JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 54, No. 5, Oct. 2019, pp. 1937–1974

doi:10.1017/S0022109019000127

Does Unusual News Forecast Market Stress?

Paul Glasserman and Harry Mamaysky*

Abstract

An increase in "unusual" news with negative sentiment predicts an increase in stock market volatility. Unusual positive news forecasts lower volatility. Our analysis is based on over 360,000 articles on 50 large financial companies, published during the period of 1996–2014. Unusualness interacted with sentiment forecasts company-specific and aggregate volatility several months ahead. Furthermore, unusual news is reflected more slowly in aggregate volatility than company-specific volatility. News measures from articles explicitly about the "market," which are more easily accessible to investors, do not forecast volatility. The observed responses of volatility to news may be explained by attention constraints on investors.

I. Introduction

Can the content of news articles forecast market stress and, if so, what type of content is predictive? Several studies have documented that news sentiment forecasts market returns. We find that a measure of "unusualness" of news text combined with sentiment forecasts stress, which we proxy by stock market volatility. The effects we find play out over months, whereas in most prior work, the stock market's response to news articles dissipates in a few days. We also find that unusual news is reflected in volatility more slowly at the aggregate level than at

*Glasserman, pg20@columbia.edu, Columbia University Business School and Office of Financial Research (OFR); Mamaysky, hm2646@columbia.edu, Columbia University Business School. This paper was produced while Glasserman was a consultant to the OFR. We thank the anonymous referees and Jennifer Conrad (the editor) for very helpful comments. We acknowledge the excellent research assistance of Il Doo Lee. We thank Geert Bekaert, Kent Daniel, Tara Sinclair, Paul Tetlock, and seminar participants at the Summer 2015 Consortium for Systemic Risk Analytics conference, Columbia University, the Office of Financial Research, the High Frequency Finance and Analytics conference at the Stevens Institute, the IAQF/Thalesians seminar, the Imperial College London Quantitative Finance Seminar, the Princeton Quant Trading Conference, the Columbia–Bloomberg Machine Learning in Finance Workshop, BNY Mellon's Machine Learning Day, the 2016 Philadelphia Fed Conference on Real-Time Data Analysis, the 2016 SIAM Financial Mathematics Conference, the 2017 Society for Quantitative Analysts Conference, and the 2018 Cornell Tech Symposium for valuable comments. We thank the Thomson Reuters Corp. for graciously providing the data that was used in this study. We use the Natural Language Toolkit in Python for all text processing applications in the paper. For empirical analysis we use the **R** programming language for statistical computing.

the company-specific level, and we explore the causes and implications of this difference.

The link between sentiment expressed in public documents and stock market returns has received a great deal of attention. At an aggregate level, Tetlock (2007) finds that negative sentiment in the news depresses returns. Tetlock, Saar-Tsechansky, and Macskassy (2008), using company-specific news stories and responses, show this relationship also holds at the individual firm level. Garcia (2013) finds that the influence of news sentiment is concentrated in recessions. Loughran and McDonald (2011) and Jegadeesh and Wu (2013) apply sentiment analysis to 10-K filings. Da, Engelberg, and Gao (2014) measure sentiment in Internet search terms. Manela and Moreira (2017) find that a news-based measure of uncertainty forecasts returns. Our focus differs from prior work because we seek to forecast market stress rather than the direction of the market.[1] We document important differences between aggregate and company-specific responses to news, a novel finding that suggests greater efficiency at the micro level compared to the macro level. We apply new tools to this analysis, going beyond sentiment word counts.

The importance of unusualness is illustrated by the following two phrases, both of which appeared in news articles from Sept. 2008:

"the collapse of Lehman,"

"cut its price target."

Both phrases contain one negative word and would therefore contribute equally to an overall measure of negative sentiment in a standard word-counting analysis. But we recognize the first phrase as much more unusual than the second, relative to earlier news stories. This difference can be quantified by taking into account the frequency of occurrence of the phrases in prior months. As this simple example suggests, we find that sentiment is important, but it becomes more informative when interacted with our measure of unusualness.

Research in finance and economics has commonly measured sentiment through what is known in the natural language processing literature as a bag-of-words approach: An article is classified as having positive or negative sentiment based on the frequency of positive or negative connotation words that it contains. As the example above indicates, this approach misses important information. The unusualness of the first phrase lies not in its use of "collapse" or "Lehman" but in their juxtaposition. We therefore measure unusualness of consecutive word phrases rather than individual words.[2]

Our analysis uses all news articles in the Thomson Reuters Corp. database between Jan. 1996 and Dec. 2014 that mention any of the top 50 global banks, insurance, and real estate firms by market capitalization as of Feb. 2015. After some cleaning of the data, we are left with 367,331 articles about these 50 firms,

[1]Kogan, Levin, Routledge, Sagi, and Smith (2009) employ support vector regressions to forecast year-ahead volatility of returns using text data from companies' annual reports. Kogan, Routledge, Sagi, and Smith (2011) interpret the goodness of fit of this technique in different subsamples as a measure of the effectiveness of the Sarbanes–Oxley Act of 2002.

[2]Routledge, Sacchetto, and Smith (2013) also use word phrases (from 10-K's) to forecast M&A activity.

with an average of 1,611 per month. We calculate measures of sentiment and unusualness from these news stories and study their ability to forecast realized or implied volatility at the company-specific and aggregate levels.

The consistent picture that emerges from this analysis is that the interaction of unusualness with sentiment yields the best predictor of future stock market volatility among the news measures we study.[3] Importantly, our analysis shows that news is not absorbed by the market instantaneously. We also find that the information in news articles relevant for future company-specific volatility is better reflected in firm-level option prices and realized volatility than macro-relevant information is reflected in the prices of S&P 500 options and S&P 500 realized volatility.

In simple forecasting regressions of company-specific implied volatility on lagged company-specific news measures, our interacted measure of unusual negative news, ENTSENT_NEG, provides a statistically and economically significant predictor of volatility.[4] To control for known predictors of volatility (as documented, e.g., in Poon and Granger (2003), Bekaert and Hoerova (2014)), we include lagged values of implied and realized volatilities and negative returns as explanatory variables in panel regressions of company-specific volatility measures on company-specific news measures. Even with the inclusion of the controls, our interacted measures of sentiment (both positive and negative) and unusualness remain economically and statistically significant at lags of up to 2 months, with positive measures forecasting a decrease in volatility and negative measures forecasting an increase. These results indicate that the information in our news measures is not fully reflected in contemporaneous option prices.

For our aggregate analysis, we extract aggregate measures of unusualness and sentiment from our full set of news articles. We estimate vector autoregressions (VARs), taking as state variables the VIX, realized volatility on the S&P 500, and several aggregate news measures. We examine interactions among the variables through impulse response functions. A shock to either negative sentiment or our interacted variable ENTSENT_NEG produces a statistically significant increase in both implied and realized volatility over several months. Once again, the effect is strongest for our interacted measure of unusual negative news. The response of implied and realized volatility to an impulse in ENTSENT_NEG (or negative sentiment, SENTNEG) is hump-shaped, peaking at around 4 months, and remaining significant even after 10 months. We find similar results (but forecasting a decrease in volatility) for the interaction between positive sentiment and unusualness, though the effects are stronger for negative news than positive news. These patterns suggest that the information in our aggregate news variables is absorbed slowly.

As a test of the economic impact of these effects, we modify a simple S&P 500 put selling strategy to take into account the information contained in our

[3]Loughran and McDonald (2011) similarly find that sentiment provides a stronger signal when they put greater weight on less frequently occurring words. However, the empirical word weights in Jegadeesh and Wu (2013) are only weakly related to word frequency.

[4]Negative sentiment and unusualness are also significant separately, but much less so.

aggregate news measures. We find that this augmented strategy meaningfully outperforms the baseline strategy, and especially so during the financial crisis.

To compare our macro results with micro results, we estimate a panel VAR for the corresponding company-specific variables; implied and realized volatility, positive and negative sentiment measures, and news measures interacted with unusualness. Impulse response functions again show that a shock to unusual negative (positive) news produces a statistically significant increase (decrease) in both implied and realized volatility. However, the responses now peak more quickly. When compared with our aggregate impulse response functions, these results indicate greater market efficiency at the micro level rather than the macro level.

Our results thus pose two questions: Why is volatility-relevant information in news not absorbed more quickly, and why is it absorbed more slowly at the aggregate level than at the company-specific level? We discuss potential answers, putting particular emphasis on the limits of investor attention. The theory of rational inattention, as developed in Sims (2003), (2015) and Maćkowiak and Wiederholt (2009), includes predictions of hump-shaped responses to news of the type we observe in our VAR analysis. A tighter constraint on information processing capacity leads to a slower response. Our results therefore suggest limits on the ability of investors to process all the information in news sources, and a greater ability to interpret company-specific articles directly than to extract aggregate information from these articles.

As further evidence of the link to inattention, we note that the aggregate signal we extract from news would only have been available to investors who read thousands of news stories about individual companies. In contrast, we conjecture that aggregate sentiment measures using only articles that are explicitly about the overall market should be more accessible to investors and thus carry less predictive value. To test for this, we construct aggregate sentiment measures using only articles that are tagged by Thomson Reuters as being about the overall market, and not about individual stocks.[5] Consistent with our conjecture, we find that explicitly market-specific news does not forecast volatility in our VAR analysis.

The contrast that we observe between single-name and aggregate responses is relevant to a related literature on micro and macro efficiency. The attention capacity constraint faced by investors forces them to specialize; even among professionals, many investors may focus on a narrow set of stocks or industries and may overlook information that becomes relevant only when aggregated over many stocks. Jung and Shiller (2005) review empirical evidence supporting what they call Samuelson's dictum, that the stock market is micro efficient but macro inefficient. The allocation of attention between idiosyncratic and aggregate information by capacity constrained agents is examined in the models of Maćkowiak and Wiederholt (2009), Peng and Xiong (2006), Kacperczyk, Van Nieuwerburgh, and Veldkamp (2016), and Glasserman and Mamaysky (2018). Maćkowiak and Wiederholt (with regard to firms) and Glasserman and Mamaysky (with regard to investors) find that economic agents favor micro over macro information.[6]

[5]We thank a referee for suggesting this test to us.

[6]Peng and Xiong (2006) reach the opposite conclusion when looking at the information choice of a representative investor.

The equilibrium effect of this, as shown in Glasserman and Mamaysky, is to make prices more informationally efficient with regard to micro than macro information.

Because the aggregate effects we find in the data play out over months, the signals we extract from news articles are potentially useful for risk monitoring purposes. Along these lines, Baker, Bloom, and Davis (2016) develop an index of economic policy uncertainty based on newspaper articles. Indicators of systemic risk (see Bisias, Flood, Lo, and Valavanis (2012)) are generally based on market prices or lagged economic data; incorporating news analysis offers a potential direction for improved monitoring of stress to the financial system. The performance of our simple S&P 500 put selling strategy modified to take into account the information contained in our aggregate news measures shows the potential of this idea.

From a methodological perspective, our work applies two ideas from the field of natural language processing to text analysis in finance. As already noted, we measure the "unusualness" of language, and we do this through a measure of entropy in word counts. Also, we take consecutive strings of words (denoted as n-grams) rather than individual words as our basic unit of analysis. In particular, we calculate the unusualness (entropy) of consecutive 4-word sequences. These ideas are developed in greater detail in Jurafsky and Martin (2008). See Das (2014) and Loughran and McDonald (2016) for overviews of text analysis in finance.

The rest of this paper is organized as follows: Section II introduces the methodology we use, and Section III discusses the data and presents some summary statistics. Section IV presents results based on company-specific volatility, and Section V examines aggregate volatility. Section VI discusses possible explanations of our results and develops the market-specific news test. Section VII concludes. Supplementary Material presents more detailed results and several robustness checks, including a subperiod analysis excluding the financial crisis.

II. Methodology

A. Unusualness of Language

A text is unusual if it has low probability, thus, measuring unusualness requires a model of the probability of language. This problem has been studied in the natural language processing literature on word prediction. Jurafsky and Martin (2008), a very thorough reference for the techniques we employ in this paper, gives the following example: What word is likely to follow the phrase *please turn your homework* …? Possibly it could be *in* or *over*, but a word like *the* is very unlikely. A reasonable language model should give a value for

$$P(\text{in} \,|\, \text{please turn your homework})$$

that is relatively high, and a value for

$$P(\text{the} \,|\, \text{please turn your homework})$$

that is close to 0. One way to estimate these probabilities is to count the number of times that *in* or *the* has followed the phrase *please turn your homework* in a large body of relevant text.

An *n-gram* is a sequence of n words or, more precisely, n tokens.[7] Models that compute these types of probabilities are called n-gram models (in the above example, $n=5$) because they give the probability of seeing the nth word conditional on the first $n-1$ words.

To use an example from our data set, up until Oct. 2011, which is around the start of the European sovereign debt crisis, the phrase *negative outlook on* had appeared 688 times, and had always been followed by the word *any*. In Oct. 2011, we observe in our sample 13 occurrences of the phrase *negative outlook on France*. We would like our language model to consider this phrase unusual given the observed history.

Consider an N-word sentence $w_1 \dots w_N$. We can write its probability as

$$(1)\quad P(w_1 \dots w_N) \quad = \quad P(w_1)P(w_2|w_1)P(w_3|w_1 w_2)\cdots P(w_N|w_1 w_2 \dots w_{N-1}).$$

Such n-gram models are used in this context to approximate conditional probabilities of the form $P(w_k|w_1 \dots w_{k-1})$ when k is so large (practically speaking, for $k \geq 6$) that it becomes difficult to provide a meaningful estimate of the conditional probabilities for most words. In the case of an n-gram model, we assume that it suffices to consider the $n-1$ previous words in estimating the conditional probability of the next words, so

$$P(w_k|w_1 \dots w_{k-1}) \quad = \quad P(w_k|w_{k-(n-1)} \dots w_{k-1}).$$

This allows us to approximate the probability in equation (1) as

$$(2)\qquad P(w_1 \dots w_N) \quad \approx \quad \prod_{k=n}^{N} P(w_k|w_{k-n+1} \dots w_{k-1}).$$

In equation (2), we have dropped first $n-1$ terms from equation (1).

