Sentiment during recessions*

Diego García[†]
Kenan-Flagler Business School
University of North Carolina at Chapel Hill

May 23, 2011

ABSTRACT

This paper studies the effect of sentiment on asset prices during the 20th century (1905–2005). As a proxy for sentiment, we use the fraction of positive and negative words in two columns of financial news from the New York Times. The main contribution of the paper is to show that, controlling for other well-known time-series patterns, the predictability of stock returns using news' content is concentrated in recessions. A one standard deviation shock to our news measure during recessions changes the conditional average return on the DJIA by twelve basis points over one day.

JEL classification: G01, G14.
Keywords: media content, stock returns, sentiment, recessions.

---

*I would like to thank Malcolm Baker, Craig Carroll, Riccardo Colacito, Jennifer Conrad, Alex Edmans, Joey Engelberg, Francesca Gino, Mark Grinblatt, Michael Halling, Pab Jotikasthira, Øyvind Norli, Chris Parsons, Chris Roush, Francesco Sangiorgi and Paul Tetlock for comments on an early draft, as well as seminar participants at UNC at Chapel-Hill and at the European Winter Finance Summit (2010). Special thanks to the CDLA staff, Kirill Fesenko, Fred Stipe, and particularly Rita Van Duinen for their help in the optical character recognition stage of this project.

[†]Diego García, Kenan-Flagler Business School, University of North Carolina at Chapel Hill, McColl Building, C.B. 3490, Chapel Hill, NC, 27599-3490. Tel: 1-919-962-8404; Fax: 1-919-962-2068; Email: `diego_garcia@unc.edu`; Webpage: `http://www.unc.edu/~garciadi`

1 Introduction

Shiller (2000) argues that the news media plays an important role in setting the stage for market moves and provoking them. His conjecture is that investors follow the printed word even though much of it is pure hype, suggesting that market sentiment is driven by news' content. More formally, Tetlock (2007) shows that the number of negative words in the "Abreast of the Market" column of the Wall Street Journal predicts stock returns at the daily frequency from 1984 to 1999.

This paper revisits both Shiller's conjecture and Tetlock's evidence by studying financial market news from the New York Times from 1905 to 2005. The literature from psychology and economics, which we review below, suggests that investors' sensitivity to news will be most pronounced when they are going through hard times. Our main result shows that the link between media content and Dow Jones Industrial Average (DJIA) returns is indeed concentrated in times of hardship, proxied empirically by NBER recessions. A one standard deviation change in our sentiment measure moves the daily average of the DJIA by 12 basis points during recessions, while the effect in expansions is only 3.5 basis points. The effect we uncover is robust to different normalizations of the media factors, volatility adjustments, outliers analysis, and controlling for autocorrelation and other known determinants of stock returns.

We construct our proxy for market sentiment by counting the number of positive and negative words from two financial columns from the New York Times.[1] Both were published daily and covered general financial news – from stock market performance to industry news and macroeconomic events. Thus, they are natural candidates as gauges of the excitement and agitation in US stock markets during the 20th century.[2]

For the majority of our sample the supply of news was much more concentrated than today. The two main media sources with regular coverage of business news were the Wall Street Journal and the New York Times. Thus, most investors would have read the columns that we study by the time the market opened. By studying a longer time series than Tetlock (2007) not only do we have more statistical power, but we can take advantage of the variation in the business cycle and the content of a media outlet that was in most investors' hands every morning.

It is plausible that part of the effect we document is related to the arrival of new information.

[1] The columns were titled "Financial Markets" and "Topics in Wall Street" for roughly half of our sample period. The "Financial Markets" column was essentially a recount of major stock market moves the previous day. The "Topics in Wall Street" column (which used other names, such as "News, Comment and Incident in the Stock Exchange," "Financial and Business Sidelights of the Day" and eventually become "Market Place") had a broader coverage of economic and financial news.

[2] It is important to emphasize that the columns we study are the type of news that Shiller (2000) and Tetlock (2007) have in mind. Annual statements (Tetlock, Saar-Tsechansky, and Macskassy, 2008; Kogan, Routledge, Sagi, and Smith, 2009; Loughran and McDonald, 2010) or earnings announcements (Engelberg, 2008) are likely to contain new information. On the other hand, a daily description of the stock market is almost an opinion piece, speculating about why the market acted as it did over the recent past, and about what it may do in the near future.

We find that the effect of news partially reverses over the following few trading days, which argues for a non-informational impact. But one cannot underestimate the role of a newspaper, such as the New York Times, as an important channel of financial news during most of our sample period. While the time-series data we use does not allow us to directly disentangle the information versus sentiment hypothesis, we conduct indirect tests that seem to give more bite to the sentiment interpretation of our data.[3]

We first study whether financial reporting changes along the business cycle. The news in our sample have a "tag-along" flavor, much as Shiller (2000) describes. They essentially report on the previous days events, giving ex-post explanations about past asset price movements. In our dataset, this comes to light in a strong correlation between media content on a given afternoon and stock returns on that day. Financial reporting does not appear to be related to the business cycle, so how journalists describe previous market movements does not drive our results.

Next, we sort stories based on the quantity of numbers that the columns include. The logic is that articles with hard information, such as earnings, dividends, sales, would include more numbers, and that the tone of the article, measure by signed words, would be more important on days when such data was released. We find that the effect we uncover is not related to the amount of hard data contained in the articles, which suggests qualitative information does not interact with hard facts.

We also document that the effect is stronger during weekends. News written on Saturdays and Sundays have a significant impact on Monday's stock returns. For a large part of our sample virtually all businesses, including the New York Times, were closed on Sundays.[4] It is not clear what type of information journalists could have gathered over the weekend that would significantly move stock prices on Monday. On the other hand, circulation of newspapers is much higher on Sundays. Following DellaVigna and Pollet (2009), we interpret this evidence as supporting the hypothesis that traders tune out of their investments at the end of the week, catching up with the markets on Mondays.

Finally, we look at intraday data on the DJIA. If the New York Times columns contain information that did not make it into the previous day closing prices, we should not be surprised about the predictability using close-to-close prices. We find that our media measures can predict stock returns after the NYSE opening (11am–close), which rules out alternative hypotheses that predict new information is quickly impounded into prices.

Our analysis is predicated on the assumption that economic recessions correspond with times of heightened sensitivity to news. The psychology literature has forcefully argued that

[3]In order to be able to separate the two one must have some instrument that affects sentiment and not fundamentals. See Engelberg and Parsons (2011) and Dougal, Engelberg, García, and Parsons (2011) for casual evidence on the link between media and asset pricing variables.

[4]Neilson (1973) discusses at length the state of business journalism in the first half of the 20th century. It was not until half-way through our sample period that the Monday edition of the Wall Street Journal was written on Sundays.

emotions affect decision making, and information processing in particular. For example, Tiedens and Linton (2001) show that emotions elicit different reliance on heuristic versus systematic processing. The literature has also found that anxiety, hope and sadness are associated with a greater sense of uncertainty (Smith and Ellsworth, 1985; Ortony, Clore, and Collins, 1988). Gino, Wood, and Schweitzer (2009) show that anxiety makes agents more receptive to advice, even if this advice is bad.[5] This literature shows that priming subjects into negative mood states changes their decision making abilities. One can reasonably argue that in periods of expansion investors feel happy and optimistic, whereas during recessions they feel fearful and anxious. The job losses and the uncertainty over the future that investors experience during recessions put the population at large in negative mood states.[6] This evidence suggests that investors will use different decision making rules in recessions than in expansions, being particularly sensitive to news in economic downturns.

Although the above experimental studies from psychology establish that agents' moods and emotions affect their individual behavior, it is the behavioral economics literature that has shown how such sentiment can move aggregate quantities. For example, Hirshleifer and Shumway (2003) show how stock returns are affected by the weather across the world, and Edmans, García, and Norli (2007) associate the outcomes of sporting events, such as the World Cup, to drops in the stock markets when the country loses a game (see Hirshleifer, 2001, for a survey on these topics). Akerlof and Shiller (2009), while discussing confidence and the Michigan Consumer Sentiment Index, state that "we conceive of the link between changes in confidence and changes in income as being especially large and critical when economies are going into a downturn, but not so important at other times." Shiller (2000) highlights the potential role of the media in creating asset bubbles and triggering market crashes. Both the existing empirical findings in finance and the experimental evidence from psychology thus suggest that human behavior is significantly different in times of anxiety and fear versus periods of prosperity and tranquility, which motivates our empirical analysis.

The paper contributes to the growing literature on the role and content of the media and its effects on asset prices and investor behavior. Given data limitations, most of the literature has studied the last twenty–five years, where news are available electronically in text format. In the journalism literature, Bow (1980) argues that there were no predictive signs in the media prior to the 1929 stock market crash, while Griggs (1963) gives a similar account in the context of the 1957–1958 recession.[7] The closest paper to ours is Tetlock (2007), who studies a column from

[5] See Forgas (1991) for an excellent survey of the earlier literature, and Bless, Clore, Schwarz, Golisano, Rabe, and Wölk (1996), Forgas (1998), Park and Banaji (2000), and Lerner and Keltner (2000, 2001) for more recent work.

[6] It is worthwhile noticing that our classification of recessions, following the NBER, was backdated for virtually our entire sample. While investors may not have gotten a formal NBER press release, clearly our classification seems the most reasonable one: the NBER chose such dates because they were times of economic hardship.

[7] Norris and Bockelmann (2000) and Roush (2006) both have extensive discussions as to the role of the media

the Wall Street Journal from 1984 to 1999. We show how the predictability he uncovered is much stronger during economic downturns. We also find that positive words help predict stock returns, whereas in his research only negative word counts have predictive power. Further, we show that the effect is particularly important over the weekend, when investors have a chance to read over the news. Since Tetlock (2007), the literature has examined the cross-section of stock returns, other types of investor behavior, or news originating from sources other than media outlets.[8] Our large sample provides an excellent laboratory where to examine the predictability of a liquid market index using media content.

The paper is also related to the literature on investor sentiment (see Baker and Wurgler, 2007, for a survey). Our study is the first to show how sentiment can play a more important role during recessionary periods. Most of the sentiment indexes that have been suggested by other authors have data requirements that restrict their implementation to the last forty years.[9] As a consequence, they are less likely to be able to take advantage of the frequency and severity of the business cycle that the US experienced during the first half of the 20th century. Finally, another important advantage of our media based measures is that they are available daily, and as such they can be used in high-frequency studies.

The rest of the paper is structured as follows. Section 2 constructs the sentiment measures we use in our study. Section 3 formally studies the relationship between daily returns and sentiment for the DJIA, and goes over other robustness checks. Section 4 looks at the effect of the news on different days of the week, as well as a function of the content of the news, and studies intraday data. Section 5 concludes.

2 The data

The paper uses several sources of data. The first is stock return information. We collect the total return index for the Dow Jones Industrial Average from Williamson (2008).[10] The Dow Jones Industrial index goes back to the turn of the 20th century, and thus allows us to have a

since the beginning of the 20th century. Shiller (2000) discusses both the 1929 and the 1987 crashes in more detail.

[8]Standard databases of news start after 1980, so there is no much room for time-series research outside the 1980-2008 window. The only other pure time-series analysis on a wide index is Larkin and Ryan (2008), who study the effect of news feeds in intraday returns for the S&P500 index. For other related papers see Cutler, Poterba, and Summers (1989), Klibanoff, Lamont, and Wizman (1998), Chan (2003), Kaniel, Starks, and Vasudevan (2007), Schmitz (2007), Barber and Odean (2008), Gaa (2008), Tetlock, Saar-Tsechansky, and Macskassy (2008), Yuan (2008), Engelberg (2008), Fang and Peress (2009), Solomon (2009), Bhattacharya, Galpin, Yu, and Ray (2009).

