Media Content and Stock Returns: The Predictive Power of Press

Nicky J. Ferguson
University of Cambridge, UK

Dennis Philip
Durham University Business School, UK

Herbert Y. T. Lam
University of China, China

Jie Michael Guo
Durham Univercity Business School, UK

This paper examines whether tone (*positive* and *negative*) and volume of firm-specific news media content provide valuable information about future stock returns, using UK news media data from 1981–2010. The results indicate that both tone and volume of news media content significantly predict next period abnormal returns, with the impact of volume more pronounced than tone. Additionally, the predictive power of tone is found to be stronger among lower visibility firms. Further, the paper finds evidence of an attention-grabbing effect for firm-specific news stories with high media coverage, mainly seen among larger firms. A simple news-based trading strategy produces statistically significant risk-adjusted returns of 14.2 to 19 basis points in the period 2003–2010. At the aggregate level, price pressure induced by semantics in news stories is corrected only in part by subsequent reversals. Overall, the findings suggest firm-specific news media content incorporates valuable information that predicts asset returns. (JEL: G1, G14, G17)

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

---

* We would like to thank Nick Baltas, Emma Black, Jinghan Cai, Changyun Wang, Luigi Zingales and conference participants at the European Financial Management Association Annual Meeting, Midwest Finance Association Annual Meeting and the UBS Quantitative Investment Conference for helpful comments and feedback.

(*Multinational Finance Journal*, 2015, vol. 19, no. 1, pp. 1–31)

I. Introduction

News media publications play an important role in providing financial market participants with valuable information and aiding investors in forming their views on the stock market. A firm's stock prices, in theory, reflect its fundamentals and are conditional on the investors' information sets. Investors receive both private and public information concerning the underlying value of a stock. Also contained in an investor's information set are qualitative descriptions of the 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 announcements. Shiller (2005) suggests that news media actively shape public opinion and play a large role in the propagation of speculative bubbles, through feedback mechanisms and attention cascades, whereby the media may exaggerate the relevance of past price movements, affecting future price movements.

The conundrum of explaining the movements in stock prices that cannot be accounted for by new fundamental or economic information is an interesting puzzle that has remained unsolved due to the difficulties of quantifying or measuring qualitative news media data (see Cutler, Poterba, and Summers, 1989). However, in recent times researchers have begun to analyse linguistic data contained in media articles using textual analysis in an attempt to capture hard-to-quantify firm-specific information in news media data and determine the impact on stock prices (for example, Tetlock, 2007; Tetlock, Saar-Tsechansky, and Macskassy, 2008; Garcia, 2013; Loughran and McDonald, 2011; among others). By using a quantitative measure of the semantics in the language used in news articles, it is possible to measure the effects of investor reaction to such news events and identify common patterns concerning the way asset prices react to news in general, whether *positive* or *negative*.

Previous research shows that the tone in newspaper columns drives investor sentiment (Tetlock, 2007; Garcia, 2013), captures information beyond fundamentals (Tetlock, Saar-Tsechansky, and Macskassy, 2008) and affects individual trading behaviour (Kelley and Tetlock, 2013). Moreover, the tone of news can be improved by increasing local advertising spending (Gurun and Butler, 2012) and hiring investor relationship firms (Solomon, 2012). Another branch of studies shows that the amount of news media coverage reduces firms' expected returns (Fang and Peress, 2009; Peress, 2014) and stimulates local trading

(Engleberg and Parsons, 2011).[1] Dougal et al. (2012) find that financial journalists have the potential to influence investor behaviour and Griffin, Hirschey, and Kelly (2011) shows that reaction to news media varies around the world according to levels of development, information quality, and information transmission mechanisms. Nearly all the studies of media interactions with financial markets predominately examine news media content in the US market.

This paper, using information from daily firm-specific newspaper articles, investigates the link between news media content and stock market activity. The study is conducted using a large news media dataset from the UK market. Existing studies mostly rely on news media content sourced from the US market, and hence this study is one of the first to provide international evidence of the effect of news media content on stock returns. Our sample consists of 264,647 firm-specific UK news media articles covering *FTSE* 100 firms over the period 1981 to 2010. The 30-year sample period of UK news media data enables us to conduct a comprehensive analysis of the effect of news media content on the distribution of UK stock returns. Our sample period is large and comparable to those considered in other media studies. The UK, as a leading global financial centre, with some of the world's oldest and most respected news publications, is a key market for analysing the role of the media in shaping public opinion and investor reaction. We source the news articles from national newspapers that are globally recognised, namely, *The Financial Times* (*FT*), *the Times*, *the Guardian* and *Mirror*.

Using this comprehensive firm-level media data, we evaluate whether stock market returns reflect information from *positive* and *negative* words in news media content. We extend the existing literature in several aspects. We first consider both *positive* as well as *negative* news media content, constructed from Loughran and McDonald's (2011) financial-news-specific word lists, to study the predictability of stock returns.[2] Previous studies, such as that of Tetlock (2007) and

1. The informational role of media content is also documented in other markets, such as the debt market. For example, Liu (2014) finds that, during the recent debt crisis, media pessimism and the volume of news provide value-relevant information not quantified by the traditional determinants of long-term sovereign bond yield spreads.

2. Previous studies, such as those of Tetlock (2007) and Tetlock, Saar-Tsechansky, and Macskassy (2008), use the Harvard psychosocial dictionary to identify words of different categories in news articles. However, Loughran and McDonald (2011) create a new word list of financial-news-specific words that have greater explanatory power over stock returns than the Harvard psychosocial dictionary categories.

Tetlock, Saar-Tsechansky, and Macskassy (2008), among others, only consider the effect of *negative* words in news stories on stock returns.[3] By studying both *positive* and *negative* measures of media content, this paper uses the overall distribution of news to gain insight into the information embedded in news articles. In addition, we consider earnings-related *positive* and *negative* words in news stories and investigate whether the linguistic tone of news stories reflects valuable information about firms' fundamentals that are not captured otherwise.

Further, we examine the combined impact of (*positive* and *negative*) news media content and the volume of media coverage on a firm's stock returns. Previous studies examine the separate effects of the tone and volume of news media on stock returns. We conjecture that if investors are shown to overreact to attention-grabbing stocks (Barber and Odean, 2008) and linguistic tone reflects investor sentiment (Tetlock, 2007; Tetlock, Saar-Tsechansky, and Macskassy, 2008), then the combined effect of the tone and quantity of news stories should magnify market reactions.

Moreover, we split our firm-specific media article sample of *FTSE* 100 stocks by size and book-to-market ratios and study the impact of news media content on the return distribution of higher and lower visibility firms. We thus explore the notion of whether investor recognition is a determinant of the cross-sectional dispersion among stock returns. Our approach substantiates the approach of Barber and Odean (2008), who proxy attention-grabbing stocks by stocks in the news, stocks experiencing high abnormal trading volume, and stocks with extreme one-day returns, and study the effect of news attention on investor buying behaviour. In order to explore the economic significance of the impact of news stories on stock returns, we build a simple news-based trading strategy using these *positive* and *negative* measures of news media content. Finally, we also provide market-level evidence of the relationship between media content and stock returns using aggregate measures of news media content.

Overall, our empirical test results show significant predictive power of firm-specific media content for stock returns, hence corroborating the US evidence using a large independent media dataset from the UK

3. Recently a few papers (executed simultaneously), such as Jegadeesh and Wu (2013) and Garcia (2013), examine the effects of *positive* and *negative* tone in newspaper columns on asset prices. In this paper, we use firm-specific information from newspaper articles rather than information from news columns to assess the impact of *positive* and *negative* tone in news media content.

market. Specifically, we find that *positive* as well as *negative* words in news stories convey valuable information about future returns. *Positive* words in firm-specific news media content significantly predict higher returns in the next trading period, while *negative* words in firm-specific news media content significantly predict lower next trading period returns. In addition, we see that earnings-related news stories associated to firms' fundamentals generate abnormal returns on the day of news publication. Further, we show that the impact of tone is significant mainly among lower visibility firms (smaller *FTSE* 100 firms and firms with high book-to-market ratio). Such firms' stock returns show a significantly *positive* (*negative*) relationship with *positive* (*negative*) words in news articles. The results indicate that firm-specific news articles provide key incremental information about less visible firms to investors.