A text or corpus is a collection of sentences.[8] Let us refer to the text whose probability (or unusualness) we are trying to determine as the *evaluation text*. Since the true text model is not known, the probabilities in equation (2) will have to be estimated from a *training corpus*, which is usually very large relative to the evaluation text. In our analysis, for evaluation text from month t, the training corpus consists of all articles appearing in months $t-27$ through $t-4$. More implementation details are given in Section II.C.

Assuming sentences are independent, the probability of an evaluation text is given by the product of the probabilities of its constituent sentences. Assume that the evaluation text consists of I distinct n-grams $\{w_1^i w_2^i \cdots w_n^i\}$, $i=1,\dots,I$, each occurring c_i times. From equation (2), we see that the evaluation text probability

[7]For example, we treat "chief executive officer" as a single token. When we refer to "words" in the following discussion, we always mean tokens.

[8]This has the effect of only counting as an n-gram a contiguous n-word phrase that does not cross a sentence boundary. Another option, as suggested by Jurafsky and Martin ((2008), p. 89), is to use a start and end of sentence token as part of the language vocabulary and then allow n-grams to lie in adjoining sentences. This would greatly increase the number of n-grams in our study, and due to data sparsity we did not pursue this option.

can be written as

$$(3) \qquad P_{\text{eval}} = \prod_{i=1}^{I} P(w_n^i | w_1^i \cdots w_{n-1}^i)^{c_i}.$$

The probabilities $P(w_n^i | w_1^i \cdots w_{n-1}^i)$ in equation (3) are estimated from the training corpus. For a 4-gram $\{w_1 w_2 w_3 w_4\}$, the empirical probability of w_4^i conditional on $w_1 w_2 w_3$ will be denoted by m_i, and is given by

$$(4) \qquad m_i = \frac{\tilde{c}(\{w_1 w_2 w_3 w_4\})}{\tilde{c}(\{w_1 w_2 w_3\})},$$

where $\tilde{c}(\cdot)$ is the count of the given 3- or 4-gram in the training corpus. In Appendix A we discuss how we handle the situation when a particular 4-gram was not observed in the training corpus.

Taking natural logs in equation (3) and dividing by the total number of n-grams in the evaluation text, $w_1 \ldots w_N$, we obtain the per word, negative log probability of this text:

$$(5) \qquad H_{\text{eval}} \equiv -\frac{1}{\sum_k c_k} \ln P(w_1 \ldots w_N) = -\frac{1}{\sum_k c_k} \sum_{i=1}^{I} c_i \ln m_i$$
$$= -\sum_{i=1}^{I} p_i \ln m_i,$$

where p_i is the fraction of all n-grams represented by the ith n-gram.

The evaluation text is *unusual* if it has low probability P_{eval}, relative to the training corpus. Equation (5) shows that, in an n-gram model, the evaluation text is unusual if there are n-grams that occur frequently in the evaluation text (as measured by p_i) but rarely in the training corpus (as measured by m_i).

The quantity in equation (5) is called the cross-entropy of the model probabilities m_i with respect to the observed probabilities p_i (see Jurafsky and Martin (2008), eq. (4.62)). We refer to H_{eval} simply as the entropy of the evaluation text. Based on this definition, unusual texts will have high entropy.

The definition of entropy in equation (5) can apply to an arbitrary list of n-grams, as opposed to all the n-grams in a text. For example, we may want to consider the list of n-grams that include the word "France," or the list of all n-grams appearing in articles about banks. For a list j of n-grams, we denote by $\{c_1^j(t), \ldots, c_I^j(t)\}$ the counts of the number of times each n-gram appears in month t. The fraction of all n-grams represented by the ith n-gram is therefore

$$p_i^j(t) = \frac{c_i^j(t)}{\sum_i c_i^j(t)}.$$

Given a list of n-grams in month t, the entropy of that list will be defined as

$$(6) \qquad H^j(t) \equiv -\sum_i p_i^j(t) \ln m_i(t),$$

where, as before, the m_i's are conditional probabilities estimated from a training corpus.

Alternative Measures

In their analysis of 10-Ks, Loughran and McDonald (2011) find that sentiment measures are more informative when individual words are weighted based on their frequency of occurrence. They use what is known in the text analysis literature as a *tf-idf* scheme because it accounts for term frequency and inverse document frequency. The weight assigned to each word in each document depends on the number of occurrences of the word in the document and the fraction of documents that contain the word. A word in a document has a greater weight if it occurs frequently in that document and rarely in other documents. This approach is less well suited to our setting because we do not analyze individual documents and because our unit of analysis is the n-gram rather than the individual word. The entropy measure allows a more direct measure of the unusualness of an entire body of text in one period relative to another.

Tetlock (2011) uses measures of similarity between news articles as proxies for staleness of news. His primary measure is the ratio of the number of words that two articles have in common to the number of distinct words occurring in the two articles combined. Although similar measures could potentially be used in our setting, Tetlock's approach seems better suited to comparing pairs of articles than to comparing large bodies of text.

B. Sentiment

The traditional approach for evaluating sentiment has been to calculate the fraction of words in a given document that have negative or positive connotations. To do so, researchers rely on dictionaries that classify words into different sentiment categories. Tetlock (2007) and Tetlock, Saar-Tsechansky, and Macskassy (2008) use the Harvard IV-4 psychosocial dictionary. Recent evidence (Loughran and McDonald (2011), Heston and Sinha (2017)) shows that the Loughran–McDonald word lists (https://sraf.nd.edu/textual-analysis/resources/##LMSentimentWordLists) do a better job of sentiment categorization in a financial context than the Harvard dictionary. We use the Loughran–McDonald dictionary in our work.

Because our core unit of analysis is the n-gram, we take a slightly different approach than the traditional literature. Rather than counting the number of positive or negative words in a given article, we classify n-grams as being either positive or negative. An n-gram is classified as positive (negative) if it contains at least one positive (negative) word and no negative (positive) words. We can then measure the tone of (subsets of) news stories by looking at the fraction of n-grams they contain which are classified as either positive or negative.

C. Variable Definitions

Throughout the paper, our empirical work is at a monthly horizon, both for our news measures and our market and volatility data. In our analysis, we use a 4-gram model. [9] Table 1 summarizes the definitions of all textual variables defined in this section.

[9] Jurafsky and Martin ((2008), p. 112) discuss why 4-gram models are a good choice for most training corpora.

TABLE 1
Summary of Textual Measures

Table 1 is a summary of textual variables defined in Section II.C. All measures are calculated at a monthly frequency. All textual measures are either aggregated across firms or reference individual firms, depending on the context. [a]Aggregate entropy computed as first principal component of single-name entropy series.

Term	Definition	Term	Definition
ENTPOS[a]	Entropy of all positive n-grams	ENTNEG[a]	Entropy of all negative n-grams
SENTPOS	Fraction of all n-grams containing a positive word, and no negative words	SENTNEG	Fraction of all n-grams containing a negative word, and no positive words
ENTSENT_POS	ENTPOS × SENTPOS	ENTSENT_NEG	ENTNEG × SENTNEG

1. Entropy

We refer to the measure of unusualness of all month t articles about company j as $\text{ENTALL}^j(t)$. The entropy of the list of n-grams classified as being negative (positive) appearing in articles that mention company j is called $\text{ENTNEG}^j(t)$ ($\text{ENTPOS}^j(t)$). The p_i^j's from equation (6) come from these lists in month t, and the m_i's are estimated in a training corpus. As mentioned in Section II.A, the training corpus for month t consists of all 3- and 4-grams in our data set that appeared in the 2-year period from month $t-27$ up to and including month $t-4$. We use a rolling window, as opposed to an expanding window from the start of the sample to $t-4$ in order to keep the information sets for all our entropy calculations of roughly the same size.[10] Note that our text data starts in Jan. 1996, but because of the entropy training window, the first entropy observation occurs in the 28th month of our sample, which is Apr. 1998. More details about these calculations are found in Appendix A.

We find that the aggregate entropy measures calculated directly from a list of all month t n-grams can be unduly influenced by a small set of frequently occurring n-grams. For example, if an n-gram i appears only in articles about one company in month t, but appears very often (i.e., has a large $p_i(t)$) and has a low model probability $m_i(t)$, this one n-gram can distort the aggregate level entropy measure. A more stable measure of aggregate entropy is the first principal component of the combined single-name entropy and aggregate entropy series. For example, we define ENTPOS as the first principal component of all the single-name ENTPOS^j series and of $\text{ENTPOS}^{\text{all}}$, which is the entropy series calculated directly from all month t positive sentiment n-grams.[11] In the rest of the paper, all

[10]We exclude the 3 months prior to month t from the training corpus because at times, a 4-gram and its associated 3-gram, in the 2 year's prior to month t, may occur for the first time in month $t-1$. Furthermore if the associated 3-gram occurred as often in month $t-1$ as the 4-gram, the training set (unmodified) probability $P(w_4|w_1w_2w_3)$ will equal 1, and the associated entropy contribution will be 0. However, this n-gram may still be "unusual" in month t if it has only been observed in month $t-1$ and at no other time in our training set. For example, the 4-gram *a failed hedging strategy* is one of the top entropy contributors (see discussion in Section III.A) in May 2012. It refers to the losses incurred in April and May 2012 by the Chief Investment Office of JPMorgan. The 3-gram *a failed hedging* also occurs for the first time in our sample in May 2012, and both occur 53 times. When this phrase appears (11 times) in June 2012, we would still like to regard it as unusual.

[11]Each principal component is roughly an average of the individual entropy series. In particular, no individual series has a disproportionate weight. Excluding $\text{ENTPOS}^{\text{ALL}}$, $\text{ENTNEG}^{\text{ALL}}$, and $\text{ENTALL}^{\text{ALL}}$ from the respective calculations yields a virtually identical series of principal components (with correlations of 0.9998, 0.9997, and 0.9998, respectively).

aggregate level entropy measures (ENTALL(t), ENTNEG(t), and ENTPOS(t)) are computed in this way.[12]

2. Sentiment

We define sentiment of a given subset of articles as the percentage of the total count of all n-grams appearing in those articles that are classified as either positive or negative. For example, we may be interested in those articles mentioning Bank of America in month t. If we denote by POS(t) (NEG(t)) the set of all time t n-grams that are classified as positive (negative), then the positive sentiment of list j is

$$\text{SENTPOS}^j(t) = \frac{\sum_{i \in \text{POS}(t)} c_i^j(t)}{\sum_i c_i^j(t)}, \tag{7}$$

with the analogous definition for $\text{SENTNEG}^j(t)$. Our aggregate measures of sentiment SENTPOS(t) and SENTNEG(t) are calculated using all month t n-grams.

3. Interacted Terms

At the single-name level, we define the interacted term between negative sentiment and negative entropy as $\text{ENTSENT\_NEG}^j(t) \equiv \text{ENTNEG}^j(t) \times \text{SENTNEG}^j(t)$. $\text{ENTSENT\_POS}^j(t)$ is defined similarly. At the aggregate level, we define $\text{ENTSENT\_NEG}(t) \equiv \text{ENTNEG}(t) \times \text{SENTNEG}(t)$ and $\text{ENTSENT\_POS}(t) \equiv \text{ENTPOS}(t) \times \text{SENTPOS}(t)$. Our intent is to capture times when news flow about a given name, or in aggregate, has both high or low sentiment, as well as elevated entropy, suggesting either unusual positive or unusual negative news. As we will show in our empirical results, both characteristics matter. There are other ways to model this interaction, for example by interacting entropy with an indicator variable of whether sentiment is at an extreme level. We do not believe this modeling choice will have a first order effect on our results, and therefore we choose the most parsimonious specification.

III. Data

Our data set consists of Thomson Reuters news articles about the top 50 global banks, insurance, and real estate firms by U.S. dollar market capitalization as of Feb. 2015. These 50 companies were our original data request to Thomson Reuters and are free from data snooping. The potential survivorship bias in this selection of companies works against the effects we find: that text measures can forecast volatility. This is because firms that disappeared during the financial crisis are not in our sample. For example, realized volatility grew substantially higher in 2008, relative to previous years, for Bear Stearns, Lehman Brothers, and Washington Mutual (not in our sample), than for Bank of America, Citigroup,

[12] Because of the need to have all data present for computing the principal component, our aggregate entropy measures use only 25 series for ENTPOS and ENTNEG, and 31 series for ENTALL. For series that have observations at the start of the sample period, but are missing some intermediate observations, we use the most recently available nonmissing value of the associated entropy measure.

Goldman Sachs, JPMorgan Chase, Morgan Stanley, and Wells Fargo (in our sample). Almost 90% of the articles are from Reuters itself, with the remainder coming from one of 16 other news services. Table 2 lists the companies in our sample. Table 3 groups our sample of companies and articles by country of domicile. The table reports the following statistics about companies domiciled in a given country: i) Average market capitalization, ii) the percentage of all articles that mention companies from that country, and iii) the number of companies. Our set of news articles leans heavily toward the English speaking countries (United States, United Kingdom, Australia, and Canada). For example, even though China has 8

TABLE 2

List of Companies in the Analysis

Table 2 reports the companies included in the Thomson Reuters news sample.

	Company		Company
1	Berkshire Hathaway	26	Australia & New Zealand Bank
2	Wells Fargo	27	AIG
3	Industrial and Commercial Bank of China	28	BNP Paribas
4	JPMorgan Chase	29	National Australia Bank
5	China Construction Bank	30	Morgan Stanley
6	Bank of China	31	Itau Unibanco
7	HSBC Holdings	32	UBS
8	Agricultural Bank of China	33	Bank of Communications
9	Bank of America	34	Royal Bank of Scotland
10	Visa	35	Prudential
11	China Life Insurance	36	Simon Property Group
12	Citigroup	37	Barclays
13	Commonwealth Bank of Australia	38	Bank of Nova Scotia
14	Ping An Insurance	39	Blackrock
15	Mastercard	40	AXA
16	Banco Santander	41	Banco Bilbao Vizcaya Argentaria
17	Westpac Bank	42	China Merchants Bank
18	American Express	43	Metlife
19	Royal Bank of Canada	44	Banco Bradesco
20	Lloyds	45	Nordea Bank
21	Goldman Sachs	46	Zurich Insurance
22	Mitsubishi UFJ	47	Intesa Sanpaolo
23	US Bancorp	48	ING
24	Allianz	49	Sumitomo Mitsui
25	TD Bank	50	Allied Irish Banks

TABLE 3

Companies Grouped by Country of Domicile

Within each country, Table 3 reports the average market capitalization of the companies in the sample as of Nov. 2015. Also reported are the percentages of all articles in the Thomson Reuters data set that mention companies from a particular country of domicile, as well the number of firms classified as being domiciled in a given country.

