Why Do Investors Disagree? The Role of a Dispersed News Flow

Michal Dzielinski* Henrik Hasseltoft*

December 2013

Abstract

Using recent advances in news analytics, we construct an empirical measure of aggregate news dispersion and study how a dispersed news flow affects investors and aggregate stock returns. Our measure reflects the polarization of news across firms, based on millions of company-specific news items. We find that news dispersion i) predicts investor disagreement positively, ii) is positively related to turnover, iii) predicts aggregate stock returns negatively, and iv) predicts realized variance positively. The effects of news dispersion are consistent with models of disagreement and short-sales constraints and support the idea that a dispersed news flow represents a fundamental reason for why investors disagree.

Keywords: disagreement, news analytics, predictability, stock returns, volatility.
JEL Classification Number: G12, G14, G17

*Dzielinski is at the Stockholm University School of Business and the Swedish House of Finance, e-mail: mdz@fek.su.se, tel: +46-8-162840. Hasseltoft is at the University of Zurich and the Swiss Finance Institute, e-mail: henrik.hasseltoft@bf.uzh.ch, tel: +41-44-6344483. We would like to thank Seong Byun, Roger Edelen, Simon Lysbjerg Hansen, Davidson Heath, Thorsten Hens, Felix Meschke, Per Stromberg, Paul Tetlock, Michela Verardo, Alexandre Ziegler, and seminar participants at the 2013 Midwestern Finance Conference, the 2013 SFS Cavalcade, the 2013 Western Finance Association Meetings, the 2013 World Finance Conference in Cyprus, the 2013 European Finance Association Meetings, the 8th Imperial College London conference on Advances in the Analysis of Hedge Fund Strategies, the University of Southern Denmark, and the University of Zurich for helpful discussions. Dzielinski would like to gratefully acknowledge the financial support from "LGT and Science". We highly appreciate comments and suggestions of relevant papers that we inadvertently have not cited. This paper circulated earlier under the title "Aggregate News Tone, Stock Returns, and Volatility".

1 Introduction

Investor disagreement has become an important concept in models seeking to explain the joint behavior of stock returns, volatility, and trading volume. On the empirical side, several proxies for disagreement have been developed and the theoretical predictions generally validated. However, there is much less evidence for *why* investors disagree in the first place. We fill this gap by providing a natural rationale for disagreement − a highly dispersed information environment.

We consider dispersion at the market level and to measure it we tap into recently developed methods of quantifying language contained in financial news. We make use of a database provided by Thomson Reuters News Analytics which contains news-tone scores, obtained through linguistic analysis, for all company-specific news announcements released during the period Jan 2003-Dec 2012. Using this information, we first construct measures of news tone at the firm level. We then construct our final measure, the aggregate dispersion of news tone, as the cross-sectional standard deviation of firm-specific news tones. We consider both daily and monthly measures of news dispersion. Intuitively, news-tone dispersion increases as news becomes more polarized across firms.

The notion of polarized news fits naturally with dynamic models of disagreement such as for example Harrison and Kreps (1978), Scheinkman and Xiong (2003), and Hong et al. (2006). In these models, different groups of investors observe the entire news flow but the groups differ in the informativeness they attach to different signals, leading to differences in beliefs about future fundamentals. Such differences are bound to be rather small in the case of a homogeneous news flow but are likely to increase as the news flow becomes more polarized, leading to substantial disagreement. In general, the only (weak) assumption needed for a dispersed news flow to generate disagreement is that investors weigh news items differently when forming their outlook.[1] Hence, we argue that dispersion in news tone represents not a

[1] Possible reasons for why investors may assign different weights to news are heterogenous priors, limited

proxy but a fundamental *reason* for investor disagreement.

Consistent with this notion, we find that higher news dispersion predicts an increase in investor disagreement next period for various categories of market participants: equity analysts, professional forecasters, financial advisors and retail investors. A one standard deviation increase in news dispersion increases disagreement next period with 0.16-0.34 standard deviations, depending on disagreement measure. Moreover, an increase in news dispersion is associated with higher trading volume, supporting the theoretical notion that changes in disagreement should induce investors to trade. Furthermore, we find that news dispersion predicts aggregate stock returns negatively and volatility positively. In monthly regressions, a one standard deviation increase in news-tone dispersion predicts a 1.91% drop in S&P500 returns the following month, an economically significant effect.

These results are consistent with models of investor disagreement and short-sales constraints along the lines of Miller (1977) and others. In these models, stock prices mainly reflect the views of optimistic investors in times of high disagreement and binding short-sale constraints, leading to over-valued assets with correspondingly low expected returns.[2] Short-sale constraints on the aggregate level are likely to exist for other reasons than the cost of shorting. For example, institutional investors such as mutual funds are generally restricted from short selling and very few of the funds that are allowed to short actually go short in practice (e.g. D'Avolio, 2002 and Almazan et al., 2004). Short-sales constraints therefore often arise for various institutional, legal, or cultural reasons. As a consequence, a large fraction of investors with negative views rarely express them with a short position.

Rather than focussing on the cost of shorting as a proxy for short-sale constraints, we investigate whether news-tone dispersion affects the relative pricing of stocks depending on their short-sale availability and sensitivity to aggregate disagreement. First, we consider

attention, and confirmation bias.

[2] See for example Diether et al. (2002) and Chen et al. (2002) for cross-sectional evidence and Park (2005) and Yu (2011) for time-series evidence.

small versus large stocks. Asquith et al. (2005) document that small-cap stocks are more likely to be short-sale constrained than large stocks which implies that small stocks should be more over-priced in times of high investor disagreement and experience lower future returns. Indeed, we find that a rise in news-tone dispersion predicts an underperformance of small stocks next period. Second, we follow Nagel (2005) and consider the institutional ownership of a stock as a proxy for short-sales constraints. Stocks with low institutional ownership are expected to be more constrained. We find that news-tone dispersion predicts returns negatively on a portfolio that is long stocks with low institutional ownership and short stocks with high institutional ownership. Third, we consider high-beta versus low-beta stocks. Hong and Sraer (2012) argue that high-beta stocks are more sensitive to aggregate disagreement and therefore likely to be over-priced in times of high disagreement and binding short-sales constraints, leading to low future returns compared to low-beta stocks. We find that news-tone dispersion predicts negative returns on a portfolio that is long high-beta stocks and short low-beta stocks. These results further suggest that news dispersion operates via investor disagreement and short-sales constraints.

Our setup also allows us to explicitly test an unexplored consequence of investor disagreement, namely that bad news is effectively "ignored" since short-sales constraints prevent it from being fully reflected in stock prices.[3] The key insight is that if some stocks receive good news and others receive bad news, the optimists will set prices for all stocks given that short-sales constraints are binding for a meaningful fraction of investors. Consequently, firms with "ignored" bad news should be more overvalued with lower expected returns than stocks with good news. As the news flow becomes more dispersed, the difference in expected returns between good-news and bad-news stocks should increase. We find support in data for this novel prediction, adding to the body of empirical evidence in support of disagreement models.

[3] We thank Paul Tetlock for this suggestion.

In addition to returns, we also consider volatility and find that news-tone dispersion predicts future realized volatility positively controlling for a range of variables. In addition, the dispersion of news tone behaves similarly to volatility, being asymmetric and countercyclical. Continuing along the lines of disagreement, these findings support the recent model of Banerjee and Kremer (2010) in which investors disagree on public information and where a rise in disagreement increases volatility.

We see our main contribution in providing an economically sound and empirically tractable basis for investor disagreement, which does not rely on specific assumptions about heterogeneous priors, limited attention, or confirmation bias. The idea of a highly dispersed information flow leading to disagreement is both intuitive and strongly supported in the data. Our findings also suggest that news dispersion can offer an information-based explanation for why volatility changes over time, contributing to the literature on the determinants of volatility. As an additional benefit, our empirical measure of news dispersion can be constructed over any desired frequency, enabling the researcher to also measure high-frequency variations in investor disagreement.

Finally, we also contribute to the recent and growing literature on how soft information in news can be quantified and linked to asset prices. Tetlock (2007) analyzes the content of a popular market commentary section in the Wall Street Journal, and finds that pessimistic words predict low stock returns.[4] Davis et al. (2006), Engelberg (2008), Tetlock et al. (2008), and Demers and Vega (2011) all examine the tone of firm-specific news items and find that the level of firm-specific news tone predicts future firm-specific earnings and returns. We add to this literature by showing that the dispersion of news tone across firms contains valuable information about investors and aggregate returns. While the aggregate dispersion of news has considerable forecasting power, we find that the aggregate level of news tone

[4]Tetlock (2007) uses the Harvard dictionary to define positive and negative words. See Loughran and McDonald (2011) for a dictionary that is more suitable for financial texts.

fails to predict returns. This result stands in contrast to Tetlock (2007) who found that the level of news sentiment predicts daily returns. Our difference in results stems from the fact that Tetlock's measure is based on a news-paper commentary section that he found mostly reflected investor sentiment while our measure is built on millions of news items, that we find mainly reflect disagreement about fundamentals.

The layout of our paper is as follows. Section 2 describes the data. Section 3 links news dispersion to turnover and investor disagreement. Section 4 studies the predictive power of news dispersion for future returns and volatility. Section 5 explores potential alternative explanations for news dispersion. Section 6 presents robustness checks and Section 7 concludes.

2 Data

2.1 News-Tone Variables

The departing point is the collection of all company-specific news announcements obtained from Thomson Reuters News Analytics. This extensive archive contains all news published either by Reuters or by the companies themselves (via direct outlets like the PR Newswire) in the period between January 2003 and December 2012. Each time a company is mentioned in the news, its identifier (Reuters Instrument Code, or RIC) is recorded, together with a precise timestamp. In particular, this means that one record per company is created whenever a news announcement mentions several companies. This is important, because the presentation of each company might be different within the same news story, e.g. good news for Company A might be bad news for its competitors etc. Similarly, the relevance of the news story for each company might be different with for example Company A being the main focus of attention, perhaps already named in the headline, while its competitors are only briefly mentioned later in the text.