Furthermore, when we consider the joint impact of tone and volume of news media content, we observe that both tone and volume (proxied by high media coverage) significantly predict next trading period abnormal returns, with the impact of volume much more pronounced than tone (for both *positive* and *negative*). We see that the effect of high media coverage on future returns is mainly driven by the largest *FTSE* 100 firms. The largest *FTSE* 100 firms attract the highest media attention and are therefore prone to market overreactions to attention-grabbing firm-specific news. More specifically, the results indicate that the market reacts to highly visible *positive* news, affecting next-period abnormal returns. This is consistent with the attention-grabbing effect of Barber and Odean (2008), whereby buying decisions are often harder than selling because investors need to choose from thousands of stocks when they decide which to buy; however, they only decide which to sell of those that they currently hold. Therefore, the attention-grabbing effect is more pronounced when investors are making buying decisions. Moreover, we also find significant market reaction to highly visible *negative* news published in the *FT*. Since *FT* publications consistently cover key news stories and are widely read to institutional investors and traders, high media coverage of *negative* news publications in the *FT* can induce *negative* pressure on prices in the market, generating *negative* next trading period abnormal returns. The results indicate that both tone and volume provide novel information about firms' future returns.

To gauge the potential economic significance of media content in stock returns, we construct a simple news-based trading strategy using firm-specific *positive* and *negative* words in news media content. For the recent period 2003 to 2010, we find that the strategy produces an average daily return of 19 basis points for trades placed using the *positive* and *negative* words published in *FT* news stories and an average daily return of 14.2 basis points for trades based on *positive* and *negative* words in the composite media content of all news articles. Finally, we show that *positive* and *negative* news media content has a significant impact on stock returns at the aggregate market-level. The evidence suggests that initial price pressures caused by the news stories does not show strong significant reversals in the subsequent trading week, and hence the linguistic media content in news articles, also at the aggregate level, conveys significant information about stock returns.

The outline of the remainder of this paper is as follows. Section II discusses the properties of the UK news media data. Section III and IV present the main results of this study, examining the effect of news media content on stock returns. Section V investigates the relationship between media content and stock returns at the market level using aggregate measures of news media content. Section VI concludes this study.

II. News Media Data Characteristics and Variable Construction

For the empirical analysis, news media articles specific to individual firms are obtained manually from LexisNexis UK. The sources of the LexisNexis UK data include the daily publications *The Financial Times*, *The Times*, *The Guardian*, and *Mirror*. The data covers UK firms listed on the *FTSE* 100 Index from 1981 through 2010. A total of 264,647 media articles were used in our analysis over the sample period considered.[4]

The content of the media articles is analysed to determine the number of *positive* and *negative* words they contain. The words in each article are compared to Loughran and McDonald's (2011) *positive* and *negative* financial word lists to identify the number of *positive* and

4. We only consider articles with a LexisNexis relevance score of 90 percent or above for each firm, to ensure the quality of firm-specific information in the articles (Fang and Press, 2009, carry out similar filtering).

negative words in a financial context.[5] Some previous studies use the Harvard psychosocial dictionary to categorize the words featured in financial news articles. Loughran and McDonald (2011) 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 of media sentiment on asset prices using the Harvard psychosocial dictionary can lead to the effect that *negative* media sentiments will be overstated. Loughran and McDonald (2011) show 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, the authors create a specialized list of words that carry a *negative* sentiment in the financial sense. This enables them to account more accurately for *negative* sentiment when reviewing financial media. Loughran and McDonald's (2011) current *positive* and *negative* lists contain 353 and 2,337 words, respectively. The measures of *positive* and *negative* news media content are determined for each individual news media article as follows:

$$Positive\ Content = \frac{number\ of\ positive\ words}{Total\ words} \tag{1}$$

$$Negative\ Content = \frac{number\ of\ negative\ words}{Total\ words} \tag{2}$$

We then average and standardize these measurements of *positive* and *negative content* for all news media articles written about each firm per day to construct the variables *Pos* and *Neg* measures per day, which provide a daily firm-specific quantitative measurement of semantic news media content.[6]

---

5. The *positive* and *negative* financial word lists can be obtained from McDonald's website at http://www.nd.edu/~mcdonald/Word_Lists.html

6. The standardization is carried out using the mean and standard deviations from the last calendar year (analogous to Tetlock, Saar-Tsechansky, and Macskassy, 2008). We also consider other measures of *positive* and *negative* news media content such as (#*Positive words*) / (#*Positive words* + #*Negative words*), (#*Negative words*) / (#*Positive words* + #*Negative words*), and Ln(1+ *Pos*) and Ln(1 + *Neg*) and find similar results, consistent with the measures selected.

TABLE 1. Summary statistics for the news media data

A. Sample statistics for raw media data

		Coverage				Average Article	Mean	Mean
Year	Total Articles	*FT*	*Times*	*Guardian*	*Mirror*	*Words*	*Positive*	*Negative*
1981–1985	15431	82%	6%	12%	0%	442	0.0098	0.0165
1986–1990	30842	47%	20%	33%	0%	435	0.0078	0.0159
1991–1995	39284	51%	20%	27%	2%	548	0.008	0.0163
1996–2000	55596	51%	22%	16%	11%	476	0.0088	0.0154
2001–2005	40391	45%	19%	24%	12%	441	0.0091	0.0185
2006–2010	83103	66%	10%	18%	6%	476	0.0090	0.0222
1981–2010	264647	56%	17%	21%	6%	475	0.0087	0.0183

B. Descriptive statistics for news media content measures

	Mean	Median	S.D.	Minimum	Maximum
Positive	0.0087	0.0079	0.0064	0.0000	0.0857
Negative	0.0176	0.0156	0.0123	0.0000	0.1600
Fund	0.1486	0.0000	0.3557	0.0000	1.0000
MC	0.2189	0.0000	0.4135	0.0000	1.0000

(*Continued*)

TABLE 1. (Continued)

Note: This table presents the summary statistics of the media data used in this study. News data are downloaded from LexisNexis UK. Coverage statistics give the proportion of media articles that came from specific publications. News articles are sourced from The *Financial Times* (*FT*), The *Times*, The *Guardian*, and *Mirror*. The data covers UK firms listed on the *FTSE* 100 from 1981 through 2010. A total of 264,647 media articles were used for constructing these variables. The variables *Positive* and *Negative* are the average proportions of *positive* and *negative words*, respectively, in firm-specific news articles published daily, determined by using textual analysis to identify words that are either *positive* or *negative* according to Loughran and McDonald's (2011) financial news word lists. The variable *Fund* is a dummy variable that is equal to one for news stories that contain the word stem 'earn' and the media coverage variable *MC* is a dummy variable that takes the value one if more than three articles covering the firm-specific news stories are published on a given day.

The news media articles are dated on the trading day on which they are published. This is appropriate, since all the news sources in our sample are daily publications. For instance, *FT*, which makes up the largest part of our sample (56%), goes to press around 1 a.m. on the day it is published. All deliveries are completed by 7 a.m., which is before the UK stock markets open. Hence it would be expected that investors would act upon the news media content on the day of the publication. Therefore we match the firm-level measures of *Pos* and *Neg* to the associated firm's daily excess stock returns. For days when there is no media coverage about a specific firm, *Pos* and *Neg* have a value of zero. This approach is similar to that of Loughran and McDonald (2011), who evaluate the proportion of words from a specific word list appearing in a firm's 10-K report. Table 1 reports the summary statistics of the news media data.