Country	Avg. Mkt. Cap. (USD$)	Percent of All Articles	No. of Firms
United States	137.47	44.25	15
Britain	82.73	19.11	5
Australia	70.45	6.35	4
Canada	72.45	6.08	3
Spain	68.28	4.68	2
France	70.59	4.63	2
Netherlands	55.70	3.19	1
China	136.00	2.70	8
Germany	80.20	2.22	1
Switzerland	57.28	1.95	2
Japan	72.26	1.69	2
Ireland	41.84	1.04	1
Italy	57.52	0.80	1
Brazil	37.46	0.68	2
Sweden	45.87	0.63	1

(of a total of 50) companies with market capitalizations on par with the U.S. companies, under 3% of our total articles mention companies from China.

The raw data set has over 600,000 news articles, from Jan. 1996 to Dec. 2014. As mentioned in Section II.C, our first entropy observation is from Apr. 1998. Many articles represent multiple rewrites of the same initial story. We filter these by keeping only the first article in a given chain.[13] We also drop any article coming from PR Newswire, as these are corporate press releases. All articles whose headlines start with `REG-` (regulatory filings) or `TABLE-` (data tables) are also excluded. This yields 367,331 unique news stories which we ultimately use in our analysis. Each article is tagged by Thomson Reuters with the names of the companies mentioned in that article. Many articles mention more than one company. Appendix B gives more details about our text processing procedure.

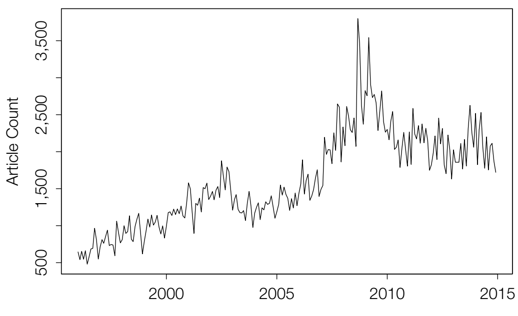
Figure 1 shows the time series of article counts in our sample. The per month article count reaches its approximate steady-state level of 1,500 or so articles in the early 2000s, peaks around the time of the financial crisis, and settles back down to the steady state level toward the end of 2014. The early years of our sample have relatively fewer articles, which may introduce some noise into our analysis.

Our market data comes from Bloomberg L.P. For each of the 50 companies in our sample, we construct a U.S. dollar total returns series using Bloomberg price change and dividend yield data. Also, for those firms that have traded options, we use 30-day implied volatilities for at-the-money options from the Bloomberg volatility surfaces. The volatility data start in Jan. 2005. Our single-name volatility series are 20-day realized volatilities of local currency returns, as calculated by Bloomberg. Our macro data series are the Chicago Board Options Exchange Volatility Index (VIX) and 20-day realized volatility for the S&P 500 Index computed by Bloomberg from daily returns.[14]

FIGURE 1

Monthly Article Count in the Thomson Reuters News Sample

Figure 1 shows the article count by month in our sample.



[13]All articles in a chain share the same *Article Id* code.

[14]Month t realized returns are realized in that month, whereas the month t VIX level is the close-of-month level.

Our S&P 500 level analysis starts in Apr. 1998 (the first month when we have entropy measures) and our single-name analysis starts in Jan. 2005 (the first month in which we have implied volatility data for firms).

A. N-grams and Their Contribution to Entropy

As an example, consider that in Jan. 2013, the 4-gram *raises target price to* appeared 491 times in the entire sample (i.e., $c^{\mathrm{ALL}}_{\{\text{raises target price to}\}}(\text{Jan. }2013)=491$, where ALL is the list of n-grams appearing in all articles). It appeared 34 times in articles that were tagged as mentioning Wells Fargo & Co., 26 times in articles that mentioned JPMorgan Chase & Co., but 0 times in articles that mentioned Bank of America Corp. If we sum across all 50 names in our data set, this 4-gram appeared 1,014 times (more than its total of 491 because many articles mention more than one company).

In each month, we focus on the 5,000 most frequently occurring 4-grams. In our 19-year data set, we thus analyze $19\times 12\times 5000=1.14$mm 4-grams, of which 394,778 are distinct. The first three tokens in the latter represent 302,973 distinct 3-grams.

By sorting n-grams on their contribution, given by $-p_i \ln m_i$, to the entropy of the overall month t corpus, we can identify for month t the most and least unusual 4-word phrases. Table 4 reports the three top and bottom phrases[15] by their contribution to entropy in the 2 months in our sample that had major market or geopolitical events: Sept. 2008 (the Lehman bankruptcy) and May 2012 (around the peak of the European sovereign debt crisis). In each case, at least one of the n-grams with the largest entropy contribution reflects the key event of that month, and does so without any semantic context. On the other hand, the n-grams with the smallest entropy contribution are generic, and have no bearing on the event under consideration.

Consider for example the n-gram *nyse order imbalance _mn_* from Sept. 2008. In our training set, the majority of occurrences of the 3-gram *nyse order imbalance* were followed by *_n_* (a number) rather than *_mn_* (a number in the millions). The frequent occurrence of *nyse order imbalance* followed by a number in the millions, rather than a smaller number, is unusual. This 4-gram has a relatively large p_i, a low m_i (and a high $-\ln m_i$), and is the top contributor to negative entropy in this month. On the other hand, the 3-gram *order imbalance _n_* is almost always followed by the word *shares*, thus giving this 4-gram an m_i of almost 1, and an entropy contribution close to 0. In May 2012, the n-gram *the euro zone crisis* is unusual because in the sample prior to this month, the 3-gram *the euro zone* is frequently followed by *'s* or *debt*, but very infrequently by *crisis*. Therefore, the relatively frequent occurrence in this month of this otherwise unusual phrase renders it a high negative entropy contributor. While anecdotal, this evidence suggests that our entropy measure sorts phrases in a meaningful way.

B. Summary Statistics

Table 5 reports the average contemporaneous correlation between the 50 individual volatility (realized and implied) and sentiment pairs. If an individual

[15]Some of the distinct 4-grams come from the same 5-gram.

TABLE 4

Top and Bottom 4-Grams in Selected Months

Table 4 reports the top and bottom three 4-grams, as determined by their contribution to ENTNEG in selected months of our sample. The "Total" column shows the number of times the given n-gram has appeared in that month, and the "Rank" column gives its rank by entropy contribution, which is lower than 5000 because we restrict analysis to those n-grams which are classified as having negative sentiment. p_i and m_i are the in-sample probability and the training sample conditional probability for the n-gram (see equation (6)). Note that some of the 4-grams come from the same 5-gram.

Month	Year	w1	w2	w3	w4	Total	Rank	p_i	m_i
9	2008	nyse	order	imbalance	_mn_	81	1	0.009	0.020
9	2008	the	collapse	of	lehman	38	2	0.004	0.004
9	2008	filed	for	bankruptcy	protection	138	3	0.016	0.245
9	2008	problem	accessing	the	internet	33	400	0.004	0.961
9	2008	imbalance	_n_	shares	on	299	401	0.034	0.999
9	2008	order	imbalance	_n_	shares	299	402	0.034	0.999
5	2012	_bn_	from	a	failed	28	1	0.008	0.009
5	2012	the	euro	zone	crisis	36	2	0.011	0.087
5	2012	declined	to	comment	on	56	3	0.017	0.258
5	2012	you	experience	problem	accessing	77	208	0.023	0.998
5	2012	experience	problem	accessing	the	77	209	0.023	0.998
5	2012	problem	accessing	the	internet	77	210	0.023	0.998

TABLE 5

Single-Name Sentiment/Entropy Correlations with Volatility

Table 5 reports the average correlation between different entropy and sentiment measures and 1 month at-the-money implied or intramonth realized volatilities for the 50 stocks in our sample. Also reported is the average correlation between the percentage of all monthly articles about a given company (ARTPERC) and realized and implied volatilities. If a stock implied volatility series is not present, and for the aggregate measures, the VIX index is used instead of single-name implied volatility. The realized volatility is available for all stocks. Cross-sectional standard errors, which assume independence, are reported.

	ARTPERC	ENTNEG	ENTPOS	ENTALL	SENTNEG	SENTPOS
Mean corr. with IVOL	−0.004	0.197	−0.004	0.095	0.306	−0.097
SE IVOL	0.036	0.026	0.019	0.024	0.024	0.017
Mean corr. with RVOL	0.085	0.208	−0.012	0.112	0.220	−0.074
SE RVOL	0.026	0.019	0.018	0.016	0.019	0.014

implied volatility series does not exist, we use the VIX as a stand-in. Cross-sectional standard errors are also calculated assuming independence of observations. We see that SENTNEG^j (SENTPOS^j) is, on average, positively (negatively) correlated with single-name volatility.

We observe a similar pattern at the aggregate level. Figure 2 shows the time series of SENTPOS and SENTNEG in our sample, as well as a scaled version of the VIX. Note that at the aggregate level, negative sentiment is contemporaneously positively correlated with the VIX, whereas positive sentiment is contemporaneously negatively correlated. These correlations are 0.458 and −0.373, respectively. Section V will study the dynamics of this relationship in depth.

Table 5 also reports the average correlation between the various single-name entropy measures and single-name implied or realized volatility. The average single-name correlations for ENTALL and ENTNEG are positive, and the ENTPOS average correlation is marginally negative, though very close to 0.

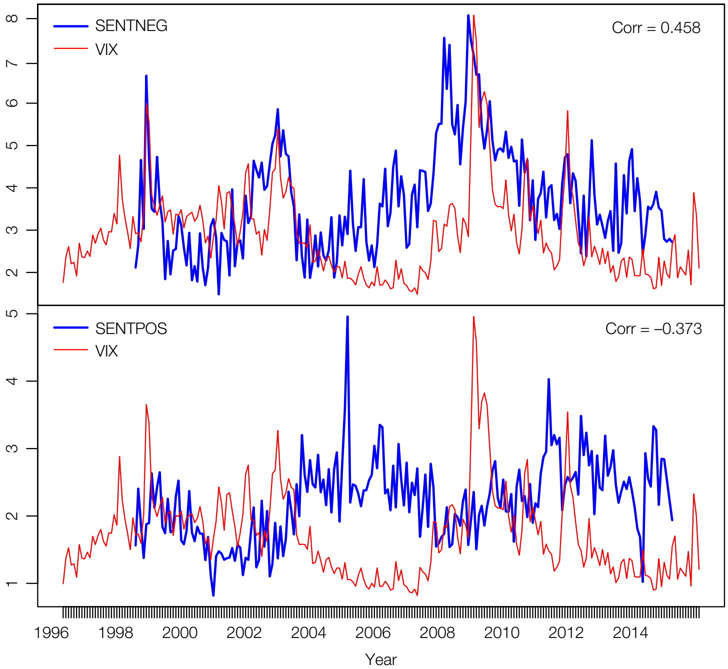
Figure 3 shows the three aggregate entropy series (ENTALL(t), ENTNEG(t), and ENTPOS(t)), with a scaled VIX superimposed.[16] All three series are

[16]ENTALL and ENTNEG, but not ENTPOS, appear to have a time trend. This is potentially due to a change in the composition of articles carried by Thomson Reuters over time (for example, the inclusion of data tables and statistical summaries lowers the average entropy). In our panel regressions

FIGURE 2

Monthly Plots of SENTNEG and SENTPOS as Defined in Equation (7)

Each series computes the proportion of all n-grams in a given month that are classified as having either positive or negative sentiment. Superimposed on each sentiment series is the scaled VIX index. VIX-sentiment correlation is shown in the upper right-hand corner.



positively correlated with the VIX. ENTPOS has the lowest correlation at 0.15, and ENTNEG has the highest at 0.48. This is in contrast to the sentiment series where negative and positive sentiment have opposite signed VIX correlations. Since entropy reflects unusualness of news, it is perhaps not surprising that all entropy series are positively correlated with the VIX, as all news (neutral, positive, and negative) may be more unusual during times of high market volatility.

IV. Single-Name Volatility

In this section, we investigate whether our company-level news-based measures contain information that is not already known to market participants. We assume that the information set of the latter contains implied volatilities, as well as other variables commonly known to forecast volatility. Poon and Granger (2003) suggest that the best performing volatility forecasting models include both implied and historical volatility as explanatory variables. They also point out that

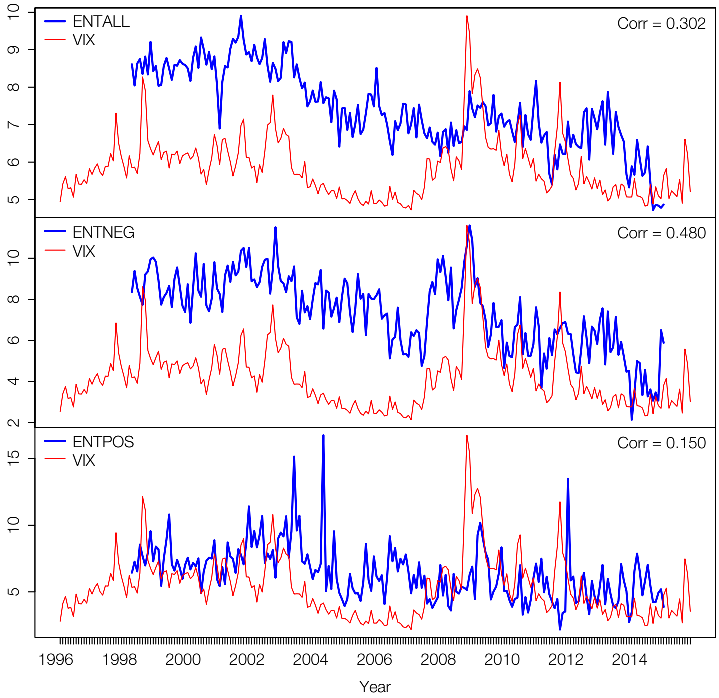
---

that include entropy as a standalone variable, the inclusion of lagged country-level article counts partially controls for the trend. However, the main focus of our analysis is on the interacted measures ENTSENT_NEG and ENTSENT_POS, and as can be seen from Figure 4, these variables show no evidence of a time trend.

