Predicting Equity Markets with Digital Online Media Sentiment: Evidence from Markov-switching Models

Steven J. Nooijen\textsuperscript{a} and Simon A. Broda\textsuperscript{a,b,c}

\textsuperscript{a}Accenture Netherlands; \textsuperscript{b}University of Amsterdam; \textsuperscript{c}Tinbergen Institute Amsterdam

\textbf{ABSTRACT}

The authors examine the predictive capabilities of online investor sentiment for the returns and volatility of MSCI U.S. Equity Sector Indices by including exogenous variables in the mean and volatility specifications of a Markov-switching model. As predicted by the semistrong efficient market hypothesis, they find that the Thomson Reuters Marketpsych Indices (TRMI) predict volatility to a greater extent than they do returns. The TRMI derived from equity specific digital news are better predictors than similar sentiment from social media. In the two-regime setting, there is evidence supporting the hypothesis of emotions playing a more important role during stressed markets compared to calm periods. The authors also find differences in sentiment sensitivity between different industries: it is greatest for financials, whereas the energy and information technology sectors are scarcely affected by sentiment. Results are obtained with the R programming language. Code is available from the authors upon request.

\textbf{KEYWORDS}

Investor sentiment; Social media; Digital news; Markov switching; GARCH; Value at risk

\textbf{INTRODUCTION}

The hypotheses of efficient markets and rational agents have long formed the basis of economic models. Lately, however, an increasing body of research focuses on market irrationality, behavioral biases, and crowd psychology. This article attempts to capture the emotions that influence investor decisions by extracting them from digital media data such as online news and social media platforms.

There is a substantial literature confirming the significance of irrationality in financial markets. Early work by De Long et al. [1990] distinguished two groups of investors: arbitrageurs and irrational noise traders. Shleifer and Vishny [1997] showed how unpredictable and long-lasting beliefs of noise traders can cause prices to diverge significantly from their fundamental values even in the absence of fundamental risk. They stressed that it is essential to understand the source of noise trading that causes the mispricing. Sentiment is one way of measuring investor mood swings. In this study, we focus on Internet sentiment specifically. Internet sentiment has a number of advantages, such as rapid availability and lack of the response bias associated with traditional surveys. Another noteworthy feature is that indicators of Internet sentiment, including those used herein, typically measure the sentiment of a broad section of the population at large, rather than just investors. Bollen, Mao and Zeng [2011] argued that “[i]t is therefore reasonable to assume that the public mood and sentiment can drive stock market values as much as news” (p. 1). In that sense, and because the information is publicly available, tests for predictive power of internet sentiment are essentially tests for semistrong market efficiency.

Table 1 provides an overview of previous studies based on sentiment data. We distinguish four different sources of online sentiment which may potentially influence financial markets in different ways: professional news such as Wall Street Journal columns (Tetlock [2007]), chatter by retail investors, for example on Twitter (Rao and Srivastava [2013]), consumer generated content on Amazon.com or eBay (Tirunillai and Tellis [2012]), and general well-being as measured with the happiness of the population (Karabulut [2013]). Predictive relationships between sentiment and stock returns are found in many studies (but not in all), but their directions are often conflicting. For example, Antweiler and Frank [2004] scanned messages in Internet stock chat rooms for buy, hold, and sell recommendations, and find that message activity does not predict returns, but rather return volatility. Other studies find
that sentiment leads volatility by investigating the implied Volatility Index VIX (see Table 1). Message posting has also been shown to correlate with trading volume, which in turn is positively correlated with volatility (Jones et al. [1994]).

The unique contribution of this study is threefold. First, we rely on the Thomson Reuters Marketpsych Indices (TRMI). The TRMI scan 50,000 professional news and 2 million social media sites for content every day, dating back to 1998. For each of these two types, they monitor 24 different emotions, rather than just bipolar positive or negative sentiment.

Second, we employ a regime-switching model to examine the impact of sentiment on both expected returns and conditional volatility. Specifically, we use a general Markov-switching model as described in Hamilton [1994], with sentiment variables in both the conditional mean and volatility equation. A two-regime setup allows us to investigate the effect of sentiment separately for calm and stressed markets. Likelihood ratio tests against simpler specifications clearly favor our model, as do information criteria and a backtesting exercise. We find that the impact of sentiment on conditional volatility is greater than the effect on expected returns, and that it plays a larger role in the regime that can be identified with high volatility periods such as the Internet bubble and the recent financial crisis.

Third, the TRMI employ entity reference algorithms to categorize web content according to the industry it concerns. This allows us to consider ten different equity sectors separately, rather than an industry-wide index as is typically done. More importantly, by using as our dependent variable the excess returns per sector over the market, we implicitly control for potential macroeconomic confounders, at least to the extent that they affect the market as a whole. Using monthly data, Stambaugh, Yu and Yuan [2012] found that the effect of sentiment is robust to including macro variables. In this study, we rely on daily data, which makes it impossible to explicitly control for macro factors due to the lower frequency with which they are published. Being able to control for them implicitly is thus a major advantage. The focus of the study is on the industrial and financial sectors, with results for other sectors available from the authors upon request.

The remainder of the study is organized as follows. In the second section we introduce the model. A description of the data is given in the third section. In the fourth section we provide estimation results, and in the fifth section we offer a conclusion.

Table 1. Existing literature on sentiment and stock market activity.

| Study                      | Type     | Source                                      | Return prediction | Volatility prediction |
|----------------------------|----------|---------------------------------------------|-------------------|-----------------------|
| Antweiler and Frank [2004] | Online   | Yahoo! Finance, Raging Bull                | No                | Yes                   |
| Tetlock [2007]             | Offline  | Wall Street Journal                         | Pessimism         | VIX Marketpsych Indices |
| Gilbert and Karahalios [2010]| Online  | LiveJournal weblogs                         | Anxiety           | VIX                   |
| Mao et al. [2011]          | Both     | Twitter, News, Google search                | Mixed             | VIX                   |
| Da et al. [2011]           | Online   | Google search                               | Yes               | —                     |
| Bollen et al. [2011]       | Online   | Twitter                                     | Calmness          | —                     |
| Baker et al. [2012]        | Offline  | Macro data                                  | Yes               | —                     |
| Finter et al. [2012]       | Offline  | Macro data                                  | No                | —                     |
| Tirunillai and Tellis [2012]| Online  | Amazon, Yahoo!, Ebay                        | Yes               | Trade volumes         |
| Rao and Svarstava [2013]   | Online   | Twitter, Google search                      | Yes               | VIX                   |
| Preis et al. [2013]        | Online   | Google search                               | Yes               | —                     |
| Karabulut [2013]           | Online   | Facebook                                    | —                 | Trade volumes         |

*Gilbert and Karahalios [2010] found that their anxiety index is an alternative to the VIX fear gauge.