[9]See Brown and Cliff (2004, 2005), Qiu and Welch (2006) and Lemmon and Portniaguina (2006) for some recent research on sentiment measures and stock returns.

[10]Historical data is available free of charge from `http://www.djaverages.com/`, including the total return for the Dow Jones Industrial Average, but as of the end of 2009 this source did not include Saturday data. For this reason, we use the data on the DJIA from Williamson (2008), see `http://www.measuringworth.org/DJA/`. Exclusion of the Saturday data does not affect our results.

metric of US stock returns prior to the coverage in the more standard Center for Research in Security Prices (CRSP), which started in 1926. We remark that during our period the Dow Jones Industrial index was composed by as few as twelve securities in 1905, and increased to thirty starting in 1928. We let R_t denote the log-return on the DJIA index on date t. The business cycle information is obtained from the NBER website `http://www.nber.org/cycles.html`. In Section 4 we use intraday data on the DJIA index, provided by Global Financial Services, which is available since 1933. The last source of data is the novel measure used in this paper based on media content, which we describe next.

The media content measures are constructed starting from the Historical New York Times Archive, which goes back to the origins of the newspaper in 1851. This dataset was built by scanning the full content of the New York Times newspaper. It is available to any subscriber of the New York Times, as well as via other media providers (i.e. ProQuest). In order to have a consistent set of articles that cover financial news, we focus on two columns that were published daily during this period: the "Financial Markets" column, and the "Topics in Wall-Street" column. The "Topics in Wall-Street" column ran daily under different titles (i.e. "Sidelights from Wall-Street", "Financial and Business Sidelights of the Day," "Market Place") until the end of our sample period.[11] The "Financial Markets" column stopped being published with such a heading in the 1950s, although the New York Times obviously continued to publish a column with the financial news for the day, which we use in our analysis. The paper studies a total of 55,307 pdf files from the Historical Archive that were associated with either of these two columns from January 1, 1905 through December 31, 2005.

Both columns under study were essentially summaries of the events in Wall Street during the previous trading day. The "Financial Markets" column was somewhat shorter, around 700 words per day, versus 900 words for a typical "Topics in Wall-Street." The latter would typically be subdivided into multiple sections that described anything from particular companies or industries to commodities and general market conditions. The themes discussed in both columns were of a business nature, with a focus on financial matters. As such, they are good candidates to measure the content of financial news in the US during the 20th century. Figures 1 and 2 present a sample of each of the columns.

To construct the media content measures, we transform the scanned images available from the New York Times Historical Archive into text documents. This is referred to in the computer science literature as "optical character recognition" (OCR). We use ABBYY software, the leading package in OCR processing, to convert the 55,307 images into text files. A sample of the output from the OCR processing for the two columns in Figures 1 and 2 is included in the Appendix. Although the quality of the transcription of the articles is high, it is important to notice that the

[11] The format of the business news of the New York Times is very stable until the 1990s, when the proliferation of TV (CNBC) and the Internet made standard news sources change their formatting more frequently.

accuracy of OCR processing may be low for some files. The quality of the scanned images in the NYT Historical Archive is particularly low prior to 1905, thus our choice of starting date. The text samples in the Appendix contain a few typographical errors, all stemming from problems in the original scanned image.[12] This only adds random noise to our media content measures, and thus it will not bias our conclusions.

In order to quantify the content of the New York Times articles, this paper takes a "dictionary approach."[13] For each column i written on date t, we count the number of positive words, g_{it}, and negative words, b_{it}, using the word dictionaries provided by Bill McDonald.[14] As argued in Loughran and McDonald (2010), standard dictionaries fail to account for the nuances of Finance jargon, thus the categorization we use has particular merits for processing articles on financial events. We let w_{it} denote the total number of words in an article. We construct these media measures dating them to the day t in which they were written, with the understanding that they are published in the morning of day $t+1$. The rationale is that the information contained in these columns clearly belongs to date t. The writing process for each article started at 2:30-3:00 pm, typically just as the market was about to close, and the final copy was turned in to be edited and typeset at around 5-6 pm.

We aggregate the media content measures to create a time-series that matches the Dow Jones index return data available. The ultimate goal is to combine all the news that were printed before the market opened, and then examine whether the content of such news, our proxy for sentiment, can predict the following days' stock returns. In essence, we are trying to measure the content of the financial news on investors' desks prior to the opening of the market. The two columns of the New York Times would be typically published the day after the market closed, and they would discuss financial events related to that day. For example, the Sunday edition would discuss Wall Street events from Saturday. On some occasions, the New York Times would not print these columns on Sunday, but on Monday.

In order to aggregate the news, and not miss columns that appeared while the market was not open, we average the measures of positive/negative content from articles that were written since the market closed until the market next opens. When the market is open on consecutive days, t and $t+1$, we define our daily measure of positive media content as $G_t = \sum_i g_{it} / \sum_i w_{it}$, where the summation is over all articles written in date t (given our news selection, there are two such articles for the majority of days in our sample). Similarly, we construct our daily measure

[12] The OCR software will try to interpret anything in the original image, from spots to actual text. Different margins, multiple columns, and page formatting issues in general present a challenge for the character recognition process.

[13] Non-dictionary approaches have gained much popularity in recent research on text content analysis, in which not just the words, but the order and their role in a sentence is taken into account (i.e. the Diction software used in Demers and Vega, 2008). Given the OCR processing issues discussed above, these types of language processing algorithms are not appropriate for our study.

[14] See `http://www.nd.edu/~mcdonald/Word_Lists.html` for details.

of negative media content as $B_t = \sum_i b_{it} / \sum_i w_{it}$. In essence, we count the number of positive and negative words in the financial news under consideration, and normalize them by the total number of words. For non-consecutive market days we follow a similar approach, including all articles published from close to open. To be precise, consider two trading days t and $t+h+1$ such that $h > 0$ and the market was closed h days, from $t+1$ through $t+h$. We define the positive media content measure as $G_t = \sum_{i,s=t}^{s=t+h} g_{is} / \sum_{i,s=t}^{s=t+h} w_{sh}$. We proceed analogously for the negative media content variable and define $B_t = \sum_{i,s=t}^{s=t+h} b_{is} / \sum_{i,s=t}^{s=t+h} w_{sh}$. We define the pessimism factor as the difference between the negative and positive media content measures, i.e. $P_t = B_t - G_t$.

For consecutive trading dates, our media measures G_t and B_t are constructed using information that was available as of the end of date t when the market is open on date $t+1$ (the bulk of our sample). It is less clear whether market prices on date t reflected the information available to the journalists writing the columns, as the deadline for turning in the article to the editor was not until roughly 5-6pm, while the NYSE closed at 3–4pm. We further remark that for non-consecutive trading dates, we use articles that may have been written on days after date t, but prior to the market opening (i.e., in the case of holidays). Thus, strictly speaking, the New York Times measures we use could contain information that the market would not have as of the close of trade.

Table 1 presents sample statistics on our media measures. Panel A shows that, over the whole sample period, the mean number of positive words in an article was 1.20%. Given a typical "Financial Markets" column with 700 words, there were on average 8.4 positive words in each article. The standard deviation of the positive words measure is 0.42%. The average number of negative words over the whole sample is 2.06%, with a standard deviation of 0.67%. The pessimism measure, as expected given the numbers just discussed, has a positive mean, i.e. a typical article has about 0.86% more negative than positive words. Panels B and C present the sample statistics broken down by the business cycle. The average positive measure is slightly higher during expansions, by six basis points. On the other hand, the average negative measure is four basis points higher during recessions. The boxplots in Figure 3, which graphically illustrate the content of the two bottom panels of Table 1, show that our negative and pessimism media measures are different during recessions and booms, as one would expect. More importantly for the purposes of our tests, there is a very large amount of variation within business cycles: the volatility of the measures is an order of magnitude larger than the mean differences across the business cycle.

For the rest of the paper we normalize our sentiment measures so they have zero mean and unit variance. This will allow us to interpret the regression coefficients in terms of one-standard deviation shocks to the sentiment measures, thus making it easier to gauge the economic magnitude of our results.

Table 2 gives some summary statistics on DJIA returns. Panel A reports that the mean return on the DJIA was 2 basis points per day over the 1905–2005 period, with a daily volatility of 107 basis points. During recessions, which comprise 6467 days out of the total of 27449 trading days in our sample, the return was -5.3 basis points, whereas during expansions it was 4.2. Equally important is the difference in the volatility across the business cycle. During recessions the daily volatility was 141 basis points, whereas it was only 94 basis points during expansions.

Panel B presents the estimates of a parsimonious time-series model, which we shall augment with our media variables in the next sections. We estimate the following model of stock returns:

$$R_t = (1 - D_t)\boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + D_t \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\eta} X_t + \epsilon_t; \tag{1}$$

where $\mathcal{L}_s$ denotes an s-lag operator,[15] the variable D_t is a dummy variable taking on the value 1 if and only if date t is during a recession, the vector X_t denotes a set of exogenous variables, and ϵ_t is a zero-mean error term with possibly time-varying volatility. The set of exogenous variables X_t includes a constant term, day-of-the-week dummies, and a dummy for whether date t belongs to a recession or an expansion. We estimate the specification in (1) letting the lag operators have $s = 5$. We report White (1980) heteroskedasticity robust standard errors.

Starting with the first five rows of Panel B, we see that there is a statistically significant positive autocorrelation in the returns of the DJIA during expansions, but not during recessions. Thus, if autocorrelation is a concern for our results, the evidence from Panel B highlights that it should be a more serious concern during expansions than during recessions. The last three rows of Panel B show that there is a strong Monday effect, as the omitted day-of-the-week indicator. Returns on Monday are in the order of 13-19 basis points lower than the returns on other days of the week.

In some of our analysis we will attempt to deal with the time-variation in volatility that is present in our data. In order to address such concern we fit a GARCH(1,1) model to the DJIA returns. In particular we estimate a model with a constant mean, $R_t = \mu + \epsilon_t$, and time-varying volatility $\sigma_{t+1}^2 = \omega + \alpha_1 \epsilon_t^2 + \beta_1 \sigma_t^2$, where $\sigma_t^2 \equiv \mathrm{var}(\epsilon_t)$. The estimated coefficients for the variance equation are given in Panel C of Table 2. As expected, there is strong evidence of time-varying volatility.

[15] For an arbitrary random process Y_t, $\mathcal{L}_s(Y_t) = \{Y_{t-1}, \ldots, Y_{t-s}\}$.

3 Sentiment and DJIA returns

In order to formally analyze the relationship between stock returns and news measures, we postulate the following model for stock returns:

$$R_t = \boldsymbol{\beta}\mathcal{L}_s(M_t) + \boldsymbol{\gamma}\mathcal{L}_s(R_t) + \boldsymbol{\psi}\mathcal{L}_s(R_t^2) + \boldsymbol{\eta}X_t + \epsilon_t, \tag{2}$$

where M_t denotes one of our media measures, i.e. $M_t = G_t$ in the case of positive news, $M_t = B_t$ in the case of negative news, and $M_t = B_t - G_t$ in the case of our pessimism factor. Throughout the paper, we set $s = 5$ for our lag operators.[16] As the set of exogenous variables X_t we include a constant term, day-of-the-week dummies, as well as a dummy for whether date t belongs to a recession or an expansion, D_t.