FIGURE 3

Monthly Plots of ENTALL(t), ENTNEG(t), and ENTPOS(t) as Defined in Section II

Each series is the first principal component of the associated single-name entropy measures, for those names with observations available in all time periods of the sample. Superimposed on each entropy series is the scaled VIX index. The correlation between entropy and VIX is shown in the upper right-hand corner of each chart.



models using short-dated at-the-money implied volatility work about as well as more sophisticated approaches that take into account the volatility term structure and skew. Bekaert and Hoerova (2014) show that, at the index level, in addition to lags of implied and realized variance, stock price jumps also matter for forecasting future realized variance. To control for these effects, we use our 30-day at-the-money implied volatility measure IVOL, 20 trading-day realized volatility RVOL, and the negative portion of monthly returns r^- (denoted as RET_MI in the tables) as explanatory variables for future realized and implied volatility.[17] Because these variables have been shown to be effective for forecasting future volatility, the bar for our news measures to add incremental value is quite high.

Our basic specification for evaluating the forecasting power of a news-based measure NEWS^j is the following panel regression:

$$\begin{aligned}(8)\qquad \text{VOL}^j(t) \;=\;& a^j + c_1'\mathcal{L}_s\text{RVOL}^j_{30\text{day}}(t) + c_2'\mathcal{L}_s\text{IVOL}^j_{1\text{mo}}(t) + c_3'\mathcal{L}_s r^{-j}(t)\\ &+ b1'\mathcal{L}_s\text{ARTPERC}^j(t) + b2'\mathcal{L}_s\text{NEWS}^j(t) + \epsilon^j(t),\end{aligned}$$

[17] $r^- \equiv \max(-r,0)$. Adding $r^+ \equiv \max(r,0)$ as an explanatory variable was not impactful in any of our specifications, hence we do not include this variable in our regression results.

where VOL is either IVOL or RVOL, a^j is an individual fixed effect term, $\mathcal{L}_s$ is an s-lag operator,[18] and ARTPERC^j is the percentage of all month t articles that mention company j. The variable ARTPERC^j is intended to control for the information content of news volume. All news measures are normalized to have unit variance.[19]

We show results for $s=2$ (those for $s=3$ are qualitatively similar and omitted to conserve space). We ran this specification in variance, log variance and volatility terms, all of which yield similar qualitative results. We show the volatility results in the paper as these are the easiest to interpret. We do not have implied volatility data for some of our companies, and these are excluded from the panel regressions.

Before turning to the forecasting regression in equation (8), we examine briefly the drivers of our news-based measures. The following is our descriptive panel specification:

$$\text{(9)} \qquad \begin{aligned} \text{NEWS}^j(t) \quad = \quad & a^j + c_1'\mathcal{L}_2\text{RVOL}^j_{30\text{day}}(t) + c_2'\mathcal{L}_2\text{IVOL}^j_{1\text{mo}}(t) \\ & + c_3'\mathcal{L}_2 r^{-j}(t) + b'\mathcal{L}_2\text{NEWS}^j(t) + \epsilon^j(t). \end{aligned}$$

This is run with NEWS^j set to each of the following:

- positive: SENTPOS, ENTPOS, ENTSENT_POS;
- negative: SENTNEG, ENTNEG, ENTSENT_NEG.

Table 6 reports the results of this descriptive regression. While lagged volatility has little effect on the positive sentiment news measures, high past realized volatility forecasts higher negative sentiment. Absence of past negative returns forecasts higher positive sentiment, whereas the presence of negative returns forecasts higher negative entropy. The positive and negative news measures are less persistent than percentage article counts, although all the news measures exhibit some persistence.

Tables 7 and 8 report the results of the specification in equation (8) for implied and realized volatility, respectively. The control variables (lagged IVOL, RVOL, and r^-) all matter for both future realized and implied volatility, and enter the panel with the expected positive sign (only $r^-(t-2)$ enters with a negative sign, though the magnitude of the effect is much smaller than that of $r^-(t-1)$).

Model 1 of Table 8, which has ARTPERC as the sole news-based measure, offers some evidence that firms that are in the news a lot, irrespective of sentiment, tend to have lower realized volatilities in future months. Along similar lines, Jiao, Veiga, and Walther (2016) find that idiosyncratic realized volatility is lower for companies that receive greater attention from the news media. Furthermore, this

[18] $\mathcal{L}_s Y(t) = \{Y(t-1), Y(t-2), \ldots, Y(t-s)\}$.

[19] For a panel VAR with fixed effects, Nickell (1981) shows that OLS estimates of the AR coefficients are asymptotically consistent as the length of the time series grows, with a bias that is inversely proportional to the length of the time series. Our panels include data from Jan. 2005 (the start of our implied volatility data) to Dec. 2014 (the end of our text data). Accounting for some missing observations, we have approximately 90 months of data per company. Any bias should therefore be very small.

TABLE 6
Descriptive Panels for News Measures

Table 6 reports the results of the panel model from equation (9). The dependent variable is shown in the column heading, with the regressors in the rows. The notation _LN indicates the variable is lagged by N months. The row labeled "Sum NM 2" reports the sum of the two bottom-most coefficients in each column. The regression is run with individual fixed effects. Residuals are clustered by time for computing standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The regressions include data from Mar. 2005 to Dec. 2014. All news measures have been normalized to have unit variance.

Variable	ARTPERC	SENTPOS	ENTPOS	ENTSENT_POS	SENTNEG	ENTNEG	ENTSENT_NEG
IVOL_L1	0.001	−0.001	0.001	0.000	0.001	0.001	0.001
IVOL_L2	0.003***	−0.001	−0.001	−0.001	0.000	−0.003**	−0.002
RVOL_L1	0.000	0.000	0.001	0.001	0.004***	0.004***	0.006***
RVOL_L2	−0.002***	0.000	0.000	0.000	−0.001	0.001	0.000
RET_MI_L1	0.007***	−0.007*	−0.007	−0.007	0.003	0.009**	0.007
RET_MI_L2	−0.001	0.000	−0.005	−0.005	0.003	0.006	0.007
ARTPERC_L1	0.367***						
ARTPERC_L2	0.217***						
SENTPOS_L1		0.146***					
SENTPOS_L2		0.095***					
ENTPOS_L1			0.150***				
ENTPOS_L2			0.140***				
ENTSENT_POS_L1				0.131***			
ENTSENT_POS_L2				0.075***			
SENTNEG_L1					0.221***		
SENTNEG_L2					0.169***		
ENTNEG_L1						0.195***	
ENTNEG_L2						0.151***	
ENTSENT_NEG_L1							0.195***
ENTSENT_NEG_L2							0.113***
Sum NM 2	0.583***	0.241***	0.29***	0.207***	0.39***	0.346***	0.307***
p-valuew NM	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Adj. R^2	0.264	0.042	0.052	0.028	0.154	0.136	0.166

finding may shed light on Fang and Peress (2009), who show that stocks receiving high media attention earn lower returns in the ensuing month; perhaps part of the story is that such stocks are also less volatile.[20]

The positive category news measures (models 2–4) all show up with negative coefficients (with the exception of one case in Table 8). This suggests positive news at time $t-1$ or $t-2$ forecast lower time t implied and realized volatility, after controlling for known forecasting variables. Summing the lag 1 and lag 2 coefficients on NEWS^j, reported in the row labeled "Sum NM 2," allows us to evaluate the importance of the different news measures.[21] We find that ENTSENT_POS has a larger effect on future volatility than either positive sentiment or entropy on their own. Furthermore, the economic significance of the effect is large. For example, as Table 7 shows, these two coefficients are −1.04 for ENTSENT_POS, suggesting that a 1-standard-deviation increase in current positive and unusual news forecasts a 1-point drop (e.g., from 20 to 19) in implied volatility next month. The results for future realized volatility in Table 8 are similar.

[20]Fang and Peress (2009) double sort by the prior month's media coverage and the prior month's idiosyncratic volatility, but do not study the effects of media coverage on future volatility.

[21]This is similar to an F-test of the joint significance of the 2 lag coefficients.

TABLE 7
Implied Volatility Panels

Table 7 reports the results of the panel model from equation (8). The dependent variable is implied volatility with the regressors in the rows. The notation _LN indicates the variable is lagged by N months. The row labeled "Sum NM 2" shows the sum of the two bottom-most coefficients in each column, or the sum of coefficients of ENTSENT_POS where multiple news measures are present. The regression is run with individual fixed effects. Residuals are clustered by time for computing standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. "Text F-test" is the p-value from an F-test that restricts all text-based measures (excluding ARTPERC) in the regression to be 0. The regressions include data from Mar. 2005 through Dec. 2014. All news measures have been normalized to have unit variance.

	IVOL								
Variable	1	2	3	4	5	6	7	8	9
IVOL_L1	0.289***	0.307***	0.305***	0.305***	0.304***	0.252***	0.251***	0.243**	0.283***
IVOL_L2	0.067	0.094	0.080	0.080	0.091	0.103	0.102	0.094	0.064
RVOL_L1	0.222***	0.209***	0.209***	0.208***	0.207***	0.226***	0.222***	0.219***	0.246***
RVOL_L2	0.119***	0.102***	0.108**	0.107**	0.098**	0.104**	0.100**	0.103**	0.086
RET_MI_L1	0.371***	0.364***	0.375***	0.374***	0.360***	0.420***	0.411***	0.400***	
RET_MI_L2	−0.165*	−0.174*	−0.165	−0.166	−0.172*	−0.173*	−0.174*	−0.163	
ARTPERC_L1	−0.681	−0.677	−0.666	−0.638	−0.732	−0.806	−0.913*	−0.670	
ARTPERC_L2	0.224	0.270	0.148	0.168	0.233	0.149	0.247	0.108	
SENTPOS_L1		−0.422						0.035	0.044
SENTPOS_L2		−0.235						0.127	0.181
ENTPOS_L1			−0.369						
ENTPOS_L2			−0.395						
ENTSENT_POS_L1				−0.635				−0.715	−0.778
ENTSENT_POS_L2				−0.406				−0.567	−0.654
SENTNEG_L1					0.802**			0.773	0.953
SENTNEG_L2					0.467			−0.473	−0.497
ENTNEG_L1						0.225			
ENTNEG_L2						0.757**			
ENTSENT_NEG_L1							0.739**	0.310	0.248
ENTSENT_NEG_L2							0.807**	1.247**	1.264**
Sum NM 2	−0.456	−0.657*	−0.764	−1.041**	1.27***	0.982*	1.546***	−1.282*	−1.432*
p-value NM	0.378	0.084	0.145	0.031	0.006	0.070	0.001	0.076	0.055
Adj. R^2	0.574	0.576	0.576	0.576	0.577	0.593	0.594	0.582	0.574
Text F-test	—	0.205	0.346	0.098	0.017	0.116	0.006	0.016	0.008

The negative category news measures (models 5–7) forecast future implied and realized volatility with a positive sign. All three news-based measures (SENTNEG, ENTNEG, and ENTSENT_NEG) are economically and statistically meaningful, with the interacted term ENTSENT_NEG having the largest economic effect. A 1-standard-deviation increase (at both lags) in the latter implies a rise of 1.546 (2.429) in next month's implied (realized) volatility (as can be seen from the "Sum NM 2" row of Tables 7 and 8), which is a very large economic effect.

In model 8 of Tables 7 and 8, we include all news based measures in the panel (except the noninteracted entropy measures).[22] The results of this regression are very stark. When adding the 2 lags for both positive and negative sentiment (SENTPOS and SENTNEG), the cumulative effect on future volatility is very close to 0, while the effects of ENTSENT_POS (on implied volatility) and ENTSENT_NEG (on both implied and realized volatility) are economically and statistically (when looking at the sum of coefficients of lags 1 and 2) very large.[23]

[22]Model 9 of Tables 7 and 8 is discussed in Section VI.A.

[23]For models 8 and 9, the row labeled "Sum NM 2" shows the sum and associated p-value for the coefficients on ENTSENT_POS.

TABLE 8

Realized Volatility Panels

Table 8 reports the results of the panel model from equation (8). The dependent variable is realized volatility, with the regressors in the rows. The notation _LN indicates the variable is lagged by N months. The row labeled "Sum NM 2" reports the sum of the two bottom-most coefficients in each column, or the sum of coefficients of ENTSENT_POS where multiple news measures are present. The regression is run with individual fixed effects. Residuals are clustered by time for computing standard errors. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. "Text F-test" is the p-value from an F-test that restricts all text-based measures (excluding ARTPERC) in the regression to be 0. The regressions include data from Mar. 2005 through Dec. 2014. All news measures have been normalized to have unit variance.

