

Essays on Investors' Sentiment and Attention

D I S S E R T A T I O N

of the University of St. Gallen,

School of Management,

Economics, Law, Social Science,

International Affairs and Computer Science,

to obtain the title of

Doctor of Philosophy in Economics and Finance

submitted by

Daniele Ballinari

from

Sagno (Tessin)

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Zhi Da, PhD

Dissertation no. 5086

Difo-Druck GmbH, Untersiemaun 2021

Essays on Investors' Sentiment and Attention

D I S S E R T A T I O N
of the University of St. Gallen,
School of Management,
Economics, Law, Social Science,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Economics and Finance

submitted by

Daniele Ballinari

from

Sagno (Tessin)

Approved on the application of

Prof. Francesco Audrino, PhD

and

Prof. Zhi Da, PhD

Prof. Dr. Michael Lechner

Dissertation no. 5086

Difo-Druck GmbH, Untersiemaun 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St. Gallen, November 19, 2020

The President:

Prof. Dr. Bernhard Ehrenzeller

“All we have to decide is what to do with the time that is given us.”

G. T. G.

*I thank my supervisor Francesco Audrino, my parents, and my friends
for their constant support.*

Summary

The first paper investigates the predictive power of investors' sentiment and attention for the stock returns' volatility. We introduce a novel and extensive dataset that combines information from social media platforms, news articles, search engine data, and information consumption. Applying a state-of-the-art sentiment classification technique, we construct measures of investors' sentiment and attention for 18 U.S. stocks and the financial market in general. We identify investors' attention, as measured by the number of Google searches on financial keywords (e.g. "financial market" and "stock market"), and the daily volume of company-specific short messages posted on the social media platform StockTwits to be the most relevant variables.

The second paper investigates a potential driver of the predictive power documented in the first paper. We focus on news releases of 360 U.S. companies from the S&P 500 universe and analyze how investors' attention affects the speed at which new information is incorporated in stock prices. Our results show that higher investors' attention around news releases is related to higher contemporaneous volatility. Further, retail investor attention increases the post-announcement volatility, whereas institutional investor attention has a small but negative impact on volatility on days following news releases.

The third paper extends the analysis of the first paper to the multivariate stock return volatility. Building on the theoretical and empirical evidence that links the price comovements with retail investors' behav-

ior, we analyze the predictive power of retail investors' sentiment and attention for the realized correlation matrix of 35 Dow Jones stocks. We propose a new model of realized covariances that allows exogenous predictors to influence the correlation dynamics while ensuring the predicted matrices' positive definiteness. Using this model, we find retail investors' attention to have predictive power for return correlations, especially for longer forecasting horizons and during the COVID-19 pandemic.

The last paper analyzes in more detail the time-series properties of the daily online investor sentiment measures used in the first two papers. We detect structural breaks in the sentiment series for most of the 360 U.S. companies considered in this paper. We illustrate the economic significance of this finding with a return prediction exercise.

Zusammenfassung

Im ersten Aufsatz der vorliegenden Dissertation wird die Vorhersagekraft der Stimmungslage und Aufmerksamkeit von Investoren für die Aktienvolatilität untersucht. Wir stellen einen neuen und umfangreichen Datensatz vor, der Informationen aus Sozialen Medien, Nachrichtentiteln und Suchmaschinen kombiniert. Unter Anwendung einer modernen Textklassifizierungstechnik konstruieren wir die Stimmungs- und Aufmerksamkeitsmasse von Anlegern für 18 U.S.-Aktien und den Finanzmarkt im Allgemeinen. Als die relevantesten Variablen identifizieren wir die Aufmerksamkeit von Investoren, gemessen an der Anzahl von Google-Suchen nach finanziellen Schlüsselwörtern (z.B. “Finanzmarkt” und “Börse”), und das tägliche Volumen von unternehmensspezifischen Kurznachrichten, die auf dem sozialen Netzwerk StockTwits veröffentlicht werden.

Der zweite Aufsatz untersucht einen potentiellen Treiber der im ersten Aufsatz dokumentierten Vorhersagekraft. Wir konzentrieren uns auf die Pressemitteilungen von 360 Unternehmen und analysieren, wie die Aufmerksamkeit der Anleger die Geschwindigkeit beeinflusst, mit der neue Informationen in die Aktienkurse einfließen. Unsere Ergebnisse zeigen, dass eine höhere Aufmerksamkeit von Privatanlegern die Volatilität nach der Bekanntmachung erhöht, während die Aufmerksamkeit institutioneller Anleger einen kleinen, aber negativen Einfluss auf die Volatilität an den Tagen nach der Pressemitteilung hat.

Der dritte Aufsatz erweitert die Analyse des ersten Aufsatzes auf

die multivariate Aktienvolatilität. Aufbauend auf der theoretischen und empirischen Evidenz, die die Bewegung der Preise mit dem Verhalten von Kleinanlegern verknüpft, analysieren wir die Vorhersagekraft der Stimmung und Aufmerksamkeit von Kleinanlegern für die Korrelationsmatrix von 35 Dow-Jones-Aktien. Zu diesem Zweck schlagen wir ein neues Modell der realisierten Kovarianzen vor. Anhand dieses Modells stellen wir fest, dass unter Berücksichtigung der Aufmerksamkeit von Privatanlegern die Vorhersagegenauigkeit von Renditekorrelationen verbessert werden kann. Dies hat sich insbesondere während der COVID-19-Pandemie bewahrheitet.

Im letzten Beitrag werden die Zeitreihen-Eigenschaften der in den beiden ersten Aufsätzen verwendeten Online-Stimmungsmaße genauer analysiert. Wir stellen strukturelle Brüche in den Stimmungsreihen für die meisten der 360 in dieser Arbeit betrachteten US-Unternehmen fest. Die wirtschaftliche Bedeutung dieses Ergebnisses wird in einer Renditeprognose veranschaulicht.

Contents

Summary	vii
Zusammenfassung	ix
1 The impact of sentiment and attention measures on stock market volatility	1
1.1 Introduction	3
1.2 Data	10
1.2.1 Realized volatility	11
1.2.2 Social media, news, information consumption, and search engine data sources	13
1.2.3 Sentiment and attention measures	15
1.2.4 Economic covariates	23
1.3 The impact of investors' sentiment on realized volatility .	29
1.3.1 Joint significance of sentiment and attention variables	32

1.3.2	A regularized HAR model for identifying the pre- dictive power of individual variables	34
1.4	Forecasting volatility with social media, news, information consumption and search engine data	45
1.4.1	Forecasting methodology	45
1.4.2	One-day ahead forecasts	47
1.4.3	Long-horizon realized volatility forecasts	53
1.4.4	A risk management application: value-at-risk fore- casting	58
1.5	Conclusion	62
	References	66
	Appendix	77
1.A	Data	77
1.B	Robustness checks	83
1.C	Supplementary tables	88
1.C.1	In-sample	88
1.C.2	Forecasting	89
1.C.3	Risk management application	92
2	When does attention matter? The effect of investor at- tention on stock market volatility around news releases	95
2.1	Introduction	97
2.2	Related literature	105
2.3	Methodology	110

2.4	Data	114
2.4.1	Attention measures	114
2.4.2	News releases	119
2.4.3	Volatility measure	126
2.5	The relation between investor attention, news, and volatility	129
2.5.1	News releases and earnings announcements	129
2.5.2	Heterogeneities in retail investors' attention	136
2.6	The economic value of accounting for attention around news releases	143
2.6.1	Evaluating the economic value of volatility predictions	145
2.6.2	The value of attention for news-event portfolios . .	149
2.7	Conclusion	152
References		154
Appendix		160
2.A	Tables	160
2.B	Results using Twitter attention	164
2.C	Results with alternative relevance scores	167
2.D	Predictive accuracy of attention-based volatility forecasts	171
2.E	Maximization of expected mean-variance utility with trans- action costs	173
3	Retail investors' trading activity and the predictability of stock return correlations	177
3.1	Introduction	179

3.2	A realized covariance model with exogenous predictors . .	182
3.2.1	Model specification	182
3.2.2	Model estimation	189
3.3	Data and forecasting procedure	192
3.3.1	Data	192
3.3.2	Forecasting procedure and model alternatives . . .	195
3.4	Empirical results	199
3.4.1	Full-sample estimation results	199
3.4.2	Out-of-sample prediction results	204
3.4.3	Prediction results during the COVID-19 pandemic	206
3.5	Conclusion	209
	References	212
	Appendix	217
3.A	Bias correction	217
3.B	Robustness	220
4	Structural breaks in online investor sentiment: A note on the nonstationarity of financial chatter	225
4.1	Introduction	227
4.2	Detecting structural breaks in the dynamics of a time series	229
4.3	Online investor sentiment data	231
4.4	Structural breaks in online investor sentiment	233
4.5	The effect of structural breaks: Predicting excess returns .	238
4.6	Conclusion	242

References	244
Curriculum Vitae	249

Chapter 1

The impact of sentiment and attention measures on stock market volatility

International Journal of Forecasting (2020), 36 (2), 334–357

Co-authors: Francesco Audrino, Fabio Sigris

Abstract

We analyze the impact of sentiment and attention variables on stock market volatility by using a novel and extensive dataset that combines social media, news articles, information consumption, and search engine data. Applying a state-of-the-art sentiment classification technique, we investigate the question of whether sentiment and attention measures contain additional predictive power for realized volatility when controlling for a wide range of economic and financial predictors. Using a penalized regression framework, we identify investors' attention, as measured by the number of Google searches on financial keywords (e.g. "financial market" and "stock market"), and the daily volume of company-specific short messages posted on StockTwits to be the most relevant variables. In addition, our study shows that attention and sentiment variables are able to significantly improve volatility forecasts, although the improvements are of relatively small magnitude from an economic point of view.

Keywords: investor sentiment; investor attention; volatility prediction; realized volatility; high-dimensional regression

1.1 Introduction

According to the efficient market hypothesis (Fama, 1970), the prediction of stock returns is not possible since market prices reflect all available information. On the other hand, since the early 1990s, there has been growing empirical evidence reported by behavioral finance researchers showing that the stock market is driven by investors' psychology; see, e.g., Daniel, Hirshleifer, and Teoh (2002) for a literature review. There are various explanations in behavioral finance for this finding (see e.g. Tseng, 2006) such as, for instance, the misattribution bias, which says that people make risky decisions depending on mood states (Johnson & Tversky, 1983).

In this article, we study the impact of sentiment and attention measures on daily stock market volatility, and to what extent these variables can be used to obtain improved volatility forecasts. We introduce a large and novel dataset of social media, news articles, information consumption, and search engine data. Volatility of stocks and a stock market index is estimated using the concept of realized volatility (Andersen, Bollerslev, Diebold, & Ebens, 2001; Andersen, Bollerslev, Diebold, & Labys, 2003; Barndorff-Nielsen & Shephard, 2002) calculated from high-frequency intra-day data. We extend the heterogeneous autoregressive (HAR) model (Corsi, 2009) by additionally including sentiment, attention, and a comprehensive set of economic variables as predictor variables. To the best of our knowledge, a similar set of economic predictors has only been employed at weekly or monthly frequencies in the context

of volatility modelling (e.g. Christiansen, Schmeling, & Schrimpf, 2012; Mittnik, Robinzonov, & Spindler, 2015; Nonejad, 2017) and has not been used jointly with sentiment and attention variables. We account for the fact that the dimension of the predictor variables is high by using a regularization technique known as the adaptive lasso. This allows us to investigate whether (i) social media, news, information consumption, and search engine data influence future volatility when controlling for other economic variables, (ii) which data source and which type of sentiment or attention measure is most relevant, and (iii) if the variables constructed from our novel dataset can help to improve volatility forecasts.

Our results show that sentiment and attention variables have significant predictive power for future volatility. The variables with the most predictive importance are measures of investors' attention, such as the posting volume on social media platforms or number of queries on search engines. We observe the largest increase in predictive accuracy due to sentiment and attention data for companies with a large market capitalization and/or a low share of institutional investors. Further, our results show that micro-blogging, search engine and news article data have particular predictive power when the volatility is large. We illustrate this in an exemplary manner in Figure 1.1 for one company, the plastic supplier A. Schulman. In the top row, the figure represents the cumulative squared error difference (CSED) for one-day ahead forecasts between the classical HAR model and a HAR model which also includes sentiment and economic data (dashed line), as well as the CSED between a HAR model which includes economic variables and a HAR model extended with both

economic and sentiment predictors (solid line); see Sections 1.2, 1.3, and 1.4 for more details on the data and methodology. The figure shows that the inclusion of sentiment and attention variables results in increased predictive accuracy compared to a HAR model which includes economic variables. Over the first year the two CSED lines are approximately identical since almost no economic and financial predictors are selected in the HAR model including only economic variables for A. Schulman over this time period; see Figure 1.6 for more details. The second row shows the realized volatility and daily log returns of A. Schulman's stock. Selected events during which sentiment variables have particular predictive power are highlighted by grey-shaded areas.

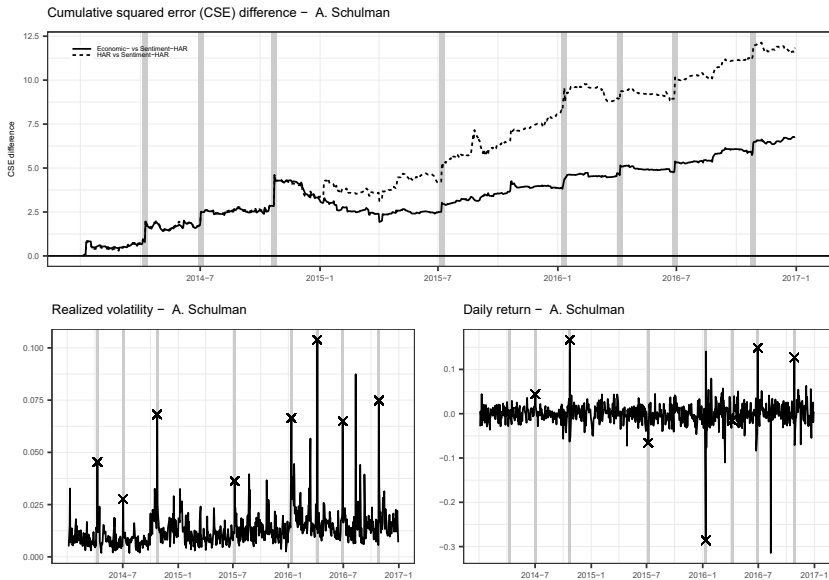


Figure 1.1: Cumulative increase in predictive accuracy of sentiment and attention variables over time for one company (A. Schulman)

NOTE: The figure in the first row represents the cumulative squared error difference between the Economic-HAR and the Sentiment-HAR (solid line) and between the baseline HAR and the Sentiment-HAR models (dashed line); the figures in the second row represent the realized volatility and the daily log returns of A. Schulman's stock. Selected events during which sentiment variables have particular predictive power are highlighted by grey-shaded areas.

As highlighted in Figure 1.1, sentiment and attention variables have particular forecasting power during periods of high volatility.

Several existing studies investigate to what extent sentiment and attention variables obtained from social media and other internet platforms can be used to predict financial market returns. Some studies conclude that public sentiment, obtained from social media such as Twitter or LiveJournal, or internet activity volumes can be used to predict stock market movements (Bollen, Mao, & Zeng, 2011; Da, Engelberg, & Gao, 2011, 2015; Gilbert & Karahalios, 2010; Joseph, Wintoki, & Zhang, 2011; Liew & Budavári, 2016; Nofer & Hinz, 2015; Oliveira, Cortez, & Areal, 2013; Porshnev, Lakshina, & Redkin, 2016; Preis, Moat, & Stanley, 2013; Renault, 2017; Sun, Najand, & Shen, 2016; X. Zhang, Fuehres, & Gloor, 2011). For instance, Da et al. (2011) and Joseph et al. (2011) show that increased investor attention predicts higher prices in the short-run and a price reversal in the long-run. In a more recent study, Sun et al. (2016) analyse the predictability of 30-minutes returns using the Thomson Reuters MarketPsych Index—a sentiment measure based on news wires, internet news sources and social media. Their results show that changes in investors’ sentiment has predictive power for intraday stock returns. In a related study, Renault (2017) shows that the first half-hour change in sentiment from messages posted on StockTwits has predictive power for the last half-hour return. Other studies are more skeptical about the predictive power of social media (Antweiler & Frank, 2004; Oliveira et al., 2013; Tumarkin & Whitelaw, 2001). Analysing messages posted on Yahoo! Finance and Raging Bull, Antweiler and Frank (2004) find that

online messages have no economically significant predictive power for stock returns. Further, Oliveira et al. (2013) find no evidence for return predictability using messages posted on StockTwits. Overall, there is no consensus on whether stock market returns are indeed predictable using sentiment analysis (Schoen et al., 2013).

Similarly, several researchers find that social media sentiment, news sentiment, or search volume can be used to predict stock market volatility (Antweiler & Frank, 2004; Aouadi, Arouri, & Teulon, 2013; Behrendt & Schmidt, 2018; Bordino et al., 2012; Caporin & Poli, 2017; Dimpfl & Jank, 2015; Hamid & Heiden, 2015; Ho, Shi, & Zhang, 2013; Mao, Counts, & Bollen, 2011; Saavedra, Duch, & Uzzi, 2011; See-To & Yang, 2017; Siganos, Vagenas-Nanos, & Verwijmeren, 2017; Y.-H. Wang, Keswani, & Taylor, 2006; J. L. Zhang, Härdle, Chen, & Bommers, 2016). For instance, Antweiler and Frank (2004) find that the sentiment in stock messages posted on Yahoo! Finance and Raging Bull can help to predict stock market volatility. Y.-H. Wang et al. (2006) outline the importance of controlling for other economic and financial variables when analysing the influence of lagged sentiment on realized volatility. Their study shows that sentiment proxies, such as the OEX pull-call trading volume ratio, have only little influence on future volatility after controlling for lagged returns. Mao et al. (2011) is one of the few studies which compare different sources of sentiment data (social media, news and search engine data) for the prediction of return, volume and implied volatility, without controlling for financial covariates, though. They find a statistically significant relation between both stock returns and implied volatility with

past Google search volume and Twitter sentiment. Using a data set of firm-specific and macroeconomic news, Ho et al. (2013) find that news sentiment has a significant impact on intraday volatility of a series of individual US stocks. Similarly, Dimpfl and Jank (2015), Hamid and Heiden (2015) and Aouadi et al. (2013) show that search query volumes can be used to predict stock market volatility. Based on sentiment scores calculated for financial articles obtained from a platform operated by NASDAQ, J. L. Zhang et al. (2016) find that increased sentiment influences volatility as well as volume. Andrei and Hasler (2015) use a theoretical approach to show that increased investor attention leads to higher stock market volatility. See-To and Yang (2017) and Siganos et al. (2017) show that high divergence in investor's sentiment is followed by increases in daily volatility. In a recent study, Caporin and Poli (2017) show that news-related variables can be used to obtain improved volatility forecasts. Using sentiment scores computed from Twitter data, Behrendt and Schmidt (2018) show that company specific sentiment has an influence on intra-day volatility, but the forecast improvements are economically insignificant.

What distinguishes our study from previous research on this subject is that we jointly (i) use a large set of sentiment and attention variables derived from diverse sources with large amounts of data, (ii) control for economic and financial variables, (iii) apply a state-of-the-art sentiment analysis method based on a multi-layer convolutional neural network, and (iv) use a high-dimensional predictive regression model that is appropriate for modeling volatility. To the best of our knowledge, all previous

studies either consider only a limited set of predictor variables, often not controlling for economic variables at all, and/or use only a limited amount of data or data sources for sentiment scores, and/or use a sentiment analysis method that is not state-of-the-art, and/or use overly simple regression models that are not adequate for modelling financial volatility. Moreover, the analyzed time period is often quite short. For instance, the studies by Mao et al. (2011) and Caporin and Poli (2017) do not control for classical economic and financial variables. However, without controlling for other economic variables such as credit spreads or option implied volatility, one cannot shed light on the question of whether social media sentiment, social media volume, search engine volume, or news sentiment provide additional predictive power.

The remainder of this paper is organized as follows. Section 1.2 presents the data and, in particular, the economic and financial variables as well as the construction of the attention and sentiment variables. Section 1.3 introduces the model and the employed variable selection technique. An extensive in-sample analysis uncovers the main economic and sentiment drivers of realized volatility. Finally, in Section 1.4 the predictive power of economic and sentiment data is investigated.

1.2 Data

We consider 18 different US companies: Intel, Microsoft, Verizon, Citigroup, BlackRock, Coca-Cola, Baxter, Exxon Mobile, Gilead Sciences, Hasbro, Nike, Caterpillar, General Electric, Walmart Inc, Ventas, A. Schulman, Parkway and Circor. The companies are all listed either on the New York Stock Exchange (NYSE) or on the Nasdaq Stock Market. The choice of the companies is motivated by the fact that sentiment and attention variables might have different influence on future realized volatility depending on the type of stock under consideration. For this reason, we select companies from different industries (e.g. Intel and Microsoft are tech-companies, Walmart and Coca-Cola are consumer defensive stocks, and Circor and General Electric are industrial companies) and different market capitalizations (e.g. Parkway and A. Schulman have a small capitalization, Microsoft and Walmart Inc are instead large-caps). We consider only 18 stocks due to computational reasons. Both the preprocessing of the data and the application of the high-dimensional regression models, in particular the choice of tuning parameters and the large number of robustness checks done, is computationally demanding. In addition, we also consider the Dow Jones Industrial Average Index. The time period for our analysis ranges from the beginning of 2012 until the end of 2016. Even though some of the raw data considered in this study is available at the intra-day frequency, e.g. social media posts and stock prices, our predictive regression models are based on daily data. More precisely, forecasts are made using information ranging up

to midnight (Eastern Time) of the day where the prediction is made. In other words, our goal is to investigate whether sentiment and attention variables collected during day t contain predictive information for the volatility on the following days, and in particular for day $t + 1$.

1.2.1 Realized volatility

In the early 2000s, the concept of realized volatility was introduced (Andersen et al., 2001, 2003; Barndorff-Nielsen & Shephard, 2002). Instead of using GARCH type models or general stochastic volatility models, the idea is to compute estimates of volatility using high-frequency data. In this way, volatility can be estimated more accurately and then assumed to be observable up to a noise component. We refer to McAleer and Medeiros (2008b) for an overview of realized volatility and a discussion of various extensions that incorporate price jumps and microstructure noise. See also Bollerslev, Patton, and Quaedvlieg (2016) for more information on the fact that finite sample realized volatility estimators are noisy estimates of the true volatility.

For the computation of realized volatilities of the company stocks, we retrieve intraday trade data from the NYSE Trade and Quote (TAQ) database. We clean the high-frequency data by following the procedure suggested by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008). Based on intraday 5 minute returns, we estimate daily realized volatilities using the *median realized volatility* estimator (MedRV) introduced by Andersen, Dobrev, and Schaumburg (2012). We use the MedRV estimator since it has been shown to have good finite-sample robustness to jumps

and “zero” returns. However, our results are robust to the choice of the realized variance estimator. In Table 1.18 in Appendix 1.B, we report the prediction accuracy results using the bipower realized variance estimator and also plain realized variances, and we observe only marginal differences. Further, MedRV estimates for the Dow Jones Industrial Average Index were collected from Oxford-Man Institute of Quantitative Finance (<https://realized.oxford-man.ox.ac.uk/>).

Over the past years, the literature has widely investigated the relation between price movements and information flows, finding empirical evidence that jumps in prices are often observed after (unexpected) firm-specific and macroeconomic news announcements (see among others Evans, 2011; Lahaye, Laurent, & Neely, 2011; Lee & Mykland, 2007). In contrast to these studies, we focus in our analysis on the mid- and long-term impact of investors’ sentiment and attention related to news on the structural component of the quadratic variation of the log-price process, namely the continuous variation. In fact, the above studies analyze the effect of news at the intraday frequency whereas our study analyses the impact of sentiment and attention measures primarily at the daily frequency, but providing also some insights on the durability of the effects up to one month. This motivates the use of a jump-robust realized volatility estimator such as the MedRV estimator.

1.2.2 Social media, news, information consumption, and search engine data sources

To measure sentiment and attention signals related to individual stocks and a stock market index, we use text data from two social media platforms (Twitter and StockTwits), financial news articles obtained through RavenPack News Analytics, and volumes of search engines and information consumption (Google Trends and Wikipedia). In this section, we describe the data sources and filters that we apply for obtaining the raw text and attention data. In the following Section 1.2.3, we then show how we process this data to obtain sentiment and attention variables.

For Twitter, we have a dataset that contains a random historical sample of approximately 1% of all short messages (tweets) as captured by the Internet Archive (<http://archive.org>). In total, this amounts to 6,468,478,453 tweets for the period between 2012 and 2016. Out of these, we search for all English language tweets that contain the company names of the above-mentioned stocks or mention the Dow Jones Index. For the companies, we additionally search for cashtag symbols. For instance, for Microsoft, we search for all tweets that either contain the term “microsoft” or the term “\$msft”. Further, we search for tweets related to the “stock market”, “interest rates”, and “financial markets”. This is done using these terms as keywords. Concerning the keyword “stock market”, we additionally include the terms “stocks” and “wall street” and join the results together under the keyword “stock market”. In total, we find 2,524,369 English language tweets that contain one of the above

keywords. We note that the Internet Archive does not provide data for January 2014 as well as for January and February 2015. Sentiments for this missing data are imputed using neutral sentiments and volumes using averages per stock, the stock market index, and the additional general financial market-related keywords. Since this is a relatively small period, 3 out of 60 months, and the reason why the data is missing is not related to any conditions on financial markets, we believe that the impact on our results is negligible.

StockTwits is a social media platform where users share information about the market and individual stocks in the form of short messages. In contrast to Twitter, StockTwits is exclusively about investing. From StockTwits, we have obtained the full historical data containing all short messages for the period considered in this study. In total, this includes 65,178,316 messages. We apply the same search criteria as for the Twitter data and find a total of 1,862,690 StockTwits tweets.

Data for search volume is obtained from Google Trends, and data for information consumption through Wikipedia is gathered by accessing the public access page view volume in Wikipedia’s open access view statistics (<https://dumps.wikimedia.org/other/pagecounts-raw/>). Similarly as for the social media platforms, we consider the above-mentioned companies plus the four additional keywords “Dow Jones”, “stock market”, “interest rates”, and “financial markets” for retrieving both Google Trends and Wikipedia page views data.

Sentiment and volume of financial news articles is obtained from RavenPack News Analytics, which maintains a database with news article

sentiment scores for a large number of stocks and for 200 economies. The RavenPack News Analytics database covers information from Dow Jones Newswires, regional editions of the Wall Street Journal, Barron's and MarketWatch as well as press releases from leading global media organizations. We collect sentiment scores for 43,274 news articles for the above-mentioned companies. Moreover, we retrieve scores for 188,568 news articles and press releases about the US economy and the US dollar. See Section 1.2.3 for more information on how we process the data obtained from RavenPack.

1.2.3 Sentiment and attention measures

Sentiments of tweets are calculated using a state-of-the-art natural language processing method denoted by *Deep-MLSA* (Deriu et al., 2016, 2017), which can cope with both the short and informal character of social media tweets. Deep-MLSA is a non-linear classifier with three states which correspond to negative, neutral, and positive sentiments. This technique won the message polarity classification subtask of task 4 “Sentiment Analysis in Twitter” 2016 SemEval competition (Nakov, Ritter, Rosenthal, Sebastiani, & Stoyanov, 2016). It can thus be considered as one of the currently best performing sentiment analysis methods for short messages. Robustness results reported in Table 1.18 in Appendix 1.B show that when a lexicon and rule-based approach is adopted, the predictive power of sentiment variables decreases.

In the following, we briefly describe the Deep-MLSA technique. See Goodfellow, Bengio, Courville, and Bengio (2016) for a comprehensive

introduction to neural networks and Deriu et al. (2016) and Deriu et al. (2017) for more details on the Deep-MLSA technique. The Deep-MLSA approach consists of a multi-layer convolutional neural network. In the first layer, every word, which is represented as a sparse vector in a high-dimensional space consisting of all words, is mapped to a low-dimensional space of dimension d . Such a mapping is called an embedding. Intuitively, an embedding captures the semantics of words in such a way that words with similar meanings are mapped to close locations. The words of every tweet are then concatenated together in a sentence matrix¹ $X \in \mathbb{R}^{d \times n}$. The next layers then consists of two linear convolutional layers each followed by a non-linear max-pooling layer. Roughly speaking, the idea is that a convolutional layer allows for detecting word patterns, which go beyond simple single words or word pairs, that are related to the sentiment of a text. Since the exact location in the sentence of such a word combination is not important for the sentiment, one takes the maximum over sequences of word combinations in the sentence. In more detail, a convolutional layer applies a total of m different filters to the sentence matrix X . If $X[i : i + h]$ denotes the concatenation of word vectors x_i to x_{i+h} , then a feature c_i is generated for a given filter matrix $F \in \mathbb{R}^{d \times h}$ by $c_i = \sum_{k,j} (X[i : i + h])_{k,j} F_{k,j}$. The concatenation of all vectors in a sentence produces a feature vector $c \in \mathbb{R}^{n-h+1}$. The vectors c of all m filters are then concatenated into a matrix $C \in \mathbb{R}^{m \times n-h+1}$. Next, max pooling is applied to the output of each filter by taking the maximum over a fixed set of non-overlapping intervals. The resulting

¹Zero padding is used to create sentence matrices of equal lengths.

pooled feature map matrix has the form $C_{pool} \in \mathbb{R}^{m \times (n-h+1)/s}$, where s is the length of each interval. Subsequently follows a fully connected layer and the final layer consists of an exponential softmax layer which results in a value in $\{negative, neutral, positive\}$. Intuitively, these last two layers allow for finding combinations of word patterns in order to predict the sentiment. Figure 1.2 illustrates this convolutional neural network. We use a pre-trained model obtained from the authors (Deriu et al., 2017).

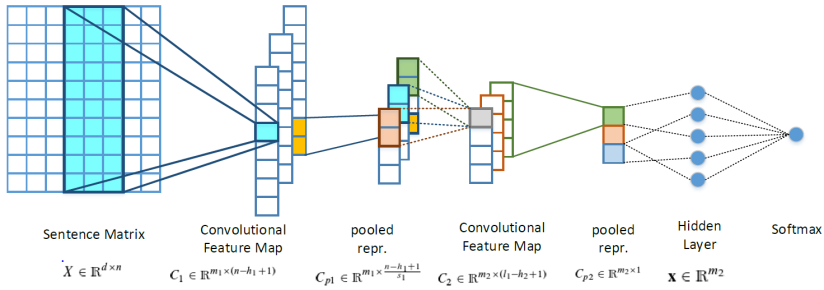


Figure 1.2: Illustration of the convolutional neural network used by the Deep-MLSA technique (figure source: Deriu et al. (2016))

The Deep-MLSA technique classifies each tweet as having either a negative, neutral, or positive sentiment. We code the sentiments as 0.5 (negative), 1 (neutral), and 1.5 (positive) and aggregate sentiments of individual tweets by day based on Eastern Time (ET). Aggregation of individual tweets is done by calculating daily empirical means and standard deviations for each company, the stock market index, and the additional keywords presented above. The daily means capture the average daily sentiment among investors (StockTwits data) and the general population (Twitter data). The standard deviation is used as

a measure of disagreement in sentiments. In Figure 1.3, we illustrate daily average sentiments as well as number of tweets on both Twitter and StockTwits for one of the companies (Verizon).

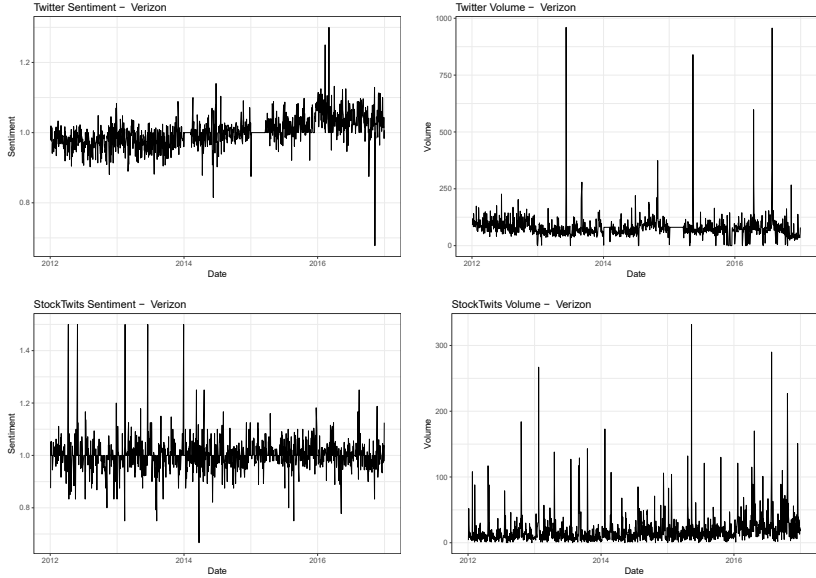


Figure 1.3: Sentiment and volume of micro-blogs posted on Twitter and StockTwits for Verizon

NOTE: The figure represents on the left-hand side the daily average sentiment, and on the right-hand side the volume of messages posted on Twitter (first row) and StockTwits (second row).

RavenPack News Analytics provides different types of sentiment and relevance scores.² For the purpose of our research, we consider the *Event Sentiment Score*, *Composite Sentiment Score*, and *News Impact Score*. The *Event Sentiment Score* is determined by using a supervised Bayes Classifier trained with news articles categorized by financial experts as having a short-term positive or negative effect on a company or

²Relevance scores measure how important a news article or press release is for a specific company.

the economy. RavenPack News Analytics computes this score only for news articles or press releases which are highly relevant for a company. The *Composite Sentiment Score* combines various sentiment analysis techniques, trained to detect the influence of financial news articles on stock prices.³ The *News Impact Score* is a sentiment measure trained to predict the influence of the wording in news articles or press releases on the short-term volatility. The latter two scores are provided only for traded companies. RavenPack News Analytics reports sentiment scores on a scale from 0 (negative) to 100 (positive). For consistency with the social media sentiment scores, we recode the scores to range from 0.5 (negative) to 1.5 (positive). Aggregation is again done by computing daily empirical means and standard deviations of the sentiment scores. Since RavenPack News Analytics reports the *News Impact Scores* and *Composite Sentiment Scores* independently from the news article's relevance for a company, we weight their scores according to the respective relevance scores. For illustration, we show in Figure 1.4 daily average *Event Sentiment Scores* as well as numbers of news articles for Verizon.

The daily sentiment measures constructed from Twitter, StockTwits, and news article data are likely affected by noise. The reason for this is twofold. First, when selecting, for instance, Twitter messages associated with a certain company, we probably include irrelevant tweets. Secondly, the existing methodologies for extracting the sentiment in a text message are only proxies for the true sentiment. Because of these issues, we are not able to directly observe the variables of interest, i.e. the daily investor

³More precisely, the *Composite Sentiment Score* combines classifiers trained on expert's opinions as well as on the market impact of the news article's wording.

sentiment and disagreement. Similarly as in Borovkova and Mahakena (2015) and Borovkova, Garmaev, Lammers, and Rustige (2017), we therefore filter out noise by applying the Kalman filter to the following random walk plus noise state-space model:

$$\begin{aligned} S_t &= \mu_t + v_t, v_t \sim N(0, \sigma_v^2), \\ \mu_t &= \mu_{t-1} + w_t, w_t \sim N(0, \sigma_w^2), \end{aligned}$$

where S_t denotes a noisy sentiment measure at time t . The smoothed sentiment, or the trend of the sentiment, is then obtained by calculating the conditional mean $E(\mu_t | S_1, \dots, S_t)$ using the Kalman filter. The parameters of the above state-space model are estimated using maximum likelihood estimation. Note that we are not assuming that sentiments follow a random walk. The random walk is a prior model for a latent variable, and by conditioning on the data one obtains a distribution $\mu_t | S_1, \dots, S_t$ that can be very different from a random walk. We observe that estimated signal-to-noise ratios (i.e. $\hat{\sigma}_w^2 / \hat{\sigma}_v^2$) are generally small (results not tabulated) and the degree of smoothing of the Kalman filter is high. In order to analyse if the random walk plus noise state-space assumptions might represent a potential issue for our analysis, we conducted a series of robustness checks. More precisely, we repeated our analysis using raw daily sentiment measures as well as sentiment measures filtered with an AR(1) state-space model, a 5-day moving average, and a 5-day exponential moving average filter. A summary of the out-of-sample results for these robustness checks can be found in Table 1.18 in Appendix 1.B. These robustness checks show that the predictive power

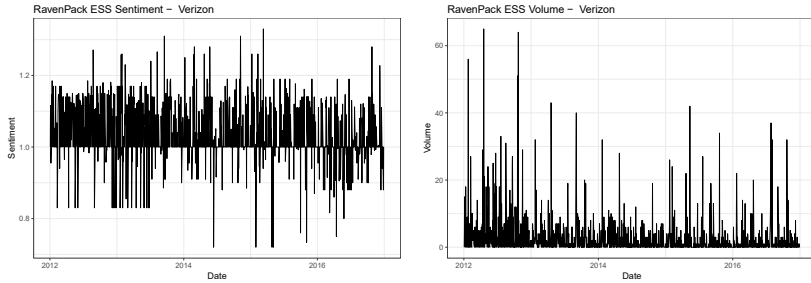


Figure 1.4: Sentiment and volume of news articles about Verizon as reported on the RavenPack database

NOTE: The figure represents on the left-hand side the daily average sentiment, and on the right-hand side the volume of news articles about Verizon, as reported on the RavenPack database. The daily sentiment is an average of news sentiments weighted by the relevance of the article for Verizon.

of sentiment and attention variables is qualitatively the same when a different specification for the latent sentiment in the state-space model is used (autoregressive process) or also when a simpler filter is used (moving average or exponential moving average filter). However, when raw sentiment measures are included, the predictive power of our model is smaller compared to all other filtering specifications. Furthermore, in-sample results (not tabulated) show that when no smoothing is applied, sentiment variables are almost never selected in a penalized regression framework.

In addition to the daily sentiments, we also include the squares of de-noised sentiments as covariates in the regression models in order to account for the fact that volatility might potentially react both symmetrically or asymmetrically to negative and positive sentiments. Further, we also count the number of tweets per day and use this as a measure of attention on the two social media platforms. In general, we observe that there are differences in the amount of data from the two platforms for the

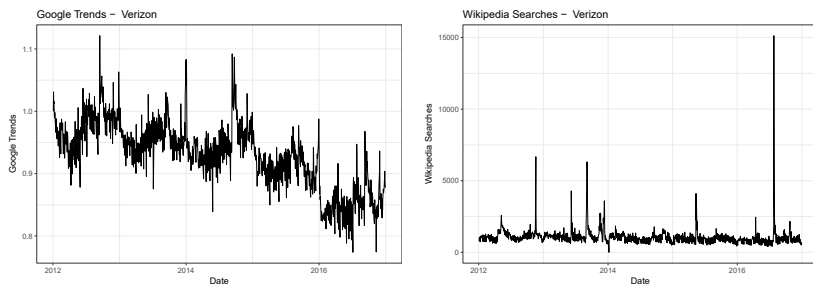


Figure 1.5: Google Trends and number of Wikipedia searches for Verizon

NOTE: The figure represents on the left-hand side Google Trends standardized to 1 at the beginning of 2012, and on the right-hand side the number of searches on the online encyclopedia Wikipedia for Verizon.

different stocks. Some stocks are more frequently discussed on Twitter whereas others are more popular on StockTwits. For instance, Nike has a relatively large number of tweets on Twitter but few on StockTwits. On the other hand, Intel has little Twitter data but more StockTwits data. The different data sources can thus potentially complement one another.

The number of search queries on Google and the number of Wikipedia page views are used as additional measures of attention. In Figure 1.5, we illustrate Google Trends data and Wikipedia page views for Verizon. Since Google does not provide the absolute number of search queries but only relative ones, we normalize the number of Google searches to one at the beginning of the study period by dividing each time series with the value at the first day. For the volume measures, we include in the predictive regression models both log volumes as well as log returns of volume.⁴

In total, we include 91 sentiment and attention variables per company.

⁴More precisely, we consider the logarithm of the raw volume plus 1. By doing so, we avoid taking the logarithm of zero.

A summary of all sentiment and attention variables included in the predictive regression can be found in Table 1.16 in Appendix 1.A. See also Table 1.1 in Section 1.2.4 where we report descriptive statistics as well as correlations with economic variables for a selection of important sentiment and attention variables. Further, in Table 1.17 in Appendix 1.A, we report average sentiments and volume numbers as descriptive statistics for the different data sources and the different companies.

1.2.4 Economic covariates

Besides the sentiment and attention variables, we include in our analysis a set of economic and financial covariates. Controlling for other economic and financial variables ensures that the estimated impact of sentiment variables is not biased by the exclusion of other relevant variables. For instance, the information enclosed in a news article about an earnings increase might already be captured by the company's earnings-price ratio.

The recent literature has shown that macroeconomic and financial data can improve standard volatility forecasting models; see e.g. Paye (2012), Mittnik, Robinsonov, and Spindler (2015), Nonejad (2017) and Christiansen et al. (2012). Christiansen et al. (2012) found that proxies for credit risk and funding liquidity consistently improve monthly volatility forecasts. Mittnik et al. (2015) identify the *CBOE Market Volatility Index* (VIX) as being one of the main risk drivers of volatility. Nonejad (2017) shows that the most important predictors of monthly realized volatility of the S&P 500 are past volatilities, proxies for risk premia, short-term interest rates and the default spread on corporate bonds.

Following Christiansen et al. (2012), Mittnik et al. (2015) and Nonejad (2017), we include financial and macroeconomic variables that can be grouped in five different categories:

- *Equity market variables*: this set includes well known equity market valuation ratios (e.g earnings-price ratio, dividend-price ratio), the Fama-French risk factors as well as implied volatility as measured by the CBOE Market Volatility Index. Moreover, we included past returns of the stock itself and of two major stock indices (MSCI and Dow Jones Industrial Average).
- *Bond market variables*: this set consists of interest rates, term spreads and bond risk premia. More precisely, we include the variables employed by Welch and Goyal (2008): T-Bill and relative T-Bill rate, Long-Term Bond return, relative Long-Term yield and a term spread. Moreover, as a measure for Bond Risk premia, we included the yield spread between Baa-rated corporate and long-term government bonds.
- *Exchange rate variables*: this category contains a set of variables which have been shown to capture risk premia and return variation in foreign exchange markets (see Lustig, Roussanov, & Verdelhan, 2011, 2014). In particular, we include returns on the exchange rates between the US dollar and four major currencies (Euro, Swiss franc, British pound and the Japanese yen), a carry trade factor and the average forward discount which measures the interest rate differential between the US dollar and a set of foreign currencies.

- *Liquidity variables*: this set includes variables that capture the liquidity in financial markets. In particular, we consider as proxies for liquidity the turnover-ratio⁵ (of the stock itself and of two major stock indices: MSCI World and Dow Jones Industrial Average), the default spread measured as the difference between the yield of Baa- and Aaa-rated corporate bonds, the average bid-ask spread of five major currencies and the TED spread (difference between the three-month LIBOR rate and the T-Bill rate).
- *Macroeconomic variables*: this category includes a broad range of macroeconomic time series. We consider inflation, measures of production, job market variables, commodity prices, money supply, and consumer and producer sentiment, as measured by surveys.

Overall, we consider a set of 56 economic covariates. A summary and detailed definition of all financial and macroeconomic variables considered in our analysis can be found in Table 1.15 in Appendix 1.A. Note that 17 of the macroeconomic time series are only available at a monthly frequency. We decided to include these variables for the in-sample analysis presented in Section 1.3 by interpolating daily values. More precisely, daily observations are imputed by a linear interpolation of the logarithmic monthly observations.⁶ Our results (not tabulated) are very similar when excluding these monthly variables. Further, when analyzing the predictive power of our models in an out-of-sample forecasting setting

⁵We define the turnover-ratio as the daily traded volume divided by the market capitalization.

⁶Note that the linear interpolation is done on the logarithm of level variables. Month-over-month or year-over-year changes are computed on the imputed values.

in Section 1.4, we exclude variables which are not available at a daily frequency to ensure true ex-ante forecasts.

Table 1.1 reports summary statistics for a selection of economic, sentiment and attention variables included in our predictive regressions.⁷ The table reports the average value, the empirical standard deviation and the minimum and maximum values. Moreover, we also report average correlation coefficients among these variables. Panel A reports the summary statistics for the volume of messages posted on StockTwits about a specific company or the Dow Jones Index and the turnover-ratio of the analysed companies and the stock market index. Panel B shows the summary statistics of five general predictors included in our models: the CBOE Volatility Index, the (relative) volume of Google searches about “financial market”, the squared value of the smoothed sentiment of StockTwits messages containing the word “stock market”, and the squared value of the smoothed sentiment (Event Sentiment Score) computed by RavenPack for news articles about the US economy. Panel C reports the average (across the 18 companies and the Dow Jones Index) Pearson correlation coefficients for the aforementioned six predictors. Moreover, to illustrate the difference between the two social media platforms considered in our study, we report average correlation coefficients for the company specific volume of messages posted on Twitter and the squared smoothed sentiment score of Twitter messages containing the word “stock market”. From Panel C we observe that the company specific attention measures, i.e. the volume of messages posted on StockTwits, is positively correlated

⁷The covariates for which summary statistics are reported, are a selection of the most important variables (see Section 1.3).

with a general measure of attention, i.e. the relative volume of Google searches for “financial market”. Moreover, the CBOE Volatility Index is positively correlated with the squared news article sentiment score. Interestingly, the squared sentiment score for the “stock market” obtained from StockTwits is negatively correlated with the squared sentiment score for the “US economy” obtained from RavenPack News Analytics.

Table 1.1: Summary statistics for a selection of economic, sentiment and attention variables

Panel A: Subset of company specific variables								
	StockTwits Volume				Turnover-ratio (in %)			
	Avg.	Std.	Min.	Max.	Avg.	Std.	Min.	Max.
Intel	31.59	50.96	0	699	0.68	0.33	0.13	2.89
Microsoft	70.91	78.44	3	896	0.48	0.25	0.11	3.01
Verizon	21.18	25.5	0	332	0.45	0.49	0.1	14.99
Citigroup	36.8	36.54	0	756	0.87	0.46	0.16	4.3
BlackRock	9.92	17.68	0	271	0.42	0.39	0.09	12.45
Coca-Cola	14.99	19.29	0	258	0.33	0.13	0.07	2.21
Baxter	3.44	6.02	0	65	0.76	0.75	0.16	14.05
Exxon Mobil	21.53	20.76	0	218	0.3	0.12	0.1	1.87
Gilead Sciences	102.21	180.13	0	2287	0.83	0.49	0.18	8.22
Hasbro	4.15	13.1	0	207	1.08	0.7	0.26	7.41
Nike	32.97	90.09	0	1617	0.49	0.28	0.14	4.71
Caterpillar	21.85	32.91	0	322	0.95	0.43	0.24	4.53
General Electric	19.78	18.4	0	163	0.41	0.27	0.13	4.58
Walmart Inc	30.65	50.18	0	1142	0.25	0.14	0.06	2.59
Ventas	1.61	2.51	0	24	0.67	0.31	0.14	3.46
A. Schulman	0.85	2.71	0	31	0.59	0.44	0.11	6.5
Parkway	0.6	1.26	0	11	0.67	0.95	0.03	15.86
Circor	0.55	1.36	0	16	0.59	0.44	0.08	5.98
Dow Jones	5.72	5.5	0	44	0.12	0.05	0.04	0.45

NOTE: The table reports summary statistics (average, empirical standard deviation, minimum and maximum values) and Pearson correlation coefficients for a subset of predictor variables. Panel A reports the summary statistics for the volume of messages posted on StockTwits about a specific company or the Dow Jones Index and the turnover-ratio of the analyzed companies and the stock market index. Panel B shows the summary statistics of five general predictors included in our models: the CBOE Volatility Index, the (relative) volume of Google searches about “financial market”, the squared value of the smoothed sentiment of StockTwits messages containing the word “stock market” (multiplied by 10,000), and the squared value of the smoothed sentiment (Event Sentiment Score) computed by RavenPack for news articles about the US economy (multiplied by 10,000). Panel C reports the average (across the 18 stocks and the Dow Jones Index) Pearson correlation coefficients of the subset of six variables and the company specific volume of messages posted on Twitter and the squared smoothed sentiment score (SS) of Twitter messages containing the word “stock market”. Note that, in order to better visualize the magnitude of posting volumes, we report summary statistics for the raw volume of messages posted on StockTwits. The average correlation coefficients reported in Panel C, however, are computed using log volumes.

Table 1.1 continued: Summary statistics for a selection of economic, sentiment and attention variables

Panel B: Subset of general variables				
	Avg.	Std.	Min.	Max.
CBOE Volatility Index	15.75	3.49	10.32	40.74
Google Volume - "financial markets"	1.04	0.08	0.59	1.37
Sq. smooth. StockTwits sent. - "stock market"	7.58	6.65	0	34.3
Sq. smooth. RavenPack ESS sent. - "US economy"	8.17	6.05	0	45.35

Panel C: Average correlation between a subset of variables								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
StockTwits vol. (1)	1.00	0.25	0.02	0.32	0.18	-0.17	0.24	0.09
Turnover-ratio (2)		1.00	0.22	0.04	-0.01	0.07	0.06	0.09
VIX (3)			1.00	0.12	0.02	0.26	-0.02	0.29
Google vol.-"fin. market" (4)				1.00	0.15	-0.03	0.12	0.1
SS SockTwits-"st. mar." (5)					1.00	-0.47	0.11	0.02
SS RavenPack-"US econ." (6)						1.00	-0.03	-0.11
Twitter vol. (7)							1.00	-0.07
SS Twitter-"stock market" (8)								1.00

NOTE: The table reports summary statistics (average, empirical standard deviation, minimum and maximum values) and Pearson correlation coefficients for a subset of predictor variables. Panel A reports the summary statistics for the volume of messages posted on StockTwits about a specific company or the Dow Jones Index and the turnover-ratio of the analyzed companies and the stock market index. Panel B shows the summary statistics of five general predictors included in our models: the CBOE Volatility Index, the (relative) volume of Google searches about "financial market", the squared value of the smoothed sentiment of StockTwits messages containing the word "stock market" (multiplied by 10,000), and the squared value of the smoothed sentiment (Event Sentiment Score) computed by RavenPack for news articles about the US economy (multiplied by 10,000). Panel C reports the average (across the 18 stocks and the Dow Jones Index) Pearson correlation coefficients of the subset of six variables and the company specific volume of messages posted on Twitter and the squared smoothed sentiment score (SS) of Twitter messages containing the word "stock market". Note that, in order to better visualize the magnitude of posting volumes, we report summary statistics for the raw volume of messages posted on StockTwits. The average correlation coefficients reported in Panel C, however, are computed using log volumes.

This suggests that the information in social media and news article data is of a different nature. The company specific volume obtained from the two social media platforms as well as the squared sentiment scores obtained from StockTwits and Twitter are both positively correlated, but the magnitude of the Pearson correlation coefficients is small. In general, the correlation coefficients are relatively small (all below 0.5 in absolute terms). When computing the correlation coefficients between all covariates considered in our dataset, we also observe higher correlations

among some predictors. For instance, the volume of Google searches for “stock market” and “Dow Jones” have a correlation of almost 88%.

1.3 The impact of investors’ sentiment on realized volatility

In this section, we investigate the impact of sentiment and attention variables on volatility using a predictive regression model while controlling for past volatilities as well as economic and financial variables.

We use the Heterogeneous Autoregressive model of realized volatility (HAR) introduced by Corsi (2009) as a baseline model. The HAR model is one of the most popular models for realized volatility due to its simplicity and good predictive performance. By superimposing different short memory frequencies, stylized facts such as long-range dependence observed in empirical volatility series can be accounted for. Various extensions of the basic HAR model have been proposed, e.g., including jumps (see, among others, Andersen, Bollerslev, & Diebold, 2007; Corsi, Pirino, & Renò, 2010), leverage effects (see, for instance, Bollerslev, Litvinova, & Tauchen, 2006; Corsi & Renò, 2012; Patton & Sheppard, 2015), or allowing for non-linearities in the volatility process (e.g. Fengler, Mammen, & Vogt, 2015; McAleer & Medeiros, 2008a). We refer to Corsi, Audrino, and Renò (2012) for a more extensive overview of different extensions of the classical HAR model.

The baseline HAR model we use in our analysis is given by

$$\begin{aligned} \log RV_{t+1}^{(d)} = \\ c + \beta^{(d)} \log RV_t^{(d)} + \beta^{(w)} \log RV_t^{(w)} + \beta^{(m)} \log RV_t^{(m)} + \epsilon_{t+1}, \end{aligned} \quad (1.1)$$

where $\log RV_t^{(w)} = \frac{1}{5} \sum_{i=1}^5 \log RV_{t-i+1}^{(d)}$ is the weekly and $\log RV_t^{(m)} = \frac{1}{22} \sum_{i=1}^{22} \log RV_{t-i+1}^{(d)}$ is the monthly average of daily log realized volatilities, respectively, and $\{\epsilon_t\}$ is a zero-mean innovation process.⁸

In the following, we extend this baseline model in two ways. First, we include lagged macroeconomic and financial data:

$$\log RV_{t+1}^{(d)} = c + (\log RV_t)' \beta_{(RV)} + M_t' \gamma_{(eco)} + \epsilon_{t+1}, \quad (1.2)$$

where $\log RV_t$ is a 3-dimensional column-vector containing daily log realized volatilities, weekly and monthly average daily log realized volatilities and M_t is a q -dimensional column-vector of economic and financial variables defined in Section 1.2.4. In the following, we refer to this model as the Economic-HAR model. Second, we extend the Economic-HAR

⁸The reason why we model logarithmic and not plain realized volatility is that plain realized volatility is skewed; see Gonçalves and Meddahi (2011) for a discussion on this topic. In a regression context, this implies that regression coefficient estimates have large variance and a few large volatilities can have an overly large influence on the estimates. Using logarithmic realized volatility thus results in more efficient coefficient estimates with lower variance. This is particularly important since our models include high-dimensional covariates. Further, the use of logarithmic realized volatility avoids the enforcement of a positivity constraint.

model by including lagged sentiment and attention variables:

$$\log RV_{t+1}^{(d)} = c + (\log RV_t)' \beta_{(RV)} + M_t' \gamma_{(eco)} + Z_t' \theta_{(sent)} + \epsilon_{t+1}, \quad (1.3)$$

where Z_t is a p -dimensional column-vector containing the sentiment and attention variables presented in Section 1.2.3. We refer to this model as the Sentiment-HAR model. Table 1.2 summarizes the number of predictor variables for the different models. The stock market index, i.e. the Dow Jones, has fewer economic and sentiment predictors (values are reported in parentheses). This is due to the fact that for each company we include company-specific variables as well as general variables on, e.g., the stock market, whereas for the index we do not include any company-specific variables.

In order to investigate whether sentiment and attention variables contain additional information that influences future realized volatility, we

Table 1.2: Summary of the model and data dimensions

Model	Lagged RV (s)	Economic var. (q)	Sentiment var. (p)	Num. predictors	Sample size
HAR	3 (3)	0 (0)	0 (0)	3 (3)	1199
Economic-HAR	3 (3)	56 (51)	0 (0)	59 (54)	1199
Sentiment-HAR	3 (3)	56 (51)	91 (77)	150 (131)	1199

NOTE: The table summarizes the dimensionality of the three considered models. We report the number of predictors and the available data points in our sample. The number of predictors for the Dow Jones Industrial Average Index are reported in parentheses. Note that the Dow Jones Index has fewer economic and sentiment predictors. This is due to the fact that, for each company we include company-specific variables as well as general variables on e.g. the stock market, whereas for the index, we do not include any company-specific variables.

proceed in two steps. First, we analyze whether the addition of sentiment variables improves significantly the fit of our model and whether a variable selection procedure selects predictors from the set of micro-blogging, news, information consumption, and search engine data. In a second step, we compare the out-of-sample forecasting accuracy of the Sentiment-HAR model with the Economic-HAR and the baseline HAR models.

1.3.1 Joint significance of sentiment and attention variables

Due to the high dimensionality and the form of the predictor variables, it is not feasible to use standard t-tests to test separately for every variable the null-hypothesis that the variable has a zero coefficient for the following two reasons. First, there is multicollinearity.⁹ Second, we have a multiple testing problem due to the potentially large number of hypotheses.

In the following, we first investigate whether all sentiment and attention variables jointly provide predictive power in addition to the standard HAR model and the economic variables. Using the model specified in Equation (1.3), we consider the hypotheses

$$\begin{aligned} H_0 : \theta_{(sent),i} &= 0, \quad \forall i = 1, \dots, p, \\ H_a : \theta_{(sent),i} &\neq 0, \quad \text{for at least one } i, \end{aligned} \tag{1.4}$$

where p is the number of sentiment variables considered in our analysis.

⁹Variance inflation factors (not tabulated) are very large and well above 10, which suggests high multicollinearity in our predictor variables.

For each stock and the Dow Jones Index, we test the null-hypothesis given in Equation (1.4) using an F-test. To overcome the multiple testing problem, we apply a Bonferroni-Holm correction (Holm, 1979) to the p-values for determining significance. The test statistics and p-values (in %) are reported in Table 1.3 with stars indicating significance (* indicates significance at the 10%, ** at the 5% and *** at the 1% level). The results show that, even when conservatively correcting p-values, sentiment and attention variables are jointly significant at the 5%-significance level for all companies. This result clearly shows that the addition of sentiment and attention variables is able to significantly improve the fit of the HAR model also when controlling for macroeconomic and financial data. What's more, it is well known that the power of

Table 1.3: Joint significance of the sentiment variables

	F-Statistic	p-value (in %)
Intel	2.34	0***
Microsoft	2.88	0***
Verizon	2.42	0***
Citigroup	1.86	0***
BlackRock	1.79	0***
Coca-Cola	1.86	0***
Baxter	2.1	0***
Exxon Mobil	2.53	0***
Gilead Sciences	2.36	0***
Hasbro	1.32	2.71**
Nike	2.74	0***
Caterpillar	2.65	0***
General Electric	1.82	0***
Walmart Inc	1.72	0.01***
Ventas	1.73	0***
A. Schulman	2.07	0***
Parkway	1.89	0***
Circor	1.9	0***
Dow Jones	2.61	0***

NOTE: The table reports the F-statistic for joint significance of the sentiment and attention variables and corresponding p-values for 18 stocks and the Dow Jones Index when the Sentiment-HAR model is estimated by ordinary least squares. The stars indicate if the null-hypothesis is rejected in favor of the sentiment and attention variables when applying the Bonferroni-Holm correction. * indicates that the result is still significant at the 10%, ** at 5% and *** at 1% significance level.

F-tests for high dimensional linear regressions is low (see, among others, Steinberger, 2016; S. Wang & Cui, 2013). Being able to reject the joint hypothesis defined in Equation (1.4) even if the F-test is likely to have low power underlines the strength of the evidence that sentiment and attention measures contain significant additional predictive information.

1.3.2 A regularized HAR model for identifying the predictive power of individual variables

In this section, we investigate which sentiment and attention variables contain the most predictive information for the realized volatility. To overcome the problems related to the high dimensionality of our data, we adopt a high-dimensional methodology known as the Adaptive Least Absolute Shrinkage and Selection Operator (adaptive lasso) introduced by Zou (2006), which is an extension of the original lasso by Tibshirani (1994). In the previous section, we have found that jointly sentiment and attention variables clearly have predictive power for realized volatility. In general, there are two possible scenarios for the significance of the F-test. First, it is possible that some of the sentiment and attention variables have relatively large non-zero coefficients while others have zero coefficients. On the other hand, it is also possible that none of the coefficients is large but several coefficients of the sentiment and attention variables are of small but non-zero magnitude, and jointly they have a significant effect. In the latter case, it can be very difficult or impossible to consistently identify the non-zero coefficients with most high-dimensional methods since the assumption that the coefficients are sparse and/or satisfy the

so-called “beta-min” condition might not hold true; see e.g. Bühlmann and Van De Geer (2011).

The adaptive lasso estimator for the Sentiment-HAR model is given by:

$$\begin{aligned}
 (\hat{c}_n^{AL}, \hat{\beta}_{(RV)}^{AL}, \hat{\phi}_n^{AL}) = \\
 \arg \min_{c, \beta, \phi} \frac{1}{n} \sum_{t=1}^n \left(\log RV_{t+1}^{(d)} - c - \beta'_{(RV)} \log RV_t - \phi' X_t \right)^2 \\
 + \frac{\lambda_n}{n} \sum_{i=1}^{q+p} \lambda_{n,i} |\phi_i|,
 \end{aligned} \tag{1.5}$$

where $X_t = (Z_t, M_t)$ and $\phi' = (\theta'_{(sent)}, \gamma'_{(eco)})$. The penalty weights are given by $\lambda_{n,i} = 1/|\hat{\phi}_{n,i}^{Ridge}|$ where $\hat{\phi}_n^{Ridge}$ are the Ridge regression estimators of our model. The choice of using a Ridge over the least square estimators is motivated by the presence of multicollinearity in our data (Zou, 2006). Due to the form of the penalty term, the adaptive lasso automatically does variable selection. Note that the HAR-components (log realized volatilities) are not penalized and are thus always included in our models.

The tuning parameter λ_n is determined by means of cross-validation. In the context of financial data, temporal dependence represents a potential issue for classical cross-validation procedures. Racine (2000) suggested a block cross-validation, where blocks of v observations are used as a test set and the remaining observations as training sets. Following Bergmeir and Benítez (2012), we employ a similar but computationally less intensive strategy. Essentially, we divide our sample into five non-overlapping

and connected intervals, choose one block as test set and estimate the model using the remaining data. To ensure independence, we remove h observations at the boundaries of the test set. We repeat this procedure for all five blocks and choose λ_n to minimize the out-of-sample mean squared error. In summary, the regularized HAR model is estimated as follows:

1. Divide the dataset into five non-overlapping and connected intervals.
2. Determine the optimal tuning parameter of the first stage of the adaptive lasso when estimating the model with an L^2 -penalty (Ridge regression): for a range of potential tuning parameters, we choose a test set consisting of one block and fit the model over the remaining four blocks (removing h data points between the two sets¹⁰). We repeat this procedure for all five blocks and measure the mean squared test error. The tuning parameter is then choose to minimize the average mean squared test error.
3. Fit the L^2 -penalised model using the above determined tuning parameter. We define the penalty weights $\lambda_{n,i} = 1/|\hat{\phi}_{n,i}^{Ridge}|$ where $\hat{\phi}_n^{Ridge}$ are the Ridge regression estimators of our model. Note that the coefficients of the HAR components have a zero penalty weight.
4. Determine the optimal tuning parameter of the second stage of the adaptive lasso defined in Equation (1.5) analogously as described in point 2. above.

¹⁰The number of observations which are removed between the test and training set are determined by finding the largest lag at which the autocorrelation in the HAR's residuals is still significant at the 5% level (Bonferroni-Holm corrected).

5. Fit the model defined in Equation (1.5) using the above determined tuning parameter.

A summary of goodness-of-fit measures of the three analysed models is reported in Table 1.4. The first row shows the average increase in R^2 between the models. We observe that the Sentiment-HAR increases on average the R^2 of the baseline HAR model by 19.09%; compared to the Economic-HAR model the increases amounts on average to 6.11%. The second and third rows report the average reduction of the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC). We observe that the largest reduction in both information criteria can be attributed to the economic predictors. Nevertheless, the addition of sentiment and attention variables on top of economic predictors can further reduce both information criteria. More detailed results about the in-sample fit of the three models are reported in Table 1.19 in Appendix 1.C.

In order to obtain p-values and to be able to test individual significance of coefficients estimated by the adaptive lasso in a time series context,

Table 1.4: Average improvement of different goodness-of-fit measures

	Sent./Baseline HAR	Eco./Baseline HAR	Sent./Eco. HAR
Avg. increase of R^2 (in %)	19.09	12.21	6.11
Avg. decrease of AIC (in %)	8.96	6.2	2.93
Avg. decrease of BIC (in %)	4.68	3.79	0.92

NOTE: The table reports the average improvements in three goodness-of-fit measures when comparing the three models of interest, i.e. Sentiment-HAR, Economic-HAR and the classical HAR model. The goodness-of-fit measures considered for the in-sample analysis are the regression R^2 , the Akaike Information Criterion (AIC) and the Bayes Information Criterion (BIC). Note that the degrees of freedom for Lasso correspond to the expected number of non-zero estimates (Zou, Hastie, & Tibshirani, 2007); a simple unbiased estimator of the degrees of freedom is the number of estimated non-zero coefficients (Bühlmann & Van De Geer, 2011). We follow the definition of Zou et al. (2007) for the information criteria of lasso.

Table 1.5: Number of selected economic and sentiment variables and number of significant sentiment variables

	Num. economic variables	Num. sentiment variables	Num. significant sentiment var.		
			p < 1%	p < 5%	p < 10%
Intel	5	4	2	3	4
Microsoft	8	10	3	5	8
Verizon	12	17	11	14	15
Citigroup	6	2	0	2	2
BlackRock	14	17	0	3	6
Coca-Cola	5	4	0	1	1
Baxter	8	8	0	2	3
Exxon Mobil	6	4	2	3	3
Gilead Sci.	8	12	2	4	5
Hasbro	7	4	2	3	3
Nike	8	7	2	3	6
Caterpillar	3	3	0	1	1
General El.	8	5	2	2	3
Walmart Inc	5	2	1	2	2
Ventas	3	3	1	2	2
A. Schulman	6	9	2	4	4
Parkway	9	8	2	3	5
Circor	9	6	1	1	2
Dow Jones	8	4	2	4	4

NOTE: The table reports summary results of adaptive lasso estimations for the Sentiment-HAR model (see Equation (1.3)) applied to 18 stocks and a stock index. The first column displays the total number of selected economic variables and the second column the number of selected sentiment and attention variables. The last three columns display the number of significant sentiment and attention variables among the selected ones for three different significance levels. The tuning parameter λ was chosen to minimize the mean-squared-error via blocked cross-validation.

we adopt the recentered bootstrap approach introduced by Audrino and Camponovo (2018). In Table 1.5, we report a summary of the selected variables and parameter significances for the Sentiment-HAR model applied to the 18 analyzed companies and the Dow Jones Index. The first and second columns display the total number of economic and sentiment variables selected by the adaptive lasso. The last three columns display the number of selected sentiment variables for which the null-hypothesis of no effect is rejected at three different significance levels (1%, 5%, and 10%). We observe that the adaptive lasso selects on average about 7 variables from the set of sentiment and attention

variables. In particular, the number of selected sentiment and attention variables is positive for all stocks. Further, we find that the number of selected economic variables is similar or slightly larger than the number of selected sentiment variables for most stocks and the Dow Jones Index. Concerning significance, the large majority of the selected sentiment and attention variables are significant at the 10%-level, and roughly half of the selected sentiment variables are significant at the 5%-level.¹¹

To better understand what source and form of sentiment and attention data is most relevant in predicting future realized volatility, we report in Tables 1.6 and 1.7 the number of selected and the number of significant variables for different sources and forms of sentiment and attention variables when aggregating the numbers over all stocks and the index considered. In Table 1.6, we group the numbers by the different data sources (Twitter, StockTwits, Wikipedia, Google, and RavenPack) and we distinguish between sentiment and attention measures for each source. Additionally, we divide the set of variables into company-specific variables denoted by “Company” and general variables denoted by “General”. The former are variables derived from keywords that are company-specific, i.e., the name of the company and its ticker symbol, whereas the latter are constructed from general keywords such as “stock market” or “interest rates”.

From Table 1.6 we observe that variables derived from the volume of Google search queries for general market keywords (“interest rates”,

¹¹Not tabulated inference results for the economic variables show that roughly 60% of the selected economic variables are significant at the 5% level. The focus being on sentiment variables, we do not report these results in detail.

“financial markets”, “stock markets” and “Dow Jones”) are the most frequently selected predictors. Note that this category includes both the logarithmic search volume and the logarithmic change in search volume. This result confirms the effect of investors’ attention on trading volume and volatility reported in previous studies (see, e.g. Aouadi et al., 2013; Bordino et al., 2012; Dimpfl & Jank, 2015; Mao et al., 2011). The effect of company-specific online searches is clearly less pronounced. This might be due to the fact that, on the one hand, investors become more informed about the stock and uncertainty is thereby reduced and, on the other hand, a high number of searches might signalize concerns and insecurities of investors. Attention and sentiment variables constructed from company-specific StockTwits data, in particular the daily volume of messages, are the second most selected variables. Aggregated over all instruments, the adaptive lasso selects 18 stock-specific StockTwits volume variables out of which 10 are significant at the 1% level. This result is in line with Antweiler and Frank (2004) who find that the number of messages posted on Yahoo! Finance has a significant impact on future volatility. Further, 14 general market sentiment variables from StockTwits data are selected with only approximately one third being significant at the 5% level, though. Interestingly, concerning both Twitter sentiment and volume, neither stock-specific nor general market-related tweets appear to be particularly relevant predictors for realized volatility. This result might be explained by the differences between Twitter and StockTwits users. The latter are generally stock market participants who react quickly to new information. Twitter, instead, is the most

widely used micro-blog platform and has a much more heterogeneous user-base, which likely increases the amount of irrelevant information in sentiment and attention measures for financial markets. Finally, we also observe that some variables constructed with news article information are selected by the lasso, especially stock-specific attention measures, i.e., the daily volume of news articles about a specific company. The number of significant variables is, however, lower compared to the attention variables derived from Google Trends and StockTwits data.

Table 1.6: Number of selected and number of significant sentiment variable grouped by their source

	Num. sentiment var.	Num. significant sentiment var.		
		p < 1%	p < 5%	p < 10%
Twitter Sent. - Company	2	1	1	2
Twitter Volume - Company	2	0	0	0
Stocktwits Sent. - Company	3	2	2	2
Stocktwits Volume - Company	18	10	13	15
Wikipedia Volume - Company	2	0	2	2
Google Volume - Company	4	2	3	4
RavenPack ESS Sent. - Company	3	0	1	2
RavenPack ESS Volume - Company	11	5	6	8
RavenPack CSS Sent. - Company	5	0	1	1
RavenPack NIP Sent. - Company	4	0	1	1
Twitter Sent. - General	8	2	5	5
Twitter Volume - General	5	0	1	2
Stocktwits Sent. - General	14	2	5	8
Stocktwits Volume - General	6	3	3	4
Wikipedia Volume - General	4	2	2	2
Google Volume - General	29	4	12	17
RavenPack ESS Sent. - General	5	1	3	3
RavenPack ESS Volume - General	4	1	1	1

NOTE: The table reports the number of selected sentiment and attention variables and the number of significant sentiment and attention variables among the selected ones of a specific category (at three different significance levels). The third row, for instance, shows that out of the three selected company-specific StockTwits sentiment variables, two are significant at the 1% level. The tuning parameter λ was chosen to minimize the mean-squared-error via blocked cross-validation.

In summary, a clear pattern is the importance of attention measures. In particular, we find that the daily volume of company-specific messages posted on StockTwits and Google searches about financial markets in

general are the most relevant ones. The volume variables for company-specific tweets on StockTwits are selected for 15 out of the 18 stocks, of which 12 are at least significant at the 5%-significance level, and the general financial market Google search volume is selected for 12 out of 18 stocks and for the Dow Jones Index, of which 7 are at least significant at the 10%-significance level (results not tabulated). Moreover, the estimated coefficients for these two variables are always non-negative, indicating that an increase in investors' attention raises future volatility. Noteworthy is also the fact that, among the economic variables, the VIX Index is selected for all companies and the stock index. This is not surprising considering the fact that VIX represents the volatility implied by option prices on the S&P 500 Index and therefore represents the investors' expectations and reflects the market participants' risk aversion. Buncic and Gisler (2016), for instance, show that including log VIX in the classical HAR model can significantly improve its out-of-sample forecasting accuracy. Besides the VIX, the adaptive lasso selects the turnover-ratio¹² for 12 out of the 18 analyzed companies.

In the following, we further investigate which form of sentiment and attention variable is most important. In Table 1.7, we report the number of selected and the number of significant sentiment variables aggregated over all stocks and sources when grouping by the specific forms of the sentiment and attention variables: denoised sentiment denoted by "Sentiment (smoothed)", the square of this, sentiment dispersion (smoothed), the volume, and the return of the volume. These results

¹²The turnover-ratio is defined as daily turnover divided by market capitalization.

Table 1.7: Number of selected and number of significant sentiment variables grouped by their form

	Num. sentiment var.	Num. significant sentiment var.		
		p < 1%	p < 5%	p < 10%
Sentiment (smoothed)	14	3	6	7
Squared (smoothed) sentiment	16	1	7	9
Sentiment dispersion (smoothed)	14	4	6	8
Attention (log volume)	66	18	32	40
Change in attention (log vol. return)	19	9	11	15

NOTE: The table reports the number of selected sentiment and attention variables and the number of significant sentiment and attention variables among the selected ones of a specific form (at three different significance levels). The first row, for instance, shows that out of the 14 selected sentiment variables, 3 are significant at the 1% level. The tuning parameter λ was chosen to minimize the mean-squared-error via blocked cross-validation.

confirm that attention, as measured by the daily posting or query volume, is clearly the most relevant type of predictor variable. Aggregated over all companies and the stock market index, the adaptive lasso selects 66 daily volume variables.¹³ When testing with the conservative inference procedure proposed by Audrino and Camponovo (2018), we observe that 32 of the selected posting volume variables are still significant at the 5% level. The second most selected predictors are the returns in daily posting volumes for which the adaptive lasso selects 19 variables. Sentiment variables in various forms are clearly less frequently selected compared to attention variables. In total, the adaptive lasso selects 16 squared sentiment variables and 14 smoothed sentiment and dispersion variables.¹⁴

¹³Note that this aggregate number of selected attention variables includes volume from all data sources (Twitter, StockTwits, Google, Wikipedia, news articles) considered.

¹⁴Not tabulated robustness checks show that, when changing the smoothing technique, the number of selected sentiment variables slightly changes, but the company-specific StockTwits volume and general market Google searches remain the most selected predictors. Moreover, when no smoothing is used, sentiment variables are almost never selected (results not tabulated).

In conclusion, we find that sentiment and attention variables constructed from micro-blogging, search engine, and news article data have a significant impact on volatility. In line with other studies (Antweiler & Frank, 2004; Da et al., 2011; Hamid & Heiden, 2015), attention measures are particularly informative for future realized volatility, also when controlling for a large set of economic and financial predictors and using a conservative test. The volume of internet search queries and messages on online platforms can be interpreted as a measure of retail investors' attention¹⁵ (Da et al., 2011), and retail investors are often regarded as noise traders who increase volatility (e.g. Black, 1986; Long, Shleifer, Summers, & Waldmann, 1990). The predictive importance of sentiment variables is clearly lower compared to attention variables. Nonetheless, several sentiment variables in various forms including dispersion and squared terms are also significant. However, compared to previous results in the literature (See-To & Yang, 2017; Siganos et al., 2017), we find less evidence for the predictive significance of sentiment dispersion. The likely reasons for this is the fact that we also control for other forms of sentiment and attention variables as well as for important financial and economic variables. More in line with our results are Antweiler and Frank (2004) who find no significant influence of sentiment dispersion on future volatility when controlling for investors' attention.

¹⁵Da et al. (2011), for instance, show that changes in the volume of Google searches are strongly related to retail investors' trading activity.

1.4 Forecasting volatility with social media, news, information consumption and search engine data

1.4.1 Forecasting methodology

In this section, we present the forecasting methodology and results for the three models considered, namely the HAR, Economic-HAR and Sentiment-HAR defined in Equations (1.1), (1.2) and (1.3). In the following, the models are estimated using a rolling window of 502 days, which corresponds to roughly two years. After the models are estimated with the adaptive lasso,¹⁶ h -steps ahead forecasts are produced. The estimation window is then rolled over by one day, and the procedure is repeated. Following Andersen et al. (2007) and Corsi and Renò (2012), we use direct forecasts for $h > 1$. Note that we use the adaptive lasso also for making predictions in order to be consistent with the methodology used for the in-sample analysis in Section 1.3.

Our main focus is on one-day ahead forecasts ($h = 1$) but we also report longer term forecasts ($h = 2, 3, 4, 5, 10$, and 22 days). The out-of-sample accuracy of the three analyzed models is assessed with the mean

¹⁶As for the in-sample analysis, the tuning parameter is chosen to minimize the mean squared cross-validation error. The blocked cross-validation is repeated for each rolling window.

squared prediction error (MSPE) defined as

$$\text{MSPE}^{(J)} = \frac{1}{T_{OS}} \sum_{t=T_{IS}}^T e_{t+h|t}^{(J)2} \quad (1.6)$$

where T , T_{OS} and T_{IS} are the total, out-of-sample and in-sample number of observations, respectively, and where J denotes the model that is used, $J \in \{\text{HAR, Economic, Sentiment}\}$. The forecasting error of model J is denoted by $e_{t+h|t}^{(J)}$ and defined as

$$e_{t+h|t}^{(J)} = \log RV_{t+h}^{(h)} - \hat{y}_{t+h|t}^{(J)}, \quad (1.7)$$

where $\log RV_{t+h}^{(h)}$ is the average log realized volatility between $t + 1$ and $t + h$ and the respective forecasts of the three HAR models are denoted by $\hat{y}_{t+h|t}^{(J)}$. Since we are comparing nested models, the testing procedure proposed by Clark and West (2007) is adopted (CW test). Due to the multiple testing problems, we control for the family-wise error rate in a conservative way by applying the Bonferroni-Holm correction (Holm, 1979) to the p-values of the CW test.

Besides reporting the mean squared prediction error and associated test statistics, we also compute the out-of-sample R^2 as reported by Campbell and Thompson (2008):

$$R^2 = 1 - \frac{\text{MSPE}}{\text{MSPE}_{BM}}, \quad (1.8)$$

where MSPE_{BM} is the mean squared prediction error of a benchmark model. Equation (1.8) can be interpreted as the percentual increase or

decrease in MSPE between a model of interest and a benchmark model.

In the previous section, we have also included daily interpolated versions of macroeconomic variables which are only available at a monthly frequency. In order to produce true ex-ante forecasts, we exclude these monthly variables from the predictor set. Moreover, the Kalman filter’s parameters are estimated using only data inside the rolling window. Robustness checks for the choice of the adaptive lasso’s tuning parameter, the rolling window’s length, the realized volatility estimator, the filtering approach, and the natural language processing methodology for the one-day ahead forecasts are reported in Appendix 1.B.

1.4.2 One-day ahead forecasts

In Table 1.8, we report the one-day ahead forecast results for the three analyzed models. The first two columns report the average number of economic and sentiment variables selected by the adaptive lasso over all rolling windows. The mean squared prediction errors of the Sentiment-HAR, Economic-HAR and baseline HAR models are reported in the third, fourth and fifth columns, respectively. The last three columns report the p-values of Clark-West test for superior predictive accuracy of the Sentiment-HAR model over the Economic-HAR (Sent./Eco.) and the baseline model (Sent./HAR), as well as the superior predictive accuracy of the Economic-HAR over the baseline HAR model (Eco./HAR). Tests that are significant at the 1%-, 5%- or 10%-level after applying a Bonferroni-Holm correction are indicated by ***, ** and *, respectively. In addition, we report in Table 1.9 the out-of-sample R^2 as defined by Campbell and

Thompson (2008). The first and second columns show the percentual decrease in MSPE achieved by the Sentiment-HAR and the Economic-HAR models with respect to the baseline HAR model, while the last column reports the change in MSPE between the Sentiment-HAR and the Economic-HAR models.

Table 1.8: Summary results of the 1-day ahead forecasts based on a rolling window over the time period 01.01.2014 - 31.12.2016

	Avg. eco.	Avg. sent.	MSPE			P-values Clark-West test (in %)		
			Sent.	Eco.	HAR	Sent./ Eco.	Sent./ HAR	Eco./ HAR
Intel	4.55	4.91	0.064	0.065	0.069	0.01***	0***	0***
Microsoft	5.58	5.84	0.066	0.07	0.074	0***	0***	0.01***
Verizon	3.83	6.87	0.066	0.069	0.073	0***	0***	0***
Citigroup	6.05	4.7	0.068	0.07	0.076	0.01***	0***	0***
BlackRock	5.38	5.29	0.078	0.079	0.085	0.02***	0***	0***
Coca-Cola	3.4	2.52	0.058	0.058	0.062	8.87	0.22**	0.15**
Baxter	4.69	5.9	0.067	0.067	0.071	1.31*	0***	0.01***
Exxon Mobil	4.07	4.35	0.054	0.056	0.058	0***	0***	0.09**
Gilead Sci.	5.01	4.03	0.076	0.077	0.078	1.51*	0***	0.01***
Hasbro	5.95	5.36	0.083	0.083	0.088	0.62**	0.02***	0.12**
Nike	6.04	5.35	0.059	0.062	0.067	0.01***	0***	0***
Caterpillar	4.18	2.86	0.067	0.067	0.069	1.6*	0***	0***
General El.	6.17	5.23	0.068	0.069	0.072	0.08**	0***	0***
Walmart Inc	4.32	4.51	0.07	0.07	0.074	4.86	0.02***	0.08**
Ventas	3.3	3.45	0.06	0.062	0.063	0.01***	0***	0.23**
A. Schulman	4.44	4.73	0.182	0.191	0.197	0.01***	0***	0.02***
Parkway	3.26	2.68	0.082	0.082	0.084	59.21	0.14**	0.07**
Circor	4.8	5.16	0.206	0.207	0.217	1.9*	0***	0***
Dow Jones	4.47	4.51	0.061	0.063	0.07	0***	0***	0***

NOTE: The table reports the results of 1-day ahead log realized volatility forecasts based on a rolling window. The first column shows the average number of selected economic variables and the second column the average number of selected sentiment variables. The third, forth and fifth columns report the mean squared prediction error of the Sentiment-, Economic- and classical HAR-models, respectively. The last three columns report the p-values of Clark-West test for superior predictive ability of the Sentiment-HAR model over the Economic-HAR (Sent./Eco.) and the baseline model (Sent./HAR), as well as the superior predictive ability of the Economic-HAR over the baseline HAR model (Eco./HAR). The stars indicate if we are still able to reject the null-hypothesis when applying the Bonferroni-Holm correction. * indicates that the result is still significant at the 10%, ** at 5% and *** at 1% significance level.

We observe that by including economic variables in the set of potential predictors in the baseline HAR model, the MSPE decreases for all companies. The CW test results show that these decreases are all significant

Table 1.9: Out-of-sample R^2 for comparing the MSPE of the three different HAR models (1-day ahead forecasts based on a rolling window over the time period 01.01.2014 - 31.12.2016)

	Sentiment/Baseline HAR	Economic/Baseline HAR	Sentiment/Economic HAR
Intel	7.63	5.81	1.93
Microsoft	11.02	5.69	5.66
Verizon	9.22	5.23	4.21
Citigroup	10.35	8.28	2.26
BlackRock	7.46	6.29	1.26
Coca-Cola	5.89	5.73	0.18
Baxter	6.17	5.82	0.37
Exxon Mobil	5.91	2.76	3.24
Gilead Sciences	2.01	1.27	0.76
Hasbro	6.36	5.49	0.93
Nike	11.59	8.29	3.6
Caterpillar	3.49	3.21	0.29
General Electric	6.36	4.56	1.88
Walmart Inc	4.47	5.04	-0.6
Ventas	4	2.03	2.01
A. Schulman	7.93	3.39	4.7
Parkway	1.78	2.81	-1.06
Circor	4.86	4.47	0.4
Dow Jones	12.7	9.19	3.87
Average R^2	6.8	5.02	1.89

NOTE: The table reports the out-of-sample R^2 as reported in Campbell and Thompson (2008), i.e. $R^2_{i,j} = 1 - MSPE_i / MSPE_j$. The results are all reported in percent and can be interpreted as the increase (or decrease) in forecasting accuracy of a model compared to a benchmark model. The first and second columns represent the decrease in MSPE achieved by the Sentiment-HAR and Economic-HAR models compared to the baseline HAR model, respectively. The last column reports the decrease (increase) in MSPE from the Economic- to the Sentiment-HAR model.

at the 5%-level. In particular, the results reported in Table 1.9 show that adding economic variables to the classical HAR model reduces the MSPE on average by 5.02%. Not surprisingly, the most selected economic variable is again the VIX, followed by the companies' turnover-ratio.

Concerning sentiment and attention variables, the last column in Table 1.9 shows that adding sentiment and attention variables to the HAR model which already includes economic and financial variables improves the out-of-sample R^2 by an additional 1.89% on average. In more detail, Table 1.8 shows that compared to the Economic-HAR model, the Sentiment-HAR model results in an MSPE that is lower for the

majority of the stocks and the Dow Jones index. Further, except for three companies, the Clark-West test rejects the null-hypothesis that the Sentiment-HAR model does not improve the predictive accuracy of the Economic-HAR in favor of the Sentiment-HAR model at the 10% level after applying a conservative Bonferroni-Holm correction. Following Clark and West (2007), these results can be interpreted in the sense that sentiment and attention variables consistently provide increased predictive accuracy. The results reported in Table 1.8 are confirmed by the Model Confidence Set (MCS) results (Hansen, Lunde, & Nason, 2011) reported in Table 1.20 in Appendix 1.C. The Sentiment-HAR model is always included in the 90% model confidence set and we find the highest MCS p-value for the Dow Jones Index and all companies, except Coca-Cola and Parkway, for the sentiment-enhanced HAR model. Note that many more parameters need to be estimated for the Sentiment-HAR model compared to the Economic-HAR model based on a relatively small sample size. This leads to an increase in noise in the forecasts, which might be the reason that for two stocks the Sentiment-HAR model does not improve the MSPE compared to the Economic-HAR model.

In line with the results from our in-sample analysis, the most selected predictors within the set of sentiment and attention variables are attention measures (results not tabulated). More precisely, company-specific investors' attention as measured by the number of micro-messages posted on StockTwits and general market attention as captured by the volume of Google searches for economic keywords are the most frequently selected variables.

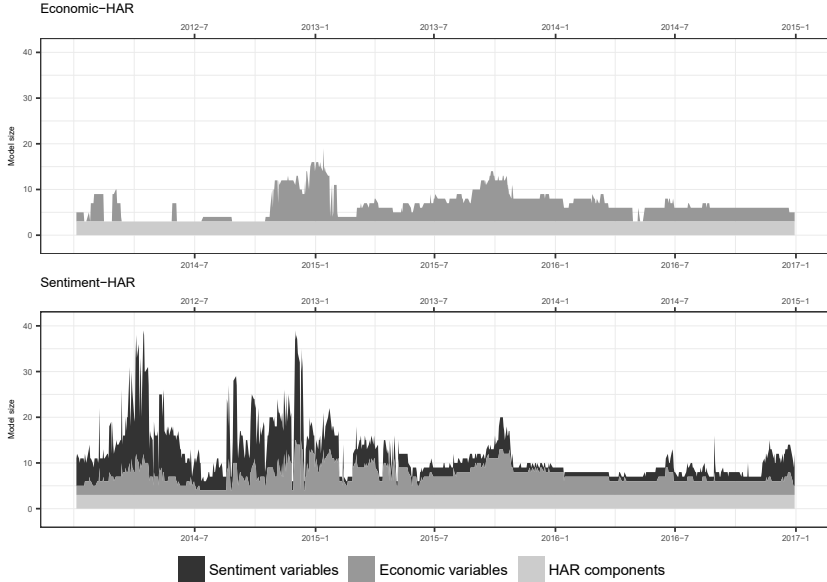


Figure 1.6: Time variation in the Economic-HAR and Sentiment-HAR models' composition for A. Schulman over the out-of-sample period 01.01.2014 - 31.12.2016

NOTE: The figure represents, the number of selected sentiment, economic and HAR components by the Economic-HAR (first row) and the Sentiment-HAR (second row) models over the out-of-sample period from 01.01.2014 to 31.12.2016 for A. Schulman. The lightest color represents the number of HAR components (lagged realized volatilities), which is always equal to three by construction. The darkest color represents the amount of sentiment predictors and the middle-tone grey the number of economic and financial predictors included in the model by adaptive lasso. The upper and lower x-axes represent the start and end points of the rolling estimation windows.

Further, we observe that both the number of selected economic variables and in particular the composition and number of selected sentiment and attention variables vary over time. To illustrate this, we show in Figure 1.6 the selected variables grouped by sentiment, economic, and lagged volatilities for A. Schulman for the different rolling estimation windows.¹⁷ The plot in the first row represents the number of selected predictors for the Economic-HAR model and the plot in the second row reports the

¹⁷Note that by construction, lagged realized volatilities are always included in all three models.

Table 1.10: Out-of-sample R^2 for comparing the MSPE of the three different HAR models for days with realized volatilities above the historical 95%-quantile (1-day ahead forecasts based on a rolling window over the time period 01.01.2014 - 31.12.2016)

	Sentiment/Baseline HAR	Economic/Baseline HAR	Sentiment/Economic HAR
Intel	9.9	7.53	2.56
Microsoft	21.27	10.41	12.12
Verizon	12.83	13.19	-0.42
Citigroup	13.88	15.56	-1.99
BlackRock	13.59	11.83	1.99
Coca-Cola	23.39	24.77	-1.82
Baxter	11.41	13.14	-2
Exxon Mobil	27.41	20.34	8.87
Gilead Sciences	12.05	12.41	-0.41
Hasbro	14.66	15.29	-0.75
Nike	25.31	20.04	6.59
Caterpillar	7.21	5.6	1.7
General Electric	17.74	19.36	-2.01
Walmart Inc	11.71	13.02	-1.51
Ventas	8.56	4.3	4.45
A. Schulman	25.88	7.98	19.45
Parkway	3.07	3.93	-0.89
Circor	12.58	7.79	5.19
Dow Jones	25.79	24.94	1.14
Average R^2	15.7	13.23	2.75

NOTE: The table reports the out-of-sample R^2 as reported in Campbell and Thompson (2008), i.e. $R^2_{i,j} = 1 - MSPE_i/MSPE_j$, for days where a company's log realized volatility was above its historical 95%-quantile. The results are all reported in percent and can be interpreted as the increase (or decrease) in forecasting accuracy of a model compared to a benchmark model on days with particularly large realized volatilities. In the first and second columns we report the decrease in MSPE achieved by the Sentiment-HAR and Economic-HAR models compared to the baseline HAR model, respectively. The last column reports the decrease (increase) in MSPE from the Economic- to the Sentiment-HAR model.

number of sentiment, economic, and lagged volatilities selected by the Sentiment-HAR model. The upper and lower x-axes represent the start and end points of the rolling windows. We observe a strong variation in the number of selected sentiment predictors over time. The time-variation of economic variables is of smaller magnitude. In particular, as mentioned in the introduction, for the first year the Economic-HAR model selects almost no economic and financial predictors.

Interestingly, we observe particularly large improvements in forecast-

ing accuracy of both the Economic- and Sentiment-HAR models for days with high realized volatility. Table 1.10 reports the out-of-sample R^2 on days where the companies' log realized volatility was above its historical 95%-quantile. The addition of economic and financial variables to the baseline HAR model reduces the MSPE by 13.23% on average. Moreover, when adding sentiment and attention variables to the HAR model that already includes economic variables, the mean squared prediction error is reduced by an additional 2.75%. Similar improvements can be observed on days with highly negative or positive returns (see Table 1.21 in Appendix 1.C).

Moreover, the largest increase in predictive accuracy due to sentiment and attention variables is observed for companies with a large market capitalization and/or a low share of institutional investors (see Table 1.22 in Appendix 1.C). This might be due to the fact that larger companies are more widely covered in social media platforms and news articles, and companies with a large share of institutional investors are less affected by noise traders.

1.4.3 Long-horizon realized volatility forecasts

The results reported in the previous paragraphs showed that sentiment variables have predictive power for realized volatility at the one-day ahead forecasting horizon. In the following, we extend our analysis to longer forecasting horizons. Figure 1.7 shows the average reduction in MSPE between the Sentiment-HAR and the baseline HAR models (solid line), between the Economic-HAR and the baseline HAR (dashed

line) and that between the Sentiment-HAR and Economic-HAR models (dotted line) for seven different forecasting horizons. More precisely, we report the average reduction in MSPE for the one-, two-, three-, four- and five-days horizons, as well as for the two-weeks and one-month ahead predictions. Figure 1.7 clearly shows that the forecasting power of sentiment and attention variables decreases with the prediction horizon. On average, micro-blogging, search engine, and news article data slightly improve the volatility predictions at the two-days horizon but cease to do so for longer prediction horizons. In contrast, economic and financial variables have predictive power for future volatility up to the two-weeks horizon. Interestingly, on average the Economic-HAR model outperforms the simple HAR model even more at the two- and three-day horizon compared to the one-day prediction horizon.

Following the literature on modelling and forecasting realized volatility (see, among others, Corsi et al., 2010; Corsi & Renò, 2012; Patton & Sheppard, 2015), we report detailed prediction results for the weekly and monthly horizons ($h = 5$ and $h = 22$). Table 1.11 reports the results for weekly and Table 1.12 for monthly forecasts. As for the one-day ahead forecasts, the tables report in the first two columns the average number of selected economic and sentiment predictors. The third, fourth and fifth columns show the MSPE of the Sentiment-HAR, Economic-HAR and baseline HAR models. The last three columns report the p-values of Clark-West test for superior predictive ability of the Sentiment-HAR model over the Economic-HAR (Sent./Eco.) and the baseline model (Sent./HAR), as well as the superior predictive ability of the Economic-

HAR over the baseline HAR model (Eco./HAR). Tests that are significant at the 1%, 5% or 10%-level after applying a Bonferroni-Holm correction are indicated with ***, ** and *.

For the weekly volatility forecasts, we notice that the Economic-HAR is still able to reduce the MSPE of the classical HAR for the Dow Jones Index and 14 out of 18 companies, of which 9 are significant at the 10%-level after the Bonferroni-Holm correction. The Sentiment-HAR model is able to reduce the MSPE of the baseline HAR for 14 out of the 18 companies and the Dow Jones Index. The results for the Dow Jones Index and 6 companies are also significant at the 5%-level after a

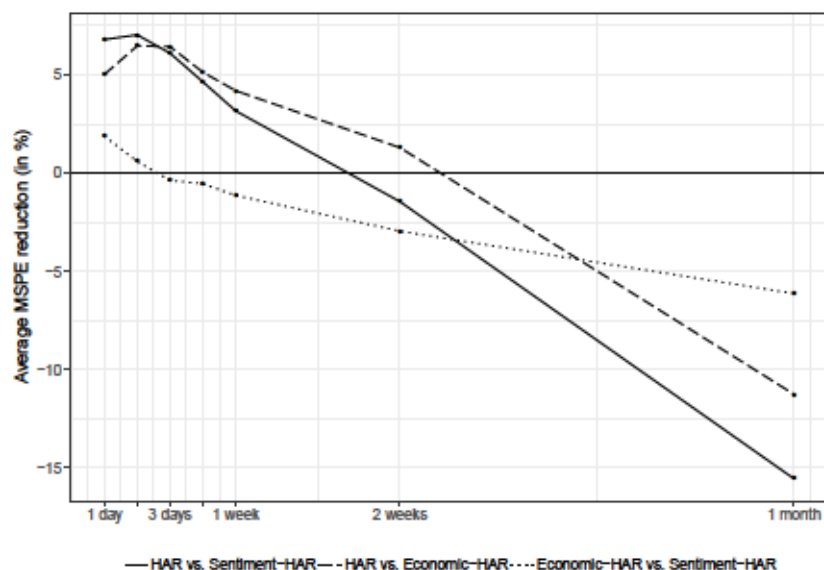


Figure 1.7: Average MSPE reduction for different forecasting horizons

NOTE: The figure represents the average reduction in MSPE between the Sentiment-HAR and the baseline HAR models (solid line), between the Economic-HAR and the baseline HAR (dashed line) and that between the Sentiment-HAR and Economic-HAR models (dotted line) for seven forecasting horizons, i.e. one day, two days, three days, four days, one week and one month.

Bonferroni-Holm correction. However, compared to the Economic-HAR model, the addition of sentiment predictors further reduces the MSPE for only 7 companies, of which none are significant at any conventional significance level. The magnitude of the reductions in MSPE is also much lower than for the one-day ahead forecasts. Enhancing the baseline HAR model with economic variables reduces on average the MSPE by 4.2% (for the one-day ahead forecasts the reduction was 5.02%); the inclusion of micro-blogging, web query and news article data even leads,

Table 1.11: Summary results of the 5-days ahead forecasts based on a rolling window over the time period 01.01.2014 - 31.12.2016

	Avg. eco.	Avg. sent.	MSPE			P-values Clark-West test (in %)		
			Sent.	Eco.	HAR	Sent./ Eco.	Sent./ HAR	Eco./ HAR
Intel	3.14	1.94	0.044	0.045	0.047	3.4	0.06**	0.04**
Microsoft	4.35	2.65	0.046	0.043	0.049	92.86	0.05**	0.01***
Verizon	2.53	2.4	0.036	0.035	0.039	43.57	0.01***	0***
Citigroup	4.1	2.3	0.048	0.047	0.051	47.14	0.02**	0***
BlackRock	4.1	3.42	0.055	0.053	0.057	37.17	0***	0.01***
Coca-Cola	1.83	0.99	0.036	0.037	0.036	1.09	1.06	0.66
Baxter	3.17	2.57	0.044	0.045	0.046	6.92	0.4	0.43
Exxon Mobil	2.11	2.2	0.047	0.047	0.046	14.14	14.61	13.7
Gilead Sci.	1.11	0.77	0.053	0.053	0.053	22.47	9.74	7.44
Hasbro	3.3	3.13	0.045	0.046	0.049	6.09	0.45	1.49
Nike	4.41	2.6	0.041	0.04	0.041	64.21	1.19	0.21*
Caterpillar	2.57	1.38	0.042	0.04	0.041	57.41	23.1	0.43
General El.	3.69	2.39	0.05	0.051	0.054	6.36	0.01***	0.05**
Walmart Inc	1.76	1.55	0.04	0.041	0.041	1.44	0.9	1.13
Ventas	2.1	2.05	0.037	0.036	0.035	12.86	25.75	3.02
A. Schulman	1.15	1.16	0.094	0.094	0.099	21.9	4.57	2.35
Parkway	2.24	2.23	0.039	0.039	0.04	16.82	2.47	2.97
Circor	1.84	0.87	0.095	0.092	0.097	52.6	0.83	0***
Dow Jones	4.45	1.91	0.061	0.06	0.069	12.86	0***	0***

NOTE: The table reports the results of 5-days ahead log realized volatility forecasts based on a rolling window. The first column shows the average number of selected economic variables and the second column the average number of selected sentiment variables. The third, fourth and fifth columns report the mean squared prediction error of the Sentiment-, Economic- and classical HAR-models, respectively. The last three columns report the p-values of Clark-West test for superior predictive ability of the Sentiment-HAR model over the Economic-HAR (Sent./Eco.) and the baseline model (Sent./HAR), as well as the superior predictive ability of the Economic-HAR over the baseline HAR model (Eco./HAR). The stars indicate if we are still able to reject the null-hypothesis when applying the Bonferroni-Holm correction. * indicates that the result is still significant at the 10%, ** at 5% and *** at 1% significance level.

Table 1.12: Summary results of the 22-days ahead forecasts based on a rolling window over the time period 01.01.2014 - 31.12.2016

	Avg. eco.	Avg. sent.	MSPE			P-values Clark-West test (in %)		
			Sent.	Eco.	HAR	Sent./ Eco.	Sent./ HAR	Eco./ HAR
Intel	1.9	3.55	0.043	0.046	0.041	1.58	1.23	7.43
Microsoft	2.04	2.79	0.039	0.038	0.04	11.88	5.36	9.1
Verizon	1.96	3.06	0.034	0.034	0.025	0.79	20.85	83.81
Citigroup	2.3	3.2	0.051	0.051	0.045	5.79	13.09	19.66
BlackRock	2.81	4.41	0.065	0.065	0.05	4.95	17.45	25.06
Coca-Cola	1.86	2.32	0.037	0.034	0.027	50.42	51.37	29.37
Baxter	2.74	5.11	0.048	0.035	0.039	47.81	6.6	4.23
Exxon Mobil	1.11	2.66	0.059	0.052	0.049	65.12	77.66	55.91
Gilead Sci.	0.83	1.35	0.05	0.06	0.051	0.57	7.35	56.13
Hasbro	3.87	5.81	0.055	0.065	0.042	5.89	4.19	6.98
Nike	2.24	4.18	0.045	0.039	0.037	82.4	30.93	18.46
Caterpillar	2	2.08	0.04	0.034	0.041	75.08	9.75	3.06
General El.	2.22	3.54	0.045	0.048	0.051	2.73	0.5	1.53
Walmart Inc	1.37	2.23	0.035	0.033	0.032	19.25	13.12	8.64
Ventas	1.87	2.61	0.031	0.044	0.025	5.78	15.84	43.84
A. Schulman	1	0.47	0.058	0.058	0.071	42.48	3.53	4.36
Parkway	2.27	3.24	0.046	0.032	0.029	36.47	7.08	7.18
Circor	1.19	0.56	0.054	0.05	0.054	20.86	7.25	5.06
Dow Jones	1.93	1.95	0.091	0.068	0.074	41.86	0.48	1.16

NOTE: The table reports the results of 22-days ahead log realized volatility forecasts based on a rolling window. The first column shows the average number of selected economic variables and the second column the average number of selected sentiment variables. The third, forth and fifth columns report the mean squared prediction error of the Sentiment-, Economic- and classical HAR-models, respectively. The last three columns report the p-values of Clark-West test for superior predictive ability of the Sentiment-HAR model over the Economic-HAR (Sent./Eco.) and the baseline model (Sent./HAR), as well as the superior predictive ability of the Economic-HAR over the baseline HAR model (Eco./HAR). The stars indicate if we are still able to reject the null-hypothesis when applying the Bonferroni-Holm correction. * indicates that the result is still significant at the 10%, ** at 5% and *** at 1% significance level.

on average, to an additional increase of 1.13%. Thus, while economic data still improve the MSPE for the majority of the analyzed companies, the additional forecasting ability of sentiment variables is limited. From Table 1.11 we observe that the average number of selected sentiment and economic variables is also lower compared to the one-day ahead forecasts.

For the one-month ahead predictions, the influence of economic and sentiment predictors becomes even smaller. The Economic-HAR model is able to reduce the MSPE for only 5 companies, none of which are

significant. In the majority of cases, enhancing the baseline HAR with economic data leads to an increase in the MSPE. The situation is similar for the Sentiment-HAR model. Compared to the classical HAR model, the Sentiment-HAR model is able to reduce the MSPE only for A. Schulman and General Electrics. For all other cases, enhancing the baseline model with both economic and sentiment data leads to a lower forecasting accuracy.

The results for longer forecasting horizons clearly show that both economic and sentiment variables have only a short-lasting effect on future volatility. While at the one-week horizon economic predictors still have some ability to improve the forecasting accuracy of the baseline HAR model, neither the addition of economic nor sentiment data appears to improve its forecasting accuracy at the one-month horizon. In the spirit of Corsi's assumption about heterogeneities among investors (Corsi, 2009), this result confirms the idea that sentiment variables are more representative of the behavior of retail (short-term trading horizon) and less of institutional investors (long-term trading horizon).

1.4.4 A risk management application: value-at-risk forecasting

In this subsection, we briefly present an application of our sentiment enhanced HAR model. Based on the one-day ahead forecasts, we compute value-at-risk forecasts for three different quantiles. More precisely, we forecast VaR_{t+1}^α such that $\Pr(r_{t+1} \leq \text{VaR}_{t+1}^\alpha) = \alpha$ for $\alpha \in \{0.5\%, 1\%, 5\%\}$, where r_t is the daily log return. The forecasts are obtained by means

of filtered historical simulation (see, e.g. Hull, Hull, White, & White, 1998) and evaluated with the asymmetric loss function proposed by González-Rivera, Lee, and Mishra (2004)

$$L = \frac{1}{T_{OS}} \sum_{t=T_{IS}}^T (\alpha - d_{t+1}^\alpha)(r_{t+1} - \widehat{\text{VaR}}_{t+1}^\alpha) \quad (1.9)$$

where $d_{t+1}^\alpha = \mathbb{1}\{r_{t+1} < \widehat{\text{VaR}}_{t+1}^\alpha\}$. The above defined loss function penalizes more heavily observations for which the daily log return falls below the forecasted quantile. We compare the value-at-risk forecasting accuracy of the Sentiment-HAR, Economic-HAR and baseline HAR by computing a *model confidence set* (Hansen et al., 2011) for each asset and each quantile level. Since for value-at-risk forecasts the total variation is more important than only the continuous part of volatility, we consider realized volatilities estimated by the sum of 5-minute squared returns instead of the MedRV estimator.¹⁸ A summary of the one-day ahead prediction accuracy for classical (log) realized volatilities of the Sentiment-HAR model can be found in Table 1.18 in Appendix 1.B.

Table 1.13 summarizes the losses of the three HAR specifications for three value-at-risk levels. More precisely, the first and second columns in Panel A show the percentual decrease in the asymmetric VaR loss achieved by the Sentiment-HAR and the Economic-HAR models with respect to the baseline HAR model, while the last column reports the change in the asymmetric VaR loss between the Sentiment-HAR and the Economic-HAR models. We observe that the Sentiment-HAR model

¹⁸We kindly thank one of the anonymous referees for pointing this out.

Table 1.13: Summary of the value-at-risk results based on the asymmetric loss function

Panel A: Average reduction in the asymmetric value-at-risk loss (in %)			
	Sentiment/Baseline HAR	Economic/Baseline HAR	Sentiment/Economic HAR
VaR 0.5%	2.62	0.48	2.14
VaR 1%	1.98	0.49	1.49
VaR 5%	1.26	1.03	0.23

Panel B: Share of lowest asymmetric value-at-risk loss (in %)			
	Sentiment-HAR	Economic-HAR	Baseline HAR
VaR 0.5%	68.42	0	31.58
VaR 1%	63.16	5.26	31.58
VaR 5%	63.16	26.32	10.53

NOTE: The table summarises the value-at-risk results based on the asymmetric loss function introduced by González-Rivera et al. (2004). The three columns in Panel A show the average reduction in the asymmetric loss achