Media Sentiment and UK Stock Returns

Nicky J. Ferguson[a], Jie Michael Guo[a], Herbert Y.T. Lam[b], Dennis Philip[a*]

[a]*University of Durham, Durham Business School, Mill Hill Lane, DH1 3LB, UK*

Email: nicky.ferguson@durham.ac.uk, jie.guo@durham.ac.uk, dennis.philip@durham.ac.uk

Tel: +44 (0) 191 33 45451, Fax: +44 (0) 191 33 45201

[b] *School of Finance, Renmin University of China, Beijing, 100872, China*

Email: herbert.yanto.lam@gmail.com Tel: +86(0)10-82500653

Abstract

This paper is the first to determine the effect that media sentiment has on stock returns for UK companies and tests whether there is any return predictability contained in the UK media sentiment data. We show that measures of positive and negative media sentiment have significant relationships with stock returns on the day news articles are published and that there is return predictability inherent in negative media sentiment the day following publication of media articles. We construct a news-based trading strategy to demonstrate the application of these results that earns significant positive abnormal returns.

JEL Classification: G10, G14, G17

Keywords: *media sentiment, stock returns, textual analysis, news-based trading strategy*

* The authors wish to thank John A. Doukas (the editor) and Alessandra Guariglia for helpful comments on an earlier draft. Any remaining errors are the responsibility of the authors alone. Correspondent Author: Herbert Y.T. Lam, herbert.yanto.lam@gmail.com Tel: +86(0)10-82500653

Executive Summary:

This paper examines the relationship between media sentiment and stock returns for FTSE 100 companies over the period 2005 – 2010 and tests whether there is any stock return predictability inherent in measures of media sentiment. We use textual analysis to determine a quantitative measure of how positive and how negative a media article is and determine the influence of these measures of media sentiment on stock returns.

Using the semantic content of media news articles, such as the fraction of positive and negative words contained therein, can provide valuable insights that quantitative data about economic fundamentals cannot. Stock valuations should be equal to discounted expected cash flows of the firms, subject to the investor's information sets. Contained in an information set however are also qualitative descriptions of expectations of a firm's future performance, such as the quality of management, talk of a merger, lawsuits or legal action being taken against the firm or new product launches.

Nearly all academic studies involving media sentiment in the field of finance have focused on the US market and US media publications. However cross-country differences in journalistic cultures and practices have been well documented, even in countries that share similar journalistic ideologies (see Weaver, 1998). Shaw (1999) documents some of the differences between media coverage in the US and the UK, highlighting specifically that US media has much greater conformity, whereas UK media has a much greater dispersion of opinion and media independence. Given the differences in characteristics between the two markets of the US and UK, we cannot assume that the effects of media sentiment on stock returns are consistent across countries. It is therefore important to investigate the influence of media sentiment in

the UK market in particular, given its leading global presence in financial markets, and the international reach of its news publications.

This is the first study that explores the effect of media sentiment on stock returns of UK companies, filling a noticeable void in the literature, as well as giving insights into the composition of company-specific news in the UK. We use 23,663 company-specific media articles in our analysis, allowing us to gauge stock market reaction to any media-worthy event.

Our main results show firstly that positive media sentiment in company-specific news articles has a significant positive relationship with stock returns and negative media sentiment has a significant negative relationship with stock returns. These significant relationships are strongly apparent on the day that the news articles are published. Further we find a strong relationship between negative media sentiment and media coverage whose interaction also has a significant effect on stock returns. When testing for stock return predictability on the day following publication of media articles however it was found that only measures of negative media sentiment have a significant effect.

Secondly, we demonstrate that the relationships between media sentiment and stock returns have significant economic implications. By constructing a simple news-based trading strategy that uses measures of media sentiment to determine trading signals, we show that the trading activity can generate substantial and significant abnormal returns.

1. *Introduction*

The analysis of media sentiment in financial research is a relatively new and exciting field. Some of the most respected and credible news publications in the world are dedicated to financial and business news, which play a key role in providing financial markets participants with information and in aiding them in forming their views. Analyzing media sentiment is important as it allows us to interpret some of the excess noise present in stock returns due to divergence of opinion.

The conundrum of explaining the excess volatility in stock prices that cannot be accounted for by fundamental or economic information is an interesting puzzle that has been devoid of a definitive answer due to the difficulties of quantifying or measuring qualitative media data (see Cutler et al. 1989). However in recent times researchers have begun to measure sentiment contained in media articles using textual analysis in an attempt to capture hard to quantify information and determine its effect on stock prices (see Tetlock, 2007; Tetlock et al. 2008; Loughran and McDonald, 2010).

Using the semantic content of media news articles, such as the fraction of positive and negative words contained therein, can provide valuable insights that quantitative data about economic fundamentals cannot. Stock valuations should be equal to discounted expected cash flows of the firms, subject to the investor's information sets (Tetlock et al. 2008). Contained in an information set however are also qualitative descriptions of expectations of a firm's future performance, such as the quality of management, perhaps a change in senior management, talk of a merger, lawsuits or legal action being taken against the firm, new product lines or advertising campaigns. By using a quantitative measure of semantics in language used in news articles, it is possible to measure the effects of such news events on stock returns. Also having a large data set containing many news events enables researchers to gauge the stock market reaction to the severity of language used within news articles, regardless of the type of news event.

Trying to predict stock returns based on news content has also turned into a popular endeavour for sophisticated financial market participants using innovations in algorithmic trading to interpret and act on news media. The New York Times and Wired Magazine have recently published articles documenting the rise of computer programs that speed read the news, with specially designed news-input coming from the Dow Jones Lexicon service and others that ‘robo-clients’ can interpret (see Bowley, 2010; Salmon and Stokes, 2010). The articles chronicle the ever increasing sophistication of linguistic algorithms being used by financial market participants, with computers now being able to read news reports, editorials, websites, blog posts and even twitter messages with instant feedback and trading signals. They have even evolved to gauge sentiment by understanding emoticons, sentence structure and unstructured data, such as social media buzz. Bowley (2010) notes that around 35% of quantitative trading firms are exploring the idea of using linguistic interpretation of unstructured data feeds, a significant increase from around 2% only two years previously. He finds that a popular use of such research and algorithms is to enable traders to close out their positions when bad news hits. Bowley (2010) somewhat boldly professes that the research into interpreting and acting on linguistic media content is at the forefront of a technological revolution on Wall Street, with information the most valuable commodity, and those who are able to interpret and act on it quickest coming out on top.