Table 3 presents the point estimates $\boldsymbol{\beta}$ from (2), with White (1980) t-statistics. From the column titled "Positive," we see that the fraction of positive words in the morning's media helps predict stock returns on a given day. The effect is statistically different from zero at standard confidence levels. Turning to the column titled "Negative," we see a similar pattern emerges: more negative words predict lower stock returns. The magnitude of the effect is along the lines of what Tetlock (2007) reports – a one standard deviation shock to the "Pessimism" metric, for example, moves the conditional average return on the DJIA by 5.5 basis points (3.9 and -4.3 for the positive and negative media metrics).

The coefficients on the media variable lagged for $t-2$ through $t-5$ given an indication as to whether the shock to stock returns caused by the media content written on date $t-1$ is permanent or temporary. For example, in the case of the pessimism factor, we see the loadings on all four lags, $t-2$ through $t-5$, are positive, suggesting at least part of the drop in asset prices associated with M_{t-1} reverses over the following four days. Panel B formally conducts a test of whether the sum of these coefficients is different than zero. Although the statistical power is not as high as the test on a single coefficient, the p-values in Table 3 suggest that there is a reversal.

The regression in (2) is essentially the same model as estimated by Tetlock (2007), so the results in this section serve as an independent confirmation of his findings, with more powerful tests. In particular, our evidence corroborates the fact that media content can indeed predict the returns on a wide market index, and that the effect of such predictability partially reverses over the following four days. We note that the fact that the positive word counts help predict stock returns is a novel result: Tetlock (2007) focuses on negative word counts for most of his analysis due to the lack of predictability using positive words (and other word categories).[17] The

[16] The choice of lags and controls in the specification (2) does not affect any of the conclusions of the paper.

[17] As mentioned previously, our study uses the Loughran and McDonald (2010) word lists. Tetlock (2007) on the other hand, uses the Harvard IV dictionary.

literature has mostly followed his lead and ignore dictionaries of positive words. The evidence in Table 4 suggests that positive words can be as important as negative words.

In order to differentiate the effect of news content on stock returns along the business cycle, the focus of our paper, we estimate the following model:

$$R_t = (1 - D_t)\left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) \\ + D_t\left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t. \qquad (3)$$

In the above specification, M_t denotes one of our media measures, and the variable D_t is a dummy variable taking on the value 1 if and only if date t is during a recession. The vector X_t denotes a set of exogenous variables, and ϵ_t is a zero-mean error term with possibly time-varying volatility. As the set of exogenous variables X_t we use a constant term, day-of-the-week dummies, and a dummy for whether date t belongs to a recession or an expansion.

Table 4 presents the estimates of the coefficients on the media variable M_t from (3). Panel A includes the coefficient estimates $\boldsymbol{\beta}_1$, which measure the effect of media content on stock returns during expansionary periods, roughly two thirds of our sample. There is evidence of predictability using both the positive and negative word counts, and as a consequence the pessimism factor. The magnitude of the effect is nonetheless small – a one standard deviation change in the pessimism factor moves the DJIA by 3.5 basis points.

Panel B presents the point estimates of $\boldsymbol{\beta}_2$, which measure the effect of our news measures on stock returns during recessions. The point estimate on the positive news measure is 0.085, with a t-statistic of 3.4. Thus, a one-standard deviation change in the counts of positive words from the two columns of the New York Times increases the returns in the DJIA by 8.5 basis points during recessions. The effect is both statistically and economically meaningful – positive word counts have a much more important effect during recessions than during expansions. A similar pattern is observed for the negative and pessimism media measures. Whereas a one standard deviation shock to the pessimism factor would move the DJIA by 3.5 basis points during expansions, the effect would be 11 basis points during recessions. The effect of the negative news metric is 2.8 basis points during expansions, and 8.7 basis points during recessions. The first row of Panel C conducts a formal test of the differences in the coefficients in Panels A and B, concluding that they are statistically different. More importantly, the economic magnitude of the differences is large: the point estimates in Panel B are anywhere from three to four times bigger than in Panel A.

By any measure, our sentiment proxies help predict stock returns the following day, with similar magnitudes to those reported by Tetlock (2007) during expansions, and significantly larger during recessions. There is some evidence of return reversals in Panels A and B, as the sum of the coefficients on lags $t-2$ through $t-5$ partially swamp the initial effect. The second and third row in Panel C conduct formal F-tests. The second row shows that the sum of these four lags is different than zero during expansions (p-value 6.6% for the pessimism factor), but the statistical power, given the evidence from Table 3, is small. The F-test in the last row of Panel C formalizes the fact that we cannot reject the null hypothesis of no reversal during recessions. While the sum of the effects of lags $t-2$ through $t-5$ of the pessimism factor is about 6 basis points, half of the coefficient at $t-1$, the F-test only yields a p-value of 13.2%.

We consider next the robustness of the results from Table 4. One potential issue is the fact that we do not explicitly model time-varying volatility in (3). Although our standard errors are adjusted, one could be concerned that periods of high-volatility could affect our results. In order to address such concern we use the estimate GARCH(1,1) model from Table 2. If we let the estimated daily volatility from the GARCH model be $\hat{\sigma}_t$, we normalize the returns of the DJIA by replacing R_t in (3) with $R_t/\hat{\sigma}_t$. This normalization essentially constructs a time-series of stock returns with their volatility normalized to unity.

Panel A in Table 5 presents the estimates of the leading terms from the model in (3) using the unit-variance DJIA index returns. We observe a similar pattern to the one found in Table 4. The effect of the positive metric is 5.1 basis points during recessions, but only 2.2 basis points during expansions. The point estimates for the negative metric are -7.0 during recessions and -2.5 during expansions. By scaling down the DJIA returns during recessions we do find slightly smaller coefficients with respect to Table 4, but the conclusions stand. In particular, the tests using the pessimism factor conclude that the coefficients in recessions and expansions are different, and that the effect only partially reverses.

The skeptical reader may also be concerned with the autocorrelation controls, in particular given the relationship between the content of a column written on date t and the stock returns on that date t reported in Tetlock (2007).[18] In order to address this issue, we replace our media measure M_t with the residuals from the following model:

$$\begin{aligned} M_t = {} & (1-D_t)\left(\lambda_1 R_t + \boldsymbol{\beta}_1 \mathcal{L}_s(R_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(M_t)\right) \\ & + D_t\left(\lambda_2 R_t + \boldsymbol{\beta}_2 \mathcal{L}_s(R_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(M_t)\right) + \boldsymbol{\eta} X_t + \upsilon_t. \end{aligned} \tag{4}$$

Thus, we strip the media content metrics of any linear relationship with returns, day-of-the-week effects, as well as their own lags. Panel B in Table 5 presents the leading terms after estimating (3) replacing M_t with this orthogonalized media measure. As with the GARCH adjustment, we find somewhat smaller coefficients on our media variables, but the magnitudes of the point estimates during recessions are 2–3 times as large as during expansionary periods.

One further concern is the possibility that outliers may be driving our results, given the fat tails that characterize stock returns. Panel C in Table 5 presents the estimates of the model in

[18] See also our analysis in Section 4, in particular Table 6.

(3) fitted by robust regression using the M-estimator of Huber (1981). This technique, standard in Statistics to deal with outliers, essentially caps the influence that single observations can have on the point estimates.[19] The values reported in Panel C show that the effect reported in Table 4 is indeed robust. The magnitude of the point estimates is very similar to that reported in Table 4, and the statistical significance is actually even higher, which suggests that, if anything, the outliers are working against our findings.

We conclude this section by presenting two more robustness tests. The first addresses the concern of running time-series regressions over a period of over hundred years. Although we adjust for heteroskedasticity in our standard errors, one could conjecture that regression coefficients could change significantly, as US markets changed dramatically during the 1905–2005 period. We estimate (2) separately during each business cycle, dropping the D_t variable and its interactions. This boils down to estimating (2) with business cycle fixed effects, and with interactions of these business cycle fixed effects with all the independent variables.

Figure 4 plots the estimates of the leading coefficients from our media variables for each business cycle in our sample. In particular, for each recession and for each expansion, we estimate the model (2) and plot the leading coefficient on the media variable M_{t-1}. Estimates from the twenty different recessionary periods are marked with an x. Estimates for the twenty-one expansionary periods are marked with crosses. The dashed lines give the time-series averages of the recessionary and expansionary coefficients. Focusing on the right figure in the graph, we see that in eighteen out of the twenty recessions the pessimism factor loads with a negative sign. Further, the largest five coefficients all occur during recessions. It is also important to note that the effect does not seem to vary through time – there is significant predictability throughout the whole 20th century. Four of the largest five point estimates for the pessimism factor fall in the second-half of the sample, indicating that the effect has not dissipated with the advent of new types of media.[20]

We also study non-parametrically the relationship between the DJIA index returns and our media measures. Figure 5 plots a non-linear estimate of the relationship between stock returns and our media measures. In particular, the graphs are lowess (locally weighted scatter-plot smoothing) plots of the residuals of a time-series regression as in (3) (dropping the media variables) against each of our media variables. The solid line presents the estimates during expansions, whereas the dashed line corresponds to recessions. The non-parametric estimates forcefully argue that the effect of the media variables on stock returns is concentrated during recessions. We note that the linear approximation is a very good one for all media metrics during recessions. It is interesting to see not just the difference in magnitude of the slopes, but the fact that during expansions the relationship is not that robust: for the pessimism metric,

[19] For the technically inclined reader, we use the routine `rlm` from the CRAN depositories for the estimation.

[20] Broadcast television started in the United States in 1928. It reached a wide audience by the 1950s, half way through our sample period.

for example, the slope is negative only for half of the support of the variable.

Overall, the data strongly supports the OLS evidence from Table 4. Normalizing the stock returns and media measures in different ways yields similar conclusions. Moreover, the results are even stronger if we control for outliers. Finally, the estimates of the coefficients for each business cycle and our non-parametric analysis both show that the effect is significantly stronger in recessions.

4 Information versus sentiment

The asymmetric response of stock returns to financial news across the business cycle is the main finding of the paper. It is consistent with a story in which media content proxies for investor sentiment (i.e., noise traders), and this sentiment is more salient during recessions. The psychology literature discussed in the introduction suggests that reactions to news will be more pronounced during periods of anxiety and fear, i.e. during economic downturns. Whereas we believe that one of the key advantages of the media measures we construct is that they are unlikely to be related to fundamental information not possessed by traders, as argued in Shiller (2000) and Tetlock (2007), a skeptical reader may interpret the counts of positive and negative words from the New York Times as new information. The question of whether sentiment or information is behind our results is the focus of this section.

The alternative explanation for our results, that readers may have in mind, starts by assuming that journalists had information that was not impounded in closing prices. Given that the market closed a few hours before they had to turn in their story, it is feasible that NYT reporters would be ahead of traders. Our findings across the business cycle would then be consistent with information production by financial intermediaries, the NYT journalists, generating more precise signals during recessions.

We first study variation in the media content itself along the business cycle. Differences in reporting style would be an indication of media processing information asymmetrically during recessions and expansions. In particular, we study the effect of the returns on the DJIA on the sentiment measures we constructed estimating the model in (4).[21] Panel A from Table 6 presents the estimated coefficients, which measure the reaction of news content to the raw stock returns of the DJIA. As expected, stock returns are indeed important predictors of the media content variables, both the positive and the negative measures. Positive returns increase the number of positive words and decrease the number of negative words, and as a consequence the pessimism measure. Given the daily standard deviation of the DJIA over our whole sample period is 107 basis points, a one standard deviation increase in stock returns increases the percentage of

[21] We note that the system (3)-(4) is not a VAR in a strict sense, since we postulate a contemporaneous term in (4).

positive words on the articles written on that day during expansions by 0.36 standard deviations, and decreases the negative words by a similar amount.[22] The pessimism factor decreases by 0.44 standard deviations for a one standard deviation move in the DJIA returns during expansions. The effect is also persistent, as the second lag of returns also has significant predictive power on the media content variables.