In Panel A we observe the characteristics of raw UK news media data and their semantic content over the last 30 years. *Positive* and *Negative* measures are average proportions of *positive* and *negative words* in firm-specific news articles published daily. We see the volume of news has been generally increasing from 1981 to 2010. News media's fascination with financial markets appears to have peaked around the time of the dot-com bubble of 1996–2000, which has the lowest mean *negative* news media content, and the recent financial crisis of 2006–2010, which has the highest mean value for *negative* news media content. In Panel B, we present the descriptive statistics for the media content variables. The variable *Fund* is a dummy variable that is equal to one for news stories that contain the word stem 'earn' and the media coverage variable *MC* is a dummy variable that takes the value one if more than three articles covering the firm-specific news stories are published on a given day. From Panel B we observe that *positive* words have a mean of 0.0087 and *negative* words have a mean of 0.0176. This indicates that the proportion of *negative* words in firm-specific news articles is almost double that of *positive* words in news articles during the sample period. The sample statistics for the *Fund* variable reveals that 15% of the new articles relate to earnings-specific news and contain the word stem 'earn'.

III. Return Predictability of News Media Content

In this section, we test the empirical hypothesis that semantic measures of news media content predict future stock returns. Tetlock, Saar-Tsechansky, and Macskassy (2008) show that the rudimentary measures capturing *negative* news stories contribute to the predictability of subsequent period stock returns. They show that there is significant qualitative information embedded in the *negative words* in news stories that is not already represented in the firms' fundamentals and stock prices. Using measures of both *positive* and *negative* news media content, we reassess the predictive power of news stories for stock returns using our independent sample of UK *FTSE* 100 firms. We hypothesize that *positive* and *negative words* in firm-specific news stories predict firms' future stock returns.

The construction of daily firm-specific *positive* and *negative* measures of news media content is detailed in Section II. We use the standardized measurements of *positive* (*Pos*) and *negative* (*Neg*) news media content in all our regressions. All news sources in our sample are daily publications of news stories from day zero, which is released before the market opens on day one (+1). We use the daily close-to-close raw stock returns ($RETURNS_{+1,+1}$) as well as the abnormal returns ($FFCAR_{+1,+1}$) from day zero to the day of the news publication to measure the impact of the media content on the closest next trading day, where we would expect the impact to be realized. We calculate the abnormal returns by subtracting the actual returns from the expected returns, which are calculated on a daily basis using the Fama and French (1993) three factor model that includes the standard risk factors MRP, SMB and HML, estimated for the UK market. We use the estimation window of [–252,–31] trading days before the day the news story takes place. In all our regressions, similar to Tetlock, Saar-Tsechansky, and Macskassy (2008), we exclude the dates with no news articles. We include in our regressions the close-to-close abnormal returns on the day the news story takes place ($FFCAR_{0,0}$), abnormal return on the previous day ($FFCAR_{-1,-1}$) and abnormal return on day –2 ($FFCAR_{-2,-2}$) to control for the recent firms' returns. We also include the cumulative abnormal return from the rest of the previous month ($FFCAR_{-30,-3}$) and the cumulative abnormal return over the previous calendar year excluding the previous month ($FFAlpha_{-252,-31}$) to control for past momentum effects and to isolate the impact of news stories. $FFAlpha_{-252,-31}$ is the intercept term from the Fama and French (1993) three factor benchmark model used in the event study methodology with the estimation window of [–252,–31] trading days before the day of the news story. Further, we include the lags of the key

TABLE 2. Predicting returns using positive and negative words

	$Return_{+1,+1}$		$FFCAR_{+1,+1}$	
	FT	*ALL*	*FT*	*ALL*
Pos	0.1474***	0.0926***	0.1574***	0.0497***
	(5.51)	(6.98)	(6.32)	(4.11)
Neg	-0.0796***	–0.0551***	–0.0923***	–0.0235***
	(–5.37)	(–6.75)	(–6.65)	(–3.39)
$FFCAR_{0,0}$	0.0384**	0.0219**	0.0268	0.0129
	(2.27)	(2.32)	(1.58)	(1.42)
$FFCAR_{-1,-1}$	–0.0156	–0.0132	–0.0302*	–0.0174*
	(–0.94)	(–1.49)	(–1.72)	(–1.85)
$FFCAR_{-2,-2}$	–0.0131	–0.0064	–0.0218	–0.0242***
	(–0.80)	(–0.73)	(–1.26)	(–2.81)
$FFCAR_{-30,-3}$	–0.0004	–0.0004	–0.0008	–0.0102***
	(–0.43)	(–0.76)	(–0.76)	(–8.62)
$FFAlpha_{-252,-31}$	–0.0002	–0.0004	–0.0091***	–0.0019***
	(–0.10)	(–0.30)	(–4.17)	(–3.19)
SIZE	–0.0419	–0.0189	–0.0167	–0.0224**
	(–1.32)	(–1.24)	(–0.60)	(–2.14)
BTM	–0.0596***	–0.0458***	–0.0227	–0.0194**
	(–2.63)	(–3.24)	(–1.26)	(–2.06)
Turnover	0.0426	0.0250	0.0077	0.0164*
	(1.58)	(1.64)	(0.32)	(1.82)
Observations	19711	60537	19711	60537
Clusters (Days)	5402	5925	5402	5925
Adjusted R^2	0.0054	0.0027	0.0083	0.0059

Note: This table reports the relationship between stock returns and the tone of firm-specific media content. The dependent variable (log returns: $Return_{+1,+1}$ or abnormal returns: $FFCAR_{+1,+1}$) is the close-to-close stock returns on the day of the news publication. Media articles were downloaded from LexisNexis UK and *Pos* and *Neg* are, respectively, the average (standardized) proportions of *positive* and *negative words* in firm-specific media articles published. We use textual analysis to identify words that are either *positive* or *negative* according to Loughran and McDonald's (2011) financial news word lists. In the regressions we control for lagged *Size* (measured as log of Equity), *BTM* (measured as log of Book-to-Market, *Turnover* (measured as log of Share *Turnover*), and past abnormal returns. *ALL* includes news articles sourced from *the Financial Times*, *the Times*, *the Guardian*, and *Mirror*. *FT* includes news articles sourced from *the Financial Times* only. We follow Froot (1989) in clustering the standard errors by trading days. Robust t-statistics are reported in parentheses below the parameter coefficients. *, **, *** denote significance at the 10, 5, and 1 percentage levels.

return predictability variables: size (measured as *Log*(*Market Equity*)), book-to-market ratio (measured as *Log*(*Book/Market*)) and trading volume (measured as *Log*(*Share Turnover*)), as in Tetlock, Saar-Tsechansky, and Macskassy (2008).

Table 2 reports the next-day predictability results for the composite media content (*ALL*) based on all news stories from *The Financial Times* (*FT*), *the Times*, *the Guardian*, and *Mirror*, as well as separately reporting results for *FT*, which constitutes a major proportion of the composite media content.

