NEWS AND STOCK PRICES: NEW INSIGHTS*

Jacob Boudoukh[a], Ronen Feldman[b], Shimon Kogan[c] and Matthew Richardson[d]

April 10, 2015

Abstract:

A basic tenet of rational finance is that asset prices change in response to unexpected fundamental information. This paper revisits this topic in a novel way. Using textual analysis, we are better able to identify fundamental information by parsing news into those that include unidentified events, identified events and complex events. Once news is correctly identified in this manner, there is considerably more evidence of a strong relationship between stock price changes and information. As applications, we revisit seminal facts from the literature, including an analysis of the relation between stock-return volatility and different types of news during trading and non-trading hours; an investigation of volatility persistence conditional on news types; and a reinvestigation of an infamous result that market model R^2s are the same on news and no news days.

*Corresponding author: Shimon Kogan, Arison Business School, IDC Herzliya, Tel: +972 (9) 966-2741, email skogan@idc.ac.il. A related version of the paper was previously circulated under the title "Which News Moves Stock Prices? A Textual Analysis." For comments and suggestions on that paper which were helpful for this paper, we would like to thank John Griffin and seminar participants at the University of Texas, Austin, NYU Stern School of Business, Wharton School of Business, University of Zurich, University of Lausanne, ESMT Berlin, the 2nd Luxembourg Asset Management Summit and the discussant Guido Baltussen, WFA 2013 and the discussant Paul Tetlock, and the *AQR Insight Award* finalists' panel.
[a] Arison School of Business, IDC Herzliya.
[b] School of Business Administration, The Hebrew University.
[c] Arison School of Business, IDC Herzliya and the University of Texas at Austin.
[d] Stern School of Business, New York University and NBER.

I. Introduction

A basic tenet of financial economics is that asset prices change in response to unexpected fundamental public information. Though early work, primarily through event studies, seemed to confirm this hypothesis [1], later empirical work has been much less supportive. Indeed, considerable attention in the finance literature has been devoted to deviations from this theory and, partly as a result, various fields of study, such as market microstructure, time-varying volatility and behavioral finance, have emerged.

Common to much of this literature, the proxy for public information has been news articles.[2] A problem with this proxy is that common news sources for companies, such as those in the Wall Street Journal stories and Dow Jones News Service, *et cetera*, release many stories that potentially contain very little relevant information about company fundamentals. The goal for the researcher is to be able to parse through which news stories are relevant and which are not. The researcher is faced with a massive computational problem since there are hundreds of thousands, possibly millions, of news stories to work through. Fortunately, advances in the area of textual analysis allow for better identification of relevant news. This paper employs one such approach based on an information extraction platform (Feldman, Rosenfeld, Bar-Haim and Fresko (2011), denote Feldman *et al*. (2011)). Specifically, for each news article, a time and date are recorded and relevant stock tickers' new stories are matched to a series of "identified" value-relevant events, or are deemed "unidentified". The dataset and relevant methodology are discussed and made available in an online appendix.[3]

---

[1] See, for example, Ball and Brown (1968) on earning announcements, Fama, Fisher, Jensen and Roll (1969)on stock splits, Mandelker (1974) on mergers, Aharony and Swary (1980) on dividend changes, and Asquith and Mullins (1986) on common stock issuance, among many others

[2] Roll (1988), Chan (2003) and Tetlock (2007).

[3] https://www.dropbox.com/s/oagtni23qcqmaqb/Appendix-Part%20I%2020150222.pdf?dl=0.

Identifying which news is relevant is important because a number of the seminal empirical results in the literature depend on showing that the distributional properties of stock prices are similar on news versus no news periods. In this paper, we add to this literature by examining stock return variation during trading and non-trading hours around specific types of news such as unidentified news (news with no identified, value-relevant, topic), identified news (news with an identified, value-relevant topic), and identified news with different degrees of complexity (to be defined precisely later on). Textual analysis allows us to identify which news is fundamental and this identification is key to our analysis. As a proof of concept, we document that stock-level volatility is similar on no-news days and unidentified news days, consistent with the idea that the intensity and importance of information arrival is the same across these days. This is economically important since over two thirds of all major-media outlet news is unidentified. In contrast, on identified news days, the volatility of stock prices is more than double that of other days.

Our paper applies the identification of news types to a reinvestigation of several major stylized facts in the literature on the relationship between stock prices and news. The bottom line from our analysis is that once news is correctly identified in this manner there is considerably more evidence of a strong relationship between stock price changes and the arrival of public information. Some of the main results are discussed below.

First, conditional on identification of relevant news, we duplicate some of the key analysis of French and Roll (1986). Notably, there is a large difference in these results if the researcher breaks the sample into trading versus non-trading hours (i.e., overnight). This is an important distinction because, like French and Roll (1986), this allows us to control for volatility induced via private information-driven trading. For example, we find that identified news during non-trading hours leads to variance ratios a magnitude higher than those of identified news during trading hours (i.e., 2.71 and 5.55 versus 1.59 and 2.11 respectively for identified and complex news). In addition, while the median daily volatility for returns during trading hours is 2.30%

versus 1.33% during non-trading hours (the French and Roll (1986) finding), for complex news, non-trading hours actually produce very similar volatility to the trading hours volatility, i.e., 2.72% versus 2.91%. These findings provide a contrast to conclusions reached by French and Roll (1986) and others.

Second, in the context of news identification, we explore the important stylized fact that volatility tends to cluster. Specifically, we reaffirm the volatility persistence literature using variance ratios by showing that conditional on large prices moves variance ratios tend to be high for many days afterwards (i.e., on day +1 to day +5, 1.91, 1.93, 1.84, 1.85, and 1.86). Remarkably however, conditional on these moves being driven by an identified news event, the variance ratios do not follow this pattern (i.e., on day +1 to +5, 1.19, 1.10, 1.07, 1.05 and 1.03). In fact, volatility is no longer persistent. This result is novel and suggests that standard results on generalized autoregressive conditional heteroskedasticity (GARCH) may need rethinking. In addition, we analyze this result in terms of the well-documented volatility-volume relation and document a surprising pattern. While an identified news day exhibits high volatility and a quick return to normal volatility the next day, the effect is amplified when also conditioning on high volume (i.e., on day +1 to +5, 0.85, 0.74, 0.72, 0.68 and 0.69). Not only is the volatility spike on the news day sharper, but also the ensuing decline.

Third, we revisit Roll's (1988) R^2 methodology and estimate the R^2 from a market model regression for no news days and for unidentified news days. Consistent with his results, R^2 levels are the same for no news days and for unidentified news days. However, when we estimate the same model over just identified news days, the R^2 drops dramatically from an overall median of 28% to 16% -- the result that Roll (1988) was originally looking for in his work but was unable to find. Moreover, by adding a measure of article tone (that is, positive versus negative) that builds on Tetlock (2007) and others, we show that the measure of tone increases the R^2 on identified news days, but not on unidentified news days, again consistent with the idea that identified news days contain price-relevant information.

Using a better proxy for relevant news, the main results of the paper put into question some of the better known stylized facts documenting the relation (or lack thereof) between news and stock prices. This is good news for efficient markets. When we can identify news, it matters. That said, the analysis goes further and shows that some of the existing anomalies are deepened once we separate out the identification of news. For example, a popular explanation for the large spread between variance ratios during trading and non-trading hours is the revelation of information through trading. This explanation has been offered for the surprisingly low R^2s on no news days (and, in our paper, also unidentified news days) of a regression of stock returns on multiple factors. We show that when one runs these regressions using close-to-open returns (i.e., during non-trading hours), and adjusting for identified news released during these hours, the puzzle still persists.

While our paper crosses into the area of the literature that focuses on using textual analysis to address the question of how prices are related to information[4], the three most closely related papers to ours, Griffin, Hirschey and Kelly (2011), Engle, Hansen and Lunde (2011) and Neuhierl, Scherbina and Schlusche (2013), actually lie outside this textual analysis research area. Griffin, Hirschey and Kelly (2011) cross-check global news stories against earnings announcements to try and uncover relevant events. Engle, Hansen and Lunde (2011) utilize the Dow Jones Intelligent Indexing product to match news and event types for a small set of (albeit large) firms. Neurhierl, Scherbina and Schlusche (2013) document significant stock price responses to a wide array of corporate press releases. While the focus of each of these papers is different, these papers provide some evidence that better information processing by researchers will lead to higher R^2s between prices and news.

This paper is organized as follows. Section II describes the data employed throughout the study. Of special interest, we describe the textual analysis methodology for identifying relevant news

[4] See, for example, Davis, Piger, and Sedor (2006), Tetlock (2007), Engelberg (2008), Tetlock, Saar Tsechansky and Macskassy (2008), Demers and Vega (2010), Feldman, Govindaraj, Livnat and Segal (2010), and Loughran and McDonald (2011), among others.

and lay out important stylized facts. Sections III and IV provide the main results of the paper, showing a strong relationship between prices and news, once the news is appropriately identified. In addition, some existing anomalies are deepened once we separate out the identification of news. Section V concludes.

II. Data Description: Stock Return Volatility and News Types

A. Textual Analysis

With the large increase in the amount of daily news content on companies over the past decade, it should be no surprise that the finance literature has turned to textual analysis as one way to understand how information both arrives to the marketplace and relates to stock prices of the relevant companies. Early work centered on document-level sentiment classification of news articles by employing pre-defined sentiment lexicons.[5] The earliest paper in finance that explores textual analysis is Antweiler and Frank (2005), who employ language algorithms to analyze internet stock message boards posted on "Yahoo Finance". Much of the finance literature, however, has focused on word counts based on dictionary-defined positive versus negative words.