The linguistic analysis we make use of is capable of grasping such differences. The algorithm developed for this purpose by Thomson Reuters works at the sentence level, identifying the subject (company) and any tone-relevant words related to it. The two procedures it is based on, Named Entity Recognition (NER) and Parts-of-Speech Tagging (POS), have both become standard tools and are widely used in content analysis (e.g., Jurafsky and Martin, 2008). Thus, we are able to track the tone of the news for each company separately, even if they are mentioned in the same text.

Another advantage of this algorithm is that it attempts to make sense of syntactic relationship in determining the tone of the news, represented as a classification variable: +1 for positive, 0 for neutral, and -1 for negative tone. This can be achieved either by defining explicit grammatical rules (*deductive* algorithms) or by supplying a training set evaluated by human "teachers", from which the algorithm *inductively* infers the relevant rules. The News Analytics algorithm belongs to the second type but in both cases the potential gains with respect to the basic "bag of words" approach are substantial. Also, contrary to what is sometimes said, syntactic approaches are not any more subjective than "bag of words". In fact, surveys of methods of content analysis assign all of them to the family of "supervised approaches", indicating human involvement in their design. This is because the dictionaries, which are behind any "bag of words" analysis have to be created by humans. Even inductive algorithms offer a fair degree of inter-subjective reliability, because the learning sets are always evaluated by more than one person and the results of learning are only accepted when the agreement between the instructors and the machine and among the instructors themselves reaches a certain threshold. In the News Analytics database this is reflected by three "tone probability" scores, which show how likely each news item was to receive one of the three tone labels: positive, neutral or negative.

The basic building block of our aggregation is the news tone for company i on day t, which is computed as follows:

$$Tone_{i,t} = \sum_{k=1}^{kpos} 1 \cdot prob\_pos_{i,t,k} + \sum_{k=1}^{kneg} (-1) \cdot prob\_neg_{i,t,k}. \tag{1}$$

That is, all positive news items (indicated by +1) for company i on day t are multiplied by the probability of being classified as positive and summed and similarly for negative news items. Adding the two sums for positive and negative news produces a measure of the difference between the positive and negative content published about company i on day t. It will be positive if there were more positive news items and/or if the positive news items had a higher probability, $prob\_pos$, attached to them and negative otherwise. The greater the number of news about company i on day t the greater the potential magnitude of the news tone in the case of a significant imbalance between positive and negative news. For this reason, companies with a lot of news flow are more likely to register very high or very low values of news tone.

When constructing firm-specific news tones, we use the period starting at 4pm on calendar day $t-1$ (i.e. yesterday's close) and ending at 3:45pm on calendar day t (i.e. 15 minutes before today's close). This ensures that our measures reflect the information available to an investor wishing to trade on news tone before the market closes on day t.[5] Having computed daily news-tone scores for each company, it is then straightforward to compute monthly news tones for each firm as:

$$Tone_{i,m} = \sum_{t=1}^{T} Tone_{i,t}, \tag{2}$$

where T denotes the number of days in a given month.

2.1.1 The Dispersion of Aggregate News Tone

The dispersion of news tone, which is our main variable of interest, can be viewed as the second moment of news and captures a different dimension of the news flow than the level of

[5] It makes little difference to our results if one instead uses close-to-close information.

news tone. To see this, consider the following two situations: (1) there is little news overall entering the market, (2) there are plenty of highly polarized news items where companies with positive news items are offset by companies with negative news. The aggregate level of news tone, constructed by summing up news tones across firms, would in both cases be close to zero. However, the two cases are admittedly very different from the viewpoint of investors. For an investor wanting to form an opinion about the aggregate market, the first case represents few signals overall while the second case represents a large number of contradictory signals.

We define the daily (D superscript below) news-tone dispersion as the cross-sectional standard deviation of daily firm-specific news tones at date t :

$$AggDisp_t^D = Std(Tone_{i,t}). \tag{3}$$

We also construct a monthly (M superscript below) news-tone dispersion measure as the cross-sectional standard deviation of monthly firm-specific news tones:

$$AggDisp_t^M = Std(Tone_{i,m}). \tag{4}$$

2.1.2 The Level of Aggregate News Tone

Having computed measures of news dispersion, we also construct measures of the aggregate level of news tone. Our construction of the level of aggregate news tone makes use of the fact that the individual company tones in Equation 1 are additive. The daily aggregate level of news tone is defined as the average across daily firm news tones:

$$AggTone_t^D = \frac{\sum_{i=1}^{N} Tone_{i,t}}{N}, \tag{5}$$

where N denotes the number of firms that had news on day t. Since firm-specific news tones are not only affected by whether news are good or bad but also by the number of news items, large companies with a large news flow will generally contribute more to the aggregate measure, which is quite similar to using value weighting when constructing price indices. This ensures that our measure of the aggregate news tone is not driven by small and less relevant stocks. We also construct a monthly level measure. The monthly aggregate level of news tone is simply the average of monthly firm-specific news tones.

$$AggTone_t^M = \frac{\sum_{i=1}^{N} Tone_{i,m}}{N}, \tag{6}$$

where N equals the number of firms with a monthly firm-specific news-tone score.

2.1.3 Basic Properties of News-Tone Variables

As the majority of news items refers to the largest stocks, we focus on news for stocks included in the $S\&P$ 500 index. Figures 1 and 2 plot the dynamics of the daily and monthly measures of the aggregate level and dispersion of news tone for our sample period January 2003 to December 2012. We have winsorized the series at the 1st and 99th percentiles in order to mitigate the influence of large outliers. We have then standardized all series by demeaning and dividing by the standard deviation, in order to ease the interpretation. Two things are apparent from the pictures. First, the level of aggregate news tone (upper panel in both figures) exhibits substantial variation over time. The aggregate news tone increased leading up to the onset of the financial crisis in mid-2007. The level of news tone then dropped sharply in 2008 after which it rebounded in 2009. The drop in news tone in mid-2010 and mid-2011 coincides with the European sovereign debt crisis. Second, the level and dispersion (lower panel in both figures) of news tone appear negatively correlated. Dispersion is generally low when tone is increasing and registers several large spikes when tone is at its lowest. This suggests that company news is highly contradictory in bad times.

Figure 3 plots the aggregate level and dispersion of news tone against the VIX index and indicates a close relationship. The apparent negative correlation between aggregate tone and the VIX and the positive correlation between aggregate dispersion and the VIX persists not only during the financial crisis but also before and after. In fact, the graph suggests that aggregate dispersion of news tone often leads the VIX, in particular for the large changes during the sample period.

2.2 Hard-Information Variables

While both the aggregate dispersion and level of news tone reflect soft-information contained in financial news, we include a broad range of control variables which reflect hard information in the form of economic and financial conditions. Our base-line results are based on monthly data wherefore this section presents our monthly data sources. We later on describe our daily data sources in Section 4.5. Our first control variable is the variance risk premium (VRP), defined as the difference between implied and expected realized variance, and which is often interpreted as a measure of economic uncertainty. Bollerslev et al. (2009) demonstrate that the variance risk premium has predictive power for stock returns. We construct VRP as the difference between $VIX^2/12$ and expected realized variance. The latter is obtained from regressing the realized variance onto lagged realized variance and lagged squared VIX. This is similar to for example Drechsler and Yaron (2011) and Bali and Zhou (2012).

As a monthly measure of economic activity we use the Chicago-Fed National Activity index which is obtained from the Federal Reserve Bank of St. Louis. We also compute growth in actual aggregate earnings using data from IBES. Data on the VIX index is downloaded from Datastream. Finally, we relate our main variable, aggregate news-tone dispersion, to measures of investor sentiment computed by Baker and Wurgler (2006, 2007), and which are obtained from the website of Jeffrey Wurgler.

Aggregate stock market returns are measured using returns on the highly liquid exchange-traded fund that tracks S&P 500, SPY, and which represents returns that are obtainable by investors in practice.[6] The returns are obtained from CRSP. Monthly excess returns are computed using Fama's one-month Treasury-bill rate obtained from CRSP. We also use monthly data on realized variance of the $S\&P$ 500 downloaded from the website of Hao Zhou (see Zhou, 2010). We also construct a measure of return dispersion, being the cross-sectional standard deviation of monthly returns for all stocks in our sample. Return dispersion controls for whether the dispersion in news tone merely reflects the fact that some companies were performing well and others poorly in the past.

2.3 Summary Statistics

Table 1 reports monthly summary statistics and correlations. The aggregate level of news tone has a negative skewness and a relatively high persistence of 0.82 in first-order autocorrelation while news-tone dispersion is positively skewed with a lower persistence of 0.56. The persistence decays to around 0.30 for both measures when using five lags. In general, the two news-tone measures display lower persistence than commonly used predictive variables such as price-earnings ratios. This implies that our predictive regressions are less subject to biases that may arise from highly persistent variables. The ADF test statistics suggest that we can reject the null hypothesis of a unit root for all variables.

The second panel of Table 1 demonstrates that the level and dispersion measures are highly negatively correlated, -0.63, and have opposite relations to stock returns and volatility. While the level of news tone is positively related to stock returns and negatively related to volatility, higher dispersion is instead related to lower returns and higher volatility. The two news-tone measures also carry significant correlations with a range of economic variables where a higher news-tone level (dispersion) is associated with higher (lower) economic activity and higher (lower) earnings growth. Simply put, higher level of news tone indicates

[6] Our results are robust to instead using a broad stock-market portfolio from CRSP.

good economic times, high returns, and low volatility while higher dispersion indicates bad economic times, low returns, and high volatility.

3 News Dispersion and Investor Disagreement

In this section, we verify our interpretation of news dispersion as representing a fundamental reason for why investors disagree about the future. First, we link news dispersion to turnover since explaining turnover is a central feature of all dynamic models of disagreement. Second, we analyze how news dispersion is related to several empirical proxies of disagreement. If it really is the case that a dispersed news flow gives rise to future disagreement among investors, then we would expect news dispersion to predict future investor disagreement with a positive sign. We also evaluate whether commonly used proxies of investor disagreement in fact can explain the trading volume we see in data.