Nearly all academic studies involving media sentiment in the field of finance have focused on the US market and US media publications. However cross-country differences in journalistic cultures and practices have been well documented, even in countries that share similar journalistic ideologies (see Weaver, 1998). Shaw (1999) documents some of the differences between media coverage in the US and the UK, highlighting specifically that US media has much greater conformity, whereas UK media has a much greater dispersion of opinion. In the US national news culture, research shows that there is a greater reluctance to challenge the sources of power in business and government and a larger infiltration of tabloid values into general press, all in aid of reaching a mass audience rather than providing a dissenting voice (see Deuze, 2002). British media however is not afraid to voice strong opinions on contentious topics and generally provide a greater independent voice than US news literature. Given the differences in characteristics between the two markets of the US and UK, we cannot assume that the effects of media sentiment on stock returns are consistent across countries. It is therefore important to investigate the influence of media sentiment in the UK market in particular, given its leading global presence in financial markets, and the international reach of its news publications.

This paper is the first to explore the effect of media sentiment on stock returns of UK companies, filling a noticeable void in the literature, as well as providing insights into the composition of company-specific news in the UK. The paper makes several unique contributions to the analysis of media sentiment in financial research. First, we incorporate aspects of the most current research in this field to build a robust methodology with which to carry out our analysis. We use as a starting point a similar methodology to Tetlock et al. (2008), and by incorporating the financial news specific dictionaries of Loughran and McDonald (2010), we are able to more accurately determine the strength of the relationship between media sentiment and stock returns, and thereby conduct a more robust test of the return predictability of media sentiment.

Second, we use both positive and negative measures of media sentiment where previous studies such as Tetlock et al. (2008) have only used negative measures of media pessimism to explain stock returns. By using both positive and negative measures of media sentiment, the overall distribution of news can then be used to gain insight into the frequency and possible biases inherent in news articles. Further, in our analysis we use continuous media coverage, not just coverage surrounding specific events such as earnings announcements. In this way it is possible to determine on average how media sentiment is incorporated into stock returns with any given news-worthy event. By determining the effects of media sentiment on stock returns, we show a possible application of this research in the form of a news-based trading strategy.

The main results obtained from this study show that positive and negative media sentiment as measured by the proportion of positive or negative words in a news article have a significant effect on stock returns on the day media articles are published. We also find significant return predictability using negative sentiment on the day following the publication of media articles.

The outline of the rest of this study is as follows. In Section 2, a literature review is conducted covering related research on media sentiment and qualitative linguistic analysis with applications to business and finance. Section 3 discusses the properties of the media data and other variables used in the study, including descriptive statistics. Section 4 outlines the methodology followed when conducting the analysis, including assumptions made and reasoning. Section 5 presents the main results concerning the relationship between media sentiment and stocks returns and Section 6 concludes this study.

2. Literature Review

Cutler et al. (1989) is one of the first empirical studies to explore the relationship between media coverage and stock prices. This research expressed difficulty in explaining the variance in stock prices, finding that only around half of the asset price volatility could be explained by news about fundamentals. After accounting for significant macroeconomic news, the authors find that news about fundamentals can explain up to one third of stock price movements and that significant world news such as political news or natural disasters does have some effect on stock prices. They also note that some of the largest market movements occur on days with no significant news.

Tetlock (2007) uses daily content from a Wall Street Journal article to examine the effect media pessimism has on market prices. By using principal component analysis on words belonging to specific categories of the Harvard psychosocial dictionary, the paper creates pessimism factors that intend to capture negative investor sentiments or risk aversion. The intertemporal links between these measures of media pessimism and stock market movements are established using vector autoregressions (VAR). Using this media pessimism factor to forecast patterns of market activity, he finds that high media pessimism predicts downward pressure on stock prices, which revert to fundamentals usually within 5 days, although this effect is much larger and noticeably slower to reverse itself in small stocks. This is consistent with models of investor sentiment and noise trading activity such as DeLong et al. (1990). He also finds that unusually high or low media pessimism predicts a temporarily high trading volume and that pessimism weakly predicts increases in market volatility. However the hypothesis that media pessimism can reflect negative fundamental information that has not been incorporated into stock prices does not receive much support from the data. This is especially true given the reversal of pessimism effects.

Tetlock et al. (2008) extends the work on pessimism to look specifically at quantifying the language used in news articles to predict firms' fundamentals and stock returns. As in Tetlock (2007), the paper uses Harvard psychosocial dictionary to classify language present in news articles and finds that the fraction of negative words in the financial press can forecast low earnings and returns. This suggests that linguistic media content can capture otherwise hard to quantify aspects of firms fundamentals. The paper also finds that there is return predictability contained in negative media sentiment on the day following publication of the media articles. The authors note that there is a strong link between news stories containing the word *earn* and stock returns, and that stories about fundamentals are a useful predictors of earnings and returns. Their results corroborate with evidence from psychology (see Baumeister et al., 2001) that negative information in news stories has more impact, also finding that the market response to negative words is up to 5 times stronger when media coverage is earnings related.

In what has received much attention from algorithmic traders, Das and Chen (2007) developed a methodology for extracting investor sentiment from stock message boards. Specifically designed to identify the slang and short hand language used in these message boards, they find that general market activity is related to investor sentiment and message board activity and that investor sentiment tracked across individual stocks could track broad market index performance.

Earnings announcements and press releases have been a popular area of study for many looking at soft information using textual analysis. Demers and Vega (2010), Sadique et al. (2008) and Davis et al. (2008) all examine the tone used in such announcements and press releases. They find that optimistic tones increase returns in the future and also have a negative effect on stock price volatility, whereas pessimistic tones decrease future returns and have a positive effect on volatility. Sadique et al. (2008) also investigate the effect of earnings announcements based on two specific earnings metrics: pro forma and GAAP. Recent studies show that firms prefer emphasising pro forma numbers before GAAP within media press releases to present investors with more optimistic earnings figures. They find that pro forma based earnings announcements have a positive effect on returns and a negative effect on volatility, and GAAP based earnings announcements have a negative effect on returns and a positive effect on volatility.

In a further extension to the research investigating the effect media coverage has on stock returns, Carretta et al. (2011) examined news that was specifically concerned with corporate governance issues. Their research gave further insight into the characteristics of media reporting that influence stock price reactions. In particular they showed that news about ownership issues or changes in the board of directors has a negative effect on stock returns unless the firm covered was unprofitable at the time.