The last five rows of Panel A show that the feedback effect is smaller during recessions. The leading coefficients are some 50% smaller than in the first five rows. Thus, it appears that journalists use more "signed words," or "tag-along" more, when writing about the day's events in Wall Street during expansions than during recessions. As the last row in Panel A shows, a formal test of equality of the coefficients can easily reject the null that the reaction of news to stock returns is equal along the business cycle ($\lambda_1 = \lambda_2$). This last conclusion should nonetheless be conditioned with the fact that the volatility of stock returns is significantly higher during recessions than during expansions. If we use the standard deviations from Table 2, we see that the effect of a one standard deviation movement in DJIA returns during expansions scales down the unconditional estimates by almost 15%, whereas a similar movement during recessions scales up the estimates by 30%. Thus the effect of a one-standard deviation movement on stock returns, conditional on the state of the business cycle, appears to be similar in magnitude.

In order to formalize this statement, we let the estimated daily volatility from the GARCH model from Panel C of Table 2 be $\hat{\sigma}_t$. We then normalize the returns of the DJIA by replacing R_t in (4) with $R_t/\hat{\sigma}_t$. This normalization essentially constructs a time-series of stock returns with their volatility normalized to unity. Table 6 gives the point estimates when the returns used in (4) are homoskedastic. The resulting effect on the media variables of a one-standard deviation movement in stock returns, reported in Panel B of Table 6, is identical during recessions and expansions, for both the positive and the negative word counts. Journalists "tag-alone" quite a bit, in the sense that their words are quite predictable given previous days' stock returns. More importantly for our purposes, there does not seem to be variation in this reporting along the business cycle. Thus, to the extent information may be reflected in the writing of the NYT columnists as a function of market conditions, such information does not seem to vary in recessions and expansions.

If journalists were producing informative signals for traders, it is not clear why the precision of these signals would increase during recessions. During economic downturns the press is hit particularly hard, as both subscriptions and advertising revenues are highly pro-cyclical. For example, during the Great Depression the subscriptions to the Wall Street Journal dropped from

[22]We remark that the effects we find is an order of magnitude larger than that reported in Tetlock (2007), as we include a contemporaneous term in (6). The inclusion of such a term has to do with the nature of the data: the columns in the New York Times were finished once the market was closed, so it is rather natural to think that the market return on that day would have an effect on the news content. Our empirical results show that indeed this is the case – although lags one through four have an effect on our media measures, as in Tetlock (2007), the return on the day the columns were written is the biggest determinant of the tone and content used by journalists.

52,000 to 28,000 (see p. 60 in Roush, 2006). It is unlikely that better coverage of financial markets would accompany staff cuts. Even if journalists were producing higher quality signals during recessions, it is also hard to explain why there is virtually no predictability during expansions.

Our next test focuses on the nature of information of the columns. Clearly many investors learned about firms and their decisions via the press during our time period, since there were no other important media outlets. Financial data itself seems to be the most relevant for investors – the NYT published regularly, in the columns under study, tables with dividends, stock prices, earnings, up to discussions on single figures. We divide the NYT articles based on the number of figures, namely the fraction of words that contain numbers.[23] The underlying idea is that articles with more "hard data" are more likely to contain information that is relevant for investors.

The NYT columns under study underwent significant changes throughout the years: some years they would include tables with stock prices, whereas others they will not have the tables as part of the column itself.[24] In order to classify our columns in terms of the amount of hard figures they would provide, we first estimate the following model:

$$N_t = \boldsymbol{\beta} Y_t + \boldsymbol{\eta} X_t + \epsilon_t; \tag{5}$$

where N_t denotes the fraction of words that contain numbers, Y_t is a matrix of year-month indicator variables, and the exogenous variables X_t include day-of-the-week dummies. We use the residuals from the above model as a proxy for the amount of hard information provided by the columns. This allows us to control for any patterns in the format of the columns through time and/or weekly conventions.

In particular, we define an indicator variable I_t that takes on the value 1 if and only if the estimated residuals from (5) are positive. We then estimate the model

$$\begin{aligned} R_t = (1-I_t)\big[&(1-D_t)\left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right)\big] \\ +I_t\big[&(1-D_t)\left(\boldsymbol{\beta}_3 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_3 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_3 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_4 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_4 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_4 \mathcal{L}_s(R_t^2)\right)\big] \\ +&\boldsymbol{\eta} X_t + \epsilon_t; \end{aligned} \tag{6}$$

which boils down to (3) with all the coefficients interacted with the figures indicator I_t. Table 7 gives the estimates of the leading coefficients on the media variables. In expansions we find that there is an effect irrespective of the amount of figures. The pessimism factor, for example, loads significantly with a coefficient of -3.1 basis points, whereas the coefficient on days low on

[23] We should note that during this time period stock prices, interest rates, dividends, ... were given as fractions. The OCR software struggles recognizing the way the NYT printed such fractions, so the text of the pdf images will typically have numbers and letters together. Clearly, in such instances the original text meant to be referring to a number. We define a "number" as any string that contains any of the ASCII characters 0-9.

[24] It should be noted that this heterogeneity is part of the classification of news that the NYT undertook when digitizing the newspaper. "Cutting up" news stories when there are hanging tables in the newspaper (i.e. above or below the actual column) clearly calls for some human judgement.

figures is only -3.8 basis points. During recessions we also find no significant differences among articles with high or low figure counts. On days that the financial columns lack hard data, the pessimism factor moves the DJIA by -10.9 basis points versus -12.7 basis points on days with more figures.

We conclude that it is not hard data which drives the relationship between our media content measures and stock returns. While this rules out some theories based on information, it could very well be that hard information and the soft information are actually substitutes, or independent from each other, and that our word counts contain facts that investors could not infer via the information contained in earnings or dividend announcements.

We next divide our sample depending on whether the previous day was a trading day or not. Our goal is to study the effect of media on DJIA returns on different days of the week, in particular on Mondays and holidays versus on other weekdays. The working hypothesis is that during holidays or weekends it is less likely that information was actually being produced, but more plausible that investors paid more attention to news. There are different motivations for this analysis. On one hand, there are distinct degrees of information production in weekdays versus weekends. Given that during our sample period most businesses were closed on weekends, it is natural to conjecture that information flow would be more intense during weekdays.[25] On the other hand, newspapers have significantly higher circulation on weekends, as investors have more time to read the news. We also conjecture that investors' attention to business news is concentrated on Mondays, as investors may be less attentive on Fridays (DellaVigna and Pollet, 2009).

Table 8 gives the estimates of the leading coefficients on the media variables from a model as in (6), where I_t is now a dummy variable that takes on the value 1 if and only if date $t-1$ was not a trading day (i.e. Mondays and days after holidays). During expansions we find that the predictability from Table 4 is concentrated on Mondays. The point estimates on Mondays or days after holidays are 5.6, -6.2 and -7.9 basis points for the positive, negative and pessimism factors, all highly significant. On weekdays the effect is much smaller – the formal test of differences of coefficients in the fifth row of Table 4 confirms this. During recessions we also find a larger effect for Mondays or days after holidays. The point estimate for the pessimism factor is -26.7 basis points, a very large effect given we are studying daily stock returns. During weekdays, this effect is also statistically significant, with a point estimate of -8.5 basis points for the pessimism factor, still large in economic terms.

Our previous conclusions are still unaltered – the effect is significantly larger during recessions than expansions. The point estimates, conditional on the day of the week, are anywhere from 2-4 times larger during recessionary periods. Although part of the effect could be due to information

[25] It should be noted that business writing on Sundays, for Monday's edition, did not start until halfway through our sample period. See Neilson (1973). Casual empiricism suggests that both businessmen and journalists work less hours during holidays.

released during the weekend, the asymmetry of the effect along the business cycle again suggests a sentiment story. We interpret the stronger effect on Mondays via higher attention paid by investor to news that come out on Sundays and Mondays. But we note that our main results also show up on weekdays: there is virtually no predictability during expansions (-2.4 basis points coefficient on the pessimism factor), whereas the point estimates during recessions are both statistically significant and economically large (-8.5 basis points).

If liquidity varied along the business cycle, an informational story would yield predictions that are consistent with our main findings. Microstructure frictions, such as lower market-depth during recessions, could potentially be driving part of our results. But note we are studying the most liquid sliver of the US stock market, and it is unclear why liquidity would exhibit the strong day-of-the-week effect we document in Table 8.

The more skeptical reader can view the tests on weekends as a robustness check. As mentioned previously, to the extent that there may be new information that journalists could include in their columns while the market was closed, it is possible that our study is contaminated by such inclusion. Holidays aside, this may be important for articles written on Saturdays, as the market closed at noon during the first half of our sample. Thus, it is possible that information that arrived in the afternoon of Saturday or during holidays could drive our results. Thus confirming that the effect is not just concentrated on weekends can be viewed as a robustness check on our main results.

We tackle directly this potential informational channel in our last set of tests. In particular, we look at the returns of the DJIA after the opening of the NYSE. If information is what is driving our results, and markets process this information quickly, we should find no predictability when looking at returns after the open. Table 9 presents the point estimates of a specification such as (3) where we replace the close-to-close DJIA returns by the returns from 11am to close.[26]

Table 9 presents the point estimates. We first see that the positive word counts have no predictive power at all during expansions, but that they do help predict DJIA post-opening returns during recessions. The point estimate on positive words is smaller than in Table 4, 5.7 basis points versus 8.5, but still economically and statistically large. Similar conclusions arise from the negative word counts and as a consequence the pessimism factor. While this analysis does not rule out an alternative hypothesis where information slowly diffuses into prices, it does dismiss informational theories where prices fully and quickly adjust to new data.

[26]Tetlock (2007) looks at returns from 10am to close, since during his sample period the NYSE opened at 9.30am. Trading hours during our sample period were 10am–3pm weekdays and 10am–noon on Saturdays (1905–1952), 10am–3.30pm (1952–1974), 10am–4pm (1974–1985), 9.30am–4pm (1985–2005), see `http://www.nyse.com/pdfs/historical_trading_hours.pdf`. The data provided by Global Financial Data has DJIA levels at hourly intervals during our sample period.

5 Conclusion

The paper constructs a measure of sentiment based on financial news from the New York Times during 1905–2005, and it studies its relationship to stock returns. The long time-series allows us to study a period that had a fairly concentrated media sector, as well as plenty of variability in the business cycle. Our main finding is that news content helps predict stock returns at the daily frequency, particularly during recessions. Whereas a one standard deviation change in our pessimism factor moves the DJIA by 12 basis points in recessions, the marginal effect during expansions is only 3.5 basis points.

We have shown how this asymmetric predictability is not driven by differences in reporting along the business cycle. The effect is specially strong on Mondays and days after holidays, when investors have time to read the news, and it persists into the afternoon of the trading day. We also find that the effect partially reverses over the following four trading days. We conclude that our results support the hypothesis that investor sentiment has a prominent effect during bad times.

The Historical Archive we use in this paper opens the door for other research questions. While our paper clearly documents the differences in impact of sentiment along the business cycle, it does not speak to the mechanism that drives the predictability. Finally, whether there are lower-frequency components to our sentiment measures is a natural avenue to explore, specially in connection to economic growth figures and long-run stock returns.

References

Akerlof, G. A., and R. J. Shiller, 2009, *Animal Spirits*. Princeton University Press, Princeton.

Baker, M., and J. Wurgler, 2007, "Investor sentiment in the stock market," *Journa of Economic Perspectives*, 21, 129–151.