Based on recent literature, we construct five different proxies of investor disagreement. We focus on proxies that are intended to capture disagreement about the aggregate outlook to match our news-dispersion variable. Two of the proxies are constructed from analyst forecasts. The first proxy is a variation of the Park (2005) measure of dispersion in forecasts of aggregate earnings of S&P500 constituents, provided by the so-called "strategist" analysts. We cast it in terms of growth rates rather than levels of earnings to eliminate the need for scaling. Specifically, for every month (t) for each outstanding forecast (i) of the level of earnings at the end of (annual) fiscal period T we define:

$$EarnGr_{i,t}^{T-1,T} = \frac{Forecast_{i,t}^{T}}{Reported^{T-1}} - 1, \tag{7}$$

as the growth rate of earnings forecasted by that analyst.[7] The disagreement, $Disp_{AggEarn}$, is then measured as the cross-sectional standard deviation of forecasted growth rates across

[7] We apply the same adjustment as in Park (2005) to maintain a fixed 12-month forecasting horizon.

all analysts in a given month. The second proxy for investor disagreement is defined as the average of the firm-level dispersion in analyst forecasts of the long-term growth rate of earnings, $Disp_{LTG}$, and has been originally proposed by Yu (2011). Notably, the second proxy is a bottom-up measure of analyst dispersion in contrast to the first proxy and to our news-dispersion measure which both are top-down measures.

The following two proxies are derived from surveys of the market outlook for two distinct groups of agents. First, we consider the American Association of Individual Investors Sentiment Survey which measures the fraction of investors that are bullish, bearish, and neutral on the aggregate stock market for the next six months. We construct a measure of disagreement as the absolute difference between the fractions of bulls and bears, normalised by the sum of bulls and bears: $Disp_{AAII} = 1 - \frac{|\%Bullish - \%Bearish|}{\%Bullish + \%Bearish}$. This ratio is constrained between 0 and 1 and a value close to zero, meaning a close to 50-50 split between bulls and bears, arguably represents large disagreement. For ease of interpretation, we take 1 minus the ratio, so that the value of the final proxy increases with disagreement. Second, we consider data from the Advisor Sentiment Survey provided by Investors Intelligence which reflects the outlook of over 120 different writers of financial newsletters. The provider of the survey classifies each newsletter as being bullish, bearish, or neutral about the stock market and computes the corresponding fractions of bulls and bears. We compute the implied disagreement in the same way as above: $Disp_{II} = 1 - \frac{|\%Bullish - \%Bearish|}{\%Bullish + \%Bearish}$.[8] Finally, we compute the dispersion of annual GDP-growth forecasts, $Disp_{GDP}$, based on data from the Survey of Professional Forecasters. This measure is only available at a quarterly frequency.

[8]Both the AAII and the II surveys are conducted weekly. In order to construct monthly measures we simply compute the monthly average of the fractions of bulls and bears before computing our dispersion measures.

3.1 Explaining Trading Volume

As mentioned earlier, explaining turnover is a key feature of all dynamic models of disagreement and therefore represents a key empirical test for any variable that is thought to represent disagreement. For example, dynamic models such as Scheinkman and Xiong (2003) and Hong et al. (2006) suggest that turnover is produced whenever the valuation of two investors cross in the sense of the optimistic investor becoming the pessimist. Together, these models imply that the change, rather than the level, in disagreement should be positively related to turnover. Hence, if news dispersion captures investor disagreement we would expect a positive relation between changes in news dispersion and turnover. We also test whether our constructed proxies for disagreement are able to explain trading volume.

We first regress abnormal turnover onto changes in news dispersion, controlling for past turnover and a range of additional variables. Table 2 reports a positive and statistically significant news-tone dispersion coefficient, suggesting that changes in news dispersion in fact is a statistically significant determinant of trading volume. To compare our measure to existing measures of disagreement in the literature, we run the same regressions using four out of the five disagreement proxies defined earlier. We leave out dispersion of GDP forecasts since it is measured on a quarterly basis. Surprisingly, none of the disagreement proxies are able to explain turnover which after all is an important theoretical aspect of and ingredient in models of disagreement. A potential reason for this is that such proxies capture the opinions of market participants already after they have realized, and presumably acted upon, their disagreement. These findings underscore the difference between our analysis of the reasons to disagree and the empirical literature on disagreement proxies.

3.2 Predicting Investor Disagreement with News Dispersion

As argued in the introduction, the only (weak) assumption needed for a dispersed news flow to generate disagreement is that investors assign different weights to news items. This can arise for various reasons such as confirmation bias, limited attention to news, or heterogenous priors. If the news spectrum ranges from very positive to very negative, as is the case when news dispersion is high, some investors will end up with a more optimistic outlook of the market, while others will adopt a more pessimistic view, leading to high disagreement among them. Our argument therefore implies that news dispersion should predict future disagreement positively.

Panel A in Table 3 reports correlation coefficients between news dispersion and future values of the five proxies of investor disagreement. We find that news dispersion is positively correlated with all measures of disagreement with correlations ranging between 0.26 and 0.73. This suggest a significant relation between news dispersion and future investor disagreement.

Next, we formally test whether news dispersion contains predictive power for future investor disagreement. Our earlier hypothesis, that a dispersed news flow is likely to generate disagreement among investors, suggest a positive predictive relation. We run predictive regressions with disagreement measures as of time $t + 1$ as dependent variables and news dispersion as of time t as independent variable, controlling for past disagreement and a range of other variables. Panel B of Table 3 reports the results and shows that a more dispersed news flow predicts an increase in disagreement for four out of our five disagreement measures. A one standard deviation increase in news dispersion predicts an increase of between 0.16 and 0.34 standard deviations in disagreement, depending on disagreement measure. These results support our earlier argument that a polarized information environment generates disagreement among investors.

4 The Impact of News Dispersion on Returns

A longstanding hypothesis dating back to Miller (1977) suggests that disagreement among investors together with limited participation should matter for future returns. In particular, if pessimistic investors are excluded from the market due to short-selling constraints, stock prices will mainly reflect the valuation of optimistic investors which is generally speaking too high provided the consensus of *all* investors is the best estimate of fair value. This optimistic overvaluation is bound to reverse as disagreement mean reverts, leading to a negative relationship between current disagreement and future returns. We therefore explore in this section whether our measure of news dispersion predicts returns and what the impact of short-sales constraints are. We also evaluate the ability of news dispersion to predict realized volatility as the recent disagreement model of Banerjee and Kremer (2010) suggests that increased disagreement may lead to higher volatility.

We draw inference on the estimated regression coefficients using two sets of standard errors. First, we use Newey and West (1987) standard errors with 10 and 5 lags for daily and monthly horizons, respectively. Second, to address small-sample issues, we compute bootstrapped standard errors using the stationary block bootstrap of Politis and Romano (1994) with 5000 repetitions and where the optimal block length is computed as in Politis and White (2004) and Patton et al. (2009). This method maintains any serial correlation and heteroscedasticity present in the data.

4.1 Predicting Aggregate Stock Returns

The dependent variable in the return regressions is the S&P 500 excess return, measured by the widely traded exchange-traded fund SPY. Table 4 presents results using a monthly forecasting horizon.[9] The first column demonstrates that news dispersion on its own pre-

[9] We consider longer horizons in the robustness section.

dicts monthly returns negatively with a statistically significant coefficient and with an R^2 of about 10%. Including control variables in fact raises both the economic magnitude of news dispersion and its statistical significance. Ultimately, when including all control variables, a one standard deviation increase in news dispersion predicts a drop of 1.91% in returns next month, an economically significant effect. An explanatory power of 31% is large for a monthly horizon and is naturally subject to sampling variability. We compute a bootstrapped standard error for the R^2 value of about 9% for the specification which includes all control variables, indicating that the R^2 is statistically significant. Interestingly, news-tone dispersion and the variance risk premium, which is a widely used predictor, predict returns with similar economic magnitudes but with opposite signs.

Figure 4 visualizes the negative relation between news-tone dispersion and one-step ahead excess returns. The figure provides a scatter plot together with a linear line of best fit and a non-linear line of best fit, where the latter is computed using the LOWESS procedure. Both lines indicate a negative relation between dispersion and ex-post returns.

It is also interesting to note that Table 4 suggests that the aggregate level of news tone fails to predict returns. This is somewhat intriguing since Tetlock (2007) found that the sentiment level of a popular news column in Wall Street Journal predicted daily index returns. The likely reason for our different results is the construction of our corresponding measures. Tetlock's measure is based on a news-paper commentary section that he found mostly reflected investor sentiment while our measure is built on millions of news items, that we find mainly reflect disagreement about fundamentals.

4.2 News Dispersion and "Ignored" Bad News

In the spirit of Miller (1977) and related models on disagreement and short-sales constraints, we consider a simple example that illustrates an unexplored consequence of investor disagreement. Consider the case of two investors observing news about two firms. Assume investor

A overweighs news about Firm 1 relative to Firm 2, while investor B does the opposite. The overweighing itself may arise due to heterogenous priors, confirmation bias, limited attention, overconfidence etc. Assume also that Firm 1 happens to announce good news while Firm 2 issues bad news, where all news is released at the same time. Hence, investor A overweighs the good news from Firm 1 and ends up being the market optimist while investor B overweighs the bad news from Firm 2 making him the market pessimist. As a result, A and B will disagree about the aggregate outlook. The larger the spread in news tones between the news items, the larger the disagreement between investors.

Given B's pessimistic view, she would like to short both firms but especially Firm 2. However, assuming binding short-selling constraints, the optimistic investor A will end up being the marginal price setter for *both* firms. This will lead to potential overvaluation of both firms but especially of Firm 2, which had some "ignored" bad news. Consequently, firms with ignored bad news should be subject to overvaluation with correspondingly lower expected returns compared to firms with good news. This argument leads to the following testable prediction:

1. A rise in news dispersion, and therefore in disagreement, should predict an underperformance next period of stocks with the worst news compared to stocks with the best news, given that short-sales constraints are binding for a meaningful fraction of investors.