Many media studies involving financial markets have used the Harvard psychosocial dictionary to categorise words featuring in financial news articles. Loughran and McDonald (2010) argue however that many words that appear in negative categories in the Harvard psychosocial dictionary are not negative in a financial sense: they are merely descriptive terms. These are words such as *depreciation*, *liability*, *foreign* and *mine*. Therefore trying to model the effects on media sentiment on asset prices using the Harvard psychosocial dictionary may lead to the effect of media articles being overstated. Their research shows that in a sample of US firms, more than half of the words in the Harvard list are not negative sentiment words in the financial sense. To overcome this problem, they created a specialist list of words that carry a negative sentiment in the financial sense. This enables them to more accurately account for negative sentiment when reviewing financial media. To test the effectiveness of each list to predict returns patterns from media data 10-K financial reports from US companies were examined. A strong relationship was found with the Loughran and McDonald list of words and announcement returns. Firms with a high proportion of negative words in these filings had subsequent lower stock returns than companies with a lower proportion of negative words. Using the negative word categories from the Harvard psychosocial dictionary they found no return pattern in the data. Significant relationships with returns were also highlighted in other word categories such as positive, litigious, and weak modal.

3. Data and variables

A study such as this requires media data specific to individual companies over a long time period to give a large enough sample. The period January 2005 - October 2010 was chosen, since the frequency of media data greatly increases after 2005 (see Table 1). This is due to the increase in user base of FT.com playing a larger role in the sample, growth in company-specific news volume, and the fact that news was not limited by newspaper space any more. Further, before 2005, company-specific news data was extremely sparse. As Tetlock et al. (2008) note, company-specific news data was mainly clustered around earnings announcements. However after 2005, with the growth of financial news channels, the increasing number of individual investors and increasing public fascination with business and the economy, the frequency of business news rose.

Media data specific to individual companies is obtained manually from LexisNexis UK. The sources of the LexisNexis UK data included, by relevance, Financial Times, FT.com, The Times, Guardian and The Mirror. The data covered UK companies from the FTSE 100 index listed on the London Stock Exchange. In total 23,663 media articles were used in our analysis covering 68 FTSE 100 companies over the sample period considered. Table 1 reports the raw news-specific statistics collected before filtering and matching articles to specific companies and days.

[Insert Table 1]

The content of the media articles was analyzed to determine the number of positive or negative words they contained. The words in each article were compared to the Loughran and McDonald (2010) positive and negative financial word lists in order to identify the number of positive and negative words in the financial context[1]. The current version of the positive list contains 353 words and the current version of the negative list contains 2,337 words. The level of positive or negative market sentiment was determined for each individual article by the following formulae:

$$P = \frac{\#pos}{\#\ldots}$$

[1] The positive and negative financial word lists can be obtained from the author's website (see http://www.nd.edu/~mcdonald/Word_Lists.html)

$$N = \frac{\#n \quad ew}{e_{i\,i} \quad g}$$

These measures of *Pos* and *Neg*, provide a quantitative measurement of media sentiment for each media article[2]. These data are then matched with stock price data of the associated companies to determine its wider effects. This approach is similar to that of Loughran and McDonald (2010) when evaluating the proportion of words from a specific word list appearing in a firm’s 10-K report.

When matching the data, in some instances there was more than one media article per company, per day. When this happened, the media source with the highest relevance was chosen, with the Financial Times having the highest relevance and The Mirror having the lowest relevance. This is consistent with Dyck et al. (2008) who found that The Financial Times has much more credibility and influence than other news sources and is therefore more effective at diffusing information to produce a significant effect. When there was more than one article per day from the same media source, a simple average of positive and negative sentiment over the articles was used.

To control for size and value effects, the log of market value and the log of book-to-market ratio are included as control variables in the regressions. The other control variables used were the log of daily turnover to control for volume effects, and media coverage, which controls for the number of media articles about a specific company on a particular day.

Daily stock prices, FTSE 100 Index levels, market capitalisation, book-to-market ratio and turnover data, were obtained from DataStream for each FTSE 100 stocks for the

---

[2] We also tested other measures of positive and negative media sentiment such as (#Positive words) / (#Positive words + #Negative words) and (#Negative words) / (#Positive words + #Negative words), and the Ln(1+ Pos) and Ln(1 + Neg) and found similar relationships which were consistent with the measure selected.

period January 2005 – October 2010. This data was then matched with the media article data by company and date.

Table 2 Panel A shows summary statistics for the data, specifically those days on which news occurs. It is clear the average negative content is much higher than that of positive content; also the variation in negative content is much higher than that of positive content. [3]

[Insert Table 2]

The mean of the daily log returns is approximately zero, as would be expected, but it is also greater than the median. The skewness of daily log returns is -2.2238. Compare this to media data, which has a skewness of -0.6839. Given that the financial crisis occupied a large part of the data set, the negative skewness in returns is not surprising. However, the fact that this skewness is not completely reflected in the media data provides an indication that media sentiment is not the only factor that explains stock returns. The negative skew in media data is expected, as news coverage during the financial crisis was more negative in nature.

Cutler et al. (1989) in their study note that some of the biggest market movements occur on days when there is no news. We investigate this issue using our data set. Table 2 Panel B and Table 2 Panel C reports the largest positive and negative FTSE 100 market movements, along with their corresponding daily average of positive and negative media sentiment.

Looking first at Table 2 Panel B, which documents the five largest negative movements of the FTSE 100 over the sample period of January 2005 – October 2010,

[3] The correlation coefficient of Pos and Neg is -0.19.

there seems to be some correlation between the average level of negative sentiment and the magnitude of the FTSE 100 daily log returns. All but one of the average negative sentiment measures for these dates when large negative market movements occurred are above the mean negative media sentiment of 0.0207 and three of the average positive sentiment values are below the mean positive media sentiment of 0.0098. So there seems to be some level of company-specific news that might justify some of the large market movements that occurred on those specific days.

Turning to Table 2 Panel C, where we consider the five largest positive moves over the sample period, here all but one of the average positive sentiment values are above the mean positive media sentiment. However all the average negative sentiment values are above the mean negative media sentiment. There seems to be no correlation with the magnitude of the daily FTSE 100 log returns for either average positive sentiment or average negative sentiment. Without further accounting for macroeconomic news, it is not possible to conclude how much effect company-specific news had on the large market movements in the FTSE 100. Many of the largest positive and negative market movements occurred around the same time in October 2008, around the same period as the collapse of Lehman Brothers when there was a sense of panic over a systemic market crash leading to large market volatility. This echoes the concerns raised by Cutler et al. (1989) in accounting for asset price volatility using information released through news media.