For example, one of the best known papers is Tetlock (2007). Tetlock (2007) employs the General Inquirer, a well-known textual analysis program, alongside the Harvard-IV-4 dictionary to calculate the fraction of negative words in the *Abreast of the Market* Wall Street Journal column. A plethora of papers post Tetlock (2007) apply a similar methodology to measure the positive versus negative tone of news across a wide variety of finance and accounting

[5] See, for example, Lavrenko, Schmill, Lawrie, Ogilvie, Jensen, and Allan (2000), Das and Chen (2007) and Devitt and Ahmad (2007), among others. Feldman and Sanger (2006) provide an overview.

applications.[6] Loughran and McDonald (2011), in particular, is interesting because they refine IV-4 to more finance-centric definitions of positive and negative words.[7]

The focus of this paper is quite different. We use a rule-based information extraction approach to identify events relevant to companies, such as new product launches, lawsuits, analyst coverage, news on financial results, mergers, *et cetera*.[8] The initial list of events were chosen to match commercial providers such as CapitalIQ but were augmented by events likely to impact stock prices. This process led to a total of 14 event categories and 56 subcategories within events. The events fall into one of the following categories: *Analyst Recommendations*, *Financial*, *Financial Pattern*, *Acquisition*, *Deals*, *Employment*, *Product*, *Partnerships*, *Inside Purchase*, *Facilities*, *Legal*, *Award*, *Stock Price Change* and *Stock Price Change Pattern*. In terms of subcategories, consider for example the *Analyst Recommendation* category. It contains nine subcategories, including *analyst expectation*, *analyst opinion*, *analyst rating*, *analyst recommendation*, *credit - debt rating*, *fundamental analysis*, *price target*, etc.

A full listing of these events and a more detailed description of the event extraction methodology are described in the online appendix. The online appendix also includes links to the ticker-date dataset used in this paper (full day, open to close and the close to open).

[6] See, for example, Davis, Piger, and Sedor (2006), Engelberg (2008), Tetlock, Saar-Tsechansky and Macskassy (2008), Kothari, Li and Short (2009), Demers and Vega (2010), Feldman, Govindaraj, Livnat and Segal (2010), and Loughran and McDonald (2011), among others.

[7] More recently, an alternative approach to textual analysis in finance and accounting has been offered by Li (2010), Hanley and Hoberg (2011), Grob-Klubmann and Hautsch (2011) and Kogan, Routledge, Sagi and Smith (2011). These authors employ machine learning-based applications to decipher the tone and therefore the sentiment of news articles

[8] Feldman *et al*. (2011) employ a proprietary information extraction platform specific to financial companies, which they denote *The Stock Sonar* (TSS), and which is available on commercial platforms like Dow Jones. TSS extracts event instances and sentiment out of the text based on a set of predefined rules. The methodology is described in the online appendix.

B. Data Description and Summary

The primary dataset used in this paper consists of all documents that pass through the Dow Jones Newswire from January 1, 2000 to December 31, 2009. For computational reasons, and in order to minimize issues related to poor tradability, we limit ourselves to the S&P500 companies with at least 20 trading days during the period. Over the sample period, the dataset therefore includes at some time or another 791 companies. To avoid survivorship bias, we include in the analysis all stocks in the index as of the first trading day of each year. We obtain total daily returns from CRSP.

In order to ensure that the analysis does not suffer from a look-ahead bias, we use the article timestamp and line it up with the trading day. Specifically, we consider date t articles those that were released between 15:31 on date $t-1$ and 15:30 on date t. Date t returns are computed using closing prices on dates $t-1$ and t.[9] Articles released on non-trading days (weekends and holidays) are matched with the next available trading day. TSS methodology processes each article separately and generates an output file in which each article/stock/day is represented as an observation.

For each of the aforementioned observations, TSS reports the total number of words in the article, the number of relevant words in the article, and any possible identified events (and sub-events). A key feature of the methodology is its ability to differentiate between relevant news for companies (defined in our context as those related to specific firm events) as opposed to unidentified firm events. For each news story, therefore, our application of TSS produces a list of relevant events connected to this company and to this particular piece of news. It is possible that multiple events may be connected to a given story. In our analysis, we ignore the *Stock Price Change* and *Stock Price Change Pattern* categories as these categories do not, on their own,

[9] We also perform an analysis using open-to-close (trading day) and close-to-open (overnight) returns. For these returns, open-close news is defined as news arriving during trading hours and close-open news is defined as news arriving after trading hours.

represent fundamental news events. We also ignore *Award*, *Facilities*, and *Inside Purchase*, since these categories do not contain a sufficient number of observations.[10] We also merged *Financial* and *Financial Patter* and are therefore left with eight main categories.

To be more precise, our goal is to analyze the difference in return patterns based on the type of information arrival. We therefore classify each stock/day into one of three categories:

1. *No news* – observations without news coverage.
2. *Unidentified news* – observations for which none of the news coverage is identified by one of the eight categories.
3. *Identified news* – observations for which at least some of the news coverage is identified as being at least one of the above categories.

Moreover, conditional on being classified as *identified news*, we provide a further breakdown of identified news into a *complex news* category, defined in one of two ways:

1. *Event complexity – identified news* days with more than two event types (either categories or subcategories).
2. *Disagreement complexity – identified news* days in which at least one event has sentiment dispersion across articles.

In addition, we consider three periods covering news and returns:

1. *Close to close* - Daily returns and news.
2. *Open to close* - Trading day returns and news.
3. *Close to open* - Overnight returns and news.

[10] Including the Award, Facilities, and Inside Purchase categories does not alter the results.

It is, of course, possible that news lag the occurrence of the event mentioned in them. In that case, using the time-stamp on the news article as a proxy for when the event occurred would reduce our power to link information with price movements.

Table 1 provides an overview of the data. The first column in panel A reports the number of observations under each of the day classifications. First, we see that most days have no news coverage, i.e., 705,430 of 1,245,709 stock/day observations contain no news reported on the Dow Jones Newswire. Second, and most important, the vast majority of the days with news coverage, 380,420 of 540,279, do not have a single topic-identified news event. Moreover, most identified news days contain only a single-identified event (i.e., 122,666 of 159,829) although these days may include several subcategories under the event.[11] Third, we also observe that identified news days contain a larger number of articles compared with unidentified news days (6.1 vs. 2.6 per stock/day). While the number of words per article does not seem to vary much by day type, the number of relevant words (as identified by TSS) is much larger on identified news days (81 vs. 49). Finally, of the 159,829 relevant news days, only 49,574 are deemed complex, with a breakdown into the following categories: (1) 37,151 *event complexity* days, and (2) 32,856 *disagreement complexity* days.[12]

Panels B and C of Table 1 report similar statistics but now broken down between periods when the market is open versus closed. The most striking result in Panel B is that there is nothing striking. For the most part, the news coverage is similar during periods when the market is open versus the market being closed. For example, the ratio of news days - unidentified, identified and complex - to total number of days is respectively 25.6%, 8.6% and 2.3% for the open-to-close trading period versus 18.7%, 6.5%, and 1.8% for the overnight period. While this finding may have something to do with when news is reported, as opposed to when it takes place, it

[11] The appendix reports the same set of statistics by event type.

[12] Not shown in table 1, but in terms of co-occurrence across these complexity measures, *event complexity* has a correlation of 0.66 with *disagreement complexity*. One notable similarity between these measures, however, is that complex days contain many more articles, on average, than non-complex days.

nevertheless suggests a continual volume of news throughout a day, irrespective of whether trading takes place. This fact will be useful when the return distributions are compared across different types of news.

C. **Return Volatility and News**

Section II.B describes a wide variety of news types from unidentified to identified, and simple to complex. What differential impact does this news assortment have on the distributional properties of returns? Identifying which news is relevant is important because a number of the seminal empirical results in the literature depend on showing that the distributional properties of stock prices are similar on news versus no news periods. In this subsection, we focus on the literature documenting the properties of stock return variance ratios based on differential news, tied to some of the more important papers in empirical finance.

The first analysis we perform is a simple comparison of variance ratios of stock returns during periods with different amounts of relevant news, starting with French and Roll's (1986) highly cited paper. The motivation of that paper was to better understand whether volatility is caused by public information, private information revealed through trading, or pricing errors by investors. (For a theoretical discussion, see Black (1986), Admati and Pfleiderer (1988), Foster and Viswanathan (1990, 1993), Madhavan, Richardson and Roomans (1997), and the survey by Madhavan (2000), among others.[13])

As a first pass at the data, Table 2 provides a breakdown of news stories by the distribution of returns. In brief, the main result is that identified news days are more likely than unidentified news to lie in the negative and positive tails of the return distribution. On the surface, this is consistent with rational models, which would suggest that information arrival should be associated with increases in volatility.

[13] For assets other than equities, see Ederington and Lee (1993), Ito, Lyons and Melvin (1998), Boudoukh, Richardson, Shen and Whitelaw (2007).

In particular, if identified news days proxy for information arrival, then we should find that news arrival would be concentrated among days with large return movements, positive or negative. To relate news arrival intensity with returns, we assign daily returns into percentiles separately for each stock and year: bottom/top 10% (i.e., extreme 20% of returns), moderate 40% of return moves, and the smallest 40% return moves. We perform the assignment for each stock separately to control for cross-sectional variation in total return volatility, and perform the assignment for each year separately to control for large time-series variations in average return volatility, e.g., 2008-9. The columns in Table 2 group observations according to this split. For each of these columns, we compare the observed intensity of different day types to the intensity predicted under the null that these distributions are independent. For example, the null would suggest that of the 700 thousand no news days, 140 thousand would coincide with returns at the bottom and top 10%, 280 thousand would coincide with returns at the following 40%, and so forth. The results in each row report the difference between the observed intensity and the null in percentage terms.

Table 2A reports the results for close-to-close daily returns. Several observations are in order. First, we find that no news days are less concentrated among days with large price changes. In particular, they are 6.6% less likely to be extreme relative to the unconditional. This is consistent with the notion that news coverage proxies for information arrival. Interestingly though, we observe very little evidence of extreme price changes on news days when we cannot identify a specific event tied to the news: only 1.6% more than the expected fraction of our defined "extreme" days. This is an important finding in the context of this paper. Ex ante, one might have imagined that large price moves would have generated “news” stories, but this result shows that this does not occur.

Second, in sharp contrast to these results, we find that identified news days are 32.5% more likely to coincide with the extremes - the bottom 10% and top 10% of return days. Thus, while we might expect under independence to have 15,983 identified news stories in the extreme tails bucket, we actually observe 21,177 news stories in that bucket. That is, identified news days, but not unidentified news days, are much more likely to be extreme return days. Third, this pattern is much more pronounced for identified news days associated with one of our two measures of complexity, i.e., multiple events or disagreement across articles. Respectively, these days are 78.3% and 89.2% more likely to coincide with extreme returns days.

The results of Table 2A suggest that our textual analysis methodology will have similar success at linking identified events to stock return variation.[14] Therefore, as a more formal look at the data, we study the link between news arrival and volatility by computing daily return variations on no news days, unidentified news days, identified news days and complex news days. Specifically, for each stock we compute the average of squared daily returns on these day types. We then calculate the ratio of squared deviations on unidentified news days to no news days, and the ratio of squared deviations on different types of identified news days to no news days.[15] For example, if both unidentified and identified news days have no additional effect on stock volatility, then we should find that these ratios are distributed around one.

The last three columns of Table 2A report the distribution of these variance ratios. Consistent with the aforementioned Table 2A results, we find that the median variance ratio of unidentified news days is close to one (i.e., 1.20) while the variance ratio of identified news days exceeds two (i.e., 2.15). That is, the median stock exhibits return variance on identified news days that is 2.15 times the variance of no news days. The result appears quite robust, with over 90% of stocks exhibiting variance ratios exceeding one on identified news days. These results are much larger for complex days, with 3.72 and 3.69 times the variance ratio respectively for multiple event and disagreement days. As additional evidence, Figure 1 depicts the distribution of these ratios across the 672 stocks for which these ratios are available (out of 791), winsorized at 10.[16] As evident,

[14] Note that, while most researchers focus on Roll's (1988) R^2 result, Roll (1988) also provided evidence that kurtosis was higher on news versus no news days, a result similar to that provided in Table 2A.
[15] We include only stocks with at least 20 observations for all day classifications.
[16] Here again we eliminate stocks with insufficiently many observations in each day type, similarly to the footnote above.

the ratios are not distributed around one for neither unidentified nor identified news days. However, the difference in distributions between unidentified and identified news days' ratios is clear: the variance ratio is much higher on identified news days compared with unidentified news days.

III. Empirical Applications

The results of Section II.C clearly demonstrate that our news classification has power to distinguish between days on which price-relevant information arrives. This section uses this news classification to revisit some well-known stylized facts in stock price-news literature, including the price impact of public versus private information during trading and non-trading hours, the persistence of volatility through time and the R^2 of daily stock returns. For each of these examples, we document a new result, which basically shows large differences between identified versus unidentified news and puts into question existing economic interpretations contained in the literature.

A. Variance Ratios During Trading and Non-Trading Hours

In their seminal paper, French and Roll (1986) study variance ratios of stock returns during trading and non-trading periods. They document considerably more variability of returns using open-to-close returns than close-to-open returns both on an absolute and hourly basis. French and Roll (1986) explore three possible explanations. First, public information may arrive more frequently during trading hours. French and Roll provide evidence against this hypothesis by showing that volatility drops over weekday exchange holidays when presumably information is still flowing. Complementary to this finding, Table 1, Panels B and C of this paper show that identified, i.e., relevant, news seems to be generated similarly during trading and non-trading hours (i.e., 8.6%, 99,959 of 1,162,221, versus 6.5%, 75,162 of 1,162,221, respectively).

Second, appealing to behavioral finance, trading itself generates noise and higher volatility. Third, private information, not public information, is the primary source for volatility. That is, private information is gradually revealed through trading, thus generating higher volatility during trading hours. French and Roll conclude that the evidence favors the latter hypothesis and strongly supports private-information rational trading models. Using more complete data and additional real-world experiments, this conclusion has generally been confirmed by, among others, Barclay, Lizenberger and Warner (1990), Ito, Lyons and Melvin (1998), Barclay and Hendershott (2003), Madhavan, Richardson and Roomans (1997), and Chordia, Roll and Subrahmanyam (2011).

In contrast to these conclusions, Jones, Kaul, and Lipson (1994) and Jiang, Likitapiwat, and McInish (2012) compare variance ratios during informationally relevant non-trading periods to trading periods. Both papers find that significant volatility occurs in the absence of trading, providing evidence that public information is a major source of volatility. Consistent with these studies, in this subsection, we reexamine the results of Table 2, but now break the returns and news type data into close-to-open and open-to-close periods. Specifically, Table 2, Panels B and C compare variance ratios of stock returns on unidentified news, identified news, and complex news days to no news days, conditional on trading versus no trading hours.

With respect to the existing literature on stock return variances during trading versus non-trading hours, Table 2B and 2C confirms the stylized fact on variance ratios for S&P500 firms – the median daily volatility for returns during trading hours is 2.30% versus 1.33% during non-trading hours, that is, 76% higher. On the surface, this result is consistent with the conclusions in French and Roll (1986) and others that the major source for return volatility is not public information, but instead either private information revealed by trading or noise trading. This conclusion is further supported by two additional facts. First, return variances are relatively higher during trading days with no news, that is, on days in which there is no discernible public information. Specifically, during trading hours versus non-trading hours, median return volatility is respectively 2.22% versus 1.14%, that is, 95% higher on no-news days compared to 76% on all days. Second, even on days with news, those typically associated with public information, return variances are 72% higher during trading hours (i.e., 2.39% versus 1.39%) if the news is unidentified.

Tables 2 Panels B and C, however, reveal a different story when the news can be identified, and especially so when the identified news is complex. Specifically, on identified news days, the median trading day volatility is 2.89% versus 2.05% for non-trading volatility, in other words, only 41% higher on identified versus 72% for unidentified news days. Equally important, the identified news median volatility of 2.05% during non-trading hours is close to the magnitude of the volatility during trading hours on no-news days (i.e., 2.22%). This latter result is important for understanding the source of volatility and illustrates the importance of public versus private information in explaining return volatility. These results are even stronger for complex news. In particular, for complex news measured by either multiple events or disagreement across articles, non-trading hours actually produce very similar volatility to the trading hours volatility, i.e., 2.72% versus 2.91%, and 2.74% versus 2.97%, respectively. These results suggest that public information, when appropriately identified, is an important source of volatility. Because by construction no trading takes place during closing hours, the existing literature's focus on private information and noise trading cannot be the source.

A corollary of these findings relates to variance ratios of returns between various news types and no news during non-trading and trading hours. Specifically, during non-trading hours, the median variance ratio of returns on unidentified news, identified news and complex news days to no news days is 1.38, 2.71, 5.55 and 5.28, respectively. This contrasts to significantly lower variance ratios during trading hours, i.e., 1.14, 1.59, 2.11 and 2.13, respectively, for the various news types.

On the one hand, this result supports the idea that private information (or noise trading) is an important determinant of stock return volatility. This is because variance ratios are lower during trading hours when private information can be revealed through trading in contrast to non-trading hours. On the other hand, on identified and complex news days, the variances are 59% to 113% higher, even during trading hours. That is, when one can identify relevant information, public information clearly plays an important role in explaining stock return volatility. This finding is amplified during non-trading hours, namely, when there is by definition no trading, the stock return variances are 171% to 455% higher on identified news and complex news days relative to no news days.

B. **Volatility Persistence**

There is a large literature devoted to understanding the phenomenon of volatility clustering, including research on autoregressive conditional heteroskedasticity (ARCH) volatility (e.g., Engle (1982), Bollerslev (1986) and Nelson (1991), among others). One explanation for ARCH-like effects is a persistent news arrival process. Indeed, Nelson (1990) shows that discrete-time versions of continuous-time exponential ARCH models can be written in terms of Clark's (1973) mixture of distributions (MOD) model (see also Lamoureux and Lastrapes (1990)). The mixture of distributions model has been appealing to financial economists because it models the joint impact that news has on prices and volume. In these models, the number of information arrivals (the mixing variable) drives the conditional distribution of returns and volume (e.g., see Tauchen and Pitts (1983), Richardson and Smith (1994) and Anderson (1996)).

The intersection of ARCH and MOD models implies a relation between volume and volatility, and provides an explanation for persistence, all driven by information arriving to the market. It should not be surprising therefore that a variety of empirical analyses have investigated the relation between volume and volatility[17]. While the methods vary and measurement of variables differs across this literature, there is broad agreement that information arrival, including public

[17] See, for example, Tauchen and Pitts (1983), Karpoff (1987), Lamoureux and Lastrapes (1990), Gallant, Rossi and Tauchen (1992), Berry and Howe (1994), Jones, Kaul and Lipson (1994), Richardson and Smith (1994), Andersen (1996), Kalev, Liu, Pham and Jarnecic (2004), Maheu and Mccurdy (2004), Fleming, Kirby and Ostdiek (2006), Engle, Hansen and Lunde (2011) and Giot, Lauren and Petitjean (2012).

news, may be a driving force behind important stylized facts such as volatility persistence and positive correlation between volatility and volume.

Some recent papers (e.g., Maheu and Mccurdy (2004), Engle, Hansen and Lunde (2011) and Giot, Lauren and Petitjean (2012)) argue that the volatility process might be a combination of a continuous and a jump component, possibly driven by private and public information, respectively. In this subsection, we exploit our measure of news types – unidentified versus identified, simple versus complex *etc*., to better understand the volatility persistence properties of stock returns. As shown below, our results are novel and raise questions about current understanding of ARCH-like effects.

Table 3 documents variance ratios of both raw and excess returns on no news days conditional on a large stock price move taking place (measured by either a 0.5 or 1.0 standard deviation shock). Obviously, because we are conditioning on large moves, the variance ratios on the day of the shock are by construction large. For example, a one-standard deviation move is associated with a variance ratio of 3.50 and 2.89 for raw and excess returns respectively on a daily basis. The interesting stylized fact, however, is that these variance ratios persist on the following days, i.e., over the week (day +1 to day +5). Table 3 documents the pattern (1.91, 1.93, 1.84, 1.85, and 1.86) for raw returns and (1.94, 1.90, 1.85, 1.79 and 1.81) for excess returns, respectively. This continued persistence in future variance ratios is just another way of documenting volatility persistence and ARCH-like effects in return volatility. Similar patterns emerge for variance ratio persistence conditional on a breakdown between trading and non-trading hours.

The interesting, and novel, finding is presented in Table 4. Comparing unidentified to identified news days mirrors other analyses performed in this paper. There is a substantive difference in the properties of variance ratio persistence. Unidentified news days behave similarly to no news days. Over the week (day 0 to day +5), Table 4 documents the pattern (1.20, 1.20, 1.15, 1.14, 1.17 and 1.16) for raw returns and (1.22, 1.22, 1.19, 1.20, 1.19 and 1.19) for excess returns, respectively. In other words, unidentified news is associated with a 20% higher volatility and this volatility continues to persist throughout the week. In contrast, identified news days are surprisingly very different. In particular, over the week (day 0 to day +5), Table 4 documents the pattern (2.15, 1.19, 1.10, 1.07, 1.05 and 1.03) for raw returns and (2.60, 1.30, 1.14, 1.09, 1.07 and 1.05) for excess returns, respectively. Here, the identified news' impact on volatility drops precipitously after the first day and is for the most part gone a week later. Apparently, return shocks due to identified news versus unidentified news (or for that matter no news) have very different volatility persistence properties. This finding is consistent with the view expressed in Maheu and Mccurdy (2004), Engle, Hansen and Lunde (2011) and Giot, Lauren and Petitjean (2012) that the return volatility process is multivariate with the components having quite different properties.

Some comments are in order. First, in comparing Tables 3 and 4, a reasonable explanation is that identified news (i.e., tied to events) produces important public information relevant to stock prices, but that this information is for the most part fully revealed on the event day. Markets efficiently incorporate the news. However, stock price shocks on no-news days or unidentified news days are more consistent with the partial revelation of private information.

Second, in comparing trading versus non-trading hours, future variance ratios conditional on current identified news actually drop significantly below one during trading hours. That is, over the week (day 0 to day +5), Table 4 documents the pattern (1.56, 1.07, 0.98, 0.96, 0.93 and 0.94) for raw returns during trading hours. One possible explanation is as follows. On any given day, information gets revealed through the trading process (i.e., privately), producing excess volatility. Based on the high variance ratios, however, an identified news event reveals large amounts of information, diminishing the value and therefore trading of privately held information on days following identified news. This hypothesis is further supported by the fact that non-trading variance ratios hover around one on days following identified news.

Third, to investigate this hypothesis more fully, Table 5 documents variance ratio persistence for different news types conditional on high volume on the news day.[18] If volume is high on the actual news day (i.e., greater than one), then we postulate that it is more likely that all the information got revealed, either publicly or through trading. The patterns confirm this hypothesis. For example, over the week (day 0 to day +5), Table 5 documents a variance-ratio pattern of 1.93, 0.85, 0.74, 0.72, 0.68 and 0.69, respectively for raw returns during trading hours. The drop in the variance ratios is startling, suggesting variances a few days after identified public information has been released are 30% lower than they are on any no news days.

The above analysis supports both the importance of public and private information. On the one hand, variances are much higher on days with publicly, identifiable information. On the other hand, variances on the following days are lower relative to a typical day, presumably because private information is more prevalent on typical days when no relevant information has recently been publicly revealed. The fact that this is amplified when conditioning the news days on high volume further supports this view.

Given the varied behavior of variance ratio patterns for different news types and across non-trading and trading hours, a follow-up analysis would suggest some documentation on the persistence of news types. Table 6 documents daily transition probabilities unconditionally and conditional on either no news, unidentified news or identified news. As documented previously in Table 1, the unconditional probability of no news, unidentified news and identified news respectively is 57%, 30% and 13%. Table 6 shows strong persistence in news types. Specifically, the probability of a news type conditional on the previous day being of that news type is much higher than the unconditional probabilities, i.e., 72%, 45% and 34% for no news, unidentified news and identified news, respectively. At least for identified news, the results provide somewhat of a puzzle in light of the aforementioned results presented in Tables 4 and 5 regarding variance ratio persistence. Even though company relevant news seems to follow other

[18] High volume occurs when the ratio of day t volume relative to the average volume during the trailing 252 days of volume exceeds 1 (see Jarrell and Poulsen (1989) and Barber and Odean (2008)).

relevant news, the large variance ratios we observe on these news days do not follow each other. This suggests that variance ratios likely drop precipitously if no additional news comes out in future days after an identifiable event.

C. R^2

How much of the variation in stocks prices is due to fundamental information has been an important topic in both the theoretical and empirical finance literature for the last quarter century (e.g., see seminal papers by French and Roll (1986), Black (1986), Admati and Pfleiderer (1988), Roll (1988) and Cutler, Poterba and Summers (1989)). In the above section, we added to this literature by examining stock return variation around specific types of news such as unidentified news, identified news, and identified news with different degrees of complexity. In this section, we explore a related angle by duplicating some of the analysis of Roll (1988) to help understand the relation between news and returns.

A seminal paper on the question of whether stock prices reflect fundamental information is Roll (1988). In that paper, Roll (1988) argues that once aggregate effects have been removed from a given stock, the finance paradigm would imply that the remaining variation of firm returns would be idiosyncratic to that firm. As a proxy for this firm-specific information, Roll (1988) uses news stories generated in the financial press. His argument is that, on days without news, idiosyncratic information is low and the R^2s from aggregate level regressions should be much higher than on days with news. Roll (1988) documents two important findings: (i) there is no discernible difference in R^2s from aggregate factor regressions between news versus no news days, and (ii) the R^2s from these regressions are in general quite low. Thus, his conclusion is that it is difficult to understand the level of stock return variation. Working off this result, a number of other papers reach similar conclusions with respect to prices and news, in particular, Cutler, Poterba and Summers (1989), and Mitchell and Mulherin (1994).

The evidence that asset prices do not reflect seemingly relevant information is not just found with equity returns. For example, Roll (1984)'s finding that, in the frozen concentrated orange juice (FCOJ) futures market, weather surprises explain only a small amount of variability of futures returns has been a beacon for the behavioral finance and economics literature. Given that weather has theoretically the most important impact on FCOJ supply, and is the focus of the majority of news stories, Roll (1984) concludes, similarly to his 1988 paper, that there are large amounts of "inexplicable price volatility". These papers by Roll (1984, 1988) have had a profound impact on the literature in terms of their understanding of the relation between asset prices and information. As some illustrations, in a heavily cited survey paper, Hirshleifer (2001), writes "little of stock price or orange juice futures price variability has been explained empirically by relevant public news," (p.1560); in his well-known book on behavioral finance, Shleifer (2000) writes "(Roll) finds the average R^2s to be only 0.35 with monthly data and 0.20 with daily data, suggesting that movements in prices of individual stocks are largely unaccounted for by public news…"; and Hong and Stein (2003) write "Roll (1984, 1988) and French and Roll (1986) demonstrate in various ways that it is hard to explain asset price movements with tangible public information" (p.487).

In contrast to Roll's (1984) FCOJ paper, Boudoukh, Richardson, Shen and Whitelaw (2007) show that when (i) the fundamental is identified -- in this case temperatures close to or below freezing, when relevant path dependencies are taken into consideration, e.g., first freeze versus second, third etc. -- and (ii) the highly nonlinear relation between weather and returns is accounted for, there is a fairly close relationship between prices and weather surprises. In this section, we make a similar argument to Boudoukh, Richardson, Shen and Whitelaw (2007). We parse out news stories into identified versus unidentified events and reevaluate Roll's (1988) finding and conclusion.

In a different context, and using a different methodology, Griffin, Hirschey and Kelly (2011) and Engle, Hansen and Lunde (2011) also provide evidence that price volatility can be partially explained by news. For example, by cross-checking global news stories against earnings announcements to try and uncover relevant events, Griffin, Hirschey and Kelly (2011) document that better information extraction can lead to higher R^2s between prices and news. Engle, Hansen and Lunde (2011) utilize the *Dow Jones Intelligent Indexing* product to match news and event types for a small set of (albeit large) firms, and show that the arrival of this public information has explanatory power for the dynamics of volatility.

The documented result in Table 2, namely, that variances are higher on days in which we can identify important events and on days with "complex" news, supports a relation between prices and fundamentals. As a more formal analysis, we reproduce the aforementioned Roll (1988) analysis for our setting. Table 7 reports results for a reinvestigation of the R^2 analysis of Roll (1988). Specifically, we estimate a one-factor pricing model and a four-factor pricing model separately for each firm and for each day classification: all, no news, unidentified news, identified and identified complex news.[19] We repeat the same analysis at the 2-digit SIC industry classification thereby imposing a single beta for all firms within a given industry and utilizing weighted least squared regressions. All R^2 are adjusted for the number of degrees of freedom.

The results in the top part of Table 7 report median R^2 across firms (column 2) and industries (column 5). Consider the median calculations for the CAPM model at the firm level. The R^2s are similar on no news and unidentified news days (i.e., 33.3% vs. 30.3%). The magnitude of the R^2s and similarity of these numbers between no news and news days (albeit unidentified) are consistent with Roll's puzzling results. However, R^2s are much lower on identified news day, i.e., 15.9%. The difference in R^2 between identified news and no-news days is striking – the ratio of median R^2 between identified news and no-news days is 2.1, in sharp contrast to Roll's results. Similar to the results from Table 2 with respect to variance ratios, the results in Table 7 are more pronounced on complex days, with R^2s even lower, i.e., 10.2% on days with more than two events, and 10.7% on days with sentiment disagreement across articles.

---

[19] We impose a minimum of 40 observations to estimate the regressions.

Roll's original theory-based conjecture, dramatically refuted empirically by his 1988 work, was that the performance of a market model, as measured by R^2, should be much worse during days on which firm-specific information arrives, compared with days when no such information arrives. In contrast to Roll's results, our results do lend support to this conjecture, since we are able to better proxy for firm-specific information arrival days using event identification.

Our results appear to be robust to the pricing model and firm/industry specification. For example, the results are analogous for the four-factor model that, along with the market, includes the book-to-market, size and momentum factors. In particular, the ratio of median R^2 between no-news and identified news days is only slightly lower (1.96 versus 2.10), and the R^2s between no-news and unidentified days is again similar. All these results change only barely when we perform the analysis at an industry level in which we constrain the betas against the 1- or 4-factor models to be the same within industry. Constraining the betas allows greater degrees of freedom for subsequent analysis when we try and understand the source of the differences between the R^2s of no-news versus unidentified days.

Assuming that firm-specific, identifiable news days are unrelated to aggregate factor return movements, Table 7 clearly demonstrates that, in contrast to a significant portion of the literature that claims otherwise, stock price movements can be related to public information. In other words, when news can be identified, news does in fact matter. Of course, Roll's (1988) second puzzle still persists. Irrespective of using a 1-factor or 4-factor model, R^2s on days with no apparent, relevant news are low, namely 33.3% and 38.6%, respectively. That is, two-thirds of the volatility needs to be explained by private information revelation or publicly revealed information not worthy of news articles.

It seems worthwhile delving more deeply into this question. In the subsection below, we ask two novel questions: (i) does an estimate of the sentiment of the news, coupled with the exact event identifier, help further explain the R^2s, and (ii) is there a large difference between the regression

R^2s when we break the sample into non-trading and trading hours, motivated by the fact that non-trading hours allow us to remove the private information hypothesis as one possibility?

As a brief preview, we find that, in contrast to our conclusion about the importance of public information, the second Roll (1988) puzzle deepens, at least with respect to the portion of unexplained variability. For example, while there is a significant link between a sentiment score and the unexplained variation from factor model regressions, the additional R^2 is not overwhelming. In addition, and more puzzling, we show that regression R^2s using close-to-open returns (i.e., during non trading hours) are not higher than R^2s using open-to-close returns on no news or unidentified news days. Given that there is no trading during the former hours, and no "relevant" news, what can possibly explain the firm's stock price movements?

i. R^2 and Sentiment Measurement

One of the main applications of textual analysis in finance has been to link sentiment scores to both contemporaneous and future stock returns. The evidence is statistically significant albeit weak in magnitude. For example, Tetlock (2007) and Tetlock, Saar-Tsechansky and Macskassy (2008), show that negative word counts of news stories about firms based on IV-4 have contemporaneous and forecast power for the firms' stock returns, though the R^2s are low. Loughran and McDonald (2011) argue that for a finance context the Harvard dictionary is not appropriate and build a sentiment score using a more finance-centric dictionary. Their application focuses on creating a dictionary appropriate for understanding the sentiment contained in 10-K reports. For their 10-K application, sentiment scores based on word counts from this alternative dictionary generally provide a better fit.

In this section, we extend the above analysis on news versus no news R^2s to include sentiment scores.[20] In the above analysis, we showed that identified news days are a good proxy for

[20] The TSS methodology that identifies events for each news article also produces a sentiment score attached to each article. The sentiment score is based on a positive and negative word count, along the lines of Tetlock (2007) and

information arrival. Below, we show that the sentiment of these articles, i.e., the directional content of this information, has some explanatory power for returns. Table 7 shows that market model regressions on news days have low R^2, that is, most of the variation of stock returns is idiosyncratic in nature. A reasonable hypothesis is that the R^2s should increase if idiosyncratic information is incorporated directly. We use the sentiment score as our proxy for this information.

To see the additional explanatory power of event-specific scores, consider the results of the last three columns of Table 7. The R^2s reported in the table are adjusted R^2s derived from industry regressions. We augment the one-factor or four-factor models with event-level scores, utilizing weighted least squared regressions estimated on the 2-digit SIC industry level (with at least 80 observations). That is, we assume that all firms within the industry have the same return response magnitude to a given event type but we allow this magnitude to vary across events and industries. Focusing on identified event days, we see that daily scores increase R^2 from a median of 16.0% to 17.8% under the one-factor model, and from 18.2% to 19.6% under the four-factor model. Most important, these increases are attained only for identified news days. In contrast, for unidentified news days, there is no increase in R^2s when sentiment scores are taken into account. In other words, to link stock prices to information, it is necessary to measure both the news event and the tone (i.e., sentiment) of this news. In order to investigate this further, we also report R^2s for complex, identified news days. The results are even stronger. For example, for days with multiple events and disagreement across articles, the increase in R^2s is 44% (i.e., from 9.1% to 13.1%), and 74% (i.e., from 7.5% to 13.1%), respectively.

Though the percentage increases are impressive, the R^2s are still quite low relative to no news days. Of course, these low R^2s may well be due to the strong assumption that *all* stocks within a given industry have the same coefficient on sentiment for *all* stories. Or it might be the case that the sentiment measures are poor proxies for the true content of the public information. It is

---

Loughran and MacDonald (2011). A full description of the particular scoring methodology of TSS is provided in the online appendix.

possible therefore that a more sophisticated model of sentiment could substantially increase the R^2s. Nevertheless, one of the puzzles of Roll (1988) still remains unexplained. On days in which there is no news on the Dow Jones wire, either identified or unidentified, the market (or four factor regression) still only explain 33.3% (38.6%) of the variation. This suggests a behavioral explanation or considerable stock return variation due to private information being impounded in prices via the trading process.

To better understand the behavioral implications of the relation between identified news types and stock returns, in the next subsection we try to partially differentiate the behavioral from the private information explanation by repeating the R^2 analysis for non-trading and trading hours. This analysis is novel to the literature and arguably deepens the excess volatility puzzle.

ii. R^2 During Trading and Non-Trading Hours

As described above, a popular explanation for the large spread between variance ratios during trading and non-trading hours is the revelation of information through trading. This explanation has been offered for the surprisingly low R^2s on no news days (and, in our paper, unidentified news days) of a regression of stock returns on multiple factors. In order to evaluate this explanation further, we run factor regressions using trading day returns and non-trading day returns, conditional on various types. These results, analogous to Table 7 albeit broken up by trading versus nontrading, are reported in Table 8, Panels A and B.

On the one hand, the results strongly support the hypothesis that when important public information is identified, this information matters for stock prices. During closing hours, that is, when no trading takes place, R^2s for identified news, complex events, and complex disagreement are 11.3%, 19.1% and 12.5% respectively compared to, during trading hours, 17.1%, 14.0% and 16.2%. That is, when we isolate to a period with highly relevant public information without either private information trading or noise trading taking place, the explanatory power of aggregate factors dramatically disappears.

Note though that, on the other hand, the results also deepen the behavioralist view that there is a large amount of unexplained stock price variability. During closing hours, when private information revelation cannot be a source for unexplained variability, conditioning on either no news or unidentified news, R^2s are only 25.2% and 22.7% respectively. More important, these R^2s are actually slightly lower than the R^2s of 26.4% and 23.8% during trading hours. This latter result fine-tunes and deepens the puzzle for rational pricing. The results in this paper show that relevant, public information is important for explaining stock price variability. The problem is that once that information is accounted for, and by construction we move away from a trading or volume-based explanation, it is not clear what rational possibilities remain.

VI. Conclusions

The bottom line from this paper is in contrast to the last 25 years of literature on stock prices and news. We find that, when information can be identified, there is a much closer link between stock prices and information. Examples of results include market model R^2s that are no longer the same on news versus no news days (i.e., Roll's (1988) infamous result), but now are 16% versus 33%; variance ratios of returns on identified news days more than double those on no news and unidentified news days, and even more so during nontrading hours; and, volatility persistence properties that are considerably different, conditional on this relevant information.

The paper, however, documents variance ratio patterns and market model R^2s that in some way deepen the excess volatility puzzle described and analyzed in the literature. The information identifier methodology described in this paper may be useful for a deeper analysis of the relation between stock prices and information, especially on the behavioral side. For example, there is a large literature that looks at stock return predictability and reversals/continuation of returns depending on under-reaction or over-reaction to news (see, for example, Hirshleifer (2000), Chan (2003), Vega (2006), Gutierrez and Kelley (2008), Tetlock, Tsaar-Tsechansky, and Macskassy (2008), and Tetlock (2010)). This paper allows the researcher to segment this news into categories likely to lead to under- or over-reaction. Moreover, there is a vast literature in the behavioral finance area arguing that economic agents, one by one, and even in the aggregate, cannot digest the full economic impact of news quickly. Given our database of identified events, it is possible to measure and investigate “complexity”, and its effect on the speed of information processing by the market. For example, “complexity” can be broken down into whether more than one economic event occurs at a given point in time, how news (even similar news) gets accumulated through time, and cross-firm effects of news. We hope to explore some of these ideas in future research.

References

Admati, A. and P. Pfleiderer, 1988, A theory of Intraday Patterns: Volume and Price Variability, *Review of Financial Studies* 1, 3–40.

Aharony, J., and I. Swary, 1980, Quarterly Dividend And Earnings Announcements And Stockholders' Returns: An Empirical Analysis, *Journal of Finance* 35, 1–12.

Andersen, T. G., 1996, Return Volatility and Trading Volume: An Information Flow Interpretation of Stochastic Volatility, *Journal of Finance*, 51, 169-204.

Antweiler, W., and M. Z. Frank, 2005, Is All That Talk Just Noise? The Information Content Of Internet Stock Message Boards, *Journal of Finance* 59, 1259–1293.

Asquith, P., and D. W. Mullins, 1986, Equity Issues And Offering Dilution, *Journal of Financial Economics* 15, 61–89.

Ball, R., and P. Brown, 1968, An Empirical Evaluation Of Accounting Income Numbers, *Journal of Accounting Research* 6, 159–178.

Barber, B. M., and T. Odean, 2008, All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, *The Review of Financial Studies* 21, 785–818.

Barclay, M.J., Litzenberger, R.H., and J.B. Warner, 1990, Private information, trading volume, and stock-return variances, *Review of Financial Studies*, 3, 233-254.

Barclay, M., and T. Hendeshott, 2003, Price Discovery and Trading After Hours, *the Review of Financial Studies*, 16(4), 1041-1073.

Berry, T. D., and K. M. Howe, 1994, Public Information Arrival, *Journal of Finance* 49, 1331–1346.

Black, F., 1986, Noise, *The Journal of Finance* 41, 529–543.

Bollerslev, T., 1986, Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307-327.

Boudoukh, J., M. Richardson, Y. Shen and R. Whitelaw, 2007, Do Asset Prices Reflect Fundamentals? Freshly Squeezed Evidence from the FCOJ Market, *Journal of Financial Economics* 83, 397–412.

Chan, W. S., 2003, Stock Price Reaction To News And No-News: Drift And Reversal After Headlines, *Journal of Financial Economics* 70, 223–260.

Chordia, T., R. Roll and A. Subrahmanyam, 2011, Recent Trends in Trading Activity and Market Quality, *Journal of Financial Economics*, 101/2, 243263.

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

Das, S. R., and M. Y. Chen, 2007, Yahoo! For Amazon: Sentiment Extraction From Small Talk On The Web, *Management Science* 53, 1375–1388.

Davis, A. K., J. Piger, and L. M. Sedor, 2012, Beyond The Numbers: An Analysis Of Optimistic And Pessimistic Language In Earnings Press Releases, *Contemporary Accounting Research* 29, 845–868.

Demers, E., and C. Vega, 2010, Soft Information In Earnings Announcements: News Or Noise?, *Working Paper, INSEAD.*

Devitt, A., and K. Ahmad, 2007, Sentiment Polarity Identification In Financial News: A Cohesion-Based Approach, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, 984–991.

Engelberg, J. E. , 2008, Costly Information Processing: Evidence From Earnings Announcements, *Working Paper, University of North Carolina.*

Engelberg, J. E., A. V. Reed, and M. C. Ringgenberg, 2012, How Are Shorts Informed? Short Sellers, News, and Information Processing, *Journal of Financial Economics* 105, 260–278.

Ederington, L. and J. H. Lee, 1993, How Markets Process Information: News Releases and Volatility, *Journal of Finance*, 48, 1161-1191.

Engle, R. F., 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica*, 50, 987-1008.

Engle, R. F, M. Hansen, and A. Lunde, 2011, And Now, The Rest Of The News: Volatility And Firm Specific News Arrival, *Working Paper.*

Fama, E. F., L. Fisher, M. C. Jensen, and R. Roll, 1969, The Adjustment Of Stock Prices To New Information, *International Economic Review* 10, 1–21.

Feldman, R., S. Govindaraj, J. Livnat, and B. Segal, 2010, Managements Tone Change, Post Earnings Announcement Drift And Accruals, *Review of Accounting Studies* 15, 915–953.

Feldman, R., B. Rosenfeld, R. Bar-Haim, and M. Fresko, 2011, The Stock Sonar Sentiment Analysis Of Stocks Based On A Hybrid Approach, *Proceedings of the Twenty-Third Innovative Applications of Artificial Intelligence Conference*, 1642–1647.

Feldman, R., and J. Sanger, 2006, The Text Mining Handbook, *Cambridge University Press.*

Fleming, J., C. Kirby and B. Ostdiek, 2006, Stochastic Volatility, Trading Volume, and the Daily Flow of Information, *Journal of Business*, 79/3, 1551-1590.

Foster, F., and S. Viswanathan, 1990, A theory of intraday variations in volumes, variances and trading costs in securities markets, *Review of Financial Studies*, 3, 593-624.

Foster, F. and S. Viswanathan, 1993, The effect of public information and competition on trading volume and price volatility, *Review of Financial Studies*, 6 (1): 23-56.

French, K. R., and R. Roll, 1986, The Arrival of Information and the Reaction of Traders, *Journal of Financial Economics* 17, 5–26.

French, K. R., and R. Roll, 1986, Stock Return Variances: the Arrival of Information and Reaction of Traders, *Journal of Financial Economics* 17, 5–26

Gallant, R., P. Rossi, and G. Tauchen, 1992, Stock prices and volume, *Review of Financial Studies*, 5, 199-242.

Giot, P., S. Laurent, and M. Petitjean, 2010, Trading activity, realized volatility and jumps, *Journal of Empirical Finance*, 17(1), 168-175.

Griffin, J. M., N. H. Hirschey, and P. J. Kelly, 2011, How Important Is The Financial Media In Global Markets?, *Review of Financial Studies* 24, 3941–3992.

Grob-Klubmann, A., and N. Hautsch, 2011, When Machines Read The News: Using Automated Text Analytics To Quantify High Frequency News-Implied Market Reactions, *Journal of Empirical Finance* 18, 321–340.

Gutierrez, R. C., and E. K. Kelley, 2008, The Long-Lasting Momentum In Weekly Returns, *Journal of Finance* 63, 415–447.

Hanley, K. W., and G. Hoberg, 2012, Litigation Risk, Strategic Disclosure And The Underpricing

Of Initial Public Offerings, *Journal of Financial Economics* 103, 235–254.

Hirshleifer, D., 2001, Investor Psychology and Asset Pricing, *The Journal of Finance*, 56, 1533–1597.

Hong, H. and J. Stein, 2003, Differences of Opinion, ShortSales Constraints, and Market Crashes, *Review of Financial Studies*, 16 (2): 487-525.

Ito, T., R. Lyons and M. Melvin, 1998, Is There Private Information in the FX Market? The Tokyo Experiment, *journal of Finance*, 53: 1111-1130.

Jarrell, G.A. and A. B. Poulsen, 1989, The Returns to Acquiring Firms in Tender Offers: Evidence from Three Decades, *Financial Management*, 18, 12-19.

Jiang, C., T. Likitapiwat, and T. McInish, 2012, Information Content of Earnings Announcements: Evidence from After-Hours Trading, *Journal of Financial and Quantitative Analysis*, Volume 47, pp 1303-1330.

Jones, C.M., Kaul, G., and M.L. Lipson, 1994. Information, trading, and volatility, *Journal of Financial Economics*, 36, 127-154.

Kalev, P., W. Liu, P. Pham, and E. Jarnecic , 2004, Public information arrival and volatility of intraday stock returns, *Journal of Banking and Finance*, 28(6), 1441-1467.

Karpoff, J., 1987, The Relation between Price Changes and Trading Volume: A Survey, *Journal of Financial and Quantitative Analysis*, 22(1), 109-125.

Kogan, S., Routledge, B. R., Sagi, J. S., and N. A. Smith, 2011, Information Content of Public Firm Disclosures and the Sarbanes-Oxley Act, *Working Paper*.

Kothari, S. P., X. Li, and J. E. Short, 2009, The Effect Of Disclosures By Management, Analysts, And Financial Press On The Equity Cost Of Capital: A Study Using Content Analysis, *Accounting Review* 84, 1639–1670.

Lamoureux, C., and W. Lastrapes, 1990, Heteroskedasticity in stock return data: volume versus GARCH effects, *Journal of Finance*, 45, 220-229.

Lavrenko, V., M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, 2000, Mining Of Concurrent Text And Time Series, *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 37–44.

Li, F., 2010, The Information Content Of Forward-Looking Statements In Corporate Filings A Naive Bayesian Machine Learning Approach, *Journal of Accounting Research* 48, 1049–1102.

Loughran, T., and B. McDonald, 2011, When Is A Liability Not A Liability? Textual Analysis, Dictionaries, And 10-Ks, *Journal of Finance* 66, 35–65.

Madhavan, A., 2000, Market microstructure: A survey, *Journal of Financial Markets*, 3/3, 205258.

Madhavan, A., Richardson, M., and M. Roomans, 1997, Why do security prices fluctuate? a transaction-level analysis of NYSE stocks, *Review of Financial Studies*, 10, 1035-1064.

Maheu, J. and T. McCurdy, 2004, News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns, *The Journal of Finance*, 59/2, 755793.

Mandelker, G., 1974, Risk And Return: The Case Of Merging Firms, *Journal of Financial Economics* 1, 303–335.

Mitchell, M. L., and J. H. Mulherin, 1994, The Impact Of Public Information On The Stock Market, *Journal of Finance* 49, 923–950.

Nelson, D. B., 1990, ARCH Models as Diffusion Approximations, *Journal of Econometrics*, 45, 7-38.

Nelson, D. B., 1991, Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347-370.

Neuhierl, A., A. Scherbina, and B. Schlusche, 2013, Market Reaction to Corporate Press Releases, *Journal of Financial and Quantitative Analysis*, forthcoming.

Richardson, M. and T. Smith, 1994, A Direct Test of the Mixture of Distributions Hypothesis: Measuring the Daily Flow of Information, *Journal of Financial and Quantitative Analysis*, 29, 101-116.

Roll, R., 1984, Orange Juice and Weather, *American Economic Review* 74, 5, 861–880.

Roll, R., 1988, R2, *Journal of Finance* 43, 541–566.

Shleifer, A., 2000, Inefficient Markets: An Introduction to Behavioral Finance, Oxford University Press, Oxford.

Tauchen, G. E. and M. Pitts, 1983, The price variability volume relationship on speculative markets, *Econometrica*, 51, 485-505.

Tetlock, P. C., 2007, Giving Content To Investor Sentiment: The Role Of Media In The Stock Market, *Journal of Finance* 62, 1139–1168.

Tetlock, P. C., 2010, Does Public Financial News Resolve Asymmetric Information?, *Review of Financial Studies* 23, 3520–3557.

Tetlock, P. C., M. Saar-Tsechansky, and S. Macskassy, 2008, More Than Words: Quantifying Language To Measure Firms' Fundamentals, *Journal of Finance* 63, 1437–1467.

Vega, C., 2006, Stock Price Reaction To Public and Private Information, *Journal of Financial Economics* 82, 103–133.

Tables

Table 1: Summary Statistics

Panel A: Close-Close

	# Obs.	# Tickers	# Articles (daily)	# Words (per art.)	# Relv. Words (per art.)
Total	1,245,709	791	3.6	325	58
No News	705,430	790	NA	NA	NA
Unid News	380,450	791	2.6	329	49
Iden News	159,829	790	6.1	316	81
Event Comp	37,151	740	11.7	320	78
Disag Comp	32,856	738	11.4	315	78

Panel B: Close-Open

	# Obs.	# Tickers	# Articles (daily)	# Words (per art.)	# Relv. Words (per art.)
Total	1,162,221	745	2.9	323	52
No News	765,278	745	NA	NA	NA
Unid News	296,984	745	2.2	329	44
Iden News	99,959	744	4.9	305	77
Event Comp	10,273	554	10.9	310	86
Disag Comp	16,182	654	9.6	309	84

Panel C: Open-Close

	# Obs.	# Tickers	# Articles (daily)	# Words (per art.)	# Relv. Words (per art.)
Total	1,162,221	745	2.4	337	65
No News	870,147	745	NA	NA	NA
Unid News	216,912	743	1.9	339	54
Iden News	75,162	730	4.0	333	97
Event Comp	7,831	539	9.1	329	88
Disag Comp	12,855	636	8.3	318	82

The table reports summary statics on the number of tickerdate observations, the number of unique tickers, the average number of articles, words per article and relevant words per article. No News days are days on which no news appeared, Unidentified News days are days on which news appeared but did not contain a corporate event, Event Complexity days are Identified News days on which more than two different corporate events (or sub-events) appeared, and Disagreement Complexity days are Identified News days on which at least one corporate event was covered in two different articles with different sentiment. Panel A includes tickerdate definitions based on a close-close window, Panel B includes tickerdate definitions based on a Close-Open window, and Panel C includes tickerdate definitions based on an Open-Close window.

Table 2: Event Frequency Across Return Ranks and Variances

Panel A: Close-Close

	Return Rank			Stock SD and Variance		
	20% Extreme	40% Moderate	40% Low	Med SD	N Tickers	Var Ratio
Total	1.0%	-0.6%	0.1%	2.64	791	1.16**
No News	-6.6%	0.5%	2.8%	2.38	781	
Unid News	1.6%	-0.4%	-0.4%	2.66	764	1.20**
Iden News	32.5%	-5.5%	-10.7%	3.51	681	2.15**
Event Comp	78.3%	-14.1%	-25.0%	4.53	401	3.72**
Disag Comp	89.2%	-16.3%	-28.3%	4.55	391	3.69**

Panel B: Open-Close

	Return Rank			Stock SD and Variance		
	20% Extreme	40% Moderate	40% Low	Med SD	N Tickers	Var Ratio
Total	1.0%	-0.5%	0.1%	2.30	745	1.04**
No News	-3.0%	0.0%	1.5%	2.22	745	
Unid News	5.4%	-1.2%	-1.5%	2.39	658	1.14**
Iden News	34.4%	-5.5%	-11.7%	2.89	546	1.59**
Event Comp	97.0%	-17.8%	-30.8%	2.91	119	2.11**
Disag Comp	89.3%	-13.5%	-31.2%	2.97	212	2.13**

Panel C: Close-Open

	Return Rank			Stock SD and Variance		
	20% Extreme	40% Moderate	40% Low	Med SD	N Tickers	Var Ratio
Total	1.0%	-0.7%	0.2%	1.33	745	1.22**
No News	-4.6%	0.4%	2.0%	1.14	739	
Unid News	4.9%	-1.5%	-1.0%	1.39	704	1.38**
Iden News	32.2%	-6.6%	-9.5%	2.05	585	2.71**
Event Comp	97.7%	-20.7%	-28.2%	2.72	126	5.55**
Disag Comp	85.5%	-18.7%	-24.0%	2.74	225	5.28**

The first three columns of the tables report the difference between the observed distribution of observations and that predicted under independence. We assign daily returns into percentiles separately for each stock and year: bottom/top 10% (i.e., extreme 20% of returns), moderate 40% of return moves, and the smallest 40% return moves. For each of these columns, we compare the observed intensity of different day types to the intensity predicted under the null that these distributions are independent. The next three columns report the median standard deviation (per day type), the number of unique tickers, and the median variance ratio (across tickers), i.e., the median ratio (across firms) of squared return deviations on each day type divided by the squared deviations on no news days. For a description of day types, see Table 1. **(*) denote p-values lower than 5% (10%) obtained from a non-parametric test of the null that the median variance ratio is equal to one.

Table 3: Variance Ratios on large-move No News days

	Cutoff	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
				Raw returns			
Close-close	0.5 SD	1.89**	1.46**	1.47**	1.44**	1.44**	1.43**
	1.0 SD	3.50**	1.91**	1.93**	1.84**	1.85**	1.86**
Close-open	0.5 SD	2.39**	1.59**	1.60**	1.55**	1.57**	1.57**
	1.0 SD	4.76**	2.05**	2.08**	1.98**	2.08**	2.02**
Open-close	0.5 SD	1.89**	1.32**	1.33**	1.32**	1.33**	1.32**
	1.0 SD	3.49**	1.77**	1.74**	1.70**	1.71**	1.70**
				Excess returns			
Close-close	0.5 SD	1.68**	1.51**	1.50**	1.47**	1.47**	1.47**
	1.0 SD	2.89**	1.94**	1.90**	1.85**	1.79**	1.81**
Close-open	0.5 SD	1.71**	1.47**	1.47**	1.44**	1.43**	1.44**
	1.0 SD	2.85**	1.98**	1.89**	1.84**	1.83**	1.84**
Open-close	0.5 SD	1.64**	1.34**	1.33**	1.30**	1.31**	1.30**
	1.0 SD	2.82**	1.78**	1.68**	1.66**	1.65**	1.64**

The table reports the variance ratios on No News days with large moves, moves greater than 0.5 (1.0) of daily standard deviation, and 1-5 days following these days. Daily standard deviations thresholds are calculated per stocks. Variance ratios are calculated for each stock by dividing the observed squared return deviation and the squared return deviations on all No News days, per stock. The median variance ratio across stocks is reported. For a description of day types, see Table 1. **(*) denote p-values lower than 5% (10%) obtained from a non-parametric test of the null that the median variance ratio is equal to one.

Table 4: Variance Ratios on Lagged Day Types

	Day type	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
			Raw returns				
Close-close	Unid News	1.20**	1.20**	1.15**	1.14**	1.17**	1.16**
	Iden News	2.15**	1.19**	1.10**	1.07**	1.05*	1.03**
Close-open	Unid News	1.42**	1.22**	1.25**	1.24**	1.25**	1.20**
	Iden News	2.80**	1.13**	1.08**	1.03	1.02	1.01
Open-close	Unid News	1.13**	1.05**	1.02	1.00	1.02	1.00
	Iden News	1.56**	1.07**	0.98	0.96**	0.93**	0.94**
			Excess returns				
Close-close	Unid News	1.22**	1.22**	1.19**	1.20**	1.19**	1.19**
	Iden News	2.60**	1.30**	1.14**	1.09**	1.07**	1.05**
Close-open	Unid News	1.19**	1.13**	1.11**	1.10**	1.11**	1.09**
	Iden News	1.86**	1.04*	1.01	0.97	0.99	0.99
Open-close	Unid News	1.16**	1.05**	1.04**	1.02	1.02	1.02
	Iden News	1.70**	1.13**	1.01	0.96*	0.95**	0.95**

The table reports the variance ratios on Unidentified and Identified News days and 1-5 days following these days. Variance ratios are calculated for each stock by dividing the observed squared return deviation and the squared return deviations on all No News days, per stock. The median variance ratio across stocks is reported. For a description of day types, see Table 1. **(*) denote p-values lower than 5% (10%) obtained from a non-parametric test of the null that the median variance ratio is equal to one.

Table 5: Variance Ratios on High Volume

	Day type	t	$t+1$	$t+2$	$t+3$	$t+4$	$t+5$
				Raw returns			
Close-close	Unid News	1.16**	0.94**	0.84**	0.83**	0.84**	0.82**
	Iden News	1.93**	0.85**	0.74**	0.72**	0.68*	0.69**
Close-open	Unid News	1.32**	1.02**	0.92**	0.88**	0.89**	0.84**
	Iden News	2.53**	0.82**	0.74**	0.71	0.66	0.67
Open-close	Unid News	1.07**	0.82**	0.78	0.73	0.74	0.72
	Iden News	1.35**	0.78**	0.67	0.64**	0.63**	0.63**
				Excess returns			
Close-close	Unid News	1.19**	0.94**	0.85**	0.83**	0.82**	0.80**
	Iden News	2.37**	0.92**	0.74**	0.70**	0.68**	0.67**
Close-open	Unid News	1.15**	0.97**	0.89**	0.87**	0.85**	0.83**
	Iden News	1.68**	0.78*	0.77	0.73	0.70	0.71
Open-close	Unid News	1.11**	0.82**	0.76**	0.73	0.73	0.71
	Iden News	1.44**	0.84**	0.69	0.65*	0.65**	0.64**

The table reports median variance ratios for a given day type, across firms, relative to the variance on No News days at time 0. Only observations with high volume are used for calculation. We define day t volume as high volume if it exceeds the average volume during the trailing 252 days (for the same stock). Excess returns are defined as raw returns minus the CRSP Value weighted index return. For a description of day types, see Table 1. **(*) denote p-values lower than 5% (10%) obtained from a non-parametric test of the null that the median variance ratio is equal to one.

Table 6: Transition Probability of Day Types

	No News	Unid News	Iden News	N
		Close-Close		
No News	72%	21%	7%	704,914
Unid News	40%	45%	15%	380,256
Iden News	28%	38%	34%	159,748
Unconditional	57%	31%	13%	1,244,918
		Close-Open		
No News	77%	18%	5%	764,630
Unid News	48%	40%	12%	296,776
Iden News	36%	39%	24%	99,877
Unconditional	66%	26%	9%	1,161,283
		Open-Close		
No News	83%	13%	5%	869,424
Unid News	52%	37%	11%	216,746
Iden News	49%	34%	17%	75,113
Unconditional	75%	19%	6%	1,161,283

The table reports the probability of transitioning from a given day type to another day type, separately for news arrival measured over the close-close period, close-open period, and open-close period. For a description of day types, see Table 1.

Table 7: R^2 regressions – Close-Close

	Returns – Firm Level			Returns – Industry Level			Score&Returns – Industry Level		
	Med R^2	Ratio (No News)	N	Med R^2	Ratio (No News)	N	Med R^2	Ratio (No Score)	N
				Single Factor Regressions					
Total	27.8%	1.19**	791	27.5%	1.20**	60	27.9%	1.01**	60
No News	33.3%	1.00	774	33.0%	1.00	60	33.0%	1.00	60
Unid News	30.3%	1.10**	721	29.3%	1.13**	58	29.4%	1.00**	58
Iden News	15.9%	2.10**	597	16.0%	2.07**	55	17.8%	1.11**	55
Event Comp	10.2%	3.25**	259	9.1%	3.65**	43	13.1%	1.44**	43
Disag Comp	10.7%	3.10**	264	7.5%	4.38**	44	13.1%	1.74**	44
				Four Factor Regressions					
Total	32.7%	1.18**	791	30.4%	1.20**	60	30.8%	1.01**	60
No News	38.6%	1.00	774	36.5%	1.00	60	36.5%	1.00	60
Unid News	35.9%	1.08**	721	33.0%	1.11**	58	33.0%	1.00**	58
Iden News	19.6%	1.96**	597	18.2%	2.01**	55	19.6%	1.08**	55
Event Comp	14.4%	2.68**	259	12.0%	3.03**	43	16.1%	1.33**	43
Disag Comp	14.9%	2.59**	264	9.0%	4.08**	44	14.7%	1.65**	44

The table reports results from firm and industry level return regressions, across a number of different specifications. In all regressions, the dependent variable is time t raw returns of the firm or the two-digit SIC code industry return. In the top panel, labeled "Single Factor Regressions", we use the CRSP Value Weighted Returns as the independent variable. In the bottom panel, labeled "Four Factor Regressions", we use the Fama-French 4 factors (including momentum) as the independent variables. Columns 1-3 report results from firm-level regressions, columns 4-6 report results from industry-level regressions, and columns 7-9 report results from industry-level regressions where the event-level sentiment scores are added as independent variables. The values reported in the table are the median R^2s, across stocks or industries, the ratio of the median R^2 relative to the R^2 on no-news days, and the number of stocks/industries. The ratio in column 8 reports the median R^2 in the regressions that include sentiment scores to the same regressions without sentiment scores (column 5). For a description of day types, see Table 1. **(*) denote p-values lower than 5% (10%) obtained from a non-parametric test of the null that the median variance ratio is equal to one.

Table 8: R^2s – Firm-level Regressions

Panel A: Close-Open

	Returns – Firm Level			Score&Returns – Firm Level		
	Med R^2	Ratio (No News)	N	Med R^2	Ratio (No Score)	N
	Single Factor Regressions					
Total	25.1%	1.05**	745	25.2%	1.05**	745
No News	26.4%	1.00	744	26.4%	1.00	744
Unid News	23.8%	1.11**	613	23.9%	1.11**	613
Iden News	17.0%	1.55**	440	17.4%	1.52**	440
Event Comp	14.0%	1.89**	30	15.5%	1.70**	30
Disag Comp	16.2%	1.63**	80	16.3%	1.62**	80

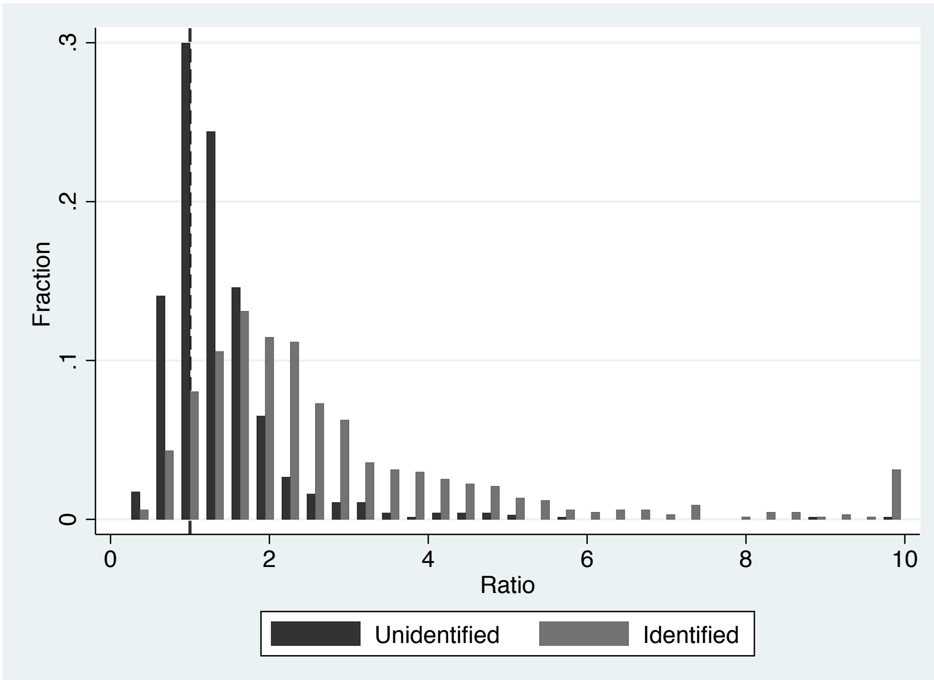
Panel B: Open-Close

	Returns – Firm Level			Score&Returns – Firm Level		
	Med R^2	Ratio (No News)	N	Med R^2	Ratio (No Score)	N
	Single Factor Regressions					
Total	20.3%	1.24**	745	20.9%	1.21**	745
No News	25.2%	1.00	734	25.2%	1.00	734
Unid News	22.7%	1.11**	648	23.5%	1.07**	648
Iden News	11.3%	2.23**	501	12.4%	2.03**	501
Event Comp	19.1%	1.32**	47	20.9%	1.21**	47
Disag Comp	12.5%	2.02**	104	14.8%	1.71**	104

The table reports results from firm level return regressions, across a number of different specifications. In all regressions, the dependent variable is time t firm return. Panel A reports the results for open-close regressions while Panel B reports the results for close-open regressions. The independent variable in all regressions is the SPY returns calculated over the relevant window. Columns 4-6 add daily sentiment scores as an independent variable. The values reported in the table are the median R^2s, across stocks, and the ratio of the median R^2 relative to the R^2 on no-news days, and the number of observations. For a description of day types, see Table 1. **(*) denote p-values lower than 5% (10%) obtained from a non-parametric test of the null that the median variance ratio is equal to one.

Figures

Figure 1: Distribution of Variance Ratios



The figure depicts the distribution of variance ratios, calculated within stocks, of unidentified and identified news days over no news days. Ratios are winsorized at 10. For a description of day types, see Table 1.