To test this prediction, we simply sort firms every month into deciles based on their individual news tone over that month. We then predict the subsequent returns on a portfolio that is long stocks with the worst news tones and short stocks with the best news tones. Results are summarized in the last column of Table 4 and confirm our prediction. The news-dispersion coefficient is negative and statistically significant and where a one standard deviation increase in news dispersion predicts an underperformance of 0.91% next period for the worst news decile. These results support the idea that as news dispersion rises, and therefore also disagreement, firms with "ignored" bad news become more overvalued than firms with good news and therefore experience lower expected returns.

4.3 News Dispersion and Short-Sale Constraints

Our results so far suggests that a more dispersed news flow predicts an increase in disagreement next period while also predicting aggregate stocks returns with a negative sign, suggesting that a more polarized news flow lowers expected returns. While these results point in the direction of models on disagreement, we need to also condition on the level of short-sales constraints.

One could argue that short-sales constraints are hardly binding on the aggregate level since the cost of shorting the overall market, for example through shorting highly liquid exchange-traded funds or futures, is low. However, as we described in the introduction, there is a large fraction of investors that despite the low costs never go short due other reasons such as institutional, regulatory, or cultural constraints. In fact, the majority of mutual funds and pensions funds are "long-only" meaning they do not express a negative market view by going short. So rather than focussing on the cost of shorting as a proxy for short-sales constraints, we instead focus on the cross-section and the relative pricing of stocks depending on their short-sale availability.

We consider three cross-sectional sorts as proxies for short-sales constraints. First, we consider small versus large stocks. Asquith et al. (2005) document that small-cap stocks are more likely to be short-sale constrained than large stocks. Small stocks should therefore be more over-priced in times of high investor disagreement and experience lower future returns. Second, we consider the documented return spread between high and low-beta stocks. Hong and Sraer (2012) document that high-beta assets are more sensitive to macro-disagreement and therefore more likely to be over-priced in times of rising disagreement and binding short-sales constraints, leading to low future returns compared to low-beta stocks. Third, we follow Nagel (2005) and consider the return spread between stocks with low versus high institutional ownership where stocks with low institutional ownership are expected to be more constrained. As a result, the latter stocks should be subject to a larger over-valuation as disagreement rises and therefore lower future returns.

We run monthly predictive regressions using the three long-short portfolios as dependent variables and news dispersion as predictive variables, using the same control variables as in our earlier return regressions. If a rise in news dispersion is associated with a larger contemporaneous over-pricing of short-sales constrained stocks, we would expect negative signs in all regressions. Table 5 reports the results and show that a rise in news dispersion predicts returns negatively on all three portfolios. That is, a rise in news dispersion predicts an underperformance of small stocks versus larger stocks, of high-beta stocks versus low-beta stocks, and of stocks with low versus high institutional ownership. These results together with the positive link between news-tone dispersion and investor disagreement further suggest that news dispersion operates via investor disagreement and short-sales constraints.

4.4 Explaining and Predicting Realized Volatility

The recent disagreement model of Banerjee and Kremer (2010) suggests that volatility should rise as disagreement among investors increases. In our earlier discussion of summary statistics, we documented a significant positive contemporaneous correlation between news-tone dispersion and both realized and implied volatility. We further test this relation by analyzing whether news dispersion is able to explain contemporaneous volatility and predict future volatility, controlling for a range of variables. We regress contemporaneous and next-period monthly realized variance of S&P 500 returns onto news dispersion, controlling for the level of aggregate news tone plus the additional five control variables we used in our earlier regressions. Table 6 presents the results and shows that the news-dispersion coefficient is positive and statistically significant in both regressions. Hence in line with models on disagreement, a more dispersed news flow is significantly associated with both higher volatility and more trading.

4.5 High-Frequency Disagreement

Recent advances in news analytics allows one to compute news dispersion on any desired frequency. As a result, news dispersion can even be measured on a daily basis, yielding a high-frequency measure of investor disagreement. In this section we provide evidence that our main results also hold at the daily horizon. All variables are defined as before except some of the controls that are not available at the daily frequency, for which we try to find close substitutes. Our daily control variables are made up of returns, realized volatility, the VIX index, return dispersion, and the Aruoba-Diebold-Scotti Business Condition Index (ADS). The latter index measures economic activity on a daily basis and is obtained from the Federal Reserve Bank of Philadelphia. We compute daily excess returns by assuming that the one-month rate is constant within each month. Daily realized volatility is obtained from the Oxford-Man Institute.

Table 7 presents results for explaining daily changes in turnover and predicting daily returns and volatility. We find that changes in news dispersion is a strongly significant determinant of daily abnormal turnover while also predicting returns negatively and volatility positively. A one standard deviation increase in news dispersion suggests a drop of around 8 basis points in returns the next day. Overall, the results suggest that news dispersion operates through investor disagreement over both daily and monthly frequencies.

5 Alternative Explanations

5.1 Market Fundamentals or Investor Sentiment?

We have so far interpreted our measure of news dispersion as reflecting heterogeneity of information about future fundamentals. However, it is possible that aggregate news dispersion instead captures the polarization of "moods" (sentiment) among investors. In fact, using linguistic analysis to quantify the content of financial news offers little *a priori* guidance as to whether the resulting indicators reflect market fundamentals or investor sentiment.

On one hand, Tetlock et al. (2008) conclude that negative words in firm-specific news stories "capture otherwise hard-to-quantify aspects of firms fundamentals" because they can predict earnings in the following quarter. The effect is driven mostly by news stories containing the word stem 'earn', suggesting they are linked to company fundamentals. On the other hand, Tetlock (2007) studies negative words in a popular Wall Street Journal column and concludes that "the hypothesis that [news] pessimism represents negative fundamental information not yet incorporated into prices receives very little support from the data". Given that our dataset consists of company-specific news but mostly not related to earnings, it is unclear which interpretation is more accurate in our case. In addition, our news-dispersion measure is inherently different from earlier news-tone measures in the literature since earlier measures have only considered the level of news tone and not the dispersion.

We try to address this issue by first regressing news-tone dispersion onto the investor sentiment measures developed by Baker and Wurgler (2006, 2007), obtained from the website of Jeffrey Wurgler. Table 8 reports the results and shows that several of the investor sentiment variables are statistically significant but the explanatory power is rather low with a R^2-value of 23%. For comparison, we subsequently regress news dispersion onto two economic variables in the form of the Chicago Fed National Activity Index (Fed) and growth in aggregate corporate earnings (EarnGrowth). The Fed variable turns out negative and strongly significant and where the explanatory power of 53% is more than twice as large as the preceding one. Finally, we include both sentiment and economic variables as independent variables and find that the explanatory power only rises modestly to 58%. Overall, the results suggest that news-tone dispersion mainly reflects the "soft" part of fundamentals rather than investor sentiment.

5.2 Aggregate Economic Uncertainty

Rather than interpreting a rise in news dispersion as a rise in disagreement about future fundamentals, it is possible that a more dispersed news flow reflects an increase in the aggregate level of informational or economic uncertainty. Such interpretation would imply that news dispersion reflects a direct measure of aggregate information uncertainty since it is based on the actual information flow that investors observe. We explore this potential interpretation by examining the relation between news dispersion and the perhaps most commonly used proxy for economic uncertainty, the variance risk premium (e.g., Bollerslev et al., 2009, and Drechsler and Yaron, 2011).

We analyze the relation between news dispersion and the variance risk premium by running a contemporaneous regression of the variance risk premium onto news-tone dispersion and a set of control variables. The results are reported in Table 9 and indicate that news dispersion, only controlling for the level of news tone, is significantly positively related to the variance risk premium, explaining around 38% of the variation. However, the same table shows that news dispersion loses its significance when including additional control variables. These results therefore ultimately suggest a rather weak relation between news dispersion and economic uncertainty.

Assuming that news dispersion would indeed reflect economic uncertainty, which are then the implications for the relation between news dispersion and expected returns? Intuitively, a rise in aggregate economic uncertainty or risk in a world populated by fully rational investors should be accompanied by higher expected returns. This is supported by models with recursive preferences in which investors dislike uncertainty and therefore demand higher expected returns in times of increasing economic uncertainty and by empirical results showing that the variance risk premium predicts returns with a *positive* sign (e.g., Bollerslev et al., 2009, and Drechsler and Yaron, 2011). Consequently, news-tone dispersion should be *positively* related to future returns if it reflects economic uncertainty. However, we find strong opposite results in the form of highly significant *negative* coefficients. This together with the previous regression results cast doubt on economic uncertainty as a potential explanation.

6 Robustness Checks

6.1 Different Types of News

Our data set allows us to separate news into different categories and we exploit this fact to test the robustness of our results. First, we sort news according to so-called headline news where the company is mentioned both in the headline and in the accompanying news text as opposed to only in the latter text. News items in which the company is mentioned in the headline are potentially more "attention-grabbing" and might therefore carry incremental information. Second, we consider earnings versus non-earnings news. This news sort sheds light on whether our results are mainly driven by the earnings season or whether dispersion of all types of news in fact matters.

Table 10 reports results from running monthly predictive regressions for returns and volatility. First, the predictive power of headline-news dispersion is very similar to our baseline measure. Second, the dispersion of both earnings and non-earnings news are associated with negative and statistically significant coefficients in the return regressions and positive coefficients in the volatility regressions. However, the coefficient for earnings news turns out statistically insignificant when predicting volatility. Interestingly, the economic magnitude of non-earnings news dispersion is substantially larger than for earnings-news dispersion, suggesting that our results are not simply driven by the earnings season.