	RVOL								
Variable	1	2	3	4	5	6	7	8	9
IVOL_L1	0.245***	0.268***	0.265***	0.264***	0.263***	0.295***	0.292***	0.281***	0.368***
IVOL_L2	0.091**	0.120***	0.128***	0.128***	0.118**	0.118**	0.118**	0.124**	0.069
RVOL_L1	0.349***	0.331***	0.312***	0.311***	0.329***	0.311***	0.306***	0.292***	0.364***
RVOL_L2	0.140***	0.122**	0.130**	0.130**	0.116**	0.116**	0.108**	0.110**	0.082
RET_MI_L1	0.865***	0.859***	0.928***	0.925***	0.850***	0.860***	0.845***	0.887***	
RET_MI_L2	−0.077	−0.092	−0.108	−0.113	−0.089	−0.105	−0.104	−0.119	
ARTPERC_L1	−1.269**	−1.246**	−1.049*	−0.973*	−1.314**	−1.321**	−1.432**	−1.277**	
ARTPERC_L2	−0.024	0.026	−0.244	−0.254	−0.048	−0.146	−0.031	−0.241	
SENTPOS_L1		−0.684*						−0.210	−0.166
SENTPOS_L2		−0.401						0.183	0.336
ENTPOS_L1			0.052						
ENTPOS_L2			−0.431						
ENTSENT_POS_L1				−0.556				−0.469	−0.661
ENTSENT_POS_L2				−0.662				−0.925	−1.182
SENTNEG_L1					1.646***			0.527	0.800
SENTNEG_L2					0.234			−0.674	−0.636
ENTNEG_L1						0.985*			
ENTNEG_L2						0.521			
ENTSENT_NEG_L1							1.977***	1.893**	1.887**
ENTSENT_NEG_L2							0.451	1.023	0.979
Sum NM 2	−1.293**	−1.085**	−0.379	−1.217*	1.88***	1.506*	2.429***	−1.394	−1.843
p-value NM	0.027	0.023	0.628	0.068	0.004	0.074	0.001	0.230	0.129
Adj. R^2	0.587	0.589	0.583	0.584	0.591	0.592	0.595	0.588	0.568
Text F-test	—	0.064	0.689	0.186	0.002	0.162	0.000	0.016	0.006

In fact the sum of the two coefficients for both variables is comparable to the results from models 2–7. Interestingly, unusual positive and unusual negative news both matter. A 1-standard-deviation increase in unusual positive (negative) news over both lags leads to a drop (increase) in future realized and implied volatilities of between 1.3 and 2.9 points (e.g., from 15 to 17). This is a very large economic effect.

The bottom row of Tables 7 and 8 shows results from an F-test that restricts all text based variables (excluding ARTPERC) in each panel to be 0. For panels that forecast implied volatility (Table 7), we can reject the restriction for all models that include SENTNEG or ENTSENT_NEG, and marginally reject for models including only ENSENT_POS or ENTNEG. For model 8, which includes all text measures except noninteracted entropy, the restriction is soundly rejected. For the realized volatility panels in Table 8, we can additionally reject the restriction for the SENTPOS regression. The restriction that our text based measures are zero in model 8 is again strongly rejected.

In summary, our panel results suggest that, even after controlling for known predictors of future volatility, our news based measures contain useful forecasting information. The coefficient estimates on lagged news-measures are statistically and economically meaningful. For both the positive and negative sentiment categories, the interacted news terms (ENTSENT_POS and ENTSENT_NEG) contain more information than either sentiment or entropy on its own. In the Supplementary Material, we investigate the performance of our panel regressions in the pre-crisis subperiod. The results are somewhat weaker, especially for the positive news sentiment measures, but are otherwise qualitatively similar to the full period estimation.

V. Aggregate Volatility

We now turn from company-specific measures of entropy, sentiment, and volatility to aggregate measures. We document evidence that unusual negative news predicts an increase in volatility as measured either by the VIX or by realized volatility on the S&P 500 index. As discussed in Section II.C, each aggregate measure of entropy is the first principal component of the corresponding measures across the financial companies listed in Table 2, whereas aggregate sentiment follows from equation (7) applied to the set of all n-grams in month t. Our aggregate news measures therefore contain information from all news flow about the 50 companies in our sample.

A. Contemporaneous Correlations

Table 9 reports descriptive statistics for the five aggregate news-based measures in Figures 2 and 3 and the interacted variables. Figure 4 shows a plot of ENTSENT_NEG and ENTSENT_POS versus the VIX index. ENTSENT_NEG and the VIX track each other quite closely, except from 2004 to 2006 and just prior to the financial crisis in 2008. ENTSENT_POS is negatively correlated with the VIX, and the correlation is considerably weaker than in the case of ENTSENT_NEG.

TABLE 9

Summary Statistics

Table 9 reports summary statistics for the aggregate news-based measures, as well as the VIX and realized volatility for S&P 500. SENTNEG and SENTPOS are aggregate negative and positive sentiment measures. ENTALL, ENTNEG and ENTPOS are the first principal components of single-name level entropy measures applied to all n-grams, and those classified as negative and positive, respectively. ENTSENT_NEG (ENTSENT_POS) interacts SENTNEG (SENTPOS) with ENTNEG (ENTPOS). All data series are monthly, and run from Apr. 1998 to Dec. 2014.

Variable	Mean	Min.	Max.	Std. Dev.
ENTNEG	7.401	2.140	11.592	1.837
ENTPOS	6.443	2.172	16.747	2.092
ENTALL	7.446	4.724	9.908	1.066
SENTNEG	3.744	1.484	8.077	1.265
SENTPOS	2.233	0.819	4.958	0.594
ENTSENT_NEG	28.156	7.881	77.348	13.948
ENTSENT_POS	14.189	4.947	45.117	5.711
VIX	21.444	10.420	59.890	8.272
SPX_RVOL	17.675	4.140	87.880	10.776

Table 10 reports that SENTPOS and ENTSENT_POS have a negative correlation with the VIX, whereas all the other measures have a positive correlation. In particular, at the aggregate level, all entropy measures increase with market

FIGURE 4

Monthly Plot of ENTSENT_NEG and ENTSENT_POS

The entropy series is the first principal component of the associated single-name entropy measures, for those names with observations available in all time periods of the sample. Superimposed in a thinner line is the scaled VIX index. The correlation between ENTSENT_NEG (ENTSENT_POS) and the VIX is shown in the upper right-hand corner.

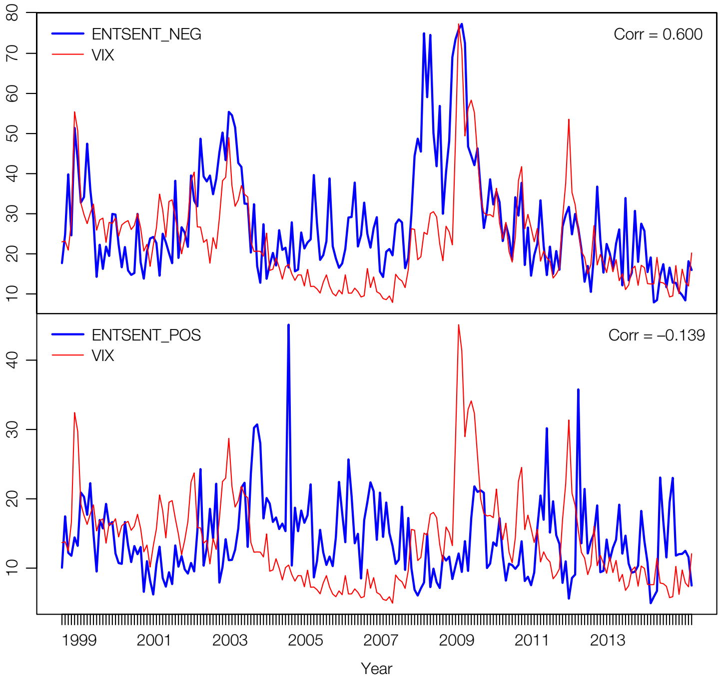


TABLE 10

Contemporaneous Correlations

Table 10 reports contemporaneous correlations among monthly levels of our news-based indicators and the VIX index. SENTNEG and SENTPOS are aggregate negative and positive sentiment measures. ENTALL, ENTNEG, and ENTPOS are the first principal components of single-name level entropy measures applied to all n-grams, and those classified as negative and positive, respectively. ENTSENT_NEG (ENTSENT_POS) interacts SENTNEG (SENTPOS) with ENTNEG (ENTPOS).

	SENTNEG	SENTPOS	ENTALL	ENTNEG	ENTPOS	ENTSENT_NEG	ENTSENT_POS	VIX
SENTNEG	1.00							
SENTPOS	−0.14	1.00						
ENTALL	−0.18	−0.42	1.00					
ENTNEG	0.19	−0.44	0.71	1.00				
ENTPOS	−0.09	−0.16	0.56	0.34	1.00			
ENTSENT_NEG	0.86	−0.32	0.19	0.64	0.08	1.00		
ENTSENT_POS	−0.15	0.54	0.16	−0.03	0.73	−0.14	1.00	
VIX	0.46	−0.37	0.30	0.48	0.15	0.60	−0.14	1.00

volatility. The entropy measures are positively correlated with one another, and negatively correlated with SENTPOS.

Note that in Table 10, although ENTNEG and SENTNEG have a low correlation of 0.19, their correlations with the VIX are 0.48 and 0.46, respectively. Thus, even though the two do not share much in common, it appears they both explain a meaningful portion of VIX variability. The interaction variable ENTSENT_NEG has the highest VIX correlation of the news based measures at 0.6. It also has a high correlation with its constituents: 0.86 with SENTNEG and 0.64 with ENTNEG. ENTSENT_POS has a −0.14 correlation with both the VIX and with ENTSENT_NEG. This correlation result, the visual evidence in Figure 4 and the descriptive statistics in Table 9 all suggest that the interacted variable ENTSENT_NEG is a closer fit to the VIX than either negative sentiment or entropy separately.

B. Impulse Response Functions

We investigate interactions among the aggregate variables through vector autoregressions (VARs). We estimate a VAR model in 6 variables, initially ordered as follows: VIX, SPX_RVOL (realized volatility), SENTNEG, ENTSENT_NEG, SENTPOS, and ENTSENT_POS.[24] As in our single-name panels, we include lagged implied and realized volatility in our aggregate VARs. Hence, we analyze the ability of our news measures to forecast future volatility only after we have stripped out the information available from lagged realized volatility and from options markets.

The Akaike information criterion selects a model with 2 lags. We analyze the model through its impulse response functions.[25] Each impulse is a 1-standard-deviation shock to the error term for 1 variable in a Cholesky factorization of the error covariance matrix. A shock to 1 variable has a direct effect on variables listed later in the order of variables, but not on variables listed earlier. Our ordering is thus stacked against finding an influence on either measure of volatility from the entropy and sentiment measures.

Graph A1 of Figure 5 shows impulse response functions in response to a shock to ENTSENT_NEG, together with bootstrapped 95% confidence intervals. Both the VIX and realized volatility have statistically significant responses to the shock. A 1-standard-deviation increase in ENTSENT_NEG increases the VIX by 1.5 points and increases realized volatility by 2 points, thus a 2- to 3-standard-deviation shock to ENTSENT_NEG has a large economic impact on volatility. Graph A2 shows corresponding results in response to a shock to SENTNEG. Here, neither VIX nor realized volatility exhibits a statistically significant response.

Next we reverse the order of ENTSENT_NEG and SENTNEG and recalculate the impulse response functions. Graph B1 of Figure 5 shows that the VIX and realized volatility now have statistically significant responses to SENTNEG, increasing by roughly 1.25 and 1.75 points, respectively. But Graph B2 shows that they still have marginally significant responses to ENTSENT_NEG following the order change. Together with Graph A, these results suggest the following conclusions: An increase in negative sentiment or its interaction with entropy each predicts an increase in volatility; the effect of negative sentiment is captured by

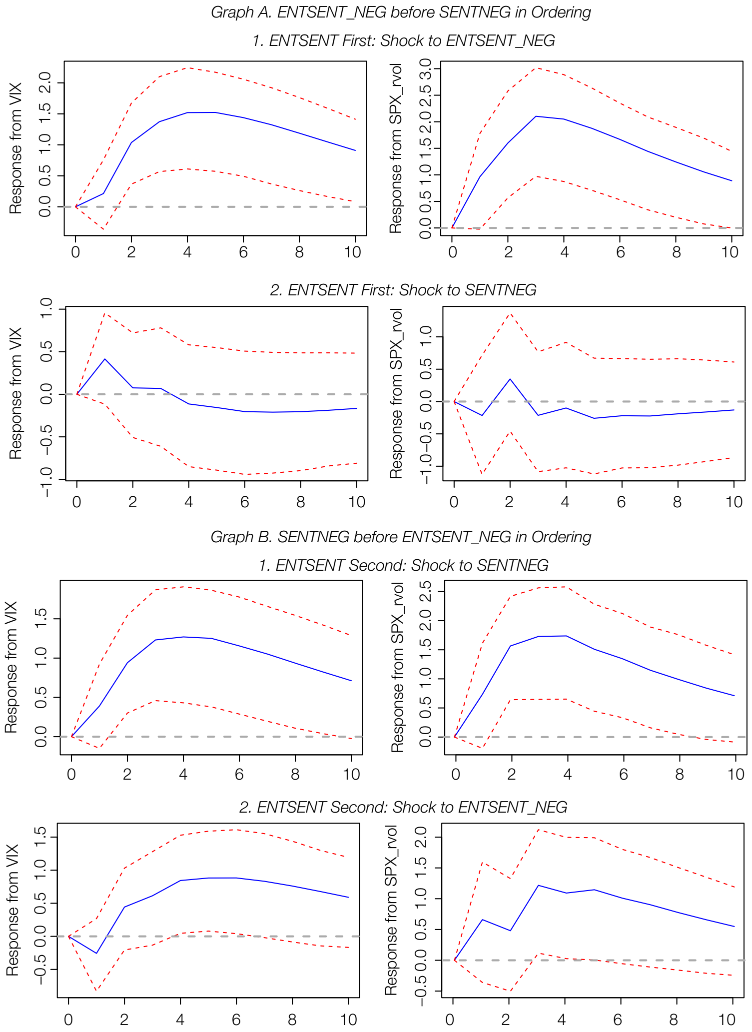
[24]Running the analysis in variance or log variance terms, with or without r^- as one of the model variables, does not change any of our results. We focus on the volatility model that excludes r^- for simplicity.

[25]We used the **R** package vars for the VAR estimation and impulse response functions (see Pfaff (2008)).

FIGURE 5

Aggregate VARs: Impulse Responses to Negative News

Impulse response functions for a shock to ENTSENT_NEG and SENTNEG. The order of the variables in the VAR model of Graph A is VIX, SPX_RVOL, ENTSENT_NEG, SENTNEG, ENTSENT_POS, and SENTPOS. Graph B interchanges the order of the 3rd and 4th, and 5th and 6th variables. Dashed lines show 95% bootstrap confidence intervals. The horizontal time axis is in months.