Barber, B., and T. Odean, 2008, "All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors," *Review of Financial Studies*, 21(2), 785–818.

Bhattacharya, U., N. Galpin, X. Yu, and R. Ray, 2009, "The role of the media in the internet IPO bubble," *Journal of Financial and Quantitative Analysis*, 44, 657–682.

Bless, H., G. L. Clore, N. Schwarz, V. Golisano, C. Rabe, and M. Wölk, 1996, "Mood and the use of scripts: does a happy mood really lead to mindlessness?," *Journal of Personality and Social Psychology*, 71(4), 665–679.

Bow, J., 1980, "The "Times's" Financial Markets column in the period around the 1929 crash," *Journalism Quarterly*, 57, 447–450.

Brown, G. W., and M. T. Cliff, 2004, "Investor sentiment and the near-term stock market," *Journal of Empirical Finance*, 11, 1 – 27.

——— , 2005, "Investor Sentiment and Asset Valuation," *Journal of Business*, 78(2), 405 – 440.

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., J. M. Poterba, and L. H. Summers, 1989, "What Moves Stock Prices?," *Journal of Portfolio Management*, 15, 4–12.

DellaVigna, S., and J. M. Pollet, 2009, "Investor inattention and Friday earnings announcements," *Journal of Finance*, pp. 709–749.

Demers, E. A., and C. Vega, 2008, "Soft Information in Earnings Announcements: News or Noise?," working paper, INSEAD.

Dougal, C., J. Engelberg, D. García, and C. Parsons, 2011, "Journalists and the stock market," working paper, UNC at Chapel Hill.

Edmans, A., D. García, and Ø. Norli, 2007, "Sports sentiment and stock returns," *Journal of Finance*, 62, 1967–1998.

Engelberg, J., 2008, "Costly information processing: evidence from earnings announcements," working paper, University of North Carolina.

Engelberg, J., and C. Parsons, 2011, "The causal impact of media in financial markets," *Journal of Finance*, 66(1), 67–97.

Fang, L. H., and J. Peress, 2009, "Media coverage and the cross-section of stock returns," *Journal of Finance*, 64(5), 2023–2052.

Forgas, J. P., 1991, *Emotion and social judgments* Pergamon Press, Oxford, chap. Affect and social judgments: An introductory review, pp. 3–29.

——— , 1998, "On being happy and mistaken: mood effects on the fundamental attribution error," *Journal of Personality and Social Psychology*, 75(2), 318–331.

Gaa, C., 2008, "Good news is no news: asymmetric inattention and the neglected firm effect," working paper, University of British Columbia.

Gino, F., A. Wood, and M. E. Schweitzer, 2009, "How Anxiety Increases Advice-Taking (Even When the Advice is Bad)," working paper, Wharton School of the University of Pennsylvania.

Griggs, H., 1963, "Newspaper performance in recession coverage," *Journalism Quarterly*, 40, 559–564.

Hirshleifer, D., 2001, "Investor psychology and asset pricing," *Journal of Finance*, 61, 1533–1597.

Hirshleifer, D., and T. Shumway, 2003, "Good day sunshine: stock returns and the weather," *Journal of Finance*, 58, 1009–1032.

Huber, P. J., 1981, *Robust Statistics*. Wiley.

Kaniel, R., L. T. Starks, and V. Vasudevan, 2007, "Headlines and bottom lines: attention and learning effects from media coverage of mutual funds," working paper, University of Texas.

Klibanoff, P., O. Lamont, and T. Wizman, 1998, "Investor reaction to salient news in closed-end country funds," *Journal of Finance*, 53(2), 673–699.

Kogan, S., B. R. Routledge, J. S. Sagi, and N. Smith, 2009, "Information Content of Public Firm Disclosures and the Sarbanes-Oxley Act," working paper, University of Texas at Austin.

Larkin, F., and C. Ryan, 2008, "Good news: using news feeds with genetic programming to predict stock prices," *Lecture Notes in Computer Science*, 4971, 49–60.

Lemmon, M., and E. Portniaguina, 2006, "Consumer Confidence and Asset Prices: Some Empirical Evidence," *Review of Financial Studies*, 19(4), 1499 – 1529.

Lerner, J. S., and D. Keltner, 2000, "Beyond valence: toward a model of emotion-specific influences on judgement and choice," *Cognition and emotion*, 14(4), 473–493.

——— , 2001, "Fear, anger, and risk," *Journal of Personality and Social Psychology*, 81(1), 146–159.

Loughran, T., and B. McDonald, 2010, "When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks," *Journal of Finance*, forthcoming.

Neilson, W., 1973, *What's News – Dow Jones*. Clinton Book Company, Radnor, PA.

Norris, F., and C. Bockelmann, 2000, *The New York Times — Century of Business*. McGraw-Hill, New York City, New York.

Ortony, A., G. Clore, and A. Collins, 1988, *The cognitive structure of emotions*. Cambridge University Press, Cambridge, England.

Park, J., and M. R. Banaji, 2000, "Mood and heuristics: the influence of happy and sad states on sensitivity and bias in stereotyping," *Journal of Personality and Social Psychology*, 78(6), 1005–1023.

Qiu, L. X., and I. Welch, 2006, "Investor Sentiment Measures," working paper, Brown University.

Roush, C., 2006, *Profits and Losses*. Marion Street Press, Oak Park, Illinois.

Schmitz, P., 2007, "Market and individual investors reactions to corporate news in the media," working paper, University of Mannheim.

Shiller, R. J., 2000, *Irrational Exuberance*. Princeton University Press, Princeton.

Smith, C., and P. Ellsworth, 1985, "Patterns of cognitive appraisal in emotion," *Journal of Personality and Social Psychology*, 48, 813–838.

Solomon, D., 2009, "Selective publicity and stock prices," working paper, University of Chicago.

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

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008, "More than words: quantifying language to measure firms' fundamentals," *Journal of Finance*, 63(3), 1437–1467.

Tiedens, L. Z., and S. Linton, 2001, "Judgement under emotional certainty and uncertainty: the effects of specific emotions on information processing," *Journal of Personality and Social Psychology*, 81(6), 973–988.

White, H., 1980, "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity," *Econometrica*, 817–838.

Williamson, S. H., 2008, "Daily Closing Values of the DJA in the United States, 1885 to Present," working paper, MeasuringWorth.

Yuan, Y., 2008, "Attention and trading," working paper, University of Iowa.

Appendix

This Appendix presents the output of the optical character recognition on two columns from the New York Times. The first one is the “Financial Markets” column from October 12th 1915, displayed in Figure 1. The second is the “Topics in Wall Street” column from June 25th, 1916, displayed in Figure 2. Positive words, using the Loughran and McDonald (2010) dictionaries, are marked in italics, whereas negative words are marked in bold.

FINANCIAL MARKETS
Oct 12, 1915; pg. 14

Another Day of *Great* Activity, with Big *Gains* for the Industrials.

Kxpcotalions that yesterday’s Stock Kxohan’ge session, sandwiched in be-hvirn two holidays, would see much less activity were quickly **disappointed**. The market opened very active and *strong*, anr] afu-r a small reaction in the forenoon resumed its upward trend with almost the same **violence** shown in the *excited* sessions of last week. The list was again **irregular**, but by far the larger number of stocks scored substantial *gains* and the upward movement of some of the war issues, which had j been checked by banks and brokers who ’foiosaw **trouble** if the advance were not i held under control, was resumed with a *great* deal of visor. The most striking *gain* among such issues was scored by Baldwin Locomotive, which, after hang-in̂for several days around 113, returned yesterday to 127M>. **closing** at 12(i, with a net advance of 11 points. This secondary stage of activity for Baldwin was accompanied by fresh merger rumors, which do not appear to have any substantial basis In fact. Even more active and relatively as *strong* was Westirighouse. of which. more than WO.Udti shares changed hands — up a range of .”’i points. It **closed** at J 13.*!. with a *gain* of points above Saturday’s close. The American Car & Foundry made n *good* recovery to X.V-S. and *gains* of from t̂o .” points were numerous. The motor issues returned to *popularity*, all three classes of Maxwell stock *advancing* on the expectation of pom©kind of an announcement Wednesday of a plan looking to the payment of the accumulated dividend on the first preferred. Studebaker advanced 2]< and General Motors 1 point.

The rails retained some of their momentum from last week, and most of j the leaders sold at new high prices for : the. year. Xev.s of the note being prepared for dispatch to tlreat Britain was received too **late** to affect the market, if indeed ûch news can have any ef-i feet on the present temper of traders, and the list **closed** pretty clofc to the top.

j Some uneasiness was caused yesterday I by a new development of **weakness** in the foreign exchange market. Demand sterling, sold down to ?4.<<7% compared with the low price of Splits1.., on Saturday. The **failure** of the conclusion of the So”>”,0(10,000 Anglo-French loan to help foreign exchange rates gave special interest to an important meeting of bankers held yesterday afternoon, which was addressed by Hr l-Zdward Holden, one of the visiting Commissioners.

TOPICS IN WALL STREET
June 25, 1916; pg E6

American munition Orders.

Until yesterday the stock market gave no indication that the war stocks derived a chance of profit from war with Mexico. To speculators in these shares it was in fact a matter of the keenest **disappointment** that they went down on war news. Over and over they have repeated the **question**: " What sort of a war stock is it that is **depressed** by a new war? " Yesterday an advance of 17 points in Bethlehem Steel held out a ray of hope and *advances* In most of the others on covering by professionals *strengthened* hopes that the next turn would be for the *better*. Officers of many of the munitions companies expected orders from the United States Government in the near future, but nowhere was it believed that these orders would be placed at terms permitting as *great* profits as those obtained in some of the contracts with the Allies. **

The Extent of the **Declines**.

From the high point of week before last to tbe low point of last week, which was the low point of Friday's market, the average price of fifty representative stocks **declined**- 55.33 a share. These stocks included many railroad shares in which the **declines** were small compared with **losses** in some of the speculative industrials. Reading, which **lost** 8% points in this period, and Norfolk & "Western, with a **loss** of 5%, were the only rails to **decline** more than the 51-3-point average of the fifty. A score of industrials sustained *greater* **losses** and many of these **losses** ran into double figures, among them being: New York Air Brake, 11; Mexican Petroleum, 12%: Baldwin Locomotive, 13; United States Smelting, 13; Tennessee Copper, W/y, American Zinc, 14%; Willys-Overland, 16; Butte and *Superior*, 10%; United States Industrial Alcohol, 2<% On the Curb Chevrolet Motors **lost** 46 points.

Sow Up, Now Down.

It is interesting to note the change in sentiment that sweeps over the floor of the Stock Exchange after a pronounced rise, or sharp **decline**. Traders who have been bearish for weeks were turning bullish yesterday morning. They figured that the **break** which had been needed had been supplied, and that, therefore, stocks were a purchase again. ***

Tbe Mexican Fuetor.

An old-time member eaid after the close that neither the Mexican war **danger**, nor the **inadequacy** of our war machinery, was really back of the slump which took place last week. Those **arguments** were advanced to support the **decline**, but in his opinion the **break** would have come had the Mexican situation continued unchanged. This man's theory is that the market had become badly congested with stocks, and had to be cleaned out by a return to lower prices. A number of pools were carrying large amounts of stock which they had not been *able* to market, and there were some large individual accounts that needed shaking out. The low prices made on Friday brought In a number of fresh buyers, and If this trader's theory works out the market will develop a much *better* tone this week, regardless of developments across the border. When the list grows stale' nothing but a sharp **setback** will attract new money. That this market had become stale was evidenced by its utter **disregard** of *good* news, such as new and Increased dividends. <<**

No Extra Holiday.

When the brokers gave up their expected extra holiday before May 3p, they looked for an extra day preceding the Fourth. The uncertainty of | the political situation appears to have **destroyed** any chance of getting it. No petition has been circulated on the floor, and it is unlikely that the situation will clear in time to allow the drafting of one before the next meeting of Governors. *>>*

Bonds) Have **Idle** Week.

The bond market **suffered** along with stocks last week, but without registering substantial **declines**. Bonds' were effected more through a let-down of buying than from the **liquidation** of securities. Some of the banks and large dealers were reported as sellers of a considerable amount of bonds which they had been carrying for a month or more, and on which they had *good* profits. If this actually did take place the offerings were rather *easily* absorbed, and inquiries among bond men **failed** to show that there had been any **urgent** selling through **fear** that the Mexican situation might wipe out profits before they could be realized. The investment demand is believed to be widening, now that supplies from Europe have begun to fall away, leaving room for other offerings, and the bankers are inclined to think that business will pick up again with the coming of definite developments south of the border.

Table 1
Sample statistics for media content variables during recessions and expansions

The table reports sample statistics for the media content measures used in the paper. These measures are constructed from the columns "Financial Markets" and "Topics in Wall-Street" published in the New York Times in the period 1905–2005. We construct the "Positive" and "Negative" measures by counting the number of positive and negative words and normalizing it by the total number of words of each article, using the Loughran and McDonald (2010) dictionaries. The "Pessimism" variable is simply the difference between the "Negative" and "Positive" measures. All numbers are given in percentages. Panel A presents the sample statistics for the whole sample period, which comprises a total of 27449 trading days. Panel B and C break it down by the business cycle. Panel B, which contains all trading days during a recession in the 1905–2005 period, has 6467 observations, whereas Panel C, which contains the expansionary dates, has 20722.

Media measure	Mean	Median	25%-quant.	75%-quant.	Stand. dev.
A. All dates					
Positive	1.20	1.16	0.90	1.46	0.42
Negative	2.06	1.99	1.59	2.45	0.67
Pessimism	0.86	0.81	0.26	1.40	0.88
B. Recessions					
Positive	1.15	1.12	0.88	1.38	0.39
Negative	2.09	2.04	1.64	2.48	0.64
Pessimism	0.94	0.90	0.38	1.46	0.84
C. Expansions					
Positive	1.21	1.17	0.91	1.48	0.43
Negative	2.05	1.98	1.57	2.45	0.68
Pessimism	0.84	0.78	0.23	1.38	0.89

Table 2
Sample statistics for daily DJIA returns, 1905–2005

The table reports sample statistics for the DJIA returns used in the paper. Panel A gives unconditional sample statistics for the daily returns of the DJIA for the period 1905–2005. The first row presents the sample statistics, whereas the following two rows break down the sample period into NBER recessions and expansions. Panel B reports the estimated coefficients from the model

$$R_t = (1 - D_t)\gamma_1 \mathcal{L}_s(R_t) + D_t \gamma_2 \mathcal{L}_s(R_t) + \boldsymbol{\eta} X_t + \epsilon_t;$$

where $\mathcal{L}_s$ denotes an s-lag operator, namely $\mathcal{L}_s(R_t) = \{R_{t-1}, \ldots, R_{t-s}\}$, and D_t is a dummy variable taking on the value 1 if and only if date t is during a recession. As the set of exogenous variables X_t we include a constant term, day-of-the-week dummies, as well as a dummy for whether date t belongs to a recession or an expansion, D_t. Panel C presents the estimates of a GARCH(1,1) model. Namely we assume the return equation has a constant mean, $R_t = \mu + \epsilon_t$, but we allow for time-varying volatility of the form $\sigma^2_{t+1} = \omega + \alpha_1 \epsilon^2_t + \beta_1 \sigma^2_t$, where $\sigma^2_t \equiv \mathrm{var}(\epsilon_t)$. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

A. Sample statistics

	Mean	Median	25%-quant.	75%-quant.	Stand. dev.
All dates	0.020	0.044	-0.450	0.526	1.071
Expansions	0.042	0.056	-0.410	0.517	0.943
Recessions	-0.053	-0.011	-0.637	0.565	1.408

B. Time-series regression

Expansions	γ_1	t-stat	Recessions	γ_2	t-stat
$(1 - D_t) \times R_{t-1}$	0.052	3.1	$D_t \times R_{t-1}$	0.024	0.8
$(1 - D_t) \times R_{t-2}$	-0.045	-2.7	$D_t \times R_{t-2}$	-0.019	-0.7
$(1 - D_t) \times R_{t-3}$	0.004	0.3	$D_t \times R_{t-3}$	0.004	0.2
$(1 - D_t) \times R_{t-4}$	0.005	0.5	$D_t \times R_{t-4}$	0.062	2.6
$(1 - D_t) \times R_{t-5}$	0.011	0.7	$D_t \times R_{t-5}$	0.022	0.9
	$\boldsymbol{\eta}$	t-stat		$\boldsymbol{\eta}$	t-stat
I_{Tue}	0.140	6.3	I_{Fri}	0.167	7.4
I_{Wed}	0.153	6.7	I_{Sat}	0.189	7.5
I_{Thu}	0.125	5.6	D_t	-0.091	-5.0

C. GARCH(1,1) estimates

	$\omega, \alpha_1, \beta_1$	t-stat
Constant, ω	0.011	11.8
Innovations term, α_1	0.081	24.9
Autoregressive term, β_1	0.910	255.2

Table 3
Feedback from news content to the DJIA

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = \boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}\mathcal{L}_s(R_t) + \boldsymbol{\psi}\mathcal{L}_s(R_t^2) + \boldsymbol{\eta} X_t + \epsilon_t.$$

The dependent variable R_t is the log-return on the Dow Jones Industrial average from 1905–2005. The dependent variable M_t is one of our media measures, i.e. $M_t = G_t$ in the case of positive words, $M_t = B_t$ in the case of negative words, and $M_t = B_t - G_t$ in the case of our pessimism factor. The media measures are constructed from the columns "Financial Markets" and "Topics in Wall-Street" published in the New York Times in the period 1905–2005. The media variables are normalized to have unit variance. As the set of exogenous variables X_t we include a constant term, day-of-the-week dummies, as well as a dummy for whether date t belongs to a recession or an expansion, D_t. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
A. Media variables ($\boldsymbol{\beta}$)						
M_{t-1}	0.039	5.2	-0.043	-5.2	-0.055	-6.3
M_{t-2}	0.003	0.4	0.003	0.3	0.001	0.2
M_{t-3}	-0.008	-1.1	0.005	0.7	0.008	1.0
M_{t-4}	-0.013	-1.8	0.008	1.0	0.013	1.6
M_{t-5}	-0.005	-0.6	0.009	1.2	0.010	1.3
B. Tests	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_1 = 0$	26.9	0.000	26.8	0.000	40.0	0.000
$\sum_{j=2}^{5} \beta_j = 0$	3.1	0.077	3.6	0.059	5.6	0.018

Table 4
Feedback from news content to the DJIA along the business cycle

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = (1 - D_t)\left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t.$$

All independent variables are as in Table 3. The dependent variable R_t is the log-return on the Dow Jones Industrial average from 1905–2005. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
A. Expansions ($\boldsymbol{\beta}_1$)						
$(1 - D_t) \times M_{t-1}$	0.024	3.3	-0.028	-3.5	-0.035	-4.2
$(1 - D_t) \times M_{t-2}$	0.004	0.6	0.004	0.5	0.001	0.1
$(1 - D_t) \times M_{t-3}$	-0.004	-0.6	0.005	0.7	0.006	0.8
$(1 - D_t) \times M_{t-4}$	-0.012	-1.7	0.006	0.8	0.011	1.5
$(1 - D_t) \times M_{t-5}$	-0.004	-0.6	0.006	0.8	0.007	0.9
B. Recessions ($\boldsymbol{\beta}_2$)						
$D_t \times M_{t-1}$	0.085	3.9	-0.087	-3.4	-0.117	-4.4
$D_t \times M_{t-2}$	0.004	0.2	-0.005	-0.2	-0.004	-0.2
$D_t \times M_{t-3}$	-0.021	-1.0	0.010	0.4	0.020	0.8
$D_t \times M_{t-4}$	-0.009	-0.4	0.016	0.7	0.019	0.8
$D_t \times M_{t-5}$	-0.005	-0.2	0.028	1.2	0.026	1.1
C. Tests	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	7.2	0.007	5.0	0.025	8.6	0.003
$\sum_{j=2}^{5} \beta_{1j} = 0$	1.6	0.205	2.6	0.109	3.4	0.066
$\sum_{j=2}^{5} \beta_{2j} = 0$	0.7	0.403	1.6	0.212	2.3	0.132

Table 5
Volatility adjustments, orthogonal media content and robust regressions

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = (1 - D_t)\left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t.$$

All independent variables are as in Table 3. In Panel A, the dependent variable R_t denotes the normalized log-returns on the DJIA. These are constructed by taking the raw log-returns on the DJIA and dividing them by the estimates $\hat{\sigma}_t$ from the GARCH(1,1) model from Panel C of Table 2. In Panels B and C, R_t denotes the log-return on the DJIA average. In Panels A and C, the variable M_t denotes one of our media measures, as described in Table 3. In Panel B, the variable M_t is the residual from the model estimated in Table 6. Estimation in Panels A and B is via OLS. The estimation in Panel C is done using robust linear regression based on Huber (1981)'s M-estimator. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
A. GARCH-adjusted returns						
	β	t-stat	β	t-stat	β	t-stat
Expansions, $(1 - D_t) \times M_{t-1}$	0.022	3.1	-0.025	-3.3	-0.033	-4.1
Recessions, $D_t \times M_{t-1}$	0.051	3.5	-0.070	-4.3	-0.087	-5.1
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	3.1	0.079	6.1	0.014	8.2	0.004
$\sum_{j=2}^{5} \beta_{1j} = 0$	1.3	0.254	1.9	0.170	3.0	0.085
$\sum_{j=2}^{5} \beta_{2j} = 0$	0.0	0.868	2.7	0.103	2.1	0.148
B. Orthogonal media measures						
	β	t-stat	β	t-stat	β	t-stat
Expansions, $(1 - D_t) \times M_{t-1}$	0.022	3.3	-0.027	-3.8	-0.032	-4.4
Recessions, $D_t \times M_{t-1}$	0.078	3.8	-0.068	-3.1	-0.094	-4.1
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	6.9	0.008	3.1	0.080	6.6	0.010
$\sum_{j=2}^{5} \beta_{1j} = 0$	0.1	0.815	0.2	0.667	0.2	0.688
$\sum_{j=2}^{5} \beta_{2j} = 0$	0.0	0.898	0.0	0.867	0.0	0.917
C. Robust regression						
	β	t-stat	β	t-stat	β	t-stat
Expansions, $(1 - D_t) \times M_{t-1}$	0.024	4.0	-0.025	-4.0	-0.034	-5.3
Recessions, $D_t \times M_{t-1}$	0.055	4.6	-0.086	-7.0	-0.101	-7.9
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11} = \beta_{21}$	5.4	0.020	19.7	0.000	22.2	0.000
$\sum_{j=2}^{5} \beta_{1j} = 0$	3.2	0.075	1.3	0.258	3.4	0.065
$\sum_{j=2}^{5} \beta_{2j} = 0$	1.2	0.282	7.7	0.005	3.5	0.063

Table 6
Feedback from stock prices to news content

The table reports the estimated coefficients λ and β from the model

$$M_t = (1 - D_t)\left(\lambda_1 R_t + \boldsymbol{\beta}_1 \mathcal{L}_s(R_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(M_t)\right) + D_t\left(\lambda_2 R_t + \boldsymbol{\beta}_2 \mathcal{L}_s(R_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(M_t)\right) + \boldsymbol{\eta} X_t + \upsilon_t;$$

The variable M_t denotes one of our media measures, as described in Table 3. The set of exogenous variables X_t includes those in the specification of Table 3. In Panel A the variable R_t denotes the log-return on the DJIA. In Panel B the variable R_t denotes the normalized log-returns on the DJIA, constructed as in Panel A of Table 5. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	λ, β	t-stat	λ, β	t-stat	λ, β	t-stat
A. Using raw returns $(\lambda_1, \boldsymbol{\beta}_1, \lambda_2, \boldsymbol{\beta}_2)$						
$(1-D_t) \times R_t$	0.335	32.3	-0.332	-30.7	-0.414	-34.2
$(1-D_t) \times R_{t-1}$	0.046	6.1	-0.056	-7.6	-0.059	-7.8
$(1-D_t) \times R_{t-2}$	-0.008	-1.1	-0.012	-1.6	-0.005	-0.6
$(1-D_t) \times R_{t-3}$	0.008	1.0	-0.001	-0.1	-0.004	-0.5
$(1-D_t) \times R_{t-4}$	-0.011	-1.6	0.020	2.7	0.024	3.3
$D_t \times R_t$	0.197	16.8	-0.221	-19.5	-0.263	-20.3
$D_t \times R_{t-1}$	0.048	5.4	-0.045	-5.4	-0.052	-5.7
$D_t \times R_{t-2}$	0.011	1.3	-0.007	-0.8	-0.009	-1.0
$D_t \times R_{t-3}$	0.003	0.3	-0.019	-2.7	-0.012	-1.6
$D_t \times R_{t-4}$	-0.007	-0.9	0.022	2.4	0.023	2.5
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\lambda_1 = \lambda_2$	28.5	0.000	20.2	0.000	28.6	0.000
B. Returns normalized by GARCH(1,1) $(\lambda_1, \boldsymbol{\beta}_1, \lambda_2, \boldsymbol{\beta}_2)$						
$(1-D_t) \times R_t$	0.345	48.7	-0.342	-50.2	-0.427	-61.7
$(1-D_t) \times R_{t-1}$	0.040	5.6	-0.044	-6.4	-0.047	-6.8
$(1-D_t) \times R_{t-2}$	-0.017	-2.4	0.001	0.2	0.010	1.5
$(1-D_t) \times R_{t-3}$	0.005	0.7	0.002	0.3	-0.000	-0.0
$(1-D_t) \times R_{t-4}$	-0.007	-1.0	0.014	2.0	0.017	2.5
$D_t \times R_t$	0.324	26.3	-0.338	-31.4	-0.413	-37.0
$D_t \times R_{t-1}$	0.065	5.5	-0.050	-4.7	-0.063	-5.6
$D_t \times R_{t-2}$	-0.003	-0.3	0.003	0.3	0.008	0.8
$D_t \times R_{t-3}$	-0.007	-0.7	-0.016	-1.5	-0.001	-0.1
$D_t \times R_{t-4}$	-0.002	-0.1	0.026	2.4	0.024	2.2
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\lambda_1 = \lambda_2$	1.9	0.086	0.9	0.474	0.8	0.564

Table 7
Hard news and the effect of media content on stock returns

The table reports the estimated coefficients β from the model

$$\begin{aligned}R_t = (1-I_t)\left[(1-D_t)\left(\boldsymbol{\beta}_1\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_1\mathcal{L}_s(R_t)+\boldsymbol{\psi}_1\mathcal{L}_s(R_t^2)\right)+D_t\left(\boldsymbol{\beta}_2\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_2\mathcal{L}_s(R_t)+\boldsymbol{\psi}_2\mathcal{L}_s(R_t^2)\right)\right]\\ +I_t\left[(1-D_t)\left(\boldsymbol{\beta}_3\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_3\mathcal{L}_s(R_t)+\boldsymbol{\psi}_3\mathcal{L}_s(R_t^2)\right)+D_t\left(\boldsymbol{\beta}_4\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_4\mathcal{L}_s(R_t)+\boldsymbol{\psi}_4\mathcal{L}_s(R_t^2)\right)\right]\\ +\boldsymbol{\eta}X_t+\epsilon_t;\end{aligned}$$

where the variable I_t is an indicator variable that takes on the value one if and only if the news on date t contain more figures (numbers) than the average day in the month. and all other independent variables are as in Table 3. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Expansion, high-information, β_{11}	0.020	1.9	-0.026	-2.1	-0.031	-2.6
Expansion, low-information, β_{21}	0.026	3.0	-0.029	-3.0	-0.038	-3.8
Recession, high-information, β_{31}	0.098	3.1	-0.094	-2.6	-0.127	-3.4
Recession, low-information, β_{41}	0.075	2.9	-0.082	-2.8	-0.109	-3.7
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11}=\beta_{21}$	0.3	0.611	0.1	0.819	0.2	0.644
$\beta_{31}=\beta_{41}$	0.4	0.532	0.1	0.779	0.2	0.662
$\beta_{11}=\beta_{31}$	5.7	0.017	3.1	0.076	6.1	0.014
$\beta_{21}=\beta_{41}$	3.2	0.075	2.9	0.086	5.2	0.022

Table 8
Mondays, post-holiday returns and media

The table reports the estimated coefficients β from the model

$$\begin{aligned} R_t = (1-I_t)\left[(1-D_t)\left(\boldsymbol{\beta}_1\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_1\mathcal{L}_s(R_t)+\boldsymbol{\psi}_1\mathcal{L}_s(R_t^2)\right)+D_t\left(\boldsymbol{\beta}_2\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_2\mathcal{L}_s(R_t)+\boldsymbol{\psi}_2\mathcal{L}_s(R_t^2)\right)\right] \\ +I_t\left[(1-D_t)\left(\boldsymbol{\beta}_3\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_3\mathcal{L}_s(R_t)+\boldsymbol{\psi}_3\mathcal{L}_s(R_t^2)\right)+D_t\left(\boldsymbol{\beta}_4\mathcal{L}_s(M_t)+\boldsymbol{\gamma}_4\mathcal{L}_s(R_t)+\boldsymbol{\psi}_4\mathcal{L}_s(R_t^2)\right)\right] \\ +\boldsymbol{\eta}X_t+\epsilon_t; \end{aligned}$$

where the dummy variable I_t takes on the value one if and only if date $t-1$ was not a trading date, and all other independent variables are defined as in Table 3. The dependent variable R_t is the log-return on the DJIA index from 1905–2005. The sample period comprises a total of 27449 trading days, of which 6467 were during recessions. The t-stats reported are computed using White (1980) standard errors.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
Expansion, Monday/holidays, β_{11}	0.056	3.3	-0.062	-3.5	-0.079	-4.1
Expansion, back-to-back weekdays, β_{21}	0.015	1.9	-0.019	-2.1	-0.024	-2.6
Recession, Monday/holidays, β_{31}	0.188	4.0	-0.208	-3.8	-0.267	-4.9
Recession, back-to-back weekdays, β_{41}	0.062	2.8	-0.061	-2.3	-0.085	-3.1
	F-stat	p-value	F-stat	p-value	F-stat	p-value
$\beta_{11}=\beta_{21}$	4.7	0.031	4.8	0.029	6.6	0.010
$\beta_{31}=\beta_{41}$	6.2	0.012	6.0	0.014	10.0	0.002
$\beta_{11}=\beta_{31}$	6.9	0.008	6.2	0.013	10.5	0.001
$\beta_{21}=\beta_{41}$	4.0	0.046	2.3	0.128	4.5	0.034

Table 9
Media content and returns after the NYSE opening

The table reports the estimated coefficients $\boldsymbol{\beta}$ from the model

$$R_t = (1 - D_t)\left(\boldsymbol{\beta}_1 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_1 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_1 \mathcal{L}_s(R_t^2)\right) + D_t\left(\boldsymbol{\beta}_2 \mathcal{L}_s(M_t) + \boldsymbol{\gamma}_2 \mathcal{L}_s(R_t) + \boldsymbol{\psi}_2 \mathcal{L}_s(R_t^2)\right) + \boldsymbol{\eta} X_t + \epsilon_t.$$

where the dependent variable R_t is the log-return on the Dow Jones Industrial average from 11am until close for the period 1933–2005. The set of independent variables are described in Table 3. The sample period comprises a total of 19184 trading days, of which 2762 were during recessions.

	Positive		Negative		Pessimism	
	β	t-stat	β	t-stat	β	t-stat
$(1 - D_t) \times M_{t-1}$	-0.001	-0.1	-0.015	-2.3	-0.012	-1.8
$D_t \times M_{t-1}$	0.057	3.1	-0.047	-2.6	-0.065	-3.5
	F-stat	p-value	F-stat	p-value	F-stat	p-value
Test $\beta_{11} = \beta_{21}$	8.9	0.003	2.9	0.088	7.3	0.007

FINANCIAL MARKETS

Another Day of Great Activity, with Big Gains for the Industrials.

Expectations that yesterday's Stock Exchange session, sandwiched in between two holidays, would see much less activity were quickly disappointed. The market opened very active and strong, and after a small reaction in the forenoon resumed its upward trend with almost the same violence shown in the excited sessions of last week. The list was again irregular, but by far the larger number of stocks scored substantial gains and the upward movement of some of the war issues, which had been checked by banks and brokers who foresaw trouble if the advance were not held under control, was resumed with a great deal of vigor. The most striking gain among such issues was scored by Baldwin Locomotive, which, after hanging for several days around 115, returned yesterday to 127½, closing at 126, with a net advance of 11 points. This secondary stage of activity for Baldwin was accompanied by fresh merger rumors, which do not appear to have any substantial basis in fact. Even more active and relatively as strong was Westinghouse, of which more than 100,000 shares changed hands on a range of 5½ points. It closed at 138, with a gain of 4½ points above Saturday's close. The American Car & Foundry made a good recovery to 85⅛, and gains of from 2 to 5 points were numerous. The motor issues returned to popularity, all three classes of Maxwell stock advancing on the expectation of some kind of an announcement Wednesday of a plan looking to the payment of the accumulated dividend on the first preferred. Studebaker advanced 2¼ and General Motors 1 point.

The rails retained some of their momentum from last week, and most of the leaders sold at new high prices for the year. News of the note being prepared for dispatch to Great Britain was received too late to affect the market, if indeed such news can have any effect on the present temper of traders, and the list closed pretty close to the top.

Some uneasiness was caused yesterday by a new development of weakness in the foreign exchange market. Demand sterling sold down to $4.67⅜ compared with the low price of $4.68¼ on Saturday. The failure of the conclusion of the $500,000,000 Anglo-French loan to help foreign exchange rates gave special interest to an important meeting of bankers held yesterday afternoon, which was addressed by Sir Edward Holden, one of the visiting Commissioners.

Figure 1
Financial Markets column published in the New York Times on October 12th, 1915.

TOPICS IN WALL STREET.

American Munition Orders.

Until yesterday the stock market gave no indication that the war stocks derived a chance of profit from war with Mexico. To speculators in these shares it was in fact a matter of the keenest disappointment that they went down on war news. Over and over they have repeated the question: "What sort of a war stock is it that is depressed by a new war?" Yesterday an advance of 17 points in Bethlehem Steel held out a ray of hope and advances in most of the others on covering by professionals strengthened hopes that the next turn would be for the better. Officers of many of the munitions companies expected orders from the United States Government in the near future, but nowhere was it believed **that** these orders would be placed at terms permitting as great profits as those obtained in some of the contracts with the Allies.