6.2 Additional Control Variables

Our base-line regressions controlled for a range of variables such as past returns, realized volatility, implied volatility, return dispersion, earnings growth, and economic activity. In this section, we consider additional variables that were used in Goyal and Welch (2008). Table 11 presents results from predicting returns and volatility using the following control variables: the default yield spread, earnings-price ratio, book-to-market ratio, t-bill rate, term spread, and the net equity expansion. The table shows that our main results go through with slope coefficients in front of news dispersion that are highly statistically significant.

6.3 Sub-Samples and Long-Horizon Returns

Finally, we consider various sub-sample regressions to evaluate the robustness of our results. Results are presented in Table 12. First, we split our sample period in half and run predictive return regressions over the two sub-periods January 2003 to December 2007 and January 2008 to December 2012. The results suggest that news dispersion is a negative and significant return predictor in both sub-samples, although the economic and statistical significance are both stronger in the second sub-period. Second, we exclude observations pertaining to the financial crisis, defined as June 2007 to June 2009. Again, news dispersion turns out to be a significant predictor, suggesting that our results are not driven by the financial crisis. Third, we predict 3-month and 6-month returns and find news dispersion to also contain a statistically and economically significant predictive power over longer horizons. A one standard deviation increase in news dispersion predicts a drop of between 4 and 5% in returns over the next 3 to 6 months.

7 Conclusions

Using recent advances in news analytics, we construct an empirical measure of news dispersion and study how a dispersed news flow affects investors and aggregate stock returns. News dispersion increases as firm news become more polarized and: i) predicts investor disagreement positively, ii) is positively related to turnover, iii) predicts aggregate stock returns negatively, iv) predicts an underperformance of stocks that are more short-sales constrained or more sensitive to aggregate disagreement, and v) predicts realized volatility positively.

We believe these results speak strongly in favor of models on disagreement and short-sales constraints and support the idea that news dispersion itself represents a fundamental reason for why investors disagree in the first place. The fact that recent advances in news analytics allows news dispersion to be measured over any chosen frequency, even daily, makes it in our view a versatile measure of investor disagreement. Furthermore, our results suggest that news-tone dispersion is strongly linked to stock market volatility, being countercyclical and asymmetric, and can therefore provide a news-based explanation for why volatility changes over time.

References

Almazan, Andres, Keith C. Brown, Murray Carlson, and David A. Chapman, 2004, Why constrain your mutual fund manager?, *Journal of Financial Economics* 73, 289–321.

Aruoba, S. Borağan, Francis X. Diebold, and Chiara Scotti, 2009, Real-Time Measurement of Business Conditions, *Journal of Business & Economic Statistics* 27, 417–427.

Asquith Paul, Parag A. Pathak, and Jay R. Ritter, 2006, Short interest, institutional ownership, and stock returns, *Journal of Financial Economics* 78, 243–276.

Baker, Malcolm, and Jeffrey Wurgler, 2006, Investor sentiment and the cross-section of stock returns, *Journal of Finance* 61, 1645–1680.

Baker, Malcolm, and Jeffrey Wurgler, 2007, Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129–151.

Bali, Turan G., and Hao Zhou, 2012, Risk, uncertainty, and expected returns, Working Paper.

Banerjee, Snehal, and Ilan Kremer, 2010, Disagreement and learning: Dynamic patterns of trade, *Journal of Finance* 65, 1269–1302.

Bollerslev, Tim, George Tauchen, and Hao Zhou, 2009, Expected Stock Returns and Variance Risk Premia, *Review of Financial Studies* 22, 4463–4492.

Chen, Joseph, Harrison Hong, and Jeremy C. Stein, 2002, Breadth of ownership and stock returns, *Journal of Financial Economics* 66, 171–205.

Davis, Angela K., Jeremy M. Piger, and Lisa M. Sidor, 2006, Beyond the numbers: An analysis of optimistic and pessimistic language in earnings press releases, Working Paper.

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Demers, Elizabeth, and Clara Vega, 2011, Linguistic Tone in Earnings Press Releases: News or Noise?, Working Paper.

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina, 2002, Differences of Opinion and the Cross Section of Stock Returns, *Journal of Finance* 57, 2113–2141.

Drechsler, Itamar, and Amir Yaron, 2011, What's vol got to do with it, *Review of Financial Studies* 24, 1–45.

Engelberg, Joseph, 2008, Costly information processing: Evidence from earnings announcements, Working Paper.

Goyal, Amit, and Ivo Welch, 2008, A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* 21, 1455–1508.

Harrison, J. Michael, and David M. Kreps, 1978, Speculative investor behavior in a stock market with heterogeneous expectations, *Quarterly Journal of Economics* 92, 323–336.

Hong, Harrison, and David Sraer, 2012, Speculative betas, Working Paper.

Hong, Harrison, Jose Scheinkman, and Wei Xiong, 2006, Asset float and speculative bubbles, *Journal of Finance* 61, 1073–1117.

Jurafsky, Martin, and James H. Martin, 2008, *Speech and Language Processing*. (Pearson Prentice Hall) 2nd edition edn.

Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang, 2002, Dynamic volume-return relation of individual stocks, *Review of Financial Studies* 15.

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

Miller, Edward M., 1977, Risk, uncertainty, and divergence of opinion, *Journal of Finance* 32, 1151–1168.

Nagel, Stefan, 2005, Short sales, institutional investors and the cross-section of stocks returns, *Journal of Financial Economics* 78, 277–309.

Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.

Park, Cheolbeom, 2005, Stock Return Predictability and the Dispersion in Earnings Forecasts, *Journal of Business* 78, 2351–2376.

Patton Andrew, Dimitris N. Politis, and Halbert White, 2009, Correction to "Automatic block-length selection for the dependent bootstrap", *Econometric Reviews* 28, 372–375.

Politis, Dimitris N., and Joseph P. Romano, 1994, The stationary bootstrap, *Journal of the American Statistical Association* 89, 1303–1313.

Politis, Dimitris N., and Halbert White, 2004, Automatic block-length selection for the dependent bootstrap, *Econometric Reviews* 23, 53–70.

Richardson, Gordon, Stephan Sefcik, and Rex Thompson, 1986, A test of dividend irrelevance using volume reactions to a change in dividend policy, *Journal of Financial Economics* 17.

Scheinkman, Jose A., and Wei Xiong, 2003, Overconfidence and speculative bubbles, *Journal of Political Economy* 111, 1183–1219.

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

Tetlock, Paul, Maytal Saar-Tsechansky, and Sofus Macskassy, 2008, More Than Words: Quantifying Language to Measure Firms' Fundamentals, *Journal of Finance* 63, 1437–1467.

Yu, Jialin, 2011, Disagreement and return predictability of stock portfolios, *Journal of Financial Economics* 99, 162–183.

Zhou, Hao, 2010, Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty, Working Paper.

Table 1: **Monthly Summary Statistics and Correlations**

	Monthly Summary Statistics						
	Mean	Std	Skewness	Kurtosis	AC(1)	AC(5)	ADF
AggTone^M	1.81	3.92	-1.04	3.75	0.82	0.28	-3.47
AggDisp^M	18.95	9.46	2.71	13.88	0.56	0.36	-5.73
SP500 Ret	0.43	4.32	-0.91	5.47	0.23	-0.03	-8.53
SP500 RV	2.74	0.93	1.09	4.13	0.79	0.44	-3.63
Fed	-0.34	1.00	-2.01	7.62	0.77	0.58	-3.94
VIX	20.90	9.14	1.78	6.78	0.86	0.45	-3.03
EarnGr	0.01	0.19	1.06	37.19	0.00	0.00	-10.86
VRP	18.18	15.66	2.28	8.60	0.62	0.35	-5.37
RetDisp	8.04	2.83	2.32	9.81	0.74	0.53	-4.06
AbTurn	0.13	0.27	0.74	3.37	0.80	0.69	-3.23

	Monthly Correlations									
	AggTone^M	AggDisp^M	SP500 Ret	SP500 RV	Fed	VIX	EarnGr	VRP	RetDisp	AbTurn
AggTone^M	1.00	-0.63	0.28	-0.77	0.64	-0.72	0.32	-0.59	-0.71	-0.51
AggDisp^M	-0.63	1.00	-0.34	0.66	-0.65	0.64	-0.15	0.52	0.56	0.53

The table reports monthly summary statistics and correlations for the sample period Jan 2003-Dec2012. The summary statistics refer to the "raw" measures of aggregate news tone. For the purpose of our empirical analysis, we then standardize and winsorize our two measures at the 1st and 99th percentiles. Fed is the Chicago Fed National Activity Index. EarnGr is the monthly growth rate in aggregate 12-month trailing earnings of companies in CRSP. VRP is the variance risk premium. RetDisp is the monthly dispersion (cross-sectional standard deviation) of returns for all stocks. AbTurn is the abnormal turnover. ADF refers to the test statistic from an augmented Dickey-Fuller test allowing for a constant term and one lag where the critical values are -3.44 and -2.58 for the 1% and 10% levels.

Table 2: **Explaining Abnormal Turnover**

	AbTurn(t)	AbTurn(t)	AbTurn(t)	AbTurn(t)	AbTurn(t)
$\Delta AggDisp^M(t)$	0.15				
	(0.07)				
	[0.07]				
$\Delta Disp_{LTG}(t)$		0.00			
		(0.06)			
		[0.06]			
$\Delta Disp_{AggGr}(t)$			-1.00		
			(0.66)		
			[0.79]		
$\Delta Disp_{AAII}(t)$				-0.05	
				(0.07)	
				[0.09]	
$\Delta Disp_{II}(t)$					-0.04
					(0.17)
					[0.14]
$\Delta AggTone_t^M$(t)	-0.02	-0.04	-0.04	-0.04	-0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Fed(t-1)	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
VRP(t-1)	0.03	0.02	0.02	0.02	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Ret(t)	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
RetDisp(t)	0.01	0.02	0.02	0.02	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
AbTurn(t-1)	0.66	0.63	0.65	0.64	0.64
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
N	119	119	119	119	119
R^2_{adj}(%)	76.5%	73.8%	74.4%	73.8%	73.8%

The table presents results from regressing abnormal turnover onto changes in news dispersion ($\Delta AggDisp^M$) and changes in the proxies for disagreement. Abnormal turnover is computed as the difference between log turnover and its moving average. Section 3 describes how the disagreement proxies are constructed. Monthly news-tone dispersion is denoted AggDisp^M. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Observations for the period Jan 2003 - Dec 2012 are used.