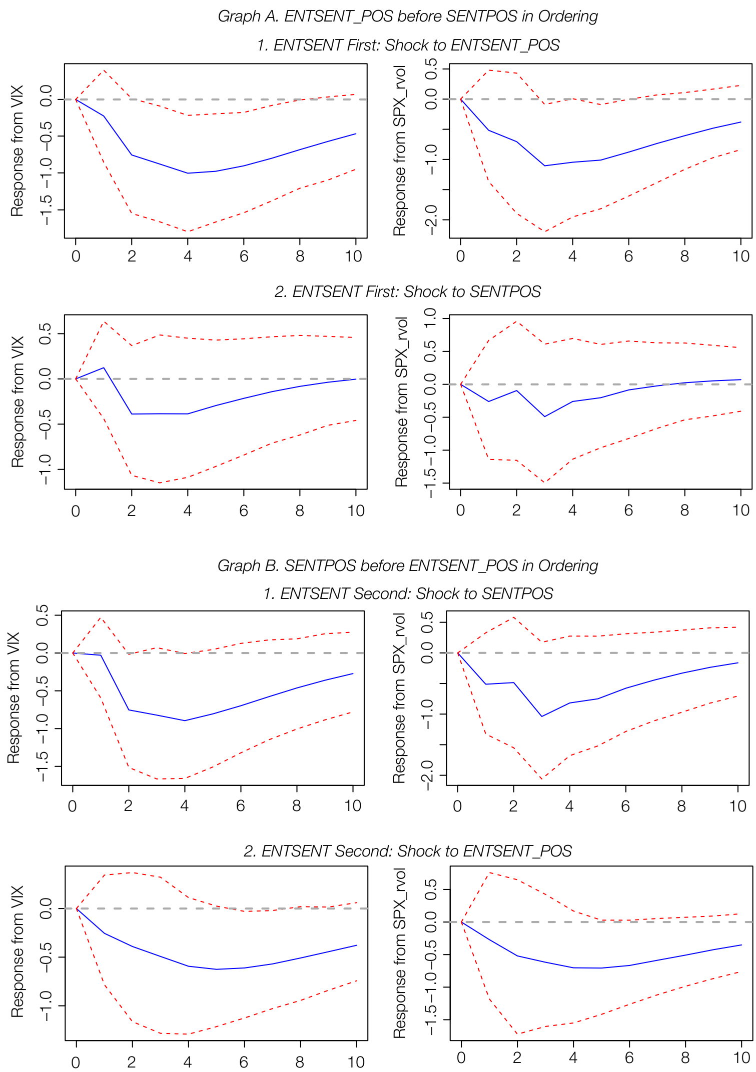
the interaction term, but there is an effect from the interaction term that is not captured by negative sentiment alone. This is consistent with our findings in the company-specific regressions of Section IV.

Figure 6 shows that a similar pattern holds for positive sentiment and its interaction with entropy. A shock to the interaction variable ENTSENT_POS has a statistically significant (negative) effect on both VIX and realized volatility when it is listed before SENTPOS (Graph A1), whereas SENTPOS does not

FIGURE 6

Aggregate VAR: Impulse Responses to Positive News

Impulse response functions for a shock to ENTSENT_POS and SENTPOS. The order of the variables in the VAR model of Graph A is VIX, SPX_RVOL, ENTSENT_NEG, SENTNEG, ENTSENT_POS, and SENTPOS. Graph B interchanges the order of the 3rd and 4th, and 5th and 6th variables. Dashed lines show 95% bootstrap confidence intervals. The horizontal time axis is in months.



(Graph A2). When the order of the variables is interchanged, SENTPOS has a statistically significant effect on VIX (Graph B1), and ENTSENT_POS has a marginally significant effect on both VIX and realized volatility (Graph B2). As one would expect, the magnitudes of the responses to the positive signals are smaller than the responses to the negative signals, but the overall pattern is similar. The pattern suggests that both positive sentiment and its interaction with entropy influence volatility, and that the interaction term captures an effect that is not present in the sentiment variable alone.

The time horizon of the impulse responses is also noteworthy. Consider, for example, the two responses in Graph A1 of Figure 5. They show that the effect on volatility of an increase in ENTSENT_NEG plays out over months, peaking around 4 months after the shock and dissipating slowly. As we will show in Section VI, single-name impulse responses estimated from a panel VAR operate over a shorter, but still multi-month time horizon. These time scales are markedly different from those in prior work using news sentiment to predict returns (including Da, Engelberg, and Gao (2014), Jegadeesh and Wu (2013), Tetlock (2007), and Tetlock et al. (2008)), where effects play out over days. In other words, directional information is incorporated into prices within days, but signals forecasting elevated volatility can remain relevant for months.

Volatility is of course much more persistent than returns are, but this property is insufficient to explain the volatility responses in Figures 5 and 6. Including implied and realized volatility in the VARs controls for persistence. Although persistence of volatility could make a predictor of high volatility in the present a predictor of high volatility in the future, the impulse responses of VIX and realized volatility to the news variables are consistently hump-shaped wherever they are statistically significant. The responses at month 4 are therefore not simply lingering effects of a larger response in month 1, as persistence by itself would predict.

Another possibility is that longer-dated options contain information about several-months-ahead volatility that the short-dated VIX does not. To test for this, we reran our VARs but included the Mid-Term VIX (VXMT) index, which tracks 6- to 9- month-ahead implied volatility for the S&P 500. The inclusion of VXMT had no impact on our hump-shaped volatility responses to news shocks. As an additional robustness check, we also estimated our VAR model in the pre-crisis subperiod and found the results to be very similar to those of the full-sample. The impulse responses from both sets of VARs (i.e., those with VXMT and those from the early subsample) are provided in the Supplementary Material.

C. Augmented Put Selling Strategy

As further evidence of the significance of our news measures, we analyze their economic value in the context of a simple put selling strategy.[26] In the baseline strategy, we sell 1-month S&P 500 puts, struck at either 90%, 95%, or 100% of the closing index level on the 3rd Friday of every month (which is when S&P 500 options expire). The account capital and sales proceeds are invested at the U.S. 3-month T-bill rate. The position is held for 1 month until the option expires. Upon expiry, if the put is out of the money, nothing happens; if it is in the money, the account realizes a capital loss on the S&P 500 index position it is forced to buy at the strike price. A position in the new 1-month put option is then established using the account's new capital base. The number of puts sold is such that, upon exercise, the cash in the trading account is exactly enough to cover the required

[26]Put selling has been widely studies in the literature (e.g., Bollen and Whaley (2004), Broadie, Chernov, and Johannes (2009)).

premium (i.e., we sell puts on an unlevered basis). Because this strategy trades very liquid S&P 500 options, and transacts only once per month (since there is no need to cover the expiring options), the associated transaction costs are low. Our S&P options data, obtained from Bloomberg, start in 2005.

To examine the economic benefit of using our text-based measures, we analyze a very simple modified strategy, where a put is sold on the 3rd Friday of month t only if the month $t-1$ ENTSENT_NEG was below the θ percentile of ENTSENT_NEG over the prior 36 months. If a put is not sold, the account stays entirely in cash in that month. In months in which the $t-1$ ENTSENT_POS was above its rolling θ percentile, we double the number of puts sold relative to the baseline (this strategy therefore uses leverage). In months where neither condition holds, we follow the baseline strategy. Note that the strategy is completely out-of-sample as it employs 3-week old, and for a given threshold θ, the strategy could have been implemented by investors.

Since we sell puts at the market price, the information content of implied volatility is embedded in the trading strategy. Therefore, if staying out of (increasing the position in) the market when aggregate unusual sentiment is negative (positive) is advantageous, it must be that implied volatility is not properly accounting for the information in our textual measure. Table 11 reports the mean excess returns, Sharpe ratios and alphas (all annualized) of the baseline strategy, as well as of the modified strategy where the out-of-market/double-exposure thresholds are set to the 70th, 80th, or 90th percentile of ENTSENT_NEG and ENTSENT_POS over months $t-36$ through $t-1$. Returns for all strategies are scaled to have a 10% annual volatility, in order to make comparison between them meaningful. The alpha is calculated using the Fama–French (2015) 5-factor model augmented with momentum.[27] Panel A reports results for the full sample, and Panel B reports

TABLE 11

Performance of Selling 90-Strike Puts

Table 11 reports the mean excess returns (MeanXR), Sharpe ratios (SRs), and alphas (all annualized) relative to the Fama–French 5-factor model, together with momentum, of the baseline and modified strategies for the 90 strike puts. Panel A reports strategy performance in the full sample, from Jan. 2005 to Dec. 2014; and Panel B reports performance from Jan. 2009 to Dec. 2014. The baseline strategy is labeled "Base," and the text-based strategies are labeled with their percentile cutoff. In "Percentile θ," the θ refers to the percentile cutoff for negative (positive) entropy-sentiment at which the strategy stays out of (doubles its exposure to) the market. The percentile is calculated using a rolling 36-month window. t-statistics are shown in parentheses (using independence for the Sharpe ratio calculation and Newey–West with auto-lag selection for the alphas).

	Panel A. Full Sample			*Panel B. Post-Crisis Sample*		
Strategy	MeanXR	SR	Alpha	MeanXR	SR	Alpha
Base	2.077	0.211	−0.791	8.291	0.844	2.822
	—	(0.649)	(−0.333)	—	(1.747)	(1.112)
Percentile 0.7	9.123	0.928	7.630	9.144	0.931	3.310
	—	(2.408)	(3.888)	—	(1.874)	(1.497)
Percentile 0.8	8.430	0.858	6.986	7.946	0.809	2.359
	—	(2.276)	(3.996)	—	(1.693)	(1.130)
Percentile 0.9	9.095	0.926	7.714	8.813	0.897	3.234
	—	(2.403)	(3.764)	—	(1.826)	(1.279)

[27] Data are from Ken French's Web site (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html).

results for the post-crisis subsample (which starts in Jan. 2009). We report results for the 90-strike puts, though the results for the other strikes are similar.

As can be seen from Table 11, the modified strategy increases the Sharpe ratios (SRs) more than fourfold (the baseline SR is 0.21 and the 90th percentile strategy SR is 0.926), and also leads to an economically significant alpha differential (of up to 8.5% per year). While some of the outperformance of the strategy comes during the financial crisis, the post-crisis (2009 and after) performance of the modified strategies is still favorable relative to the baseline, though the differences are less stark. Given the lack of volatility in the post-crisis period, this lack of differentiation is perhaps not so surprising.

The trading strategy we examine is intentionally simple, and perhaps there are more subtle strategies than can better use the textual information. The main takeaway from our analysis is that the information content of aggregated news measures would have been extremely useful for investors over the 10-year period from the start of 2005 to the end of 2014. One explanation for this result is that the prices of puts, and thus implied volatilities, did not properly account for the information content of our news measures. But there is an alternative explanation. If the variance risk premium (VRP) systematically falls in response to bad news, then months characterized by high negative sentiment mark a particularly inauspicious time to sell put options. By staying out of the market in these months, the put selling strategy meaningfully outperforms the unconditional strategy. We comment further on the VRP in Section VI.C and in the Supplementary Material.

VI. Interpreting the Results

As we have shown in Sections IV and V, text based measures are useful for forecasting future implied and realized volatility at the single-name and macro levels, even after controlling for known predictors of volatility, such as current and lagged implied and realized volatilities, as discussed in the introduction to Section IV. If our text-based measures contain information that is useful for forecasting future volatility, why is this information not fully reflected in option prices?

The literature provides several possible hypotheses that are consistent with this observation, and we discuss alternatives at the end of this section. Before doing so, we present evidence that limits on investor attention or information processing capacity may explain the that we observe. In particular, investors may find it easier to interpret company-specific news, which requires reading relatively few articles, than to extract signals about the aggregate market from articles on individual companies, which requires reading an order of magnitude more articles. If so, then news should be reflected in prices more quickly at the single-name level than at the aggregate level. Furthermore, news that is explicitly about the overall market, again available from a relatively small set of articles that explicitly discuss market-wide events, should be absorbed more quickly than aggregate information that is dispersed across many company-specific articles. We present evidence supporting these two predictions.

Several studies have found evidence that the limits of human attention affect market prices. Dellavigna and Pollet (2009) find a less immediate response to earnings announced on Fridays than other days and explain the differences through reduced investor attention. Barber and Odean (2008) find that retail investors tend to be net buyers of high-attention stocks. Ehrmann and Jansen (2017) document changes in the comovement of international stock prices during World Cup soccer matches, when traders are presumably distracted. Huberman and Regev (2001) and Tetlock (2011) document striking stock market responses to "news" that was previously made public, and Solomon, Soltes, and Sosyura (2014) find that media coverage affects investors through salience rather than information. Hirshleifer, Hou, Teoh, and Zhang (2004) explain stock return predictability from accounting data through limited investor attention. Corwin and Coughenour (2008) find that attention allocation by market specialists affects transaction costs. Sicherman, Lowenstein, Seppi, and Utkus (2016) document patterns of investor attention in response to market conditions. Daniel, Hirshleifer, and Teoh (2002) explain a broad range psychological effects on markets through limited attention.

Even if investors had used natural language processing techniques to extract aggregate information from articles about individual companies,[28] which would have ameliorated their capacity constraint, they are likely to have extracted a different subset of the relevant information than what is captured by our aggregate unusual news measure.[29] However, at the single-name level, an analyst who read all the articles about a given company would likely have extracted virtually all relevant information, and is likely to have been aware, in a qualitative sense, of the information content of our unusual sentiment measure. In this case, we would still expect that the information content of i) stock-specific news and ii) market-specific news, with the latter also obtainable from a small set of articles, to be incorporated into prices more quickly than the specific information captured by our aggregate news measure.