[Insert Figure 1]

Figure 1, which shows a 30-day rolling average of the data, aptly demonstrates the differences between the two quantitative measures of the media sentiment. Firstly, the positive media sentiment seems to stay at a fairly consistent level, with little variation, its lowest point coming in the latter half of 2008, at the heart of the financial crisis. In contrast the level of negative media sentiment varies widely, having its highest point in the latter half of 2008. It is also apparent from Figure 1 that the average level of negative media sentiment may have risen mid 2007 in response to the financial crisis and the fear that has been subsumed in economic media since that time period.

The difference in averages of positive and negative media sentiment can partially be explained by the fact that there are significantly more words on Loughran and McDonald's (2010) negative word list, so the probability of finding a higher proportion of negative words would be much greater.

4. Methodology

To determine the relationship between the levels of sentiment expressed in media articles and its effects on stock returns and to test whether there is any stock return predictability in media data, we run the following two regressions:

$$\begin{aligned} ER_{i,t} = {} & \beta_1 + \beta_2 Pos_{i,t} + \beta_3 Neg_{i,t} + \beta_4 ER_{i,t-1} + \beta_5 ER_{i,t-2} + \beta_6 Ln(Size)_{i,t} \\ & + \beta_7 Ln(BTM)_{i,t} + \beta_8 Ln(Turnover)_{i,t} + \beta_9 MC_{i,t} + \beta_{10} MC_{i,t} * Neg_{i,t} + u_{i,t} \end{aligned} \tag{4.1}$$

$$\begin{aligned} ER_{i,t+1} = {} & \beta_1 + \beta_2 Pos_{i,t} + \beta_3 Neg_{i,t} + \beta_4 ER_{i,t} + \beta_5 ER_{i,t-1} + \beta_6 ER_{i,t-2} + \beta_7 Ln(Size)_{i,t} \\ & + \beta_8 Ln(BTM)_{i,t} + \beta_9 Ln(Turnover)_{i,t} + \beta_{10} MC_{i,t} + \beta_{11} MC_{i,t} * Neg_{i,t} + u_{i,t} \end{aligned} \tag{4.2}$$

In the above equations, $Pos_{i,t}$ and $Neg_{i,t}$ are the proportion of positive and negative words in company-specific media articles on day *t*, calculated before market opening on day *t*. As the media articles are published before market opening on day *t*, their effect on stock returns should be realised during day *t*. They are determined by using textual analysis to identify words that were either positive or negative in nature according to the Loughran and McDonald (2010) financial news word lists. *ER* is the excess returns over the market, *Ln(Size)* is the log of the market capitalisation, *Ln(BTM)* is the log of the book-to-market ratio and *Ln(Turnover)* is the daily trading volume. *MC* is the daily per stock media coverage and *MC*Neg* is an interaction term to account for the relationship between media coverage and negative media sentiment. It is calculated as the daily stock specific media coverage multiplied by the stock specific daily measure of negative media sentiment.

Equation (4.1) examines the relationship between stock returns and media sentiment on the day the media articles are published and equation (4.2) tests for return predictability of media sentiment on the day following the publication of the media articles.

5. Results

5.1 The relationship between media sentiment and stock returns

We first examine the relationship between media sentiments (positive and negative) about a given firm and its excess returns on the day the media articles are published. The media sentiment is measured by the fraction of positive and negative words in firm-specific media articles -published -before the market opens on that day. Table 3 displays the results of the OLS regression constructed in equation (4.1).

[Insert Table 3]

We find the signs of the coefficients are as we would expect, positive for positive sentiment and negative for negative sentiment. That is, positive sentiment is associated with positive daily excess returns, and negative sentiment is associated with negative daily excess returns. Both positive and negative sentiment variables are individually significant. Accounting for the interaction between negative media sentiment and media coverage, we see that positive media sentiment has a much more significant effect on stock returns. The interactive term has a higher level of significance than the negative media sentiment variable, showing that this relationship has a significant effect on excess stock returns.

We see significant continuation and reversal effects on the coefficients of ER_{t-1} and ER_{t-2} respectively, showing persistence in positive excess returns for a day, before a reversal two days after. The reversal effect is much more significant however. Media coverage was also found to have a significant relationship with excess stock returns, again highlighting its importance and influence. Other significant relationships were found with *Ln(Book-to-Market)* and *Ln(Turnover)*. *Ln(Size)* was not found to be significant but this is not surprising given that all stocks in the sample are large market capitalisation stocks.

Given the extremely low R-Squared values obtained from the regression, (0.32%) these results should not be overstated as a predictor or a strong explanation of stock returns. As Loughran and McDonald (2010, p.22) concluded in their study, "Textual analysis is not the ultimate key to the returns cipher." One explanation for the very low value for R-squared could be that no variable for macroeconomic news was included. Cutler et al. (1989) found that up to one third of stock price movements could be explained by macroeconomic events and the news surrounding them. With only company-specific news, the regression analysis fails to incorporate a large amount of news that could effect stock returns. These results support those of Loughran and McDonald (2010) who found significant relationships with stock returns for positive and negative words in company 10-k filings.

5.2 Using measures of media sentiment to predict stock returns

Having established that measures of positive and negative media sentiment have significant relationships with excess stock returns on the day of publication of the media articles, we now test whether these measures of media sentiment on day zero can predict firms' close-to-close returns on day *t+1*, as in Tetlock et al. (2008). For US firms in the S&P 500, Tetlock et al. (2008) found that negative words in firm-specific news articles predict lower returns on the day following the news articles publication. We investigate whether these results obtained for the S&P 500 firms, also hold true for the FTSE 100 UK firms, given differences in media cultures and practices. The OLS regression is conducted using equation (4.2) and the results are displayed in Table 3.

The main result of this regression is that negative media sentiment, as measured by the fraction of negative words in a media article, robustly predicts lower stock returns on the day following publication of the media articles. However measures of positive media sentiment do not have any predictive power the day following publication of media articles. The significance however of the relationship between negative media sentiment and stock returns is much weaker the day following publication than on the day the media articles are published. The magnitude of the effects of negative media sentiment on excess returns also decreases between day *t* and *t+1*. The media coverage variable, and the interactive term between media coverage and negative media sentiment were both found to be insignificant on day *t+1*. Therefore we see that the effects of media coverage and its interaction with negative media sentiment are incorporated rapidly into prices on day *t*. This is as expected; an event with a high level of media coverage would encourage more informed investors to act on this new information. After controlling for media coverage, we suspect that the predictability of negative media sentiment on day *t+1* is due to the cognitive dishonesty of investors when reacting to bad news. This is supported by behavioural models that suggest investors tend to hold on to losing stocks longer than they should (see Shefrin and Statman, 1985; Frazzini 2006).