Table 3: **Predicting Investor Disagreement**

	$Disp_{AggEarn}(t+1)$	$Disp_{LTG}(t+1)$	$Disp_{AAII}(t+1)$	$Disp_{II}(t+1)$	$Disp_{GDP}(t+1)$
			Panel A: Correlations		
AggDisp^{M}(t)	0.46	0.37	0.26	0.47	0.73
			Panel B: Predictive regressions		
AggDisp^{M}(t)	0.16	0.03	0.34	0.23	0.26
	(0.07)	(0.05)	(0.09)	(0.10)	(0.10)
	[0.08]	[0.07]	[0.11]	[0.11]	[0.17]
$Disp_{AggEarn}(t)$	0.85				
	(0.06)				
$Disp_{LTG}(t)$		0.89			
		(0.05)			
$Disp_{AAII}(t)$			0.36		
			(0.13)		
$Disp_{II}(t)$				0.56	
				(0.08)	
$Disp_{GDP}(t)$					0.39
					(0.13)
AggTone^{M}(t)	0.06	0.05	0.33	0.07	0.08
	(0.05)	(0.06)	(0.11)	(0.10)	(0.13)
Fed(t)	0.03	0.11	-0.03	0.04	-0.19
	(0.06)	(0.06)	(0.13)	(0.11)	(0.12)
VRP(t)	0.00	0.16	-0.02	0.09	0.02
	(0.07)	(0.07)	(0.09)	(0.10)	(0.11)
Abturn(t)	0.37	0.16	0.43	0.24	0.77
	(0.16)	(0.14)	(0.47)	(0.26)	(0.38)
N	119	119	119	119	39
$R^2_{adj}(\%)$	84.5%	85.8%	24.8%	46.6%	72.7%

The table presents two panels. Panel A presents correlation coefficients between news-tone dispersion and next-period values of investor disagreement. Panel B presents results from predicting future values of investor disagreement with news dispersion, controlling for a range of variables. Section 3 describes how the five disagreement measures are constructed. Monthly news-tone dispersion is denoted AggDisp^{M}. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used except when predicting $Disp_{GDP}(t+1)$ for which quarterly observations are used.

Table 4: **Predicting Monthly Stock Returns**

	Ret(t+1)	Ret(t+1)	Ret(t+1)	Ret(t+1)	Ret(t+1)	WorstRet(t+1)-BestRet(t+1)
AggDisp^M(t)	-1.40	-1.83		-1.87	-1.91	-0.91
	(0.62)	(0.47)		(0.69)	(0.69)	(0.43)
	[0.60]	[0.58]		[0.77]	[0.76]	[0.54]
AggTone^M(t)		-0.67			-0.32	-0.96
		(0.52)			(0.65)	(0.49)
Fed(t)			2.12	1.53	1.52	0.73
			(1.00)	(0.87)	(0.88)	(0.51)
EarnGrowth(t)			2.31	2.66	2.91	4.63
			(1.66)	(1.44)	(1.66)	(3.11)
VRP(t)			2.21	2.27	2.25	0.10
			(0.49)	(0.57)	(0.57)	(0.04)
RV(t)			-0.19	0.55	0.40	-0.25
			(0.68)	(0.54)	(0.66)	(0.55)
Ret(t)			0.23	0.19	0.19	0.06
			(0.10)	(0.09)	(0.09)	(0.14)
RetDisp(t)			0.06	0.11	0.08	-0.01
			(0.11)	(0.12)	(0.15)	(0.18)
N	119	119	119	119	119	118
R^2_{adj}(%)	9.74%	10.44%	23.77%	32.03%	31.59%	12.30%

The table presents results from predicting monthly excess returns on the S&P 500 index (columns 1-5) as well as on a portfolio that is long the decile of stocks with the worst news and short the decile of stocks with the best news in month t. Monthly news-tone dispersion is denoted AggDisp^M. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used.

Table 5: **News Dispersion and Short-Sale Constraints**

	Ret(t+1) Size [lo-hi]		Ret(t+1) β [hi-lo]		Ret(t+1) IO [lo-hi]	
AggDisp^M(t)	-1.08	-1.08	-3.49	-3.75	-0.78	-1.04
	(0.46)	(0.44)	(0.90)	(1.21)	(0.38)	(0.48)
	[0.48]	[0.49]	[1.21]	[1.46]	[0.38]	[0.53]
AggTone^M(t)	-0.50	-0.14	-1.48	-0.36	-1.02	-0.60
	(0.46)	(0.50)	(1.03)	(1.32)	(0.49)	(0.53)
Fed(t)		0.12		2.51		0.05
		(0.43)		(1.52)		(0.37)
EarnGrowth(t)		0.24		1.98		3.54
		(2.20)		(3.54)		(1.51)
VRP(t)		0.05		0.31		-0.04
		(0.02)		(0.09)		(0.02)
RV(t)		-0.69		-0.19		-0.07
		(0.46)		(1.78)		(0.59)
Ret(t)		0.24		0.39		-0.11
		(0.09)		(0.22)		(0.08)
RetDisp(t)		0.37		0.40		0.55
		(0.15)		(0.32)		(0.13)
N	119	119	119	119	119	119
R^2_{adj}(%)	4.95%	22.16%	8.53%	26.24%	5.64%	15.23%

The table presents results from predicting returns on three long-short portfolios sorted according to short-sales constraints and sensitivity to aggregate disagreement. The three portfolios are: small minus large firms, high-beta minus low-beta firms, and stocks with low institutional ownership minus stocks with high institutional ownership. Monthly news-tone dispersion is denoted AggDisp^M. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used.

Table 6: **Explaining and Predicting Realized Variance**

	RV(t)	RV(t+1)
$\text{AggDisp}^{M}(t)$	0.12	0.23
	(0.05)	(0.12)
	[0.05]	[0.13]
$\text{AggTone}^{M}(t)$	-0.12	0.03
	(0.05)	(0.07)
Fed(t)	0.07	-0.13
	(0.08)	(0.11)
VRP(t)	0.31	0.14
	(0.08)	(0.08)
Ret(t)	-0.08	-0.03
	(0.01)	(0.01)
RetDisp(t)	0.02	0.00
	(0.02)	(0.02)
RV(t)		0.41
		(0.08)
RV(t-1)	0.37	
	(0.06)	
N	119	119
$R^2_{adj}(\%)$	84.4%	71.2%

The table presents results from regressing contemporaneous and future realized variance onto news-tone dispersion as of time t. Realized variance is used as a model-free measure of volatility of S&P500 index returns. Monthly news-tone dispersion is denoted AggDisp^{M}. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Observations for the period Jan 2003 - Dec 2012 are used.

Table 7: **Daily Horizon Regressions**

	AbTurn(t)	AbTurn(t)	Ret(t+1)	Ret(t+1)	Ret(t+1)	RV(t+1)	RV(t+1)
$\Delta AggDisp^D$	0.11	0.09					
	(0.01)	(0.01)					
	[0.01]	[0.01]					
$AggDisp^D$			-0.09		-0.08	0.17	0.01
			0.03		(0.04)	0.03	(0.01)
			[0.04]		[0.04]	[0.02]	[0.01]
$\Delta AggTone^D$	0.01	0.01					
	(0.01)	(0.01)					
$AggTone^D$			-0.02		0.00	-0.47	0.00
			(0.03)		(0.04)	(0.04)	(0.02)
AbTurn(t-1)	0.62	0.57					
	(0.03)	(0.03)					
ΔADS(t)		0.07		-1.57	-1.58		1.51
		(0.24)		(2.94)	(2.98)		(0.78)
RV(t)		0.14		-0.11	-0.10		0.26
		(0.01)		(0.05)	(0.05)		(0.03)
RV(t-1)		-0.07		0.02	0.03		0.25
		(0.01)		(0.05)	(0.05)		(0.02)
VIX(t)		-0.01		0.01	0.01		0.04
		(0.00)		(0.01)	(0.01)		(0.00)
Ret(t)		-0.01		-0.09	-0.09		-0.08
		(0.00)		(0.03)	(0.04)		(0.01)
Ret(t-1)		-0.01		-0.09	-0.09		-0.03
		(0.00)		(0.05)	(0.05)		(0.01)
RetDisp(t)		0.06		-0.07	-0.04		0.06
		(0.01)		(0.07)	(0.07)		(0.02)
N	2482	2482	2482	2482	2482	2471	2471
$R^2_{adj}(\%)$	40.8%	60.1%	0.3%	1.7%	1.9%	27.4%	72.5%

The table presents results from running daily regressions, explaining contemporaneous abnormal turnover and predicting next-period excess returns and volatility on the $S\&P$ 500 index. Daily news-tone dispersion is denoted AggDisp^D. Newey-West (1987) standard errors with 10 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Daily observations for the period Jan 2003 - Dec 2012 are used.

Table 8: **News-Tone Dispersion: Investor Sentiment or Fundamentals?**

	AggDisp^M(t)	AggDisp^M(t)	AggDisp^M(t)
$Sent^{\perp}$(t)	2.39		1.44
	(0.66)		(0.43)
$Sent$(t)	-1.69		-0.78
	(0.67)		(0.35)
$\Delta Sent^{\perp}$(t)	-0.49		-0.18
	(0.15)		(0.05)
$\Delta Sent$(t)	0.46		0.18
	(0.20)		(0.12)
Fed(t)		-0.70	-0.62
		(0.09)	(0.09)
EarnGrowth(t)		-0.02	0.03
		(0.15)	(0.11)
R^2_{adj}(%)	22.73%	53.48%	58.40%

The table presents results from regressing the aggregate news-tone dispersion onto the four investor sentiment factors computed by Baker and Wurgler (2006, 2007) and two economic variables in the form of the Chicago Fed National Activity Index (Fed), and growth in aggregate corporate earnings (EarnGrowth). The investor sentiment data is taken from the website of Jeffrey Wurgler which also contains descriptions of the variables. Monthly news-tone dispersion is denoted AggDisp^M. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Monthly observations for the period Jan 2003 - Dec 2010 are used as the sentiment variables are only available up until 2010.