A. Single-Name versus Aggregate Responses to News

If investor inattention is indeed the mechanism underlying our results, we should find that when relevant information is harder to gather and interpret, and thus requires more investor attention, text based measures should contain more incremental forecasting ability. To investigate this further, we compare our single-name results (where gathering and interpreting relevant information may be easier) from Section IV to our aggregate results (where information is more voluminous and has implications that are harder to interpret) from Section V. To make the analyses from these two sections comparable, we replicate our VAR specification for implied volatility (VIX) and realized volatility (SPX_RVOL) in our single-name panels. The full regression model is given by

$$
\begin{aligned}
(10)\quad \mathrm{VOL}^j(t) \;=\; & \gamma_0^j + \gamma_1'\mathcal{L}_2\mathrm{RVOL}^j(t) + \gamma_2'\mathcal{L}_2\mathrm{IVOL}^j(t) \\
& + \gamma_3'\mathcal{L}_2\mathrm{SENTPOS}^j(t) + \gamma_4'\mathcal{L}_2\mathrm{ENTSENT\_POS}^j(t) \\
& + \gamma_5'\mathcal{L}_2\mathrm{SENTNEG}^j(t) + \gamma_6'\mathcal{L}_2\mathrm{ENTSENT\_NEG}^j(t) + \epsilon^j(t),
\end{aligned}
$$

[28]We thank a referee for suggesting this thought experiment.

[29]Unless they happened to have used a very similar methodology.

where t indicates months. The left-hand side variable $\text{VOL}^j(t)$ is either realized or implied volatility for single names, and either VIX or SPX_RVOL for the aggregate series. Model 9 from Tables 7 and 8 shows estimates of equation (10) for the single-name panels with implied and realized volatilities as the left-hand side variables, respectively.[30] To complete the panel VAR model, we also run equation (10) with each of the 4 news variables on the left. Additional details about model estimation are available in the Supplementary Material.

1. Impulse Response Functions

Graph A in Figure 7 shows the impulse responses to a shock to ENTSENT_NEG (Graph A1) and SENTNEG (Graph A2) in the panel VAR. As in the aggregate VAR (see Figure 5), we see significant responses of implied and realized volatility (Graph A1) to a shock to ENTSENT_NEG. However, the response now peaks much sooner, at 2 months for implied volatility and at 1 month for realized volatility. In fact, if we omit time 0 (the time of the initial shock, at which our variable ordering forces the volatility responses to be 0), the response of realized volatility is better described as declining rather than hump-shaped. These results suggest that the information in our news measures is absorbed more quickly at the company-specific level than at the aggregate level.

Comparing Graph A2 of Figure 7 with Graph A1, we see that a shock to SENTNEG does not produce a significant response in either implied or realized volatility. In other words, consistent with what we saw in the aggregate case, it is the interaction of unusualness with negative sentiment that yields an increase in volatility.[31]

Graph B in Figure 7 shows the impulse response to a shock to ENTSENT_POS (Graph B1) and SENTPOS (Graph B2) in the panel VAR. Comparing this to the aggregate VAR (see Figure 6), we again see the same pattern: single-name realized and implied volatility react more quickly to firm-level news innovations than aggregate volatility responds to aggregate news. The volatility responses for negative and positive news shocks are both consistent with Samuelson's dictum (Jung and Shiller (2005)): Investors react more quickly to stock-specific information than to macro information.

The theory of rational inattention, as developed in Sims (2003), (2015) and Maćkowiak and Wiederholt (2009), predicts that when agents are constrained in their information processing capacity, a variable of interest (e.g., a price) will show a hump-shaped impulse response to information. The tighter the constraint faced by agents, the slower the rise in the hump; without a constraint, the response declines monotonically. Viewing our aggregate and panel VARs through this lens suggests that investors face constraints in processing volatility-relevant

[30]Comparing this to model 8, we see that dropping ARTPERC and r^- from this regression has very little effect on the remaining coefficient estimates (the latter observation is consistent with our finding from footnote B that r^- has very little impact when added to our VARs).

[31]If we switch the order of the variables ENTSENT_NEG and SENTNEG, as we did in Graph B of Figure 5, the response to a shock in SENTNEG becomes significant, but the response to a shock in ENTSENT_NEG remains significant, reinforcing the importance of the interaction. Those results are omitted for brevity, and are available in the Supplementary Material.

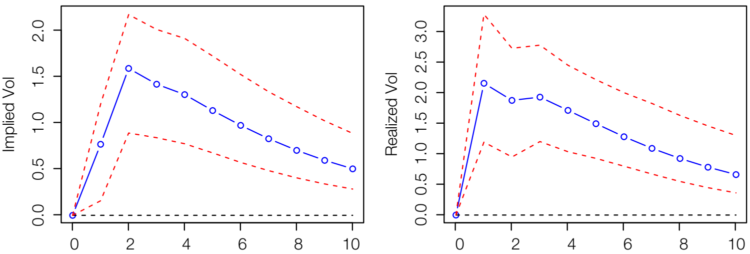
FIGURE 7

Company-Level Panel VAR

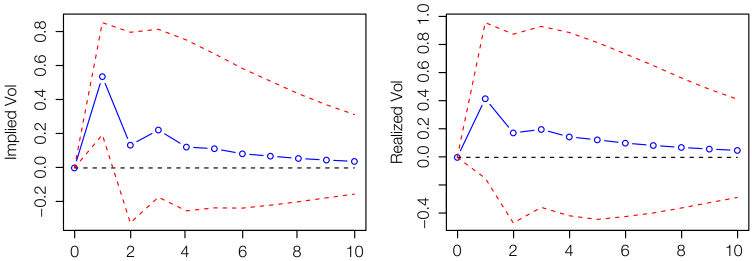
Impulse response functions for a shock to ENTSENT (Graphs A1 and B1 of Figure 7) and sentiment (Graphs A2 and B2) in the company-level panel VAR. Negative news shocks are shown in Graph A and positive news shocks in Graph B. The order of the variables in the VAR model is implied volatility, realized volatility, ENTSENT_NEG, SENTNEG, ENTSENT_POS, and SENTPOS. Dashed lines show 95% bootstrap confidence intervals. The horizontal time axis is in months.

Graph A. Impulse Responses to Negative News

1. ENTSENT_NEG Shock: ENTSENT First

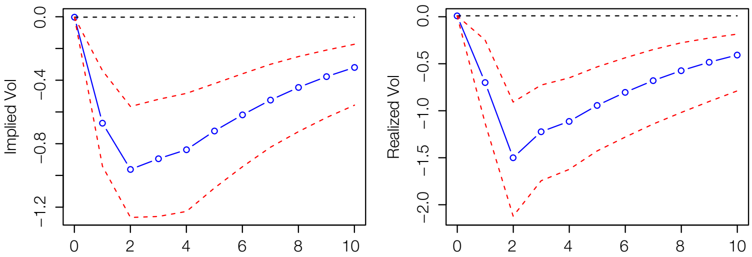


2. SENTNEG Shock: ENTSENT First

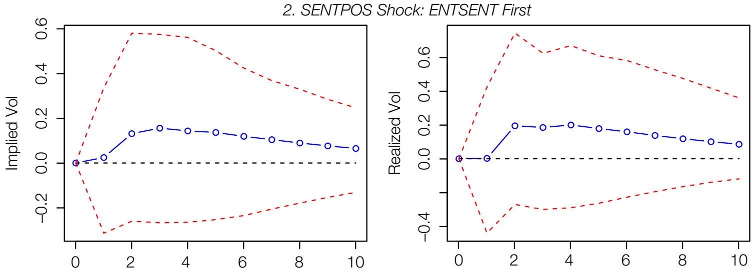


Graph B. Impulse Responses to Positive News

1. ENTSENT_POS Shock: ENTSENT First



2. SENTPOS Shock: ENTSENT First



information, and that extracting information about aggregate volatility requires more capacity than interpreting news about company-specific volatility.[32]

B. Market-Specific News

Capacity constraints faced by investors may prevent the market from quickly incorporating macro news when such news is dispersed among news stories about many different companies. However, what if there is a source of macro news that summarizes all relevant, perceived by the journalist, macro information in a single article? Then capacity constraints are much less likely to prevent the timely absorption of such market-specific news into the market.[33]

To test for this, we select all articles from the Thomson Reuters news archive that satisfy the following criteria: English language, about the United States, about stocks, and that reference the S&P 500 or the Dow Jones Industrials Index.[34] Because we are interested in market-specific news, we only accept articles that mention at most 4 tickers, that is, the SP and the Dow and at most 2 others. We therefore exclude articles that are largely about individual stocks. This yields roughly 38,000 articles, or 191 per month. Following the financial crisis, the average number of monthly articles falls to 90. While it is still unlikely that any one investor reads this many market-specific articles per month, this is a much smaller set of articles than in our company-specific data set. Furthermore, relative to company-specific articles, there may be more commonality across the market-specific articles, which means investors would not need to read all of them to extract the needed information.[35]

We calculate the average monthly positive and negative sentiment and entropy across all market-specific articles in a given month.[36] We calculate ES_NEG_MKT (ES_POS_MKT) as the product of the monthly negative (positive) sentiment series with entropy, all from our market-specific news set. We then run the same VAR (with 2 lags) as before, but include the series VIX, SPX_RVOL, ES_NEG_MKT, ENSENT_NEG (our bottom-up series), ES_POS_MKT, and ENSENT_POS (our bottom-up series), in that order. By putting the market-specific entropy-sentiment series ahead of our bottom-up measures in the VAR, we give the former the best chance of producing significant impulse response functions. Figure 8 shows the resulting impulse response functions.

Even when they are second in the VAR ordering, our bottom-up entropy-sentiment measures have the same hump-shaped response as in our core analysis. The market-specific series, though first in the ordering, elicit non-significant responses from implied and realized volatilities, with the response to ES_POS_MKT being (counter-intuitively) weakly, though (not significantly)

[32]An earlier version of this paper developed a more formal connection between our empirical results and the theory of rational inattention. The details of that model are available from the authors.

[33]We thank a referee for suggesting this test to us.

[34]These filters correspond to Thomson Reuters subject codes `N2:LEN`, `N2:US`, `N2:STX`, `R:.SPX`, and `R:.DJI`.

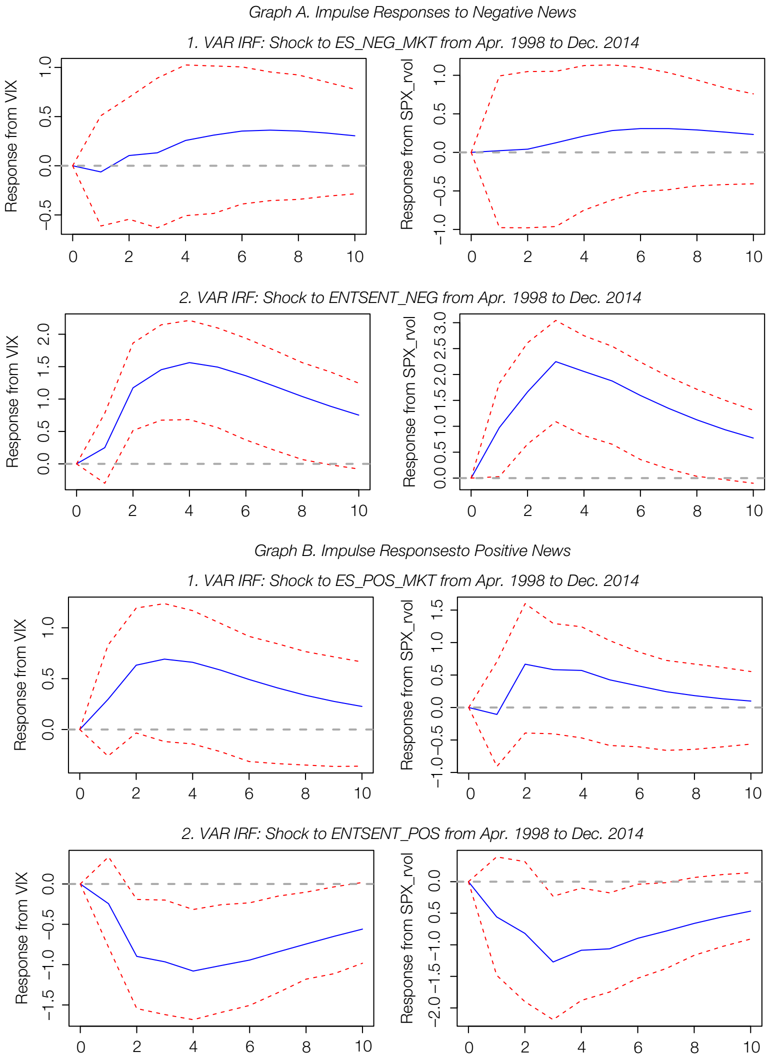
[35]We experimented with keeping a smaller set of articles filtered on word count, but found that this did not meaningfully affect our results.

[36]Our article-level negative (positive) sentiment measure is the number of negative (positive) words divided by the total words using the Loughran–McDonald dictionary. The article entropy calculation mirrors our n-gram based measure.

FIGURE 8

Impulse Response from VAR Including Market-Specific News

Impulse responses functions for a shock to market-specific news (Graphs A1 and B1 of Figure 8) and ENTSENT measures (Graphs A2 and B2), with negative news shocks in Graph A and positive news shocks in Graph B. Dashed lines show 95% bootstrap confidence intervals. The horizontal time axis is in months.



positive. We conclude from this analysis that our bottom-up series constructed from 50 large global financial firms capture information that would not have been readily available to investors who only followed (even an extraordinarily large number of) market-specific news. This finding is consistent with capacity constrained investors who are unable to quickly process all macro-relevant news flow from disaggregated sources.

Another potential explanation for this phenomenon is that there are important spillover effects from news about large financial firms, captured by our aggregate news measures, to the returns of all other firms in the economy because of the unique role played by financial intermediaries.[37] Therefore the information content of our aggregate news measure may be distinct from that of market-specific news, and aggregating news specifically originating from the finance sector may be particularly important. We believe this is potentially part of the story, and leave a resolution of this important question to future work.

C. Alternative Explanations

Although we find the patterns in our impulse response functions suggestive of an explanation through limited attention, the evidence is not conclusive, so we briefly comment on other potential explanations. These fall into two broad categories: behavioral explanations or explanations based on a time-varying risk premium.

Beyond inattention, potential behavioral explanations for underreaction include conservatism or salience bias (Daniel et al. (2002)), which cause investors to either undervalue or not observe new information. Investors may also be subject to noninformational biases, such as the disposition effect (Frazzini (2006)), which causes them to delay trading due to factors (e.g., whether they have unrealized losses in options) which we do not observe.