Continuation and reversal effects were again found with lags of excess returns, consistent with the results of the regression on ER_t. Control variables for value and turnover were found to be insignificant in the regression, the control variable was size though was found to be significant.

It is evident therefore that stocks in the FTSE 100 over the time period they were studied incorporated nearly all media information into stock prices on the day it was released. It was found that there is a small amount of return predictability using measures of negative media sentiment on the day following the publication of the media articles. Hence these results provide some support the efficient markets hypothesis of Fama (1970) which states that in an efficient market, stock prices rapidly represent all available information.

These results show that although there is contrasts in media cultures and reporting practices between the US and UK, the effect media sentiment has on stock prices is fairly consistent across the two markets on day *t+1* (see Tetlock et al. 2008).

As a robustness test, we regress lagged values of negative media sentiment on excess returns on the day following publication of media articles. The results displayed in Table 4 for this regression confirm those of Table 3 that negative sentiment has some predictive ability on the day following publication of media articles. All other lags of negative sentiment were found to be insignificant.

[Insert Table 4]

5.3 The relationship between media sentiment and media coverage

Further, we investigate how media sentiment is related to media coverage. The results in the previous section allude to significant interactions between media coverage and negative media sentiment. Here we investigate this relationship further. Table 5 documents our results. We find conclusively that the amount of media coverage is strongly related to negative media sentiment, or bad news. This is intuitive: as the amount of negative sentiment expressed in news articles rises, so does the amount of media coverage. Positive media sentiment also has a significant relationship with media coverage. However, the strength of this relationship is much weaker than for negative media sentiment. This result supports those in the psychology literature, which find that bad news has a much stronger effect than good news (see Baumeister et al. 2001). However, it should be borne in mind that the effect of bad news is magnified by the greater coverage it receives.

[Insert Table 5]

5.4 News-based trading strategy

Algorithmic trading strategies involving news-reading bots have been popularised in the news in recent times. The determination of trading signals using news media has obviously not been made public by the institutions using them but an attempt has been made using the media sentiment data to produce a simple trading strategy. Figure 2 displays the performance of a simple trading strategy using media sentiment from news articles compared to the FTSE 100, which has been rebased to 100 as of January 2005.

[Insert Figure 2]

This news-based trading strategy uses average media sentiment calculated as the average of media sentiment of all stocks that have news published about them on a particular day before trading opens, taken on a daily basis to determine whether to take a long or short position in the FTSE 100 Index. To determine the most accurate trading signal, daily average positive media sentiment is weighted by a factor of 2.11. We do this as the characteristics of the data show that average negative media sentiment is 2.11 times greater than average positive media sentiment. The strategy takes a long or short position on a daily basis, reinvesting all continuously compounded returns to date. It takes a long position in the FTSE 100 Index if the weighted average positive media sentiment is greater than average negative media sentiment. It takes a short position if average negative media sentiment is greater than weighted average positive media sentiment. It is assumed that the execution of such a strategy could be performed using an ETF or futures contracts with no transactions costs being accounted for.

Table 6 shows the risk-adjusted returns for the news-based trading strategy for each year of the sample. The strategy is adjusted for the Fama-French (1993) three-factor model to account for contemporaneous market, size and book-to-market factors.

[Insert Table 6]

The results in Table 6 show that over the sample period of 2005-2010 the news-based trading strategy earns significant positive abnormal returns of 0.08% per day, not taking into account transactions costs of market frictions. This translates to an approximate annual abnormal return of 22.3%, that can have substantial economic implications. By visually inspecting the continuously compounded returns of the trading strategy in Figure 2, the performance of the trading strategy improves significantly after 2007, this also corresponded to an increase in the volume of financial news data (see Table 1) allowing more accurate trading signals to be determined.

6. Conclusions

This study examines the relationship between measures of media sentiment and stock returns for FTSE 100 companies over the period 2005 – 2010 and tests whether there is any stock return predictability inherent in measures of media sentiment.

Our main results show firstly that positive media sentiment in company-specific news articles has a significant positive relationship with UK stock returns and negative media sentiment has a significant negative relationship with UK stock returns. This is consistent with the findings of Carretta et al. (2011) related to the effect of tone on stock returns. These significant relationships are strongly apparent on the day that the news articles are published, also apparent is a strong relationship between negative media sentiment and media coverage whose interaction also has a significant effect on stock returns. When testing for stock return predictability on the day following publication of media articles however it was found that only measures of negative media sentiment have a significant effect, consistent with the results of Tetlock et al. (2008), showing that although there may be differences in media reporting cultures between the US and UK, there are strong similarities with the way media sentiment is incorporated into stock returns.

Secondly, we demonstrate that the relationships between media sentiment and stock returns has significant economic implications by constructing a simple trading strategy using measures of media sentiment to determine trading signals that earns substantial and significant abnormal returns.

These results suggest that the UK market is fairly efficient at incorporating information and sentiment contained in media articles into stock prices. Most media sentiment was incorporated into stocks on the day the articles were released. Negative media sentiment was found to have a much less significant relationship with excess returns on day following publication, which was also smaller in magnitude. However the fact that there is some return predictability due to negative media sentiment on day following publication indicates some cognitive dishonesty towards bad news by investors, resulting in some underreaction on the day of media publication. These results find support from the literature in Grossman and Stiglitz (1980), with this underreaction to negative news providing motivation for market participants to monitor financial news releases, other results from behavioural finance also find evidence of this underreaction to negative news (see Shefrin and Statman, 1985; Frazzini 2006). Future research should investigate more thoroughly the application of media data to financial market participants and its wider impact concerning the efficiency of markets.

References

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., and Vohs, K. D., 'Bad Is Stronger Than Good', *Review of General Psychology* , Vol. 5, 2001, pp. 323-370.

Bowley, G., 'Computers That Trade on the News', *New York Times*, 22/12/2010

Carretta, A., Farina, V., Martelli, D., Fiordelisi, F., and Schwizer, P., 'The Impact of Corporate Governance Press News on Stock Market Returns', *European Financial Management* , Vol. 17 (1), 2011, pp. 100-119.

Cutler, D. M., Poterba, J. M., and Summers, L. H., 'What moves stock prices?', *The Journal of Portfolio Management* , Vol. 15 (3), 1989, pp. 4-12.

Das, S. R., and Chen, M. Y., ‘Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web’, *Management Science* , Vol 53 (9), 2007, pp. 1375–1388.