We observe that *positive* and *negative* words in news stories significantly predict returns on the day of the news publication. In all cases the signs of the coefficients associated with *Pos* and *Neg* are consistent with our predictions that firm-specific news stories with *positive* words predict higher returns in the next trading period and firm-specific news stories with *negative* words predict lower returns in the following trading period. Strong significance is seen for *Pos* and *Neg* in the case of news publications in *ALL* and *FT* and for both log return and abnormal return regressions. The larger magnitude of *Pos* and *Neg* coefficients for results based on *FT* indicate that news stories published in *FT* have a greater impact on abnormal returns than the other news publication sources. The results are driven by the fact that the news stories published in *FT* focus on large firms that attract greater media attention. In the case of *ALL*, we see that next-period abnormal returns experience an increase of 4.9 basis points after a one standard deviation increase in *positive words* and a decrease of 2.3 basis points after a one standard deviation increase in *negative words*. The magnitude of the coefficient on *Pos* in absolute value is almost double that of *Neg*. A formal test for the equality of *Pos* and *Neg* coefficients ($\beta Pos = -\beta Neg$) provides a Chi-square test statistic of 3.738 (p-value = 0.053). The test results reveal that the impact of *Pos* is economically and statistically (at 5% significance level) greater than the impact of *Neg*. Similar statistical significance for the difference in coefficients is found for the other regressions. The results indicate that media content, both *positive* and *negative*, strongly predicts next-period stock returns, with the impact being stronger for news story publications with *positive words*. Barber and Odean (2008) find that investors are more likely to buy, rather than sell, stocks that are in the news. Hence according to their findings, if a stock is in the news there is an inherent demand pressure for the stock, pushing next-period returns up. This underlying bias towards increased returns for any stock in the news could explain the fact that the *positive* impact of *positive* news media content on stock returns is more pronounced than the *negative* impact of *negative* news

TABLE 3. News about fundamentals, media coverage and firms' stock returns

	FT			*ALL*		
Pos	0.1483***	0.1285***	0.1302***	0.0392***	0.0331***	0.0239*
	(5.48)	(4.76)	(4.48)	(2.94)	(2.61)	(1.71)
Neg	–0.0955***	–0.0717***	–0.0771***	–0.0230***	–0.0204***	–0.0219**
	(–6.68)	(–4.86)	(–4.89)	(–2.92)	(–2.71)	(–2.57)
Fund	0.0001		0.0004	–0.0003		–0.0002
	(0.08)		(0.23)	(–0.42)		(–0.29)
*Pos*Fund*	0.0812		–0.0103	0.1313**		0.1134**
	(0.80)		(–0.09)	(2.57)		(2.19)
*Neg*Fund*	–0.0215		0.0047	–0.0527*		–0.0512*
	(–0.35)		(0.07)	(–1.93)		(–1.86)
MC		0.0015	0.0015		–0.0007	–0.0009*
		(1.13)	(1.06)		(–1.31)	(–1.66)
*Pos*MC*		0.1750*	0.1762*		0.1409***	0.1429***
		(1.77)	(1.68)		(3.52)	(3.36)
*Neg*MC*		–0.1727***	–0.1767***		–0.0191	–0.0067
		(–3.48)	(–3.43)		(–0.99)	(–0.33)
$FFCAR_{0,0}$	0.0231	0.0254	0.0215	0.0104	0.0127	0.0102
	(1.29)	(1.43)	(1.15)	(1.07)	(1.39)	(1.05)
$FFCAR_{-1,-1}$	–0.0298	–0.0293	–0.0288	–0.0177*	–0.0175*	–0.0177*
	(–1.62)	(–1.57)	(–1.48)	(–1.74)	(–1.86)	(–1.75)
$FFCAR_{-2,-2}$	–0.0213	–0.0254	–0.0249	–0.0254***	–0.0243***	–0.0254***
	(–1.18)	(–1.39)	(–1.31)	(–2.75)	(–2.82)	(–2.76)

(Continued)

TABLE 3. (Continued)

	FT			*ALL*		
$FFCAR_{-30,-3}$	−0.0009	−0.0009	−0.0010	−0.0019***	−0.0019***	−0.0019***
	(−0.85)	(−0.77)	(−0.85)	(−3.05)	(−3.19)	(−3.04)
$FFAlpha_{-252,-31}$	−0.0093***	−0.0091***	−0.0093***	−0.0106***	−0.0103***	−0.0107***
	(−3.97)	(−4.03)	(−3.80)	(−8.21)	(−8.68)	(−8.25)
SIZE	−0.0091	−0.0161	−0.0058	−0.0183	−0.0224**	−0.0181
	(−0.30)	(−0.52)	(−0.18)	(−1.58)	(−2.14)	(−1.55)
BTM	−0.0159	−0.0204	−0.0123	−0.0156	−0.0195**	−0.0159
	(−0.83)	(−1.07)	(−0.60)	(−1.53)	(−2.07)	(−1.57)
Turnover	0.0042	0.0083	0.0049	0.0138	0.0159*	0.0131
	(0.16)	(0.33)	(0.17)	(1.34)	(1.75)	(1.26)
Observations	18134	17893	16331	53495	60537	53495
Clusters (Days)	5206	5185	4985	5903	5925	5903
Adjusted R^2	0.0084	0.0099	0.0099	0.0064	0.0061	0.0066

Note: This table reports the relationship between abnormal returns, tone of firm-specific news about fundamentals and media coverage. The dependent variable is firms' close-to-close abnormal returns on the day of the news publication ($FFCAR_{+1,+1}$). Media articles were downloaded from LexisNexis UK and *Pos* and *Neg* are, respectively, the average (standardized) proportions of *positive* and *negative words* in firm-specific media articles published. We use textual analysis to identify words that are either *positive* or *negative* according to Loughran and McDonald's (2011) financial news word lists. *Fund* is a dummy variable that takes on the value 1 when a news story contains the word 'earn' and 0 otherwise. *Pos*Fund* (*Neg*Fund*) is the interaction between *positive* (*negative*) *words* and the *Fund* dummy. *MC* is a dummy that takes on the value 1 if more than 3 articles are published on a given day and 0 otherwise. *Pos*MC* (*Neg*MC*) is the interaction between *positive* (*negative*) words and the *MC* dummy. In the regressions we control for lagged *Size* (measured as log of Equity), *BTM* (measured as log of Book-to-Market, *Turnover* (measured as log of Share *Turnover*), and past abnormal returns. *ALL* includes news articles sourced from *The Financial Times*, *The Times*, *The Guardian*, and *Mirror*. *FT* includes news articles sourced from *The Financial Times* only. We follow Froot (1989) in clustering the standard errors by trading days. Robust *t*-statistics are reported in parentheses below the parameter coefficients. *, **, *** denote significance at the 10, 5, and 1 percentage levels.

content.[7] Further, the *positive* coefficients on $FFCAR_{0,0}$ show evidence of return continuations from the day of the news story to the next-day returns, while *negative* coefficients on abnormal returns on the previous two trading days ($FFCAR_{-1,-1}$ and $FFCAR_{-2,-2}$) show return reversal effects. The patterns observed in our regressions are in line with the predictions in Chan (2003) and analogous to the evidence found in Tetlock, Saar-Tsechansky, and Macskassy (2008).[8]

Next, in table 3 we examine whether news stories focusing on firms' fundamentals have a pronounced impact on firms' returns. In addition, we investigate whether tone and volume of news media content (proxied by high media coverage) jointly impact firms' future returns.

Columns 1 and 4 report the results for the model specification examining the next-period effect of *positive* and negative words in news stories that focus on firms' fundamentals. We predict that the next-period effect on firms' returns should be pronounced for news stories about firm fundamentals. We use the variable *Fund*, a dummy variable that is equal to one for news stories that contain the word stem 'earn', and interact it with tone variables *Pos* and *Neg* (as defined previously) in order to measure directly the impact *positive* and *negative* earnings-related news stories will have on stock returns. The dependent variable in the regressions is the next-period abnormal return $FFCAR_{+1,+1}$ and we augment the regressions with all the control variables as in table 2. We find that the coefficients associated to *Pos* and *Neg* remain strongly significant with the expected signs. This shows that both *positive* and *negative* news, over and above the earnings-specific news stories, have significant return predictability. For the case of earnings-specific *positive* and *negative* news stories, we find the predictability relationship is statistically significant and stronger for news publications in the composite media content, *ALL*.

---

7. To understand whether the effects persist or reverse over the next few days, we test the predictability of abnormal returns on days +2 and +3 and find that *Pos* and *Neg* retain their signs, but no longer have a significant effect. Hence, we observe that markets efficiently incorporate the initial price pressures from the day of the news stories and there is not significant evidence of reversals.

8. Note that the significance of the *FFCAR* variables in the regressions can be driven by the relation between the abnormal returns and the alpha term in the expected return calculations of the event study methodology. For robustness, we ignore the alpha term in the expected return calculations and re-estimate the regressions. We find that, although the *FFCAR* variables that were previously significant are now insignificant, the results for the key variables, *Pos* and *Neg*, are almost identical. Hence we confirm that the *Pos* and *Neg* results are not driven by any spurious correlations generated by the event study methodology.

This is evidenced by the magnitude difference of the coefficients *Pos* and *Pos*Fund* (0.0392 and 0.1313) and *Neg* and *Neg*Fund* (–0.0230 and –0.0527). We do not find a significant relationship for earnings-related news stories published in *FT*. This result may be driven by the fact that news stories in *FT* contain words about fundamentals most of the time anyway, and hence focusing on such a subsample is not associated with a significant impact.

Columns 2 to 5 report the results for the model specification examining whether firm-specific news stories receiving higher levels of media attention amplify investor reaction (Barber and Odean, 2008) and hence impact returns. To assess the impact of media attention on a firm's stock returns, we define the media coverage variable MC, which is a dummy variable that takes the value one if more than three articles covering the firm-specific news stories are published on a given day. Using this variable and interacting it with *positive* and *negative* news media content (*Pos* and *Neg*), we examine whether higher visibility of *positive* and *negative* news events have a greater effect on stock returns. The results indicate that high-attention *positive* news publications in *ALL* and *FT* have a significant effect on the next-period abnormal returns. This evidence is consistent with the attention-grabbing effects noted by Barber and Odean (2008), where highly visible *positive* news drives investors' buying decisions. For the case of high-attention *negative* news, we find strong significance only for news publications in *FT* (with *Neg*MC* significant at 1% level). Since *FT* publications consistently cover key news stories and are widely read to institutional investors and traders, high media coverage of *negative* news publications in *FT* can induce *negative* pressure on prices (short-selling) in the market, generating *negative* next-period abnormal returns. Hence we see that highly visible good news and bad news have a significant impact on the subsequent trading period. Further, when we include *Pos*MC* and *Neg*MC* variables in our regressions, we find that the coefficients associated to *Pos* and *Neg* measures remain strongly significant. The magnitude difference between the coefficients associated to the tone variables (*Pos* and *Neg*) and the volume variables (*Pos*MC* and *Neg*MC*) indicate that the impact of volume is much more pronounced than tone (for both *positive* and *negative* media content). Hence the results show that both tone and volume provide novel information about firms' future returns. When we consider the overall model specification with both *Fund* and *MC* variables, the main conclusions drawn above remain. In summary, the table 3 results indicate that news media content is a strong predictor of future stock returns.

Next, we analyse whether the impact of media content is influenced by firm characteristics. Large firms tend to receive more media attention than small firms and hence, for smaller firms, a lower degree of investor recognition of the stock is compensated by higher returns. Other firms that have high investor recognition include growth firms with low book-to-market ratio (also called 'glamour' firms). We predict that the effect of media content on abnormal returns is stronger for low visibility firms (such as smaller firms and firms with high book-to-market ratio). For our empirical investigation, we classify our sample of *FTSE* 100 firms into terciles created in terms of firm size and book-to-market ratio based on the preceding year.[9]

Table 4 reports the regression results for the predictive relationship between media tone and stock returns for the three groups of firms. Columns 1 to 3 report the regression results for firms classified according to firm size (market capitalization) and Columns 4 to 6 report the regression results for firms classified according to book-to-market. The results indicate that both *positive* and *negative* news have a significant predictive relationship with next-period abnormal returns and in line with our predictions, we see that the results are driven by less visible firms (smaller *FTSE* 100 firms and firms with high book-to-market ratios).[10] When we consider the news stories that focus on fundamentals, we see a larger subsequent period impact for earnings-related news media content in the case of medium market capitalization firms and firms with medium to low book to market ratios. For larger *FTSE* 100 firms, the earnings-related news does not have a significant effect on next-period abnormal returns. This result corroborates the findings of Bernard and Thomas (1990) that large firms, due to high investor recognition, tend to have less post-announcement drift. Further, when we consider the relationship between media coverage (*MC*) and next-period abnormal returns, we see the significant impact of highly visible good news on next-period returns (seen in table 3 for *ALL* stories) is driven by larger *FTSE* 100

9. Note that since our sample consists of the largest 100 UK firms listed on *FTSE*, the firms in the smallest size tercile are still relatively large.

10. These results are for the smaller *FTSE* 100 firms; one might expect even stronger results for the non-*FTSE* 100 stocks.

TABLE 4. Stock returns and news media content for different firm size and book-to-market classifications

	MV (low)	MV (Medium)	MV (High)	*BTM* (Low)	*BTM* (Medium)	*BTM* (High)
Pos	0.0535**	0.0139	0.0018	–0.0074	0.0310	0.0512*
	(2.09)	(0.59)	(0.08)	(–0.34)	(1.30)	(1.90)
Neg	–0.0431**	–0.0127	–0.0107	–0.0089	–0.0187	–0.0366**
	(–2.53)	(–0.90)	(–0.94)	(–0.71)	(–1.56)	(–2.01)
Fund	–0.0005	0.0001	–0.0002	–0.0011	0.0016	–0.0008
	(–0.27)	(0.07)	(–0.15)	(–0.94)	(1.25)	(–0.50)
*Pos*Fund*	0.1834	0.1141	0.0386	0.1414*	0.0487	0.1207
	(1.60)	(1.29)	(0.55)	(1.84)	(0.62)	(0.99)
*Neg*Fund*	–0.0614	–0.0882*	–0.0215	–0.0098	–0.0813*	–0.0742
	(–1.05)	(–1.77)	(–0.60)	(–0.23)	(–1.79)	(–1.43)
MC	–0.0013	0.0004	–0.0021**	–0.0010	–0.0013	–0.0003
	(–1.10)	(0.38)	(–2.22)	(–0.91)	(–1.55)	(–0.30)
*Pos*MC*	0.0414	0.1256*	0.2853***	0.1389*	0.0750	0.1774**
	(0.51)	(1.75)	(4.10)	(1.82)	(1.16)	(2.29)
*Neg*MC*	0.0427	–0.0426	–0.0162	0.0123	0.0294	–0.0338
	(1.02)	(–1.17)	(–0.52)	(0.29)	(1.00)	(–0.93)
$FFCAR_{0,0}$	0.0016	0.0209	0.0096	0.0260**	–0.0122	0.0107
	(0.10)	(1.36)	(0.73)	(2.22)	(–0.79)	(0.59)
$FFCAR_{-1,-1}$	–0.0083	–0.0065	–0.0393***	–0.0278**	–0.0322***	–0.0029
	(–0.54)	(–0.41)	(–2.79)	(–2.40)	(–2.78)	(–0.15)

(*Continued*)

TABLE 4. (Continued)

	MV (low)	MV (Medium)	MV (High)	*BTM* (Low)	*BTM* (Medium)	*BTM* (High)
$FFCAR_{-2,-2}$	0.0267*	–0.0129	–0.0366***	–0.0218*	–0.0488***	–0.0155
	(–1.80)	(–0.80)	(–2.88)	(–1.88)	(–3.94)	(–0.97)
$FFCAR_{-30,-3}$	–0.0016	–0.0021**	–0.0023***	–0.0015**	–0.0019**	–0.0026**
	(–1.53)	(–2.17)	(–2.76)	(–2.21)	(–2.37)	(–2.02)
$FFAlpha_{-252,-31}$	–0.0113***	–0.0082***	–0.0124***	–0.0105***	–0.0119***	–0.0117***
	(–5.05)	(–3.95)	(–6.11)	(–5.51)	(–5.90)	(–5.43)
SIZE	–0.0361	0.0387	–0.0304	–0.0268	–0.0715***	0.0148
	(–1.11)	(0.75)	(–1.47)	(–1.54)	(–3.80)	(0.61)
BTM	–0.0467**	0.0093	–0.0259	0.0289*	–0.0797	–0.0707
	(–2.18)	(0.51)	(–1.61)	(1.82)	(–0.94)	(–1.15)
Turnover	0.0206	–0.0149	0.0323**	0.0281*	0.0578***	–0.0143
	(1.18)	(–0.73)	(2.13)	(1.72)	(3.39)	(–0.74)
Observations	17034	17240	19221	19202	16637	17656
Clusters (Days)	5589	5290	4183	4932	5341	5480
Adjusted R^2	0.0064	0.0048	0.0124	0.0072	0.0099	0.0079