Table 9: **Explaining Variance Risk Premia**

	VRP(t)	VRP(t)
AggDisp^M(t)	0.31	0.03
	(0.11)	(0.10)
	[0.13]	[0.13]
AggTone^M(t)	-0.38	-0.07
	(0.16)	(0.10)
Fed(t)		-0.24
		(0.16)
EarnGrowth(t)		1.11
		(0.40)
VRP(t-1)		0.40
		(0.16)
VRP(t)		
RV(t)		0.11
		(0.15)
Ret(t)		-0.07
		(0.02)
RetDisp(t)		0.01
		(0.04)
N	120	120
R^2_{adj}(%)	38.61%	59.17%

The table presents results from a contemporaneous regression of the monthly variance risk premium onto news-tone dispersion and a set of control variables. The variance risk premium (VRP) is measured as the difference between $VIX^2/12$ and expected realized variance on the $S\&P$ 500. The expected variance is obtained by regressing realized variance for time $t : t+1$ onto its own lag and $VIX_t^2/12$. Monthly news-tone dispersion is denoted AggDisp^M. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used.

Table 10: **Robustness Check: Different Types of News**

	Headline news		Earnings news		Non-earnings news	
	Ret(t+1)	RV(t+1)	Ret(t+1)	RV(t+1)	Ret(t+1)	RV(t+1)
AggDisp^M(t)	-1.77	0.28	-1.17	0.12	-1.70	0.22
	(0.45)	(0.08)	(0.50)	(0.08)	(0.67)	(0.11)
	[0.57]	[0.10]	[0.61]	[0.09]	[0.69]	[0.13]
AggTone^M(t)	0.38	-0.03	-0.44	0.09	-0.29	0.02
	(0.59)	(0.07)	(0.43)	(0.06)	(0.62)	(0.08)
Fed(t)	1.51	-0.10	2.03	-0.19	1.59	-0.13
	(0.83)	(0.10)	(0.93)	(0.11)	(0.90)	(0.11)
EarnGrowth(t)	2.65	-0.03	2.59	0.00	3.27	-0.07
	(1.47)	(0.18)	(1.48)	(0.22)	(1.83)	(0.19)
VRP(t)	2.36	0.12	2.10	0.17	2.24	0.14
	(0.54)	(0.08)	(0.52)	(0.10)	(0.56)	(0.08)
RV(t)	0.36	0.40	-0.34	0.50	0.45	0.39
	(0.69)	(0.08)	(0.65)	(0.09)	(0.68)	(0.08)
Ret(t)	0.20	-0.03	0.16	-0.02	0.20	-0.03
	(0.08)	(0.01)	(0.08)	(0.01)	(0.10)	(0.01)
RetDisp(t)	0.10	0.01	0.22	0.00	0.04	0.01
	(0.17)	(0.02)	(0.15)	(0.03)	(0.15)	(0.02)
N	119	119	119	119	119	119
R^2_{adj}(%)	34.04%	74.54%	26.98%	69.18%	30.29%	70.90%

The table presents robustness results using news-tone dispersion (AggDisp^M) that is based on different types of news covering $S\&P$ 500 stocks. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used.

Table 11: **Robustness Check: Using Additional Control Variables**

	Ret(t+1)	RV(t+1)
AggDisp^{M}(t)	-2.08	24.17
	(0.59)	(7.89)
	[0.72]	[8.89]
AggTone^{M}(t)	-0.31	-14.47
	(0.49)	(7.47)
dfy(t)	1.09	-10.56
	(1.66)	(22.38)
e/p(t)	-1.94	9.17
	(1.53)	(9.86)
b/m(t)	-0.10	118.20
	(10.66)	(65.54)
tbl(t)	-1.29	16.12
	(0.41)	(6.93)
tms(t)	-1.55	21.72
	(0.53)	(10.44)
ntis(t)	51.21	-664.94
	(31.71)	(568.13)
R^2_{adj}(%)	17.29%	46.84%

The table presents robustness results predicting returns and volatility with news-tone dispersion (AggDisp^{M}), using a different set of control variables than before. All control variables are obtained from Goyal and Welch (2008). dfy is the default yield spread, e/p is the earnings-price ratio, b/m is the book-to-market ratio, tbl is the t-bill rate, tms is the term spread, and ntis is the net equity expansion. Newey-West (1987) standard errors with 5 lags are reported in parentheses. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used.

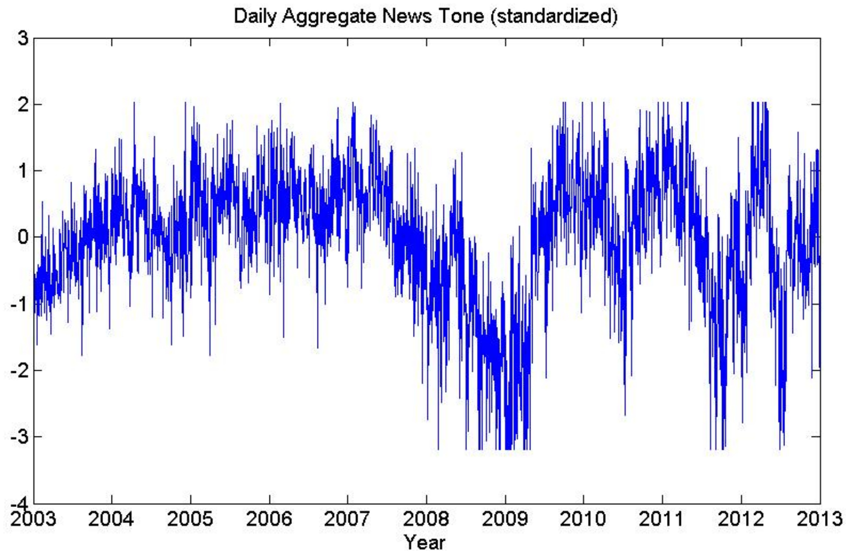
Table 12: **Robustness Check: Sub-Samples and Long-Horizon Returns**

	Ret(t+1) 2003-2007	Ret(t+1) 2008-2012	Ret(t+1) excluding Jun2007-Jun2009	3-month return	6-month return
$\text{AggDisp}^M(\text{t})$	-1.69	-2.82	-1.92	-4.16	-4.74
	(0.68)	(0.79)	(0.76)	(0.64)	(1.12)
	[1.26]	[0.98]	[0.75]	[0.93]	[1.51]
$\text{AggTone}^M(\text{t})$	-0.16	-0.98	-0.76	-0.83	-0.12
	(0.79)	(0.69)	(0.61)	(0.93)	(1.78)
Fed(t)	1.72	1.78	-0.21	3.13	5.04
	(0.81)	(1.10)	(0.83)	(1.54)	(3.06)
EarnGrowth(t)	5.49	3.07	1.83	-2.22	1.60
	(5.47)	(1.80)	(1.03)	(3.52)	(5.73)
VRP(t)	-0.36	2.73	2.27	5.94	10.13
	(0.97)	(0.45)	(1.05)	(1.22)	(2.61)
RV(t)	1.59	0.12	-0.67	-0.02	0.05
	(0.85)	(1.18)	(0.75)	(0.02)	(0.05)
Ret(t)	0.21	0.16	-0.02	0.25	0.70
	(0.14)	(0.14)	(0.08)	(0.19)	(0.35)
RetDisp(t)	-0.17	0.16	-0.01	0.16	-0.35
	(0.26)	(0.23)	(0.15)	(0.35)	(0.54)
$R^2_{adj}(\%)$	8.03%	38.19%	22.89%	37.22%	32.29%

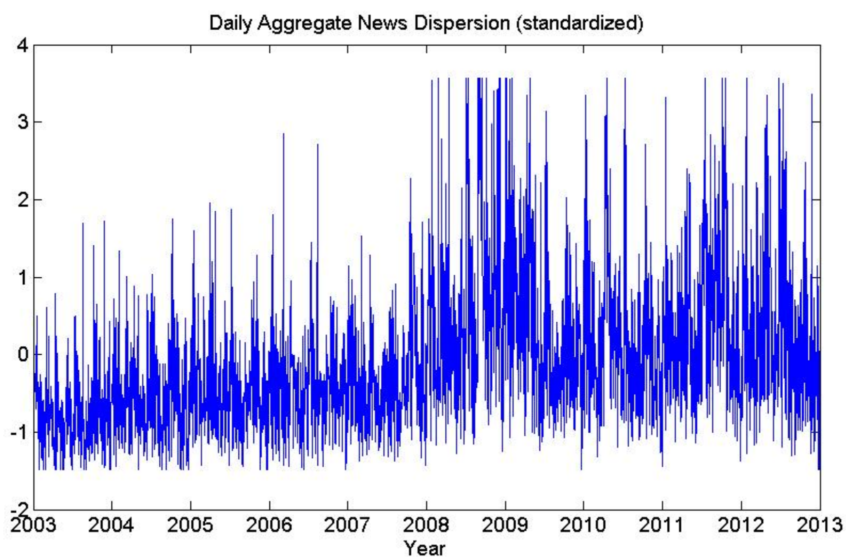
The table presents results from predicting excess returns on the S&P 500 index over sub-samples and over 3 to 6-month horizons. Monthly news-tone dispersion is denoted AggDisp^M. Newey-West (1987) standard errors with 5 lags are used. Bootstrapped standard errors are in square brackets. Monthly observations for the period Jan 2003 - Dec 2012 are used.