Davis, A. K., Piger, J. M., and Sedor, L. M., ‘Beyond the Numbers: Managers’ Use of Optimistic and Pessimistic Tone in Earnings Press Releases’, *SSRN*, 2008.

DeLong, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J., ‘Noise trader risk in financial markets’ *Journal of Political Economy* , Vol. 98, 1990, pp. 703–738.

Demers, E., and Vega, C., ‘Soft Information in Earnings Announcements: News or Noise?’ *INSEAD Working Paper*, 2010.

Deuze, M., ‘National News Cultures: A Comparison of Dutch, German, British, Australian and US Journalists’, *Journalism and Mass Communication Quarterly*, Vol. 79(1), 2002, pp. 134–49.

Dyck, A., and Zingales, L., ‘The Media and Asset Prices’, *Unpublished manuscript*, 2003.

Dyck, A., Volchkova, N., and Zingales, L., ‘The Corporate Governance Role of the Media: Evidence from Russia’ *Journal of Finance,* Vol. 63 (3), 2008, pp.1093-1135.

Fama, E. F., ‘Efficient capital markets: A review of theory and empirical work’ *Journal of Finance*, Vol. 25 (2), 1970, pp. 383-417.

Fama, E. F., and French, K. R., ‘The cross-section of expected stock returns’, *Journal of Finance*, Vol. 47 (2), 1992, pp. 427-465.

Fama, E. F., and French, K. R., ‘Common risk factors in the returns of stocks and bonds’, *Journal of Financial Economics*, Vol. 33, 1993, pp. 3–56.

Fang, L., and Peress, J., ‘Media Coverage and the Cross-section of Stock Returns’, *Journal of Finance*, Vol. 64 (5), 2009, pp. 2023–2052.

Frazzini, A., ‘The Disposition Effect and Underreaction to News’, *Journal of Finance,* Vol. 61 (4), 2006, pp. 2017–2046.

Grossman, S.J., and Stiglitz, J.E., ‘On the Impossibility of Informationally Efficient Markets’, *AER*, Vol. 70 (3), 1980, pp. 393-408.

Loughran, T., and McDonald, B., ‘When is a Liability not a Liability? Textual Analysis, Dictionaries, and 10-Ks’, *Forthcoming - Journal of Finance*, 2010.

Sadique, S., In, F., and Veeraraghavan, M., ‘The Impact of Spin and Tone on Stock Returns and Volatility: Evidence from Firm-issued Earnings Announcements and the Related Press Coverage’, *SSRN*, 2008.

Salmon, F., and Stokes, J., ‘Algorithms Take Control of Wall Street’, *Wired Magazine http://www.wired.com/magazine/2010/12/ff_ai_flashtrading/all/1*, 27/12/2010.

Shaw, C., ‘Deciding what we watch: taste, decency, and media ethics in the UK and the USA’, *Oxford University Press,* 1999.

Shefrin, H. and Statman, M., ‘The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence’, *Journal of Finance,* Vol. 40 (3), 1985, pp. 777-790.

Tetlock, P.C., ‘Giving Content to Investor Sentiment: The Role of Media in the Stock Market’, *Journal of Finance*, Vol. 62 (3), 2007, pp. 1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S., ‘More Than Words: Quantifying Language to Measure Firms' Fundamentals’, *Journal of Finance*, Vol. 63 (3), 2008, pp. 1437–1467.

Weaver, D. H., ‘The Global Journalist: News People around the World’, *New Jersey: Hampton Press*, 1998.

Table 1 Summary statistics for raw media data.

News data was downloaded from LexisNexis UK. Coverage statistics give the proportion of the media articles of which came from specific publications. Many of these articles cover the same events and the same companies on the same day. POS and NEG are the proportion of positive and negative words in news articles, determined by using textual analysis to identify words that were either positive or negative in nature according to the Loughran and McDonald (2010) financial news word lists. (FT) is the Financial Times * which also includes FT.com, (Times) is The Times newspaper, (Guardian) is The Guardian newspaper and (Mirror) is the Mirror newspaper. **The data set only covers up until 31st October 2010.

Year	Total Articles	Coverage				Average Article Words	Average POS	Average NEG
		FT*	Times	Guardian	Mirror			
2005	6424	42.17%	29.56%	20.22%	8.05%	445	0.95%	1.64%
2006	9431	60.28%	25.83%	6.21%	7.68%	441	0.95%	1.70%
2007	16976	69.59%	17.87%	8.97%	3.56%	478	0.93%	1.96%
2008	26497	71.49%	13.83%	10.19%	4.48%	484	0.86%	2.46%
2009	25926	67.81%	16.49%	11.36%	4.33%	487	0.94%	2.32%
2010**	20752	65.56%	18.92%	7.56%	7.96%	469	0.88%	2.35%

Table 2 Panel A Summary statistics for days with news.
Media articles were downloaded from LexisNexis UK, Pos and Neg are the proportion of positive and negative words in company-specific media articles on day zero, determined by using textual analysis to identify words that were either positive or negative in nature according to the Loughran and McDonald (2010) financial news word lists. Size is the daily market equity of firms, Book-to-market is the daily book value divided by market equity and turnover is the daily trading volume for each company.*For All News, this includes positive and negative news put together in one data set, however negative news is given a negative sign. Daily log returns are the company-specific daily log returns, the daily excess returns are the company-specific daily log returns – FTSE 100 daily returns.

Variable	Mean	Median	Standard Deviation	Minimum	Maximum	Skewness
Words						
Pos	0.0098	0.0090	0.0067	0	0.0602	1.0229
Neg	0.0207	0.0187	0.0133	0	0.1268	1.2704
All News*	-0.0053	0	0.0185	-0.1268	0.0602	-0.6839
Control Variables						
Ln(Size)	9.5185	9.6468	1.2099	5.3245	11.7942	0.4272
Ln(Book-to-market)	-0.7236	-0.7655	0.7437	-3.1202	2.9957	-0.8109
Ln(Turnover)	8.4445	8.6266	2.0337	0	14.6139	-1.9253
Daily Log Returns	0.0121%	0%	3.25%	-49.81%	54.95%	-2.2238
Daily Excess Returns	0.0002%	0%	2.02%	-108.64%	51.16%	-1.7513

Table 2 Panel B Largest negative market movements. The FTSE 100 is used to proxy market movements over the sample period. Average positive and average negative media sentiment is calculated for companies who had news articles published about them that day.