Model framework

Before discussing our general regime-switching model, we introduce the mean and variance specifications to be used therein in the single regime setting. Let

\[ r_t = \mu_t + \alpha_1 e_t, \quad \alpha_1 = \sigma_1 v_t, \]

denote the log return of an equity sector at time t. We model the conditional mean and variance respectively as

\[ \mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{k} \delta_i x_{i,t-1}, \tag{1} \]

\[ \log \sigma_t^2 = \omega + \alpha_1 e_{t-1} + \gamma_1 \left( e_{t-1} | - \bar{E} | e_{t-1} \right) + \beta_1 \log \sigma_{t-1}^2 + \sum_{i=1}^{k} \phi_i v_{i,t-1}, \tag{2} \]

where \( p \) and \( k \) are non-negative integers and \( e_t \) is i.i.d. with mean 0 and variance 1. Equation 1 is referred to as an ARX Model; an AR(P) model, augmented by the exogenous explanatory variables \( x_{i,t} \). The AR structure is designed to remove any linear dependence from the asset return series \( \{ r_t \} \), though market efficiency would dictate that such dependence be weak.
The volatility dynamics given in Equation 2 will be referred to as an EGARCHX specification. It is based on the exponential GARCH (EGARCH) model proposed by Nelson [1991]. As before, the “X” refers to the inclusion of the exogenous variables $v_{t-1}$ as in Liu, Margaritis and Wang [2012]. There are two advantages to using an EGARCH specification as opposed to the GARCH($m, s$) structure of Bollerslev [1986] as in Chebbi, Louafi and Hedhli [2013] and Han and Kristensen [2012]. First, it allows for an asymmetric news impact curve: due to the leverage effect, one expects a large negative shock to increase volatility more than a positive shock of similar absolute magnitude. This will be true if $\alpha_1 < 0$. Second, in a GARCH specification, it is difficult to ensure positivity of the conditional variance if exogenous regressors are included. Han and Kristensen [2012] solved this problem by squaring the exogenous regressors and imposing a non-negativity constraint on the parameters. The drawback of this approach is that negative and positive values have the same impact on volatility, so that any asymmetry is removed. Taking the exponential as in Equation 2 ensures that the conditional variance is always positive without any parameter restrictions. The model is easily extended to higher lag orders, but we constrain ourselves to the EGARCHX(1,1) model. The model is weakly stationary if $|\beta_1| < 1$ and easily estimated using quasimaximum likelihood, yielding consistent and asymptotically normal estimates under weak regularity conditions.

The previous ARX-EGARCHX model assumes the relationships between variables to be constant over the entire sample, and in particular to be identical in bear and bull markets. Gray’s [1996] generalized regime-switching model lifts this restriction, by allowing the model parameters in Equations 1 and 2 to take two different values, depending on a latent regime indicator $S_t \in \{1, 2\}$. Every period, the system either remains in the regime it is currently in, or transitions to the other. The transition probabilities are assumed constant$^1$ and can thus be collected in a matrix

$$
(S_t = i \mid S_{t-1} = j) = \begin{pmatrix} P & 1 - Q \\ 1 - P & Q \end{pmatrix} = : \Pi.
$$

where $P$ and $Q$ are model parameters to be estimated. The fact that $S_t$ is unobserved complicates estimation: the log-likelihood is

$$
\mathcal{L} = \sum_{t=1}^{T} \log f(r_t \mid F_{t-1})
$$

where

$$
f(r_t \mid F_{t-1}) = \phi(r_t, \mu_{it}, \sigma_{it}^2) \pi_{it} + \phi(r_t, \mu_{2t}, \sigma_{2t}^2) \pi_{2t},
$$

and the ex ante probabilities $\pi_{it} = \mathbb{P}(S_t = i \mid F_{t-1})$ can be evaluated by a recursive updating procedure known as the Hamilton filter [Hamilton, 1994]: Letting $\pi_t = (\pi_{1t}, \pi_{2t})'$, one iterates $\pi_t = \Pi P_{t-1}$ and

$$
P_{it} = \frac{\phi(r_t, \mu_{it}, \sigma_{it}^2) \pi_{it}}{f(r_t \mid F_{t-1})},
$$

where $p_t = (p_{1t}, p_{2t})'$, with $p_{it} = \mathbb{P}(S_t = i \mid F_t)$. The $p_{it}$ are known as the filtered probabilities. For later reference, the smoothed probabilities $\mathbb{P}(S_t = i \mid F_t)$ incorporate information from the entire dataset.

To complete the model, it remains to specify $\mu_{it}$ and $\sigma_{it}^2$ in Equation 4. It appears natural to assume the general form of Equations 1 and 2. This would imply however that the volatility at time $t$ depends on the regime at time $t - 1$, which is unsolved. Similarly, the volatility at time $t - 1$ in turn depends on the regime at $t - 2$, and so on. This problem of full path dependence is illustrated graphically in Gray [1996]. His solution is to modify the EGARCH equation as

$$
\log \sigma_{it}^2 = \omega_i + \alpha_i \epsilon_{t-1} + \gamma_i (| \epsilon_{t-1} | - \mathbb{E} | \epsilon_{t-1} | ) + \beta_i \log \sigma_{it-1}^2 + \sum_{j=1}^{k} \theta_{ij} v_{jt-1},
$$

where $\sigma_{it}^2 = p_{it}(\mu_{1t}^2 + \sigma_{1t}^2) + p_{2t}(\mu_{2t}^2 + \sigma_{2t}^2) - (p_{it}\mu_{1t} + p_{2t}\mu_{2t})^2$, $\epsilon_t = \alpha_i / \sigma_i$, $\alpha_i = r_t - (p_{it}\mu_{it} + p_{2t}\mu_{2t})$, and $p_{it} = \mathbb{P}(S_t = i \mid F_t)$. The single regime EGARCH Process (2) is stationary if $\beta < 1$. For the model in Equation 5, this condition is sufficient, but it is unclear whether it is necessary. In our empirical application, we will impose the condition nevertheless, and it rarely appears to bind. We refer to the full model as the MS-EGARCHX model.

**Data**

Our data are collected by Marketpsych Data and are provided by Thomson Reuters. The TRMI are sentiment indices, updated every minute and comprising time series of human emotions derived from online media sources, dating back to 1998. Web content crawled from the Internet is screened for its financial relevance, and emotions are extracted that are specific to several financial markets.

The TRMI distinguish between content derived from news and content derived from social media. This allows us to compare the impact of professional news with that of content associated with retail investment. For the first category, more than 50,000 Internet news sites are scraped every day, including leading news sources such as The New York Times, The Wall Street Journal, and Financial Times. Less influential news sources are
captured through crawler content from Yahoo! and Google news aggregators. We will refer to this category as news.