(*Continued*)

TABLE 4. (Continued)

Note: This table reports the relationship between abnormal returns and the tone of media content for firms classified according to market capitalization (*MV*) and book-to-market (*BTM*) based on the preceding year. Firms are classified into terciles and results are reported for low, medium and high classifications. The dependent variable is firms' close-to-close abnormal returns on the day of the news publication ($FFCAR_{+1,+1}$). Media articles were downloaded from LexisNexis UK and *Pos* and *Neg* are, respectively, the average (standardized) proportions of *positive* and *negative words* in firm-specific media articles published. We use textual analysis to identify words that are either *positive* or *negative* according to Loughran and McDonald's (2011) financial news word lists. *Fund* is a dummy variable that takes on the value 1 when a news story contains the word 'earn' and 0 otherwise. *Pos**Fund* (*Neg**Fund*) is the interaction between *positive* (*negative*) *words* and the *Fund* dummy. *MC* is a dummy that takes on the value 1 if more than 3 articles are published on a given day and 0 otherwise. *Pos**MC* (*Neg**MC*) is the interaction between *positive* (*negative*) *words* and the *MC* dummy. In the regressions we control for lagged *Size* (measured as log of Equity), *BTM* (measured as log of Book-to-Market, *Turnover* (measured as log of Share *Turnover*), and past abnormal returns. News articles are sourced from *The Financial Times, The Times*, *The Guardian*, and *Mirror*. We follow Froot (1989) in clustering the standard errors by trading days. Robust *t*-statistics are reported in parentheses below the parameter coefficients. *, **, *** denote significance at the 10, 5, and 1 percentage levels.

firms. The results are consistent with the attention-grabbing effects documented by Barber and Odean (2008). Overall, the results in table 4 indicate that the predictive nature of *positive* and *negative words* in news stories is less pronounced for more visible firms with higher investor recognition.

IV. Can News-Based Trading Strategies Provide Economic Gains?

In this section, we explore the economic significance of the relation between news media content and returns by constructing a trading strategy using firm-specific *positive* and *negative* measures of news media content that determine the buy and sell signals. Our simple news-based trading strategy takes a long position in an equal-weighted portfolio made up of firms that have their news stories reported with average net *positive* tone and simultaneously holds a short position in an equal-weighted portfolio of firms that have their news stories reported with average net *negative* tone. The tone in a news article is net *positive* (*negative*) when the difference between the number of *positive* and *negative words* deflated by the total number of words is above (below) zero. We hold our position throughout the day and rebalance every trading day based on the news media content published before the market opens on that day.

We calculate the risk-adjusted daily returns of this news-based trading strategy, broken down over eight-year time periods from 1987 to 2010. The period 1981–1986 was excluded from the trading strategy since there were too many days with no firm-specific media articles and hence trading signals could not be determined. We use the Carhart's (1997) four-factor model to adjust the trading strategy returns for contemporaneous market, size, book-to-market and momentum factors.[11]

Table 5 reports the estimates of daily risk adjusted returns (*alpha*) and the factor loadings from the news-based trading strategy. We report results based on the composite media content (*ALL*) in Columns 1 to 4,while Columns 5 to 8 report results based on media content

11. Using the Fama and French (1993) three-factor model provides similar results and hence we do not report them here.

TABLE 5. Risk-adjusted news-based trading strategy results

	FT				*ALL*			
	1987 –1994	1995 –2002	2003 –2010	1987 –2010	1987 –1994	1995 –2002	2003 –2010	1987 –2010
Alpha	–0.0210	0.1117	0.1903***	0.1164**	0.0304	0.0150	0.1419***	0.0676***
	(–0.23)	(1.25)	(2.61)	(2.38)	(0.95)	(0.38)	(3.85)	(3.21)
Market	–0.0544	–0.1177	–0.0449	–0.0591	–0.0041	0.0717	–0.0781	–0.0339
	(–0.35)	(–0.43)	(–0.42)	(–0.70)	(–0.07)	(0.65)	(–1.15)	(–0.83)
SMB	–0.0645	0.0198	–0.0153	–0.0056	–0.0216	0.0332	–0.0440	–0.0261
	(–0.63)	(0.14)	(–0.22)	(–0.11)	(–0.58)	(0.53)	(–0.87)	(–0.97)
HML	–0.0088	0.0652	0.0386	0.0533	0.0269	0.0293	–0.0386	–0.0011
	(–0.06)	(0.68)	(0.57)	(1.08)	(0.51)	(0.72)	(–0.90)	(–0.04)
UMD	–0.3332	–0.0763	–0.0727	–0.0819	0.1687*	0.1060*	–0.0013	0.0678*
	(–1.15)	(–0.49)	(–0.77)	(–1.05)	(1.91)	(1.69)	(–0.02)	(1.69)
Trading Days	467	946	1067	2485	1514	1843	1842	5229
Adjusted R^2	–0.0019	0.0021	–0.0009	0.0015	0.0025	0.0014	0.0010	0.0009

Note: This table shows the daily abnormal returns *Alpha* (Jensen's) from the news-based trading strategy. The regressions use the Fama and French (1993) three-factor model and the Carhart (1997) momentum factor to adjust the trading strategy returns for the impact of contemporaneous market (Market), size (*SMB*), book-to-market (*HML*), and momentum (*UMD*). The *Alpha* and four factor loadings from the time-series regression of the long-short news-based portfolio return have been reported. The strategy forms two equal-weighted portfolios based on the proportion of *positive* and *negative words* used in each news article for each firm during the previous trading day. The strategy takes a long position in a portfolio of firms that have their news stories reported with net *positive* tone (where the difference between the number of *positive* and *negative words* deflated by the total number of words on a particular news story is above 0), and shorts the portfolio of firms with net *negative* tone (where the difference is below 0). The strategy holds both the long and short portfolios for one full trading day and rebalances at the end of the trading day. *ALL* includes news articles sourced from *The Financial Times*, *The Times*, *The Guardian*, and *Mirror*. *FT* includes news articles sourced from The *Financial Times* only. Robust t-statistics in parentheses are based on White (1980) heteroskedastic-consistent standard errors. *, **, *** denote significance at the 10, 5, and 1 percentage levels.

exclusively from *FT*, which constitutes a major proportion of the composite media content. Ignoring transaction costs, we observe that the news-based trading strategy produces a statistically significant alpha of 19 basis points per day for *FT*-based news stories and 14.2 basis points per day for *ALL*-based news stories in the recent period 2003 to 2010.[12] We also find a significant alpha for the whole period 1987 to 2010, which is driven by the results found after 2003. The significant excess returns from the trading strategy in the recent period may be due to improved signalling, resulting from an increase in the volume of news articles published in recent times. Our news-based trading strategy results are similar to, but weaker than, those of Tetlock, Saar-Tsechansky, and Macskassy (2008), who constructed a strategy returning a significant *positive* alpha in every time period from 1980 to 2004. However, unlike our study, which uses daily newspaper publications, Tetlock, Saar-Tsechansky, and Macskassy (2008) use intraday news from the Dow Jones News Service to determine their long and short positions.[13] Further, we see that the Carhart (1997) four-factor loadings are mostly insignificant (except for momentum, which is *positive* and significant for *ALL*). Since we employ a firm-level news-based trading strategy, the results do not load heavily on the market variables. Our results are analogous to the US evidence of Tetlock, Saar-Tsechansky, and Macskassy (2008).

V. Market-Level Return Predictability of Aggregate News Media Content

In this section, we investigate whether the relationship between media content and stock returns is evident at the aggregate market level. We construct the aggregate measures of news media content *AggPos* and *AggNeg* as the average of all firm-specific measures of *positive* (*Pos*) and *negative* (*Neg*) news media content per day. These measures

12. For the period 2006-2010 with the financial crisis, the *ALL*-based trading strategy produces an alpha of 23.5 bps while the *FT*-based trading strategy produces an alpha of 20.9 bps (with both being significant).

13. An important caveat to note is that the trading strategy generates close-to-close returns and hence we assume that one can trade at the closing prices. When we consider the open-to-close returns on the day of the news publication, our results, although weaker, generate an *alpha* of 7 bps (and significant at 10% level) for the 2003–2010 period and 4 bps (and significant at 5% level) for the 2006–2010 period, for the case of *ALL* news stories.

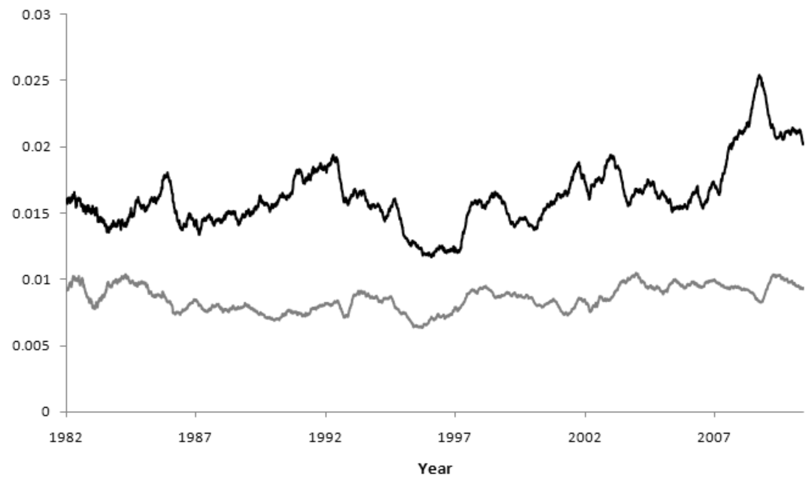

FIGURE 1.— Aggregate measures of the proportions of positive and negative words in media articles: 1981–2010.

Note: This figure shows the rolling 100-day averages of the aggregate measures of news media content *AggPos* and *AggNeg*, which are constructed as the average of all firm-specific measures of *positive* (*Pos*) and *negative* (*Neg*) news media content per day. The black line represents the *AggNeg* measure and the grey line represents the *AggPos* measure. Media articles were downloaded from LexisNexis UK and media tone is determined by using textual analysis to identify words that are either *positive* or *negative* according to Loughran and McDonald's (2011) financial news word lists.

capture the overall *positive* and *negative* media information production by newspapers in the UK on a daily basis. Figure 1 shows the rolling 100-day averages of *AggPos* and *AggNeg*.

We see that the *negative* news media content has significantly more variation than *positive* news media content. Moreover, we observe that the movements in the *AggNeg* measure accurately correspond to the market-level economic shocks experienced during the sample period. For example, the first pronounced peak in *AggNeg* occurs early in 1986 when the stock market experienced high uncertainty periods. The aggregate *negative* media content then decreases for the rest of 1986 and reaches a minimum around the time of the 'Big Bang', so termed for the sudden deregulation of British financial markets in October 1986. The next significant peak in *negative* news media content occurs in autumn 1992. This corresponds to the withdrawal of the UK from the European exchange rate mechanism. The UK economy then turned around in early 1993 and produced a strong recovery, which also corresponds to the gradual fall in *negative* news media content to its lowest point in the sample period, in early 1997. The next notable spikes in aggregate *negative* news media content appear in 2002 and 2003, as the UK economy faltered and global stock markets began to tumble, while an impending war with Iraq weighed on the UK stock market. This then brings us to the financial crisis that began in 2007. The level of *negative* news media content rose sharply throughout 2008, especially after the bankruptcy of Lehman Brothers in the US, reaching a hiatus in February and March 2009, when concerns about the strength of the UK's financial institutions were at their gravest. The steep rise and eventual high point in aggregate *negative* news media content was made more pronounced due to the unprecedented level of media coverage during the global financial crisis.

Using these *AggPos* and *AggNeg* measures, we test the stock return predictability of media content at the aggregate level. In our regressions, we use the close-to-close returns on the *FTSE* 100 on the day of the news publications as the dependent variable and consider the lags of media content measures *AggPos* and *AggNeg* up to five trading days prior to the day of the news story. The regressions also include an intercept term and the following control variables: lagged returns up to five trading days to control for past returns, past volatility proxied by five lags of detrended squared *FTSE* 100 residuals,[14] lagged volume[15] up to five trading days to capture liquidity effects, day-of-the-week dummies, and a dummy variable capturing the January effect. All regression results report White (1980) heteroskedastic-consistent standard errors.

Table 6 reports the regression results for news stories published by *FT* in Column 1, while the regression results for the case of the composite news media content *ALL* are reported in Column 2. For the case of *AggPos*, we find that *positive* news stories on an aggregate level

---

14. Similar to Tetlock (2007), we square the demeaned *FTSE* returns and then subtract the past 30-day moving average of the squared returns to obtain the proxy for volatility. Using the past 60-day moving average provides similar results.

15. We use the detrended log of turnover as a measure of volume. We use the methodology of Campbell, Grossman and Wang (1993) to detrend the log turnover series using the past 30-day moving average. Using the past 60-day moving averages gives similar results.

TABLE 6. Market-level returns and aggregate news media content

	FT	*ALL*
$AggPos_0$	0.0434***	0.0347**
	(3.0065)	(2.3104)
$AggPos_{-1}$	0.0056	0.0217
	(0.3865)	(1.4773)
$AggPos_{-2}$	–0.0275**	–0.0248*
	(–1.9073)	(–1.7941)
$AggPos_{-3}$	0.0103	–0.0057
	(0.7103)	(–0.4051)
$AggPos_{-4}$	0.0027	0.0137
	(0.1854)	(0.9415)
$AggPos_{-5}$	0.0087	–0.0162
	(0.6111)	(–1.1350)
$AggNeg_0$	–0.0824***	–0.0575***
	(–5.5399)	(–3.9694)
$AggNeg_{-1}$	0.0215	0.0124
	(1.4389)	(0.8389)
$AggNeg_{-2}$	–0.0076	–0.0090
	(–0.5187)	(–0.6281)
$AggNeg_{-3}$	–0.0088	–0.0023
	(–0.6046)	(–0.1575)
$AggNeg_{-4}$	0.0049	0.0109
	(0.3367)	(0.7460)
$AggNeg_{-5}$	0.0416***	0.0310**
	(2.8264)	(2.1489)
Test H_0: $Sum(AggPos_{-1} : AggPos_{-5}) = 0$	–0.0002	–0.0113
Chi-square(1) *test statistic*	[0.0007]	[0.2137]
Test H_0: $Sum(AggNeg_{-1} : AggNeg_{-5}) = 0$	0.0514	0.0430
Chi-square(1) *test statistic*	[2.5703]	[3.0194]

Note: This table presents the ordinary least squares (OLS) estimates for the coefficients *AggPos* and *AggNeg* in the regression equation for log returns (dependent variable). The regressions include *AggPos*t and $AggNeg_t$ (for t = 0, –1,…, –5), an intercept term and the following control variables: lagged returns up to five trading days to control for past returns, past volatility proxied by five lags of detrended squared *FTSE* 100 index return residuals, lagged volume (detrended log of *Turnover*) up to five trading days to capture liquidity effects, day-of-the-week dummies and a dummy variable capturing the January effect. The variable *AggPost* ($AggNeg_t$) is the standardized aggregate measure of *positive* (*negative*) *words* in firm-specific media articles on day *t*, constructed by taking the average of the *Pos* (*Neg*) measure across all firms with news articles published about them on each day. These measures are determined by using textual analysis to identify words that are either *positive* or *negative* according to Loughran and McDonald's (2011) financial news word lists. Media articles were downloaded from LexisNexis UK. *ALL* includes news articles sourced from *The Financial Times*, *The Times*, *The Guardian*, and *Mirro*r. *FT* includes news articles sourced from *The Financial Times* only. *t*-statistics are based on White (1980) heteroskedastic-consistent standard errors. *, **, *** denote significance at the 10, 5, and 1 percentage levels.

have a strong *positive* effect on stock returns, with the impact more significant for the case of *FT*. We see that a one standard deviation change in the *AggPos* measure of *FT* (*ALL*) news stories increases returns by 4.3 (3.4) basis points. Some of this initial *positive* impact on stock returns shows a reversal effect later in the trading week, with *negative* significance (at a maximum level of five percent) seen at lag two. The results are consistent for the *FT* as well as the composite news stories, *ALL*.

A similar pattern is observed in the aggregate *negative* news media content. For the case of *AggNeg*, we see that *negative* news stories on an aggregate level exert significant downward price pressure on next-period returns. For the case of *FT*, we see a decrease of 8.2 basis points in returns and for the case of *ALL*, we observe a decrease of 5.8 basis points in returns. These results find support from the literature, as in Grossman and Stiglitz (1980), who find that the underreaction to *negative* news provides motivation for market participants to monitor financial news releases. The evidence of this underreaction to *negative* news also has a behavioural explanation (see Shefrin and Statman, 1985; Barberis, Shleifer, and Vishny, 1998; and Frazzini 2006). As with the case of *positive* news, we see significant return reversals in the subsequent trading days, with all of the effects coming from a significant reversal in lag 5. We test whether the shock to returns caused by media content is permanent or temporary by conducting a formal Chi-square test on the lag coefficients associated to *AggPos* and *AggNeg*. The test results show that there is some reversal to the initial *negative* price pressure in the subsequent trading week, although the statistical significance is only marginal. Our results are analogous to Tetlock (2007) and Garcia (2013), who find evidence of initial declines and subsequent partial return reversals for pessimistic media information from newspaper columns. For the aggregate positive news media content, the test results show no significant reversal of the initial upward price pressure in the subsequent trading week. The results suggest that linguistic media content in news articles provides important information that significantly influences stock returns.

VI. Conclusion

Using a large panel of UK firm-specific news media data over the period 1981 to 2010, this paper provides international evidence for the predictive power of news media content for future stock returns. Unlike previous research, this paper studies the combined impact of tone and volume of firm-specific news stories on a firm's stock returns. We construct *positive* and *negative* measures of news media content based on *positive* and *negative* financial words contained in leading UK newspaper publications – *The Financial Times*, *the Times*, *the Guardian*, and *Mirror*. Our main findings show that both tone (*positive* as well as *negative*) and volume of news media content provide investors with valuable information that impacts future stock returns, with the impact of volume more pronounced than that of tone.

Specifically, we find that *positive words* in firm-specific news media content forecast higher returns next trading period, while *negative words* in firm-specific news media content forecast lower returns next trading period. In addition, we show that *positive* and *negative* news stories related to firms' fundamentals are strong predictors of returns. Further, we observe that the predictive relationship between media content and firms' returns is significant for lower visibility *FTSE* 100 firms with lower market capitalizations and higher book-to-market ratios. Furthermore, we find that high-attention news (both *positive* and *negative*) affects subsequent trading period returns. The results show that investors tend to react to highly visible news, whether *positive* or *negative*, indicating that both visibility and tone are key factors in determining how investors respond to news. Implementing a simple news-based trading strategy, we demonstrate the economic significance of *positive* and *negative* media content and stock returns. We observe that the news-based trading strategy produces statistically significant risk-adjusted returns of 14.2 to 19 basis points per day in the recent period 2003 to 2010. At the aggregate market-level, we also find significant interactions between the *positive* and *negative* aggregate media content and stock returns. The initial price pressures caused by *positive* and *negative words* in news stories do not show strong significant reversals in the subsequent trading. The overall findings of this paper shed light on the importance of *positive* and *negative* semantic information in news media publications in predicting asset returns and demonstrate that both tone and volume of firm-specific news media content embody otherwise hard-to-quantify information about asset prices.

Accepted by: *P.C. Andreou, PhD, Editor-in-Chief* (*Pro-Tem*), *November* 2014

References

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

Barberis, N.; Shleifer, A.; and Vishny, R. 1998. A model of investor sentiment. *Journal of Financial Economics* 49:307-43.

Bernard, V. L., and Thomas, J. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13:305-40.

Campbell, J. Y.; Grossman, S. J.; and Wang, J. 1993. Trading volume and serial correlation in stock returns. The Quarterly *Journal of Economics* 107:905–939.

Carhart, M. 1997. On persistence in mutual fund performance. *Journal of Finance* 52:57–82.

Chan, W. S. 2003. Stock price reaction to news and no-news: Drift and reversal after headlines. *Journal of Financial Economics* 70:223–260.

Cutler, D. M.; Poterba, J. M.; and Summers, L. H. 1989. What moves stock prices? *Journal of Portfolio Management* 15:4-12.

Dougal, C.; Engelberg, J.; Garcia, D. and Parons, C. A. 2012. Journalists and the stock market. *Review of Financial Studies* 25:639-79.

Engleberg, J., and Parsons, C. A. 2011. The causal impact of media in financial markets. *Journal of Finance* 66:67-97.

Fama, E. F., and French, K. R. 1993. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* 33:3-56.

Fang, L., and Peress, J. 2009. Media coverage and the cross-section of stock returns. *Journal of Finance* 64:2023-2052.

Frazzini, A. 2006. The disposition effect and underreaction to news. *Journal of Finance* 61:2017-2046.

Garcia, D. 2013. Sentiment during recessions. *Journal of Finance* 68:1267-1300.

Griffin, J. M.; Hirschey, N. H.; and Kelly, P. J. 2011. How important is the financial media in global markets? *Review of Financial Studies* 24:3941-3492.

Grossman, S. J., and Stiglitz, J. E. 1980. On the impossibility of informationally efficient markets. *American Economic Review* 70:393-408.

Gurun, U. G., and Butler, A. W. 2012. Don't believe the hype: Local media slant, local advertising, and firm value. *Journal of Finance* 67:561-598.

Jegadeesh, N., and Wu, D. 2013. Word power: A new approach for content analysis. *Journal of Financial Economics* 110:712-729.

Kelley, E. K., and Tetlock, P. C. 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68:1229-1265.

Liu, S. 2014. The impact of textual sentiment on sovereign bond yield spreads: Evidence from the Eurozone crisis. *Multinational Finance Journal* 18:215-248.

Loughran, T., and McDonald, B. 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66:35-65.

Peress, J. 2014. The media and the diffusion of information in financial markets: Evidence from newspaper strikes. *Journal of Finance* 69:2007-2043.

Shefrin, H., and Statman, M. 1985. The disposition to sell winners too early and ride losers too long: theory and evidence. *Journal of Finance* 40:777-790.

Shiller, R. 2005. Irrational Exuberance, Princeton NJ: Princeton University Press, 2nd Edition.

Solomon, D. H. 2012. Selective publicity and stock prices. *Journal of Finance* 67:599-638.

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

Tetlock, P. C.; Saar-Tsechansky, M.; and Macskassy, S. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 63:1437-1467.