Date	FTSE 100 Daily Log Return	Average Positive Sentiment	Average Negative Sentiment
10/10/2008	-9.27%	0.009	0.029
06/10/2008	-8.18%	0.007	0.025
15/10/2008	-7.43%	0.013	0.021
06/11/2008	-5.87%	0.010	0.024
21/01/2008	-5.64%	0.006	0.020

Table 2 Panel C Largest positive market movements. The FTSE 100 is used to proxy market movements over the sample period. Average positive and average negative media sentiment is calculated for companies who had news articles published about them that day.

Date	FTSE 100 Daily Log Return	Average Positive Sentiment	Average Negative Sentiment
24/11/2008	9.38%	0.010	0.025
19/09/2008	8.47%	0.011	0.028
13/10/2008	7.94%	0.011	0.023
29/10/2008	7.74%	0.009	0.028
08/12/2008	6.01%	0.011	0.021

Figure 1 Rolling 30 day averages of positive and negative media sentiment

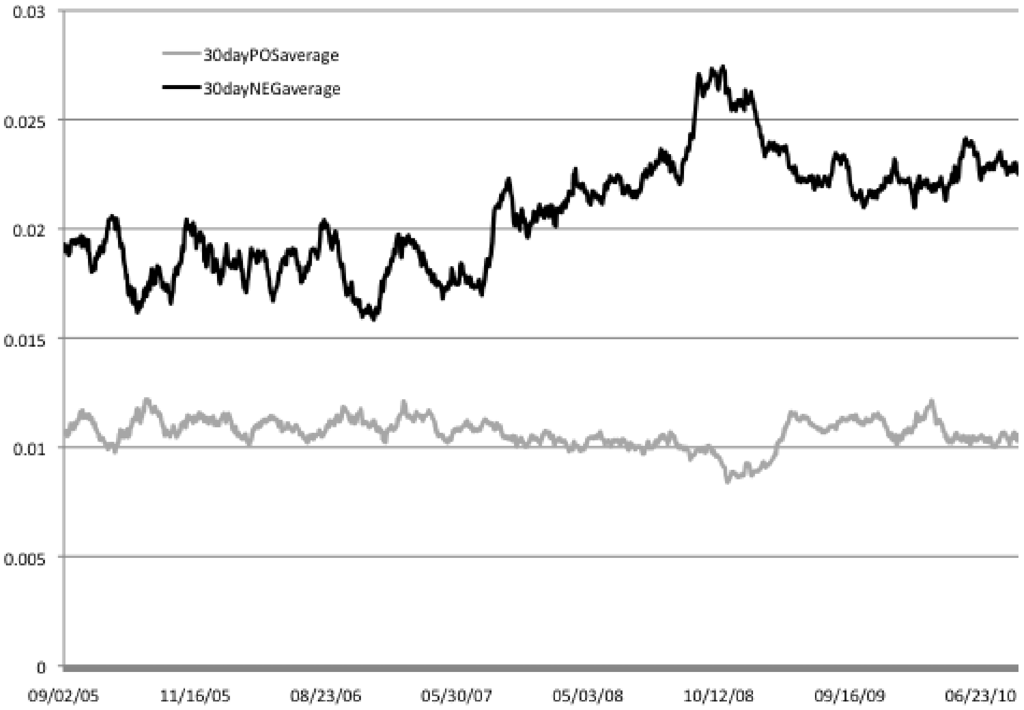

Figure 1 shows the rolling 30 day averages of positive and negative media sentiment. These are constructed by taking the daily average of positive and media sentiment across all firms, then using a rolling 30 day average of all firms to smooth the data. This aptly demonstrates the differences between the two quantitative measures of the media sentiment. Firstly, the positive content seems to stay at a fairly consistent level, with little variation, its lowest point coming in the latter half of 2008, at the heart of the financial crisis. In contrast the level of negative content varies widely, having its highest point in the latter half of 2008. It also appears from the figure that the average level of negative content may have risen mid 2007 in response to the financial crisis and the fear that has been subsumed in economic media since that time period.

Table 3 Relationship between media sentiment and stock returns. The dependent variables are excess returns on day *t* and *t+1*. Excess returns are calculated by subtracting the daily market return from the daily stock return. Media articles were downloaded from LexisNexis UK, Pos_t and Neg_t are the proportion of positive and negative words in company-specific media articles on day *t*, calculated before market opening on day *t*. As the media articles are published before market opening on day *t*, their effect on stock returns should be realised on day *t*. They are determined by using textual analysis to identify words that were either positive or negative in nature according to the Loughran and McDonald (2010) financial news word lists. ER_t, ER_{t-1} and ER_{t-2} are the daily excess returns on days *t*, *t-1* and *t-2* respectively. The regressions include control variables for market equity, book to market equity, trading volume and media coverage. Size is the daily market equity of firms, book-to-market is the daily book value divided by market equity and turnover is the daily trading volume for each company and media coverage is defined as the number of media articles about a particular company on a given day. $MC*Neg_t$ is an interaction term to account for the relationship between media coverage and negative media sentiment. It is calculated as the daily stock specific media coverage multiplied by the stock specific daily measure of negative media sentiment. Robust t-statistics are reported in parentheses below the parameter coefficients. The equations used to construct these regressions are:

$$ER_{i,t} = \beta_1 + \beta_2 Pos_{i,t} + \beta_3 Neg_{i,t} + \beta_4 ER_{i,t-1} + \beta_5 ER_{i,t-2} + \beta_6 Ln(Size)_{i,t} + \beta_7 Ln(BTM)_{i,t}$$
$$+ \beta_8 Ln(Turnover)_{i,t} + \beta_9 MC + \beta_{10} MC * Neg_{i,t} + u_{i,t}$$
$$ER_{i,t+1} = \beta_1 + \beta_2 Pos_{i,t} + \beta_3 Neg_{i,t} + \beta_4 ER_{i,t} + \beta_5 ER_{i,t-1} + \beta_6 ER_{i,t-2} + \beta_7 Ln(Size)_{i,t} + \beta_8 Ln(BTM)_{i,t}$$
$$+ \beta_9 Ln(Turnover)_{i,t} + \beta_{10} MC + \beta_{11} MC * Neg_{i,t} + u_{i,t}$$

	Excess Returns	
Variable	ER t	ERt+1
<u>Words</u>		
Pos_t	0.0991	0.0110
	(6.17)	(0.68)
Neg_t	-0.0306	-0.0213
	(-3.55)	(-2.47)
<u>Control Variables</u>		
ER_t		0.0119
		(3.83)
ER_{t-1}	0.0097	-0.0251
	(3.12)	(-8.04)
ER_{t-2}	-0.0263	-0.0090
	(-8.44)	(-2.88)
Ln(Size)	-0.00004	-0.0002