The TRMI social media content stems from more than 2 million social media sites. Primary sources include Stocktwits, Yahoo! Finance, Blogger, and other common chat rooms, forums, and blogs. The collection starts in 1998 with a number of small Internet forums, and the content analyzed grows dramatically with the rise of large social media platforms in the second half of the previous decade. Marketpsych claims to capture the top 30% of blogs, microblogs, and other social media sources. Mentions associated with retail consumption are excluded, the idea being that, for example, content from a forum on how to repair a Dell laptop will not aid in forecasting technology stocks. We refer to this type of data as social media.

The entire content base includes over 2 million new articles and posts every day. Minutes after publication, new content is processed into the TRMI feed, after which linguistic software scores the content specific to companies, currencies, commodities, and countries. The algorithm searches for keywords around the topic mentioned and looks for up words and down words in a dictionary. For example, terrible will yield a negative score, but terribly good should yield a (doubly) positive score. Machine learning algorithms are employed to resolve ambiguities and deal with variations in data sources and changes in a word’s meaning over time. These linguistic techniques allow The TRMI to score along a number of dimensions, including specific emotions and so-called buzz metrics, unlike the simple bipolar sentiment variables derived from the psycho-social categories based on the Harvard General Inquirer lexicon as used in Tetlock [2007] and Mao, Counts and Bollen [2011]. Buzz is the term used by Marketpsych for message volumes, so that, for example, the variable litigations gives an indication on the amount of discussion on people being fired. In total there are 24 variables available for each equity sector index: see Table A1 in Appendix A. More details on the TRMI can be found in the Marketpsych white paper (Peterson [2013]).

The TRMI track a broad range of entities including 29 currencies, 34 commodities, and 119 countries. We focus on 10 of 41 equity indices that correspond to the 10 MSCI U.S. equity sectors. The TRMI daily readings come in every day at 20.30, with the first observations on January 1, 1998, for both the news and social media data type. The last reading in our sample is on June 30, 2013, totaling 5,660 observations. There are 809 weekends in this 15.5-year period, leaving 4,042 weekdays. We use the social media sentiment series from the launch of Twitter in August 2006 onward, as this caused a major structural break in the series. This results in a total of 1,804 readings.

In the following we describe the transformations that are applied to the data. There are structural breaks in the data that coincide with the addition of new data sources to the news and social media feeds (Figure 1). As the dates of these breaks are known, their existence can be confirmed with a variation of Chow’s breakpoint test. For testing the break in the mean (variance), one regresses the levels (squares) of the series on an intercept and a post-break dummy. Table 2 shows the results for news sentiment in the Industrials sector. Clearly the breaks are highly significant. We correct for them by standardizing each period, using the mean and standard deviation per period.

An increasing trend in news buzz is removed by dividing buzz by its four week moving average. We refer to the resulting variables as buzzweights, defined as

\[
\text{buzzwgt},_t = \frac{\text{buzz}}{} \\
\%
\]

In order for the regression in Equation 1 to be valid, the \(x_{it}\) need to be stationary. This is tested by applying augmented Dickey-Fuller (ADF) tests to buzzweight and the sentiment series. In all cases, the null hypothesis of

\[ \text{Figure 1. News sentiment (SNTMENT) specific to the industrials sector (ID). The left panel of this figure shows the existence of two structural breaks in the raw sentiment data for the Industrials sector. The right panel has boxplots of the standardized sentiment series.} \]
nonstationarity is rejected. However, some of the sentiment series exhibit strong daily seasonality. For example, the top panel of Figure 1 shows that sentiment is structurally more positive during the week than on weekends for the Industrials sector. The box plots in the lower panel reveal additional intra-week differences: sentiment on Industrials equity depresses over the week from Monday to Friday. We correct for the daily seasonality using the simple additive model

\[ Y_t = S_t + I_t, \]

where \( S_t \) denotes the daily seasonality. The deseasonalized time-series \( I_t \) is obtained by subtracting the seasonal effects from \( Y_T \). We estimate \( S_t \) as the averages of each day of the week over the entire time frame.

Missing observations occur in sentiment series when there were no conversations on a specific topic on a given day. This differs from a zero reading in the sense that the latter did involve buzz, but people were neutrally disposed on a subject. We impute missing observations by carrying the previous observation forward. As shown in Ryan and Giles [1998], this imputation method leads to ADF tests with superior power compared to other standard methods. The amount of missing values depends on the type of variable and the industry it measures. Emotional TRMI such as joy and fear are almost always available, while buzz metrics such as market forecast and layoffs have the most missing values. Missing observations tend to cluster, possibly indicating the existence of tranquil periods with no major events. To test the hypothesis that missing values predict a lower volatility, we include a dummy that is equal to 1 when the observation for volatility regressor \( v_t \) is missing at time \( t \):

\[ D_t^{NA} = 1 \text{ if } v_t = \text{NA} \]

Each observation of every series is multiplied by the buzzweight for that particular day, thus giving more weight to a sentiment that is based on high message volume. Another advantage is that extreme observations caused by only a few overly enthusiastic sources is downweighted.

Despite the fact that returns are only observed on weekdays, we cannot exclude the possibility that they are affected by weekend sentiment, especially regarding news data; for example, Pettengill [2003] argued that a large amount of bad news published over the weekend could be the cause of the “Monday effect.” For social media, weekend data might be less relevant, because people tend to discuss their private life when markets are closed (Peterson, [2013]). To account for the release of important news over the weekend (China data, EU summits), we include a dummy that is 1 or \(-1\) on Mondays whenever the buzz-weighted sentiment of the weekend before is significantly different compared to the prior twelve weekends (a three-month period):

\[
 p^{Wknd}_{\text{Mon}} = \begin{cases} 
 1, & \text{if } x^{Wknd} > \mu_{\text{Prior}12} + 2\sigma_{\text{Prior}12} \\
 -1, & \text{if } x^{Wknd} < \mu_{\text{Prior}12} - 2\sigma_{\text{Prior}12} \\
 0, & \text{otherwise.} 
\end{cases}
\]

In Equation (6), \( \mu_{\text{Prior}12} \) and \( \sigma_{\text{Prior}12} \) denote the 12-week rolling mean and standard deviation, respectively. A weekend’s sentiment value \( x^{Wknd} \) is calculated as the Sunday minus the Friday observation. The dummy is added in both the return and volatility equations of our models.

After the above data transformations, we difference each regressor, yielding one-day, one-week, and four-week differences,

\[
 x_t^{1D} = x_t - x_{t-1} \\
 x_t^{1W} = x_t - x_{t-7} \\
 x_t^{4W} = x_t - x_{t-28}.
\]

**MSCI U.S. Equity Sector Indices**

As dependent variable we use daily closing prices of the MSCI U.S. Equity Sector Indices from Thomson Reuters. Unlike weekends, public holidays are included, but the observation is the same as on the previous day. We create excess returns by subtracting the general MSCI U.S. log return from each individual sector’s log return:

\[
 r_{jt} = (\log P_{j,t} - \log P_{j,t-1}) - (\log P_{\text{US},t} - \log P_{\text{US},t-1}),
\]

because our interest lies in the effect of sentiment on individual sector returns, rather than movements of the entire market. Descriptive statistics are given in Table 3.
As is common for financial returns data, the data exhibit strong non-normality.