* * *

The Extent of the Declines.

From the high point of week before last to the low point of last week, which was the low point of Friday's market, the average price of fifty representative stocks declined $5.33 a share. These stocks included many railroad shares in which the declines were small compared with losses in some of the speculative industrials. Reading, which lost 8⅝ points in this period, and Norfolk & Western, with a loss of 5¾, were the only rails to decline more than the 5 1-3-point average of the fifty. A score of industrials sustained greater losses and many of these losses ran into double figures, among them being: New York Air Brake, 11; Mexican Petroleum, 12⅞; Baldwin Locomotive, 13; United States Smelting, 13; Tennessee Copper, 14½; American Zinc, 14⅝; Willys-Overland, 16; Butte and Superior, 19¾; United States Industrial Alcohol, 26½. On the Curb Chevrolet Motors lost 46 points.

* * *

Now Up, Now Down.

It is interesting to note the change in sentiment that sweeps over the floor of the Stock Exchange after a pronounced rise, or sharp decline. Traders who have been bearish for weeks were turning bullish yesterday morning. They figured that the break which had been needed had been supplied, and that, therefore, stocks were a purchase again.

* * *

The Mexican Factor.

An old-time member said after the close that neither the Mexican war danger, nor the inadequacy of our war machinery, was really back of the slump which took place last week. Those arguments were advanced to support the decline, but in his opinion the break would have come had the Mexican situation continued unchanged. This man's theory is that the market had become badly congested with stocks, and had to be cleaned out by a return to lower prices. A number of pools were carrying large amounts of stock which they had not been able to market, and there were some large individual accounts that needed shaking out. The low prices made on Friday brought in a number of fresh buyers, and if this trader's theory works out the market will develop a much better tone this week, regardless of developments across the border. When the list grows stale nothing but a sharp setback will attract new money. That this market had become stale was evidenced by its utter disregard of good news, such as new and increased dividends.

* * *

No Extra Holiday.

When the brokers gave up their expected extra holiday before May 30, they looked for an extra day preceding the Fourth. The uncertainty of the political situation appears to have destroyed any chance of getting it. No petition has been circulated on the floor, and it is unlikely that the situation will clear in time to allow the drafting of one before the next meeting of Governors.

* * *

Bonds Have Idle Week.

The bond market suffered along with stocks last week, but without registering substantial declines. Bonds were effected more through a let-down of buying than from the liquidation of securities. Some of the banks and large dealers were reported as sellers of a considerable amount of bonds which they had been carrying for a month or more, and on which they had good profits. If this actually did take place the offerings were rather easily absorbed, and inquiries among bond men failed to show that there had been any urgent selling through fear that the Mexican situation might wipe out profits before they could be realized. The investment demand is believed to be widening, now that supplies from Europe have begun to fall away, leaving room for other offerings, and the bankers are inclined to think that business will pick up again with the coming of definite developments south of the border.

Figure 2
Topics in Wall Street column published in the New York Times on June 25th, 1916.

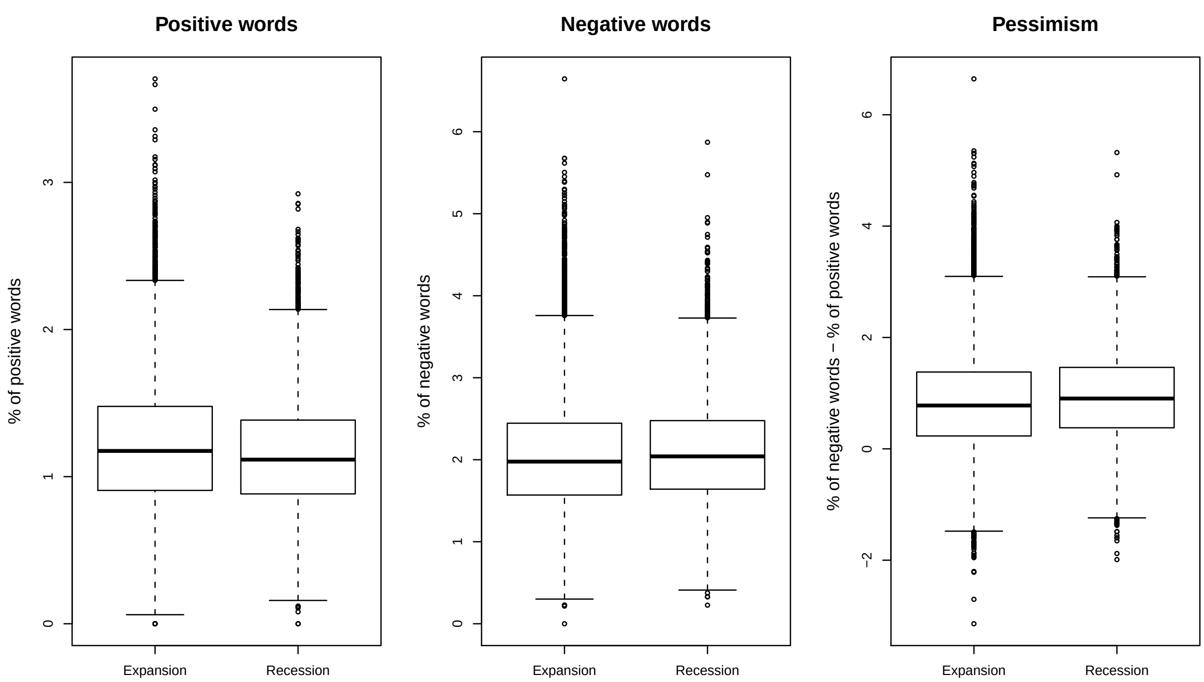

Figure 3
Boxplots of the media measures used in the paper, as a function of the business cycle. These measures are constructed from the columns “Financial Markets” and “Topics in Wall-Street” published in the New York Times in the period 1905–2005.

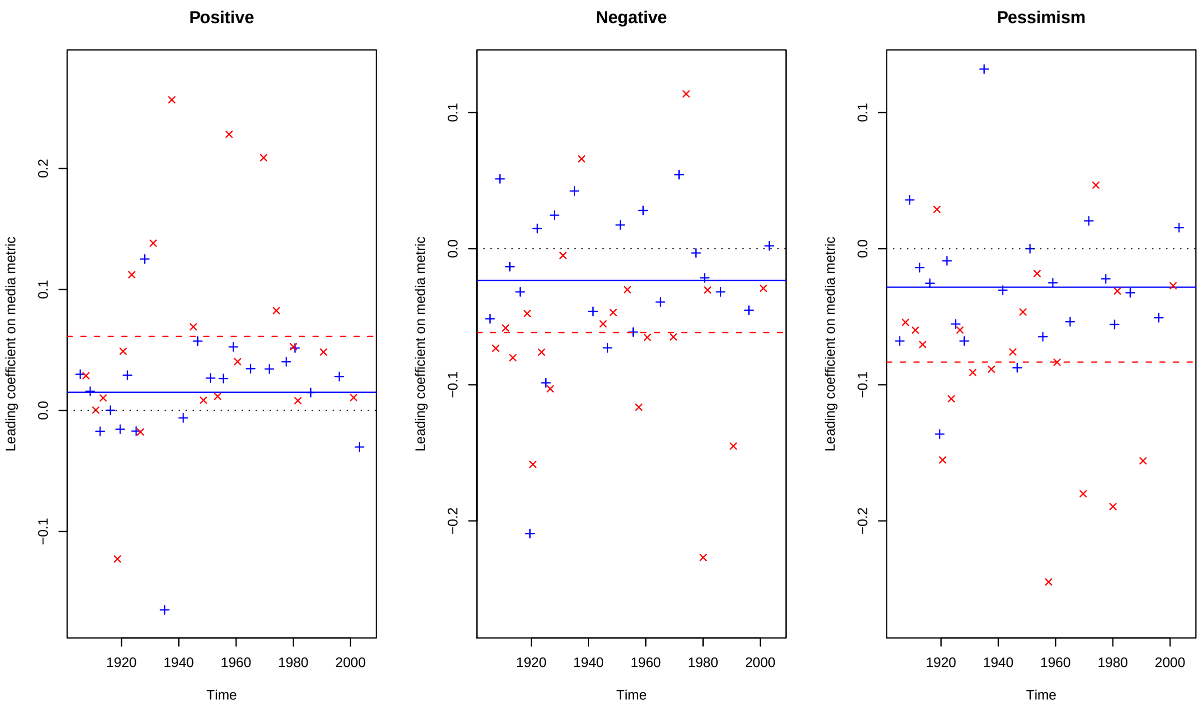

Figure 4
Plot of the coefficients on the media measures as predictors of stock returns for each business cycle during our sample period (1905–2005). The crosses corresponds to expansionary periods, according to the NBER, whereas the x's correspond to recessions. The dashed line is the time-series average of the estimates during recessions, whereas the solid line corresponds to expansions.

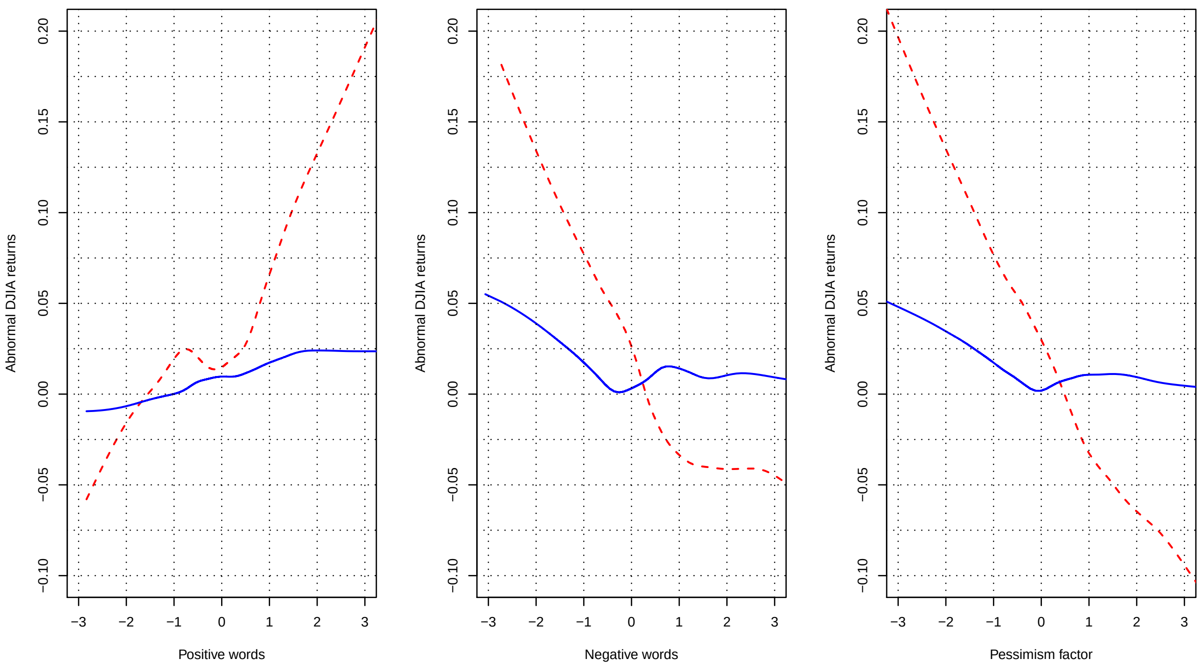

Figure 5
Non-parametric estimates of the conditional average DJIA returns as a function of the media measures constructed in the paper. The solid line is the estimate during expansionary periods, whereas the dashed line corresponds to NBER recessions.