N/A	N/A	1.0973	1.0973

**Table 5 - Trading Strategy Performance Metrics**

	<b>Active Strategy</b>	<b>Buy &amp; Hold</b>
Cumulative Return	1.34720	-0.09291
Return (p.a.)	0.06081	-0.00672
Standard Deviation (p.a.)	0.18413	0.19582
Sharpe Ratio (p.a.)	0.33027	-0.03434
Average Daily Return (on days in the market)	0.00066	0.00005



Figure 1 - Distribution of Daily News Articles ( $\diamond$  indicates the mean daily number of news articles)

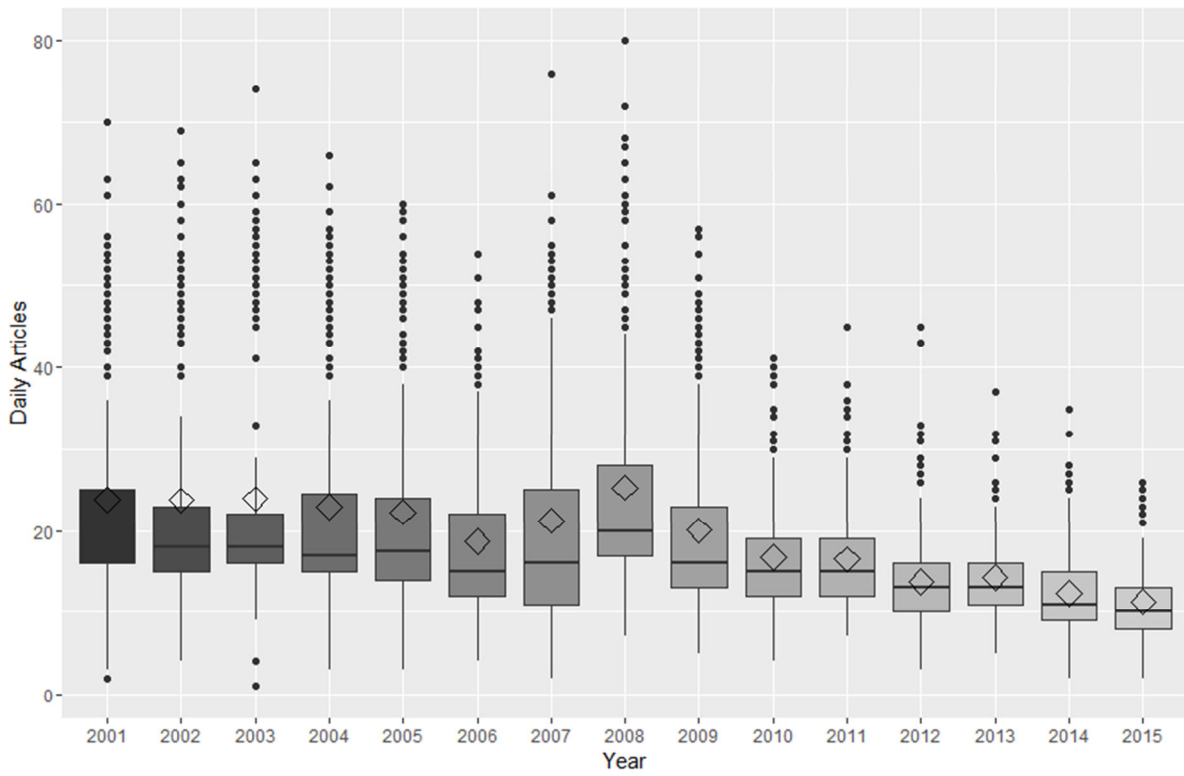
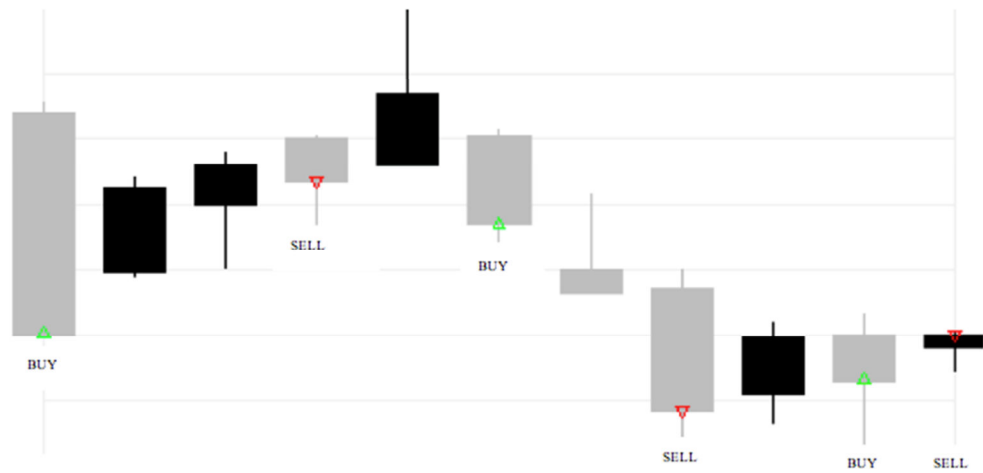


Figure 2 - Example Trades from June 2008



Date	11/06	12/06	13/06	16/06	17/06	18/06	19/06	20/06	23/06	24/06	25/06
Std_Negative	0.930	1.106	0.889	0.674	0.487	1.057	1.330	-0.312	0.422	1.004	-0.886
Threshold	0.694	0.694	0.694	0.694	0.694	0.694	0.694	0.694	0.694	0.694	0.694
Close	570.00	581.25	583.00	581.75	588.50	578.50	573.25	564.25	569.75	566.50	570.00
Trade	Buy			Sell		Buy		Sell		Buy	Sell

Grey (or black) bars symbolise days the FTSE 100 moved up (or down), Std\_Negative is calculated using news from the previous trading day's close to today's open, Threshold is the 70% threshold and Close refers to the closing price

Figure 3 - Trading Strategy Performance Graph