**Results**

To avoid computationally expensive frequent re-estimation of the full model, we conduct a preliminary analysis to determine the most relevant variables. A specific-to-general two-pass estimation approach is applied to assess each variables’ predictive power individually. In the first step, the three lags of a variable and its dummies are included in a basic ARX model (Equation 1) per sector. If more than two variables are significant at the 1% level, then the two most relevant ones are retained in a combined ARX model including the other sentiment covariates. We find that some variables lose their significance when other regressors are included.

In the second step, an EGARCHX model (Equation 2) is fit to the residuals of the combined first step ARX model. Again, a variable is retained if it is significant at the 1% level, after which the best two are selected through a combined EGARCHX estimation. The estimated signs of regression coefficients have to satisfy the prior beliefs in Table A1 to prevent hindsight bias.

The results of the preliminary analysis reveal five immediate findings. First, sentiment appears to play a more important role in predicting volatility than returns; more variables were found to be significant in the second step estimation compared to the first step. This is also true for the social media data. Second, news sentiment appears to be more powerful than social media sentiment: fewer TRMI were found to be significant in both the mean and volatility equation for the latter. Third, the financials sector appears to be the most sentiment-sensitive in this data set, especially when compared to, for example, Telecom or Energy, for which rarely any sentiment is significant. Fourth, differences between industries also exist with respect to which particular sentiment variables matter. News optimism and social media fear are the only recurring emotions across sectors. Social media apparently picked up the fear of defaults during the Lehman and Euro Crises, whereas professional news optimistically focused on the positive signs of a solution or recovery.

Lastly, the no-talk dummies are rarely significant and inconsistent with our prior beliefs. For example, we found a significant positive effect on volatility of no talk about layoffs, contradicting our hypothesis that silence on digital media corresponds to periods of low volatility. We therefore exclude the dummies from further analysis. The dummies indicating extraordinary weekend sentiment however do matter and are therefore retained. For every sector the two most significant TRMI are chosen for further analysis in the dual regime context. A table with the results across all sectors can be obtained from the authors upon request.

**News: Markov switching without conditional heteroskedasticity**

Before combining Markov switching and conditional heteroskedasticity, we compare the single regime ARX-EGARCHX model with a Markov-switching model that assumes constant variances. This will allow us to disentangle the improvements offered by these two aspects of the full model. The MS-constant variances model has the same log-likelihood as in Equation 3, with $\sigma_i^2$ replaced by $\sigma_{i2}^2$; the variance in each regime is no longer time varying. In order to keep the exposition concise, we will focus on the industrials sector for the remainder of the article. Sentiment has played a larger role in the financials industry, but the large industry specific shocks that this particular sector has witnessed lead to exceptional estimation results. We will highlight some of them throughout and provide detailed estimation results in the Appendix, but focus on the more representative industrials sector here.

The results of both models for industrial specific news sentiment are shown in Table 4. The parameters in the first column match the notation used in the second section. An additional subscript indicates the regime of a parameter in the dual state scenario. Following our findings from the preliminary analysis, we add trust and market forecast as regressors to the mean equation in both models. The EGARCHX process of the single

---

Table 3. Descriptive statistics of MSCI U.S. sector excess returns.

|        | CD  | CS  | E1  | FN  | HC  | ID  | IT  | M1  | T1  | U1  |
|--------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Mean   | 0.01| 0.00| 0.02| −0.01| 0.01| 0.00| 0.01| −0.00| −0.02| −0.01|
| Min    | −5.66| −9.70| −8.90| −11.30| −8.87| −5.35| −6.46| −6.13| −7.59| −10.30|
| Max    | 3.83| 5.31| 11.38| 9.77| 4.98| 4.22| 10.75| 7.20| 7.67| 7.12 |
| SD     | 0.61| 0.93| 1.27| 1.12| 0.89| 0.58| 1.06| 1.03| 1.06| 1.10 |
| Skewness| −0.22| −0.43| −0.12| 0.58| −0.27| −0.23| 0.25| 0.09| 0.01| −0.36|
| Kurtosis| 5.71| 8.34| 4.99| 16.03| 6.64| 6.71| 7.93| 4.87| 5.51| 5.09 |

*Note. From left to right: consumer discretionary, consumer staples, energy, financials, healthcare, industrials, information technology, materials, telecom, utilities.*
regime model is driven by buzzweight and optimism (recall that the variances of the Markov-switching model are constant for now). $P$ and $1 - Q$ are the estimated transition probabilities.

The estimated EGARCHX model is stationary ($\beta < 1$), and the model captures a significant leverage effect as $\alpha$ is negative. The news regressors market forecast and buzzweight predict respectively the returns and volatility of the industrials sector. The weekend dummy of trust is no longer significant.

The volatilities of the two regimes of the Markov-switching model are estimated at 0.3575 and 0.8887, corresponding to distinct low and high volatility regimes, respectively. Based on the estimated transition probabilities, we determine the average duration of each regime $i$ as $1 / (1 - P(S_i - i | S_{i-1} = i))$. With a duration of 148.85 days, we find that the low volatility regime is more persistent than the high volatility regime (67.28 days). This is consistent with the findings of Gray [1996]. The persistence of regimes varies considerably across sectors; for financials, the low and high volatility regime have an average duration of only 70.55 and 31.99 days, respectively.

Considering the parameter estimates of the regime switching model, it is apparent that the single regime model potentially “averages out” the effect of a variable. For example, the weekend dummy for trust is now significant at the 1% level in the high-volatility regime. This pattern is repeated across all industries and both data types: sentiment is more relevant for predicting returns in periods of high volatility. This confirms our hypothesis that emotions play a larger role in stressed markets.

It is interesting to note that the autoregressive terms are significant in both models, which is at odds with weak-form market efficiency. We nevertheless exclude the autoregressive terms from the estimation of our full model, because they entail a considerable computational cost. Due to their small estimated coefficients we expect the effect of their omission to be small.

Comparing the single regime EGARCHX model with the two-regime model with constant volatilities, it becomes apparent that the latter is too restrictive: the model has a lower log-likelihood, and its standardized residuals do not resemble white noise. The conditional volatility, shown in the middle panel of Figure 2, closely mimics the smoothed probability of being in the high volatility regime. Nevertheless, the smoothed probabilities capture important volatile periods, such as the Internet bubble and the 2008 financial crisis.

### News: Markov switching with heteroskedasticity

We now turn to the estimation of our full model, combining the Markov-switching and EGARCHX models of the previous section. For comparison, we also estimate the MS-EGARCH model without exogenous sentiment regressors. The estimates for the three models applied to the industrials sector are shown in Table 5.

The EGARCHX estimates without AR terms are almost identical to those in Table 4. The Markov-switching models still display distinct low and high volatility regimes, characterized in particular by the highly significant constants $\omega$, in the volatility specifications. The regimes have become considerably more persistent, with transition probabilities $P$ and $Q$ approaching unity. Within regimes however, volatility is less persistent in low volatility periods, as evidenced by the GARCH parameter $\beta_1$ being smaller than $\beta_2$. As $\beta_1 < 1$ for both states, the model is stationary overall. Another interesting finding is that the leverage effect is very much present in the high volatility regime, but is not even significant when volatility is low. This result could not have been captured by a single regime model. This is a recurring finding across many sectors. Likelihood ratio tests confirm that Markov switching is an improvement upon the single regime model (likelihood ratio statistic of 72.85), and adding sentiment to it improves the model fit even more (LR statistic of 51.19). Similar results are found for most sectors; Table A2 shows this for financials, with results for other sectors available upon request.

The signs of all significant coefficients are as expected (Table A1); for example, we find that predictions of...
Figure 2. Markov-switching model with constant variances for Industrial specific news sentiment. This figure shows excess returns (top panel), the conditional volatility (middle panel), and smoothed transition probabilities (bottom panel). Areas in which the latter exceeds 0.5 are shaded gray.

Table 5. Comparison of three models for industrial specific news sentiment.

| Parameter          | EGARCHX | SE        | MS-EGARCH | SE        | MS-EGARCHX | SE        |
|--------------------|---------|-----------|-----------|-----------|------------|-----------|
| $\phi_{0,1}$       | 0.0020  | (0.0074)  | 0.0052    | (0.0071)  | 0.0092     | (0.0074)  |
| TRUST wknd$_1$     | 0.0648  | (0.0524)  | —         | —         | 0.0340     | (0.0565)  |
| MKTFCST 4W$_1$     | 0.0166  | (0.0085)  | —         | —         | 0.0002     | (0.0089)  |
| $\phi_{0,2}$       | —       | —         | —         | —         | —          | —         |
| TRUST wknd$_2$     | —       | —         | —         | —         | —          | —         |
| MKTFCST 4W$_2$     | —       | —         | —         | —         | —          | —         |
| $\omega_1$         | $-0.0000$ | (0.0001) | $-0.4461$ | (0.1150)  | $-0.4192$  | (0.0885)  |
| $\alpha_1$         | $-0.0201$ | (0.0087)  | $-0.0078$ | (0.0137)  | $-0.0141$  | (0.0190)  |
| $\epsilon_1$      | 0.9973  | (0.0012)  | 0.8319    | (0.0481)  | 0.8442     | (0.0370)  |
| $\gamma_1$         | 0.0848  | (0.0231)  | 0.0778    | (0.0314)  | 0.0675     | (0.0322)  |
| BUZZWGT 1D$_1$     | 0.1526  | (0.0488)  | —         | —         | 0.1853     | (0.0430)  |
| OPTIMSM 1D$_1$     | $-0.0619$ | (0.0386) | —         | —         | 0.0060     | (0.0355)  |
| $\omega_2$         | —       | —         | $-0.1154$ | (0.0195)  | $-0.0856$  | (0.0141)  |
| $\alpha_2$         | —       | —         | $-0.0787$ | (0.0164)  | $-0.0685$  | (0.0120)  |
| $\epsilon_2$      | 0.9423  | (0.0192)  | 0.9752    | (0.0101)  | 0.1065     | (0.0165)  |
| $\gamma_2$         | —       | —         | 0.1365    | (0.0222)  | 0.0847     | (0.0400)  |
| BUZZWGT 1D$_2$     | —       | —         | —         | —         | 0.0911     | (0.0296)  |
| OPTIMSM 1D$_2$     | —       | —         | —         | —         | 0.0991     | (0.0006)  |
| $p$                | —       | —         | —         | —         | 0.9991     | (0.0006)  |
| $1 - Q$            | —       | —         | 0.9990    | (0.0006)  | 0.9991     | (0.0006)  |
| Log-likelihood     | $-2804.62$ |         | $-2790.35$ |         | $-2764.78$ |         |
| AIC                | 1.400   |         | 1.394     |         | 1.385      |         |

***, ***, and † indicate statistical significance at 0.1%, 1%, 5%, and 10%, respectively.
industrial asset prices made in professional news (market forecast) drive prices upward during high volatile periods. Increasing optimism in professional news is a good indicator of decreasing volatility. Higher volume of news messages leads to increased volatility in both regimes.

The thick gray (red online) line in the lower part of Figure 3 shows the smoothed probability of being in State 2. It is seen that the regimes have become more persistent compared to the homoskedastic Markov-switching model: essentially the model now identifies two main stressed periods. The ex ante probabilities are shown as the thinner black line in the same graph, with periods in which it exceeded 0.5 shaded gray. It is observed that the probability of being in the high volatility regime gradually builds up in the years before the financial crisis, but remains rather volatile itself. Therefore it is not clear whether the statistic was a good predictor of 2008’s steep market decline. Figure A2 in the Appendix contains the corresponding graphs for the Financials sector. Here, a sudden regime switch occurs at the end of 2007. Comparing the conditional volatility in the middle panel with the middle panel of Figure 2 shows that because of the increased flexibility of the model, the conditional volatility of the MS-EGARCHX model is smoother than that of the MS model with constant variances.

We conclude our comparison of models with a series of out-of-sample tests. We reserve the last 2.5 years of data starting in January 2011 for this purpose. The out-of-sample period contains 650 observations and includes the Euro crisis. As before, we focus on the Industrials sector for this exercise. Results for the financial sector are presented in Table A3 in the Appendix.

Return prediction is evaluated based on mean squared error (MSE), mean absolute error (MAE), and the hit ratio (HR). The latter statistic indicates how often the sign of the return is predicted correctly. Table 6 contains the results. We find that the differences between models are small, implying that neither adding sentiment nor multiple regimes help in forecasting returns, despite the coefficients being significantly different from zero. This finding is consistent with semistrong market efficiency.

Figure 3. MS-EGARCHX model for industrial specific news sentiment. This figure shows the MSCI Industrial sector index (top panel), the conditional volatility (middle panel), and in the bottom panel ex ante (thin black line) and smoothed (thick gray [red online] line) transition probabilities. Shaded areas indicate time periods in which the ex ante probability of being in the high volatility regime exceeded 0.5.
The second set of backtests comprise value-at-risk (VaR) statistics. For computing 1% and 5% VaR forecasts, we fit a scaled student’s $t$ distribution to the standardized residuals of each model to allow for fat tails. Two hypotheses will be tested: the unconditional coverage hypothesis is violated if the average frequency of VaR exceedances differs from the nominal rate of 1% or 5%. The independence hypothesis is violated if the average frequency of VaR exceedances is clustered. Both hypotheses can be tested by defining the hit series

$$Hit_t(\alpha) = \begin{cases} 1 - \alpha & \text{if } r_t < -\text{VaR}_{1|t-1}(\alpha), \\ -\alpha & \text{otherwise,} \end{cases}$$

and regressing it on a constant and its first lag, that is,

$$Hit_t(\alpha) = b_0 + b_1 Hit_{t-1}(\alpha) + \varepsilon_t.$$  

Unconditional coverage and independence can respectively be tested with $t$-tests on $b_0$ and $b_1$. This is known as the Engle-Manganelli test (Engle and Manganelli [2004]). Figure A1 shows the VaR forecasts for industrials. The test results in Table 6 show that the independence hypothesis is rejected if exceedances are clustered. Both hypotheses can be tested by defining the hit series

$$Hit_t(\alpha) = \begin{cases} 1 - \alpha & \text{if } r_t < -\text{VaR}_{1|t-1}(\alpha), \\ -\alpha & \text{otherwise,} \end{cases}$$

and regressing it on a constant and its first lag, that is,

$$Hit_t(\alpha) = b_0 + b_1 Hit_{t-1}(\alpha) + \varepsilon_t.$$  

Unconditional coverage and independence can respectively be tested with $t$-tests on $b_0$ and $b_1$. This is known as the Engle-Manganelli test (Engle and Manganelli [2004]). Figure A1 shows the VaR forecasts for industrials. The test results in Table 6 show that the independence hypothesis is strongly rejected for the single regime models for the 1% VaR level, whereas no test rejects at the 5% level for the three Markov-switching models. We observe that for both the single- and dual-regime models, the inclusion of sentiment regressors increases the estimated degrees of freedom of the $t$ distribution fitted to the residuals, indicating that the fat-tailedness of the data are slightly reduced by taking sentiment into account. Similar conclusions can be drawn across all sectors.

### Social media

We repeat the analysis of the preceding section for the social media TRMI. As with the news data, we focus on the Industrials sector and relegate the results for the Financials sector to the Appendix. As discussed in the third section (Thomson Reuters Marketpsych Indices), the sample for social media starts in August 2006 with the launch of Twitter, leaving only 1,804 observations compared to 4,020 for the news data.

From the preliminary two-step analysis the four sectors energy, healthcare, materials, and telecom do not appear to be driven by social media at all. Table 7 shows the estimates for the full model and its special cases fitted to Industrials. No sentiment regressors were included in the mean equation, but we find that a four-week increase in fear causes higher volatility in the single regime model. However, in the two regime case this significance is lost and the sign reversed. This is likely due to the small sample size (note the large standard errors). The insufficient sample size is particularly apparent for the financials sector, because the sample is largely dominated by financial crises. Therefore, the model remains in the high volatility regime for the entire period between the housing bubble and early 2013. Consequently, there exist only very few data points from which to estimate the low volatility regime parameters. Likelihood ratio tests indicate that the Financials sector is the only one for which including social media sentiment significantly increases the likelihood (see Appendix Table A2).

Certain other irregularities of the estimates can be attributed to the small sample size as well. For example, the estimates for $\beta_i$—and hence the persistence in volatility—are quite low, and the constants $\omega_i$ in the volatility equations differ markedly between regimes. In fact, for the information technology (IT) sector, the difference between $\omega_1$ and $\omega_2$ is so pronounced that the MS-EGARCHX model essentially behaves like a MS model with constant variances.

Another problem associated with the small sample size is that for some sectors, the estimators converge to different local optima depending on the starting values used. Some of these local optima are associated with very small transition probabilities $P$ or $Q$, essentially making one regime redundant. In these cases, one of the regimes essentially serves to capture an outlier. A similar

### Table 6. Out-of-sample test results for industrial specific news sentiment.

|               | EGARCH | EGARCHX | MS-const. | MS-EGARCH | MS-EGARCHX |
|---------------|--------|---------|-----------|-----------|------------|
| Parameters    | 5      | 9       | 10        | 12        | 20         |
| MSE           | 0.1314 | 0.1322  | 0.1320    | 0.1314    | 0.1318     |
| MAE           | 0.2716 | 0.2731  | 0.2721    | 0.2715    | 0.2724     |
| HR (%)        | 48.46  | 45.23   | 45.54     | 47.38     | 46.46      |
| df t-dist     | 6.35   | 6.54    | 13.94     | 7.12      | 7.27       |
| VaR 1% b0    | -0.002 | -0.001  | 0.009†    | 0.008     | 0.009†     |
| VaR 1% b1    | 0.194*** | 0.159*** | 0.058     | 0.066†    | 0.058      |
| VaR 5% b0    | -0.004 | -0.008  | -0.002    | -0.005    | 0.001      |
| VaR 5% b1    | 0.021  | 0.034   | 0.018     | 0.025     | 0.042      |

Note. This table shows out-of-sample test results for 5 models based on industrial specific news content. $b_0$ and $b_1$ correspond to tests for unconditional coverage and independence, respectively. *** and † indicate statistical significance at 0.1% and 10%, respectively.
phenomenon occurs in the estimation of mixture GARCH models. Broda et al. [2013] proposed estimation methods that alleviate the problem in the latter context. Here, we deal with such degenerate estimates by choosing alternative starting values.

We refrain from conducting an out-of-sample exercise for social media data, as removing the most recent 2.5 years of data would leave even fewer observations for estimation.

**Conclusions**

We have considered regime-switching models augmented by exogenous regressors to determine the predictive capabilities of investor sentiment for various equity markets. The estimates demonstrate the existence of a low- and a high-volatility regime, allowing us to analyze whether sentiment plays a more important role in stressed markets compared to calm periods. This is found to be true for the Industrials news data, for which sentiment is a significant predictor of returns during high volatility periods. Across sectors, there is only weak evidence for this theory. For volatility, there is no apparent difference in sentiment’s predictive power between regimes. We find, however, that the leverage effect is more significant during turbulent markets.

Substantial differences in sentiment sensitivity were found across sectors. Among the 10 industries, the TRMI were the most predictive for financial equity. Other sectors, such as energy and IT, are barely affected by sentiment. Marketwide, we find that sentiment predicts volatility better than returns.

One limitation of the present study is the short sample available for the social media data, resulting in large standard errors. Estimating the single regime EGARCHX model on subsamples of the data, we also found more evidence of predictability for more recent periods. Future research based on an extended data set might be able to take advantage of this. Additionally, a larger data set would make the estimation of extended models feasible; for example, one might consider estimating an encompassing model that includes both news and social media, although this presents a major computational challenge. One could also entertain the idea of making the transition probabilities dependent on sentiment as in Gray [1996] and Liu, Margaritis and Wang [2012].

**Acknowledgments**

This article has received an Honorable Mention for the Hillcrest Behavioral Finance Award. The authors would like to thank an anonymous referee whose constructive comments have helped greatly improve the study. Steven J. Nooljen would like to thank NN Investment Partners for providing access to the data set used herein, and for the opportunity to conduct this research during an internship there. All views expressed in the article are those of the authors and do not necessarily reflect those of NN Investment Partners or Accenture Netherlands.

**Notes**

1. They could be made time varying and dependent on external variables through a probit link function as in Gray [1996] and Ozoguz [2009], or a logit function as in Liu, Margaritis and Wang [2012].

2. Here we depart slightly from Gray [1996], who defined $\sigma^2_t$ based on the ante probabilities $\pi_t$, rather than the filtered $p_t$ so that $\sigma^2_t = \text{E}[r^2_t | F_{t-1}] = \text{E}[r_t | F_{t-1}]^2$. The differences between the resulting estimates are, however, small.

---

**Table 7. Comparison of three models for industrial specific social media sentiment.**

| Parameter | EGARCHX | SE | MS-EGARCH | SE | MS-EGARCHX | SE |
|-----------|---------|----|-----------|----|------------|----|
| $\theta_{b_1}$ | 0.0125 (0.0100) | 0.0176 (0.0095) | 0.0173 (0.0095) |
| $\theta_{b_2}$ | — | — | — | — | — |
| $\sigma_{b_1}$ | — | — | — | — | — |
| $\sigma_{b_2}$ | — | — | — | — | — |
| $\beta_1$ | 0.9887 (0.0065) | 0.6895 (0.0943) | 0.7333 (0.1148) |
| $\gamma_1$ | 0.1121 (0.0375) | 0.1281 (0.0492) | 0.1207 (0.0471) |
| FEAR 4W1 | 0.0328 (0.0121) | — | — | — | — |
| LAYOFFS 1W1 | 0.0210 (0.0157) | — | — | — | — |
| $\alpha_2$ | 0.1382 (0.1098) | 0.1366 (0.1146) | — | — | — |
| $\beta_2$ | — | — | 0.0930 (0.0687) | 0.0863 (0.0696) | — |
| $\gamma_2$ | — | — | — | — | — |
| FEAR 4W2 | — | — | — | — | — |
| LAYOFFS 1W2 | — | — | — | — | — |
| $\rho$ | — | — | — | — | — |
| $1 - Q$ | — | — | 0.9965 (0.0019) | 0.9790 (0.0018) | — |

| Log-likelihood | -1055.64 | -1041.28 | -1040.92 |
| AIC | 1.178 | 1.168 | 1.172 |

***, **, * indicate statistical significance at 0.1%, 1%, 5%, and 10%, respectively.
References

Antweiler, W. and M. Z. Frank. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance*, 59, (2004), pp. 1259–1294.

Baker, M., J. Wurgler, and Y. Yuan. "Global, Local, and Contagious Investor Sentiment." *Journal of Financial Economics*, 104, (2012), pp. 272–287.

Bollen, J., H. Mao, and X. Zeng. "Twitter Mood Predicts The Stock Market." *Journal of Computational Science*, 2, (2011), pp. 1–8.

Bollerslev, T. "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 31, (1986), pp. 307–327.

Broda, S. A., M. Haas, J. Krause, M. S. Paolella, and S. C. Steude. "Stable Mixture GARCH Models." *Journal of Econometrics*, 172, (2013), pp. 92–306.

Chebbi, A., R. Louafi, and A. Hedhli. "Financial Fluctuations in the Tunisian Repressed Market Context: A Markov-switching-GARCH Approach." *Macroeconomics and Finance in Emerging Market Economies*, (2013), pp. 1–19.

Da, Z., J. Engelberg, and P. Gao. "In Search of Attention." *The Journal of Finance*, 66, (2011), pp. 1461–1499.

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann. "Noise Trader Risk in Financial Markets." *The Journal of Political Economy*, 98, (1990), pp. 703–738.

Engle, R. F. and S. Manganelli. "Caviar: Conditional Autoregressive Value-At-Risk by Regression Quantiles." *Journal of Business and Economic Statistics*, 22, (2004), pp. 367–381.

Finter, P., A. Niessen, and S. Ruenzi. "The Impact of Investor Sentiment on the German Stock Market." *Zeitschrift für Betriebswirtschaft*, 82, (2012), pp. 133–163.

Gilbert, E. and K. Karahalios. "Widespread Worry and the Stock Market." AAAI Conference on Weblogs and Social Media (ICWSM), 4, (2010), pp. 58–65.

Gray, S. F. "Modeling The Conditional Distribution of Interest Rates As A Regime-Switching Process." *Journal of Financial Economics*, 42, (1996), pp. 27–62.

Hamilton, J. D. *Time Series Analysis*. Princeton: Princeton University Press, 1994.

Han, H. and D. Kristensen. "Asymptotic Theory for the QMLE in GARCH-X Models with Stationary and Non-stationary Covariates." Creates Research Papers, 2012–25, (2012).

Jones, C. M., G. Kaul and M. L. Lipson. "Transactions, Volume, and Volatility." *Review of Financial Studies*, 7, (1994), pp. 631–651. Karabulut, Y. "Can Facebook Predict Stock Market Activity." Available at SSRN: *Http://Ssrn.Com/Abstract=1919008*, 2013.

Liu, X., D. Margaritis, and P. Wang. "Stock Market Volatility and Equity Returns: Evidence from a Two-State Markov-switching Model with Regressors." *Journal of Empirical Finance*, 19, (2012), pp. 483–496.

Mao, H., S. Counts, and J. Bollen. "Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data." Arxiv:1112.105[V], 2011.

Nelson, D. B. "Conditional Heteroskedasticity in Asset Returns: A New Approach." *Econometrica*, 59, (1991), pp. 347–370.

Ozoguz, A. "Good Times or Bad Times? Investor’s Uncertainty and Stock Returns." *The Review of Financial Studies*, 22, (2009), pp. 4377–4422.

Peterson, R. L. "Thomson Reuters Marketpsych Indices (TRMI)." *White Paper* (2013).

Pettengill, G. N. "A Survey of the Monday Effect Literature." *Quarterly Journal of Business and Economics*, 42, (2003), pp. 3–28.

Reis, T., H. S. Moat, and E. Stanley. "Quantifying Trading Behavior in Financial Markets Using Google Trends." *Scientific Reports*, 3, (2013), pp. 1684.

Rao, T. and S. Srivastava. "Using Twitter Sentiments and Search Volumes Index to Predict Oil, Gold, Forex and Markets Indices." *Annual ACM Web Science Conference*, 5, (2013), pp. 336–345.

Ryan, K. F. and D. E. A. Giles. "Testing For Unit Roots in Economic Time-series With Missing Observations." In T. B. Fomby and R. C. Hill (Eds.), *Advances In Econometrics*, (pp. 203–242.) Greenwich: JAI Press, 1998.

Shleifer, A. and R. W. Vishny. "The Limits of Arbitrage." *The Journal of Finance*, 52, (1997), pp. 35–55.

Stambaugh, R. F., J. Yu, and Y. Yuan. "The Short of It: Investor Sentiment and Anomalies." *Journal of Financial Economics*, 104, (2012), pp. 288–302.

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

Tirunillai, S. and G. J. Tellis. "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance." *Marketing Science*, 31, (2012), pp. 198–215.
Appendix A.

Table A1. Thomson Reuters Marketpsych Indices (TRMI) Specific to Equity Markets.

| Acronym | Description | Type | Prior Return | Prior Volatility |
|---------|-------------|------|--------------|-----------------|
| BUZZ    | Sum of entity-specific words and phrases used in computations | Buzz metric | + | + |
| SENTIMENT | Overall positive references, net of negative references | Emotion | + | - |
| OPTIMISM | Optimism, net of references to pessimism | Emotion | + | - |
| FEAR    | Fear and anxiety | Emotion | - | + |
| JOY     | Happiness and affection | Emotion | + | - |
| TRUST   | Trustworthiness, net of references connoting corruption | Emotion | + | - |
| VIOLENCE | Violence and war | Emotion | - | + |
| CONFLICT| Disagreement and swearing, net of agreement and conciliation | Emotion | - | + |
| URGENCY | Urgency and timeliness, net of references to tardiness and delays | Emotion | + | + |
| UNCERTN | Uncertainty and confusion | Emotion | - | + |
| PRICEUP | Price increases, net of references to price decreases | Buzz metric | + | - |
| MKTFCST | Predictions of asset price rises, net of references to predictions of asset price drops | Buzz metric | + | 0 |
| MKTRISK | Positive emotionality and positive expectations net of negative emotionality and negative expectations. Includes factors from social media found characteristic of speculative bubbles - higher values indicate greater bubble risk. also known as the “Bubbleometer” | Buzz metric | + | + |
| GLOOM   | Gloom and negative future outlook | Emotion | - | 0 |
| ANGER   | Anger and disgust | Emotion | - | + |
| INNOVAT | Innovativeness | Buzz metric | + | 0 |
| STRESS  | Distress and danger | Emotion | - | + |
| FUNDSTR | Positivity about accounting fundamentals, net of references to negativity about accounting fundamentals | Buzz metric | + | - |
| EARNEXP | Expectations about improving earnings, less those of worsening earnings | Buzz metric | + | 0 |
| MRGBUZZ | Merger or acquisition activity | Buzz metric | + | + |
| LAYOFFS | Staff reductions and layoffs | Buzz metric | - | + |
| LITIG   | Litigation and legal activity | Buzz metric | - | + |
| UPGDWNGL | Upgrade activity, net of references to downgrade activity | Buzz metric | + | - |
| VOLATIL | Volatility in market prices or business conditions net of stability | Buzz metric | - | + |

We expect the effect of violence and conflict on energy sector returns to be positive (+).

This Table provides descriptions of the Thomson Reuters Marketpsych Indices (TRMI) used for this research. The last two columns present our prior beliefs concerning the relationship of the metric in question with returns and volatility: positive (+), negative (−), or none (0).

Table A2. Log-likelihood Values and LR Tests Corresponding to Five Different Models for Both Data Types on the Industrial and Financial Sectors.

| Statistic | Financials (FN) | Industrials (ID) |
|-----------|-----------------|-----------------|
|           | News data | Social media | News data | Social media |
| Model parameters: | | | | |
| nr. X mean | 5 | 0 | 2 | 0 |
| nr. X vol. | 2 | 2 | 2 | 2 |
| Model log-likelihoods: | | | | |
| EGARCH   | -4191.95 | -2127.77 | -2826.78 | -1063.45 |
| EGARCHX  | -4154.67 | -2111.21 | -2804.62 | -1055.64 |
| MS-Const. | -4555.42 | -2293.99 | -2790.35 | -1041.28 |
| MS-EGARCH | -4152.87 | -2094.42 | -2764.76 | -1040.92 |
| MS-EGARCHX | -4128.37 | -2080.52 | -2764.76 | -1040.92 |
| Likelihood ratio (LR) tests: | | | | |
| EGARCHX-EGARCH | 74.56*** | 33.12*** | 66.71*** | 15.64*** |
| MSEGARCH-EGARCHX | 78.17*** | 27.89*** | 72.85*** | 44.34*** |
| MSEGARCHX-MSEGARCH | 48.99*** | 27.18*** | 51.19*** | 0.73 |

Signif. codes: ‘***’<0.001 ‘**’<0.01 ‘*’<0.05 ‘.’<0.1 ‘ ’ 1

This table’s rows “nr. X mean” and “nr. X vol.” correspond to the number of sentiment regressors in the mean and volatility equations, respectively. Their sum multiplied by the number of regimes in the model in question amounts to the degrees of freedom in the null distribution of the LR test.
This table shows out-of-sample test results for 5 models based on Financial specific news content. $b_0$ and $b_1$ correspond to tests for unconditional coverage and independence, respectively.

|                | EGARCH | EGARCHX | MS-Const. | MS-EGARCH | MS-EGARCHX |
|----------------|--------|---------|-----------|-----------|------------|
| LL             | -4191.95 | -4154.67 | -4555.42  | -4152.87  | -4128.37   |
| Parameters     | 5      | 9       | 10        | 12        | 20         |
| MSE            | 0.4287 | 0.4292  | 0.4437    | 0.4291    | 0.4297     |
| MAE            | 0.4602 | 0.46    | 0.4623    | 0.4596    | 0.4596     |
| HR (%)         | 50.31  | 51.23   | 51.54     | 49.69     | 52.46      |
| df t-dist      | 5.48   | 5.97    | 6.84      | 6.01      | 6.25       |
| VaR 1% $B_0$   | -0.008*** | -0.008*** | 0.001     | -0.003    | -0.004     |
| VaR 1% $B_1$   | -0.002 | -0.002  | 0.134***  | 0.245***  | 0.33***    |
| VaR 5% $B_0$   | -0.02** | -0.022*** | 0.007     | -0.007    | -0.012     |
| VaR 5% $B_1$   | 0.024  | 0.029   | 0.054     | 0.03      | 0.048      |

*Signif. codes: ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Figure A1. MS-EGARCHX VaR out-of-sample forecasts for industrial news content. This figure shows returns (black), 1% (solid gray [red online] line) and 5% (dashed gray [red online] line) out-of-sample Value at Risk forecasts for Industrials, based on the MS-EGARCHX model driven by news content.
Figure A2. MS-EGARCHX model for financials specific news sentiment. This figure shows the MSCI Financial Sector Index (top panel), the Conditional Volatility (Middle Panel), and in the bottom panel ex ante (thin black line) and Smoothed (thick gray [red online] line) transition probabilities. Shaded areas indicate time periods in which the ex ante probability of being in the high volatility regime exceeded 0.5.