	(-0.64)	(-3.86)
Ln(Book-to-market)	-0.0005	0.0001
	(-6.17)	(1.51)
Ln(Turnover)	0.00007	-0.00006
	(2.00)	(-1.84)
Media Coverage	0.0002	-0.00002
	(4.77)	(-0.05)
(Media Coverage)*Neg_t	-0.0147	0.0013
	(-8.31)	(0.7459)
Constant	-0.0005	0.0029
	(-1.05)	(6.09)
Observations	103224	103224
R-Squared	0.0032	0.0013

Table 4 Return predictability of negative media sentiment
The dependent variable is excess returns on day *t+1*, the day after the media articles are published. Excess returns are calculated by subtracting the daily market return from the daily stock return. Media articles were downloaded from LexisNexis UK, Neg_t is the proportion of negative words in company-specific media articles on day *t*, determined by using textual analysis to identify words that were negative in nature according to the Loughran and McDonald (2010) financial news word lists. Neg_{t-1}, Neg_{t-2}, Neg_{t-3}, Neg_{t-4}, are the previous 4 days values of negative media sentiment. Robust t-statistics are reported in parentheses below the

parameter coefficients. The equation used to construct this regression is: $ER_{i,t+1} = \beta_1 + \beta_2 Neg_{i,t} + \beta_3 Neg_{i,t-1} + \beta_4 Neg_{i,t-2} + \beta_5 Neg_{i,t-3} + \beta_5 Neg_{i,t-4} + u_{i,t}$

Variable	
Neg,t	-0.0171
	(-2.54)
Neg,t-1	0.0050
	(0.72)
Neg,t-2	-0.0056
	(-0.80)
Neg,t-3	-0.0006
	(-0.09)
Neg,t-4	-0.0105
	(-1.56)
Constant	0.0004
	(4.89)
Observations	103224
R-Squared	0.0001

Table 5 Regression to test whether good or bad news affects the amount of media coverage.

Media coverage, as measured by the amount of company-specific news articles published about a particular stock on a particular day is regressed against measures of positive and negative media sentiment. Media articles were downloaded from LexisNexis UK, Pos_t and Neg_t are the proportion of positive and negative words in company-specific media articles on day *t*, calculated before market opening on day *t*. As the media articles are published before market opening on day t, their effect on stock returns should be realised on day *t*. They are determined by using textual analysis to identify words that were either positive or negative in nature according to the Loughran and McDonald (2010) financial news word lists. The regression includes control variables for market equity and trading volume. Size is the daily market equity of firms, turnover is the daily trading volume for each company. Robust t-statistics are reported in parentheses below the parameter coefficients. The equation used to construct this regression is:

$$MediaCoverage_i = \beta_1 + \beta_2 Pos_i + \beta_3 Neg_i + \beta_4 Ln(Size) + \beta_5 Ln(Turnover) + u_i$$

Variable	
Pos	5.4004
	(1.17)
Neg	40.6340
	(17.55)
Ln(Size)	0.1027
	(3.57)
Ln(Turnover)	0.7110
	(37.59)
Constant	-4.8848
	(-18.52)
Observations	23663
R - Squared	0.0890

Figure 2 Continuously compounded returns of a news-based trading strategy that takes long or short positions in the FTSE 100 Index according to trading signals determined by average media sentiment. The continuously compounded returns of the strategy are plotted against the FTSE 100 Index for comparison.

This news-based trading strategy uses average media sentiment calculated as the average media sentiment of all stocks who have news published about them on a particular day before trading opens, taken on a daily basis to determine whether to take a long or short position in the FTSE 100 Index. To determine the most accurate trading signal, daily average positive media sentiment is weighted by a factor of 2.11. We do this as the characteristics of the data show that average negative media sentiment is 2.11 times greater than average positive media sentiment. The strategy takes a long or short position on a daily basis, reinvesting all continuously compounded returns to date. It takes a long position in the FTSE 100 Index if the weighted average positive media sentiment is greater than average negative media sentiment. It takes a short position if average negative media sentiment is greater than weighted average positive media sentiment. It is assumed that the execution of such a strategy could be performed using an ETF or futures with no transactions costs being accounted for.

Table 6 Risk adjusted news-based trading strategy results.

This table shows daily risk adjusted returns from the news-based trading strategy as the dependent variable. The regressions use the Fama-French (1993) three-factor model to adjust the trading strategy returns for the impact of contemporaneous market (Market), size (SMB), and book-to-market (HML) factors. Alpha (Jensen's) is abnormal returns. This news-based trading strategy uses average media sentiment calculated as the average media sentiment of all stocks who have news published about them on a particular day before trading opens, taken on a daily basis to determine whether to take a long or short position in the FTSE 100 Index. To determine the most accurate trading signal, daily average positive media sentiment is weighted by a factor of 2.11. We do this as the characteristics of the data show that average negative media sentiment is 2.11 times greater than average positive media sentiment. The strategy takes a long or short position on a daily basis, reinvesting all continuously compounded returns to date. It takes a long position in the FTSE 100 Index if the weighted average positive media sentiment is greater than average negative media sentiment. It takes a short position if average negative media sentiment is greater than weighted average positive media sentiment. It is assumed that the execution of such a strategy could be performed using an ETF or futures with no transactions costs being accounted for.

	2005	2006	2007	2008	2009	2010	Full Period
Alpha	-0.0006	0.00005	0.00008	0.0014	0.0018	0.0010	0.0008
	(-2.07)	(0.13)	(0.11)	(1.14)	(2.06)	(1.43)	(2.47)
Market	0.3741	0.4292	0.2584	-0.4842	0.0718	-0.4226	-0.1070
	(4.09)	(5.25)	(2.94)	(-3.92)	(0.69)	(-3.72)	(-2.40)
SMB	-0.0651	-0.0956	0.0682	0.0201	0.1935	-0.1104	0.1252
	(-1.04)	(-1.79)	(1.12)	(0.28)	(3.81)	(-1.17)	(4.80)
HML	-0.0074	-0.0240	-0.1194	-0.0105	-0.0088	0.0801	0.0054
	(-0.13)	(-0.43)	(-1.79)	(-0.23)	(-0.27)	(0.87)	(0.28)
Trading Days	260	260	261	262	261	216	1519
R-Squared	0.2009	0.3013	0.0391	0.2667	0.1083	0.1082	0.0955