Figure 1: Quantitative Measures of Company News Flow - Daily Frequency

(a) Aggregate news-tone level



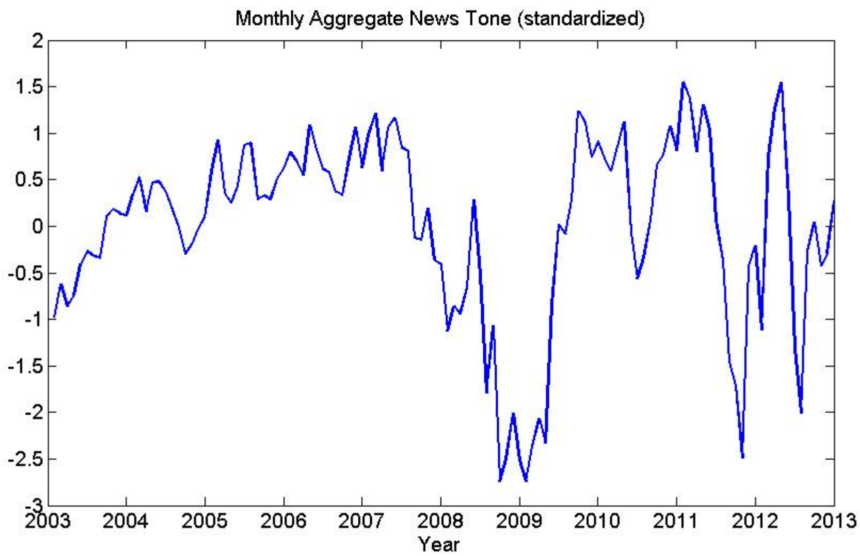
(b) Aggregate news-tone dispersion



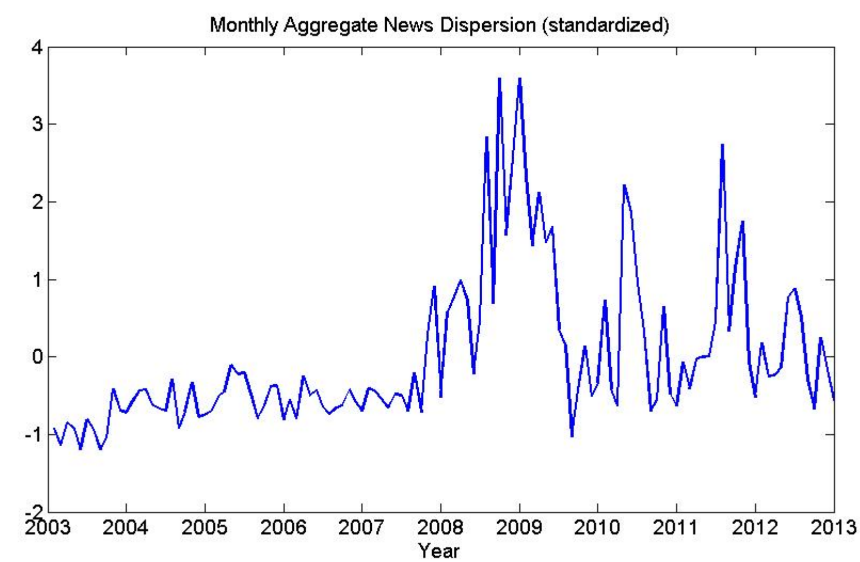
Plotted are the daily aggregate level and dispersion of news tone for $S\&P$ 500 firms for the period Jan 2003 - Dec 2012. Definitions of daily aggregate news-tone level and aggregate news-tone dispersion are discussed in Section 2 and are explicitly given in Eq. 5 and Eq. 3 respectively. Both time series are winsorized at the 1st and 99th percentiles and then standardized by demeaning and dividing by the standard deviation.

Figure 2: Quantitative Measures of Company News Flow - Monthly Frequency

(a) Aggregate news-tone level



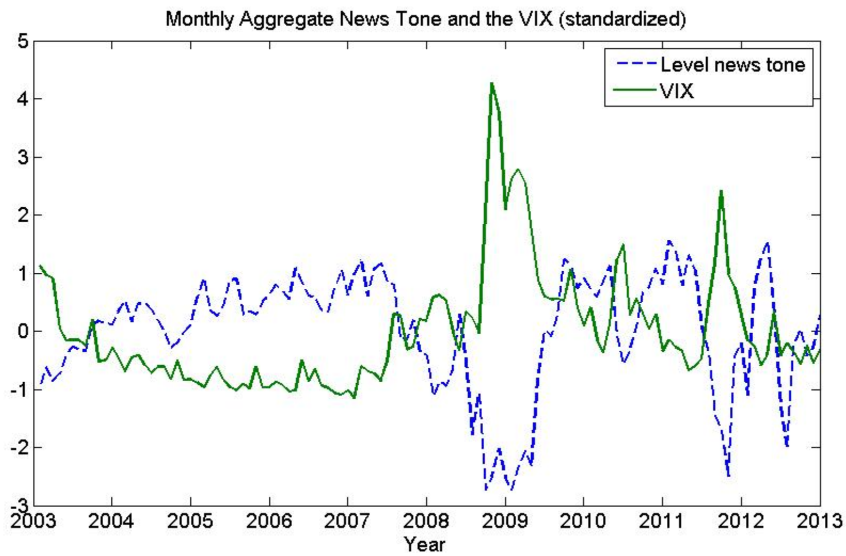
(b) Aggregate news-tone dispersion



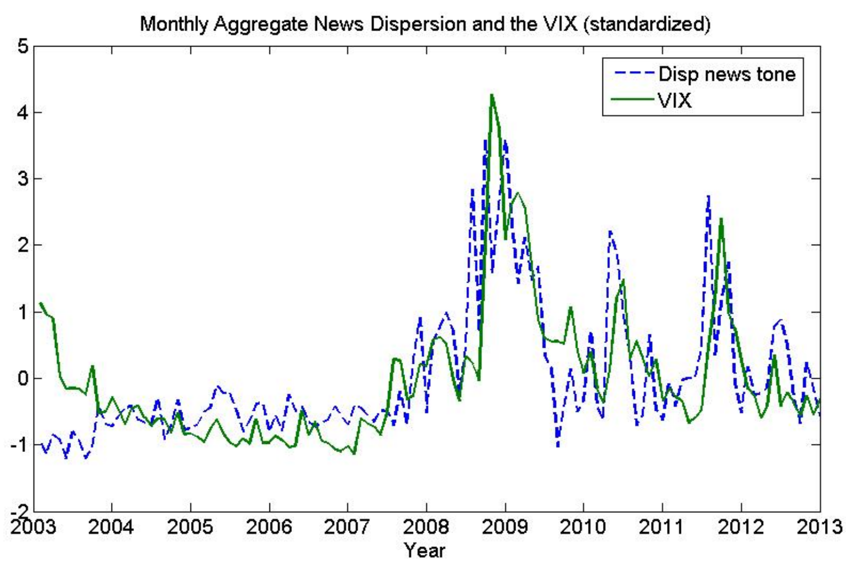
Plotted are the monthly aggregate level and dispersion of news tone for $S\&P$ 500 firms for the period Jan 2003 - Dec 2012. Definitions of monthly aggregate news tone and aggregate dispersion of news tone are discussed in Section 2 and are explicitly given in Eq. 6 and Eq. 4 respectively. Both time series are winsorized at the 1st and 99th percentiles and then standardized by demeaning and dividing by the standard deviation.

Figure 3: Monthly Measures of Company News Flow and the VIX

(a) Aggregate news-tone level and the VIX

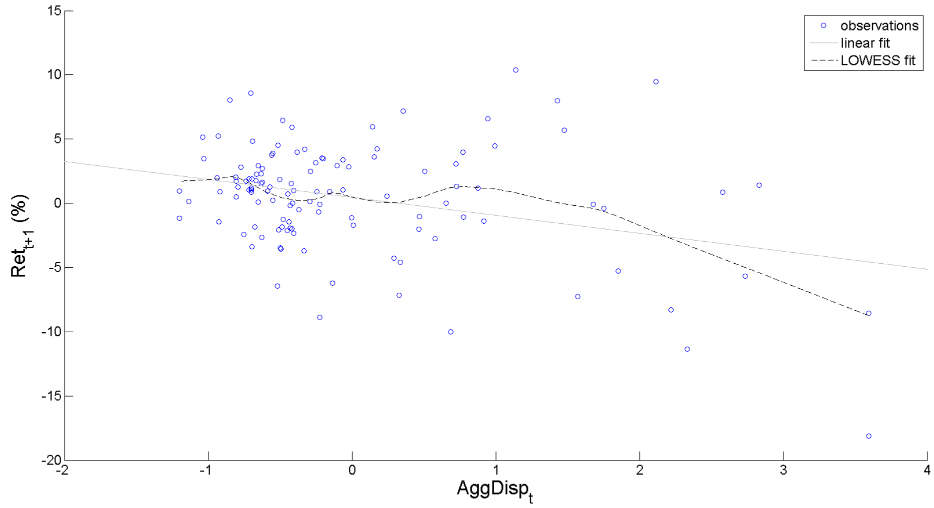


(b) Aggregate news-tone dispersion and the VIX



Monthly measures of aggregate news tone from Figure 2 are plotted (in blue, dashed line) against end-of-month levels of the VIX (in green, solid line) for the period 2003 - 2012. News-tone measures are winsorized at the 1st and 99th percentiles while all series are standardized by demeaning and dividing by the standard deviation.

Figure 4: Plotting Dispersion of News Tone against Ex-Post Stock Returns



Plotted are the dispersion of news tone against ex-post stock excess returns, using monthly data. The grey (dashed) line depicts a linear line of best fit and the black (solid) line depicts a non-linear line of best fit computed using the LOWESS procedure. The time series for news-tone dispersion is winsorized at the 1st and 99th percentiles and then standardized by demeaning and dividing by the standard deviation. Monthly observations for the period Jan 2003 - Dec 2012 are used.