It is also possible that the predictability we observe in implied and realized volatility is really a manifestation of a time-varying VRP which our text measures are able to forecast. In the Supplementary Material, we run our VAR analysis replacing implied and realized volatility with the difference between implied volatility in month t and future realized volatility in month $t+1$. This difference is a rough measure of the variance (or volatility) risk premium. We find that the response of this difference to an increase in ENSENT_NEG is a sharp decrease. This would be consistent with our news measures forecasting a change in the variance risk premium. Moreover, Bollerslev, Tauchen, and Zhou (2009) document that higher values of the VRP forecast higher returns, so the drop we see is consistent with negative news forecasting lower returns.[38] This pattern may also help explain the performance of the trading strategy in Section V.C.

Also suggestive of a time-varying risk premium is that our text-based forecasting variables have similar predictive power for future volatility to variables that have been found to forecast the equity risk premium. The incremental drop in the R^2 of the aggregate and panel versions of equation (10) from removing our text measures, on the order of 0.5% to 3.5%, is comparable to the predictability documented in Fama and French (1988), Campbell and Thompson (2008), and Welch and Goyal (2008). See the Supplementary Material for more details.

VII. Conclusion

Using techniques from natural language processing, we develop a methodology for classifying the degree of "unusualness" of news. Applying our measure

[37] We thank the editor for suggesting this interpretation.

[38] We thank a referee for suggesting that we investigate this link.

of unusualness to a large news data set that we obtain from Thomson Reuters, we show that unusual negative and positive news forecast volatility at both the company-specific and aggregate level. News shocks are impounded into volatility over the course of several months. This is a much longer time horizon than previous studies have documented, which have focused on returns rather than volatility.

Across multiple analyses, we find that interacted measures of unusualness and sentiment provide the best predictors of future volatility among the news measures we study:

i) Our interacted measures remain significant when we control for other predictors of single-name volatility (lagged volatility and negative returns), indicating that the information in these news measures is not fully reflected in contemporaneous option prices or realized volatility. In panels that include the interacted and noninteracted news measures, only the interacted news measures are economically and statistically significant.

ii) At the aggregate level, we run vector autoregressions of the VIX and realized market volatility with several aggregate news variables. Impulse response functions show that a shock to our interacted measure of unusual negative (positive) news predicts an increase (decrease) in implied and realized volatility over several months. The effect is stronger for our interacted variable than for negative or positive sentiment alone.

When comparing our single-name and macro results, we show that our news measures have more incremental information content at the macro than the micro level. This suggests that market participants have a harder time incorporating the macro component of company-specific news than company-specific news themselves. Furthermore, we find that news shocks affect realized and implied volatilities in a hump-shaped manner over time, with the peak occurring later at the macro level than the micro. Finally, we find that market-specific aggregate news, which is more readily accessible to investors than our bottom-up aggregate news measures, does not similarly forecast future realized or implied volatilities. The pattern of responses we find indicates that news is not absorbed by the market instantaneously, and the macro component of company-specific news is absorbed more slowly than the micro component. We argue that this type of response is consistent with investors who face constraints on the rate at which they can process information and who process micro information more easily than macro information.

Appendix A. Details of Entropy Calculation

It is possible that a given 4-gram that was observed in month t never occurred in our sample prior to month t. In this case, $m_i(t)$ is either 0 (and thus its log is infinite) or undefined if its associated 3-gram also has never appeared in the training sample. To address this problem, we modify our definition of $m_i(t)$ in equation (4) to be

$$m_i(t) \equiv \frac{\tilde{c}(\{w_1 w_2 w_3 w_4\}) + 1}{\tilde{c}(\{w_1 w_2 w_3\}) + 4}.$$

This means that a 4-gram/3-gram pair that has never appeared in our sample prior to t will be given a probability of 0.25. The value 0.25 is somewhere between the 25th percentile and the median $m_i(t)$ among all our training sets. For frequently occurring 4-grams, this modification leaves the value of m_i roughly unchanged. Jurafsky and Martin (2008) discuss many alternative smoothing algorithms for addressing this sparse data problem, but because of the relatively small size of our training corpus, many of these are infeasible.

We approximate $\tilde{c}(\{w_1 w_2 w_3 w_4\})$ in a training window by only counting the occurrences of those 4-grams which are among the most frequently occurring 5,000 in every month. We therefore underestimate 4-gram counts, especially for less-frequently occurring n-grams, and therefore the m_i's associated with low p_i's are biased downward. However, because $p \ln p$ tends to 0 for small p, this is unlikely to have a meaningful impact on our entropy measure. Indeed, across the 228 months in our sample, the maximum least-frequently observed n-gram empirical probability is only 0.012%. Rerunning the analysis using the top 4,000 or 10,000 n-grams in each month leaves our results largely unchanged, suggesting the analysis is not sensitive to this issue.

Our results are not very sensitive to the following modeling assumptions, discussed here and in Section II.C: setting unobserved m_i's to 0.25; having the rolling window for training as 2 years; the choice of 3 months for the training window offset; and the number of n-grams we use to estimate $\tilde{c}$.

Appendix B. Data Cleaning

Appendix B summarizes our data cleaning methodology. Further details are available from the authors.

Articles whose headlines begin with `REG-` (regulatory filings) and `TABLE-` (data tables) are deleted. The `reuters` tag at the start of an article and in the end-of-article disclaimer is removed, as is any additional post article information identifying the author of the article. Punctuation characters (`,` or `;` and so on) and quotation marks are deleted, as are prefixes and suffixes that are followed by a period (e.g., `mr`, `corp`, etc.). All known references to any of the 50 companies in our sample are replaced with the string `_company_`.[39] Different references to the same, multi-word entity are replaced with a unique string. For example, all variations of `standard & poor's` are replaced with `snp`, references to `new york stock exchange` are replaced with `nyse`.

References to years, of the form `19xx-xx` or `20xx-xx` or similar forms, are replaced with `_y_`. We replace all numbers identified as being in the millions (billions) with `_mn_` (`_bn_`). Other numbers or fractions are replaced with `_n_`. The symbols `&` and `$` are deleted. All references to percent (e.g., `%` or `pct` or `pctage` etc.) are replaced with `pct`. We make an attempt to delete all references to e-mail addresses or Web sites, though we do not have a systemic way of doing so.

Following this text processing step, we use the NLTK package from Python to convert the raw text into n-grams. First `sent_tokenize()` segments the text into sentences. Then `word_tokenize()` breaks the sentence into single words. In this step, standard contractions are split (e.g., `don't` becomes `do` and `n't`). Finally `ngrams()` is used to create 3- and 4-grams from the post-processed, tokenized text.

Our n-grams undergo a further data-cleaning step. All company names (and known variations) are replaced with the string *_company_*. Phrases such as *Goldman Sachs reported quarterly results* and *Morgan Stanley reported quarterly results* are replaced with *_company_ reported quarterly results*, thus reducing two distinct 4-grams into a single one that captures the semantic intent of the originals. In this way we reduce the number of n-grams in our sample, which will allow us to better estimate conditional probabilities in our training corpus. In another example, we replace *chief executive officer* with *ceo* because we

[39] It is likely that we have not identified all possible references to companies in our sample.

would like the entity referred to as *ceo* to appear in n-grams as a single token, rather than a 3-word phrase.

Supplementary Material

Supplementary Material for this article is available at https://doi.org/10.1017/S0022109019000127.

References

Baker, S.; N. Bloom; and S. Davis. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics*, 131 (2016), 1593–1636.

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

Bekaert, G., and M. Hoerova. "The VIX, the Variance Premium and Stock Market Volatility." *Journal of Econometrics*, 183 (2014), 181–192.

Bisias, D.; M. D. Flood; A. W. Lo; and S. Valavanis. "A Survey of Systemic Risk Analytics." *Annual Review of Financial Economics*, 4 (2012), 255–296.

Bollen, N., and R. Whaley. "Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?" *Journal of Finance*, 59 (2004), 711–753.

Bollerslev, T.; R. Tauchen; and H. Zhou. "Expected Stock Returns and Variance Risk Premia." *Review of Financial Studies*, 22 (2009), 4463–4492.

Broadie, M.; M. Chernov; and M. Johannes. "Understanding Index Option Returns." *Review of Financial Studies*, 22 (2009), 4493–4529.

Campbell, J., and S. Thompson. "Predicting Excess Returns Out of Sample: Can Anything Beat the Historical Average?" *Review of Financial Studies*, 21 (2008), 1509–1531.

Corwin, S., and J. Coughenour. "Limited Attention and the Allocation of Effort in Securities Trading." *Journal of Finance*, 63 (2008), 3031–3067.

Da, Z.; J. Engelberg; and P. Gao. "The Sum of All FEARS Investor Sentiment and Asset Prices." *Review of Financial Studies*, 28 (2014), 1–32.

Daniel, K.; D. Hirshleifer; and S. H. Teoh. "Investor Psychology in Capital Markets: Evidence and Policy Implications." *Journal of Monetary Economics*, 49 (2002), 139–209.

Das, S. "Text and Context: Language Analytics in Finance." *Foundations and Trends in Finance*, 8 (2014), 145–261.

Dellavigna, S., and J. M. Pollet. "Investor Inattention and Friday Earnings Announcements." *Journal of Finance*, 64 (2009), 709–749.

Ehrmann, M., and D.-J. Jansen. "The Pitch Rather Than the Pit: Investor Inattention, Trading Activity, and FIFA World Cup Matches." *Journal of Money, Credit and Banking*, 49 (2017), 807–821.

Fama, E., and K. French. "Dividend Yields and Expected Stock Returns." *Journal of Financial Economics*, 22 (1988), 3–25.

Fama, E., and K. French. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics*, 116 (2015), 1–22.

Fang, L., and J. Peress. "Media Coverage and the Cross-Section of Stock Returns." *Journal of Finance*, 64 (2009), 2023–2052.

Frazzini, A. "The Disposition Effect and Underreaction to News." *Journal of Finance*, 61 (2006), 2017–2046.

Garcia, D. "Sentiment During Recessions." *Journal of Finance*, 68 (2013), 1267–1300.

Glasserman, P., and H. Mamaysky. "Investor Information Choice with Macro and Micro Information." Working Paper, Columbia University (2018).

Heston, S., and N. Sinha. "News versus Sentiment: Predicting Stock Returns from News Stories." *Financial Analysts Journal*, 73 (2017), 67–83.

Hirshleifer, D.; K. Hou; S. H. Teoh; and Y. Zhang. "Do Investors Overvalue Firms with Bloated Balance Sheets?" *Journal of Accounting and Economics*, 38 (2004), 297–331.

Huberman, G., and T. Regev. "Contagious Speculation and a Cure for Cancer: A Nonevent That Made Stock Prices Soar." *Journal of Finance*, 56 (2001), 387–396.

Jegadeesh, N., and D. Wu. "Word Power: A New Approach for Content Analysis." *Journal of Financial Economics*, 110 (2013), 712–729.

Jiao, P.; A. Veiga; and A. Walther. "Social Media, News Media and the Stock Market." Working Paper, University of Oxford (2016).

Jung, J., and R. J. Shiller. "Samuelson's Dictum and the Stock Market." *Economic Inquiry*, 43 (2005), 221–228.

Jurafsky, D., and J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd ed. Upper Saddle River, NJ: Prentice Hall (2008).

Kacperczyk, M. T.; S. Van Nieuwerburgh; and L. Veldkamp. "A Rational Theory of Mutual Funds' Attention Allocation." *Econometrica*, 84 (2016), 571–626.

Kogan, S.; D. Levin; B. Routledge; J. Sagi; and N. Smith. "Predicting Risk from Financial Reports with Regression." In *Proceedings of the North American Association for Computational Linguistics Human Language Technologies Conference* (May/June 2009), 272–280.

Kogan, S.; B. Routledge; J. Sagi; and N. Smith. "Information Content of Public Firm Disclosures and the Sarbanes-Oxley Act." Working Paper, University of Texas (2011).

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

Loughran, T., and B. McDonald. "Textual Analysis in Finance and Accounting: A Survey." *Journal of Accounting Research*, 54 (2016), 1187–1230.

Manela, A., and A. Moreira. "News Implied Volatility and Disaster Concerns." *Journal of Financial Economics*, 123 (2017), 137–162.

Maćkowiak, B., and M. Wiederholt. "Optimal Sticky Prices Under Rational Inattention." *American Economic Review*, 99 (2009), 769–803.

Nickell, S. "Biases in Dynamic Models with Fixed Effects." *Econometrica*, 49 (1981), 1417–1426.

Peng, L., and W. Xiong. "Investor Attention, Overconfidence, and Category Learning." *Journal of Financial Economics*, 80 (2006), 563–602.

Pfaff, B. "VAR, SVAR and SVEC Models: Implementation within R Package Vars." *Journal of Statistical Software*, 27 (2008), 1–32.

Poon, S.-H., and C. Granger. "Forecasting Volatility in Financial Markets: A Review." *Journal of Economic Literature*, 41 (2003), 478–539.

Routledge, B.; S. Sacchetto; and N. Smith. "Predicting Merger Targets and Acquirers from Text." Working Paper, Tepper School of Business, Carnegie Mellon University (2013).

Sicherman, N.; G. Lowenstein; D. J. Seppi; and S. Utkus. "Financial Attention." *Review of Financial Studies*, 29 (2016), 863–897.

Sims, C. "Implications of Rational Inattention." *Journal of Monetary Economics*, 50 (2003), 665–690.

Sims, C. "Rational Inattention and Monetary Economics." *Handbook of Monetary Policy*. Elsevier (2015).

Solomon, D. H.; E. Soltes; and D. Sosyura. "Winners in the Spotlight: Media Coverage of Fund Holdings as a Driver of Flows." *Journal of Financial Economics*, 113 (2014), 53–72.

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

Tetlock, P. "All the News That's Fit to Reprint: Do Investors React to Stale Information?" *Review of Financial Studies*, 24 (2011), 1481–1512.

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

Welch, I., and A. Goyal. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction." *Review of Financial Studies*, 21 (2008), 1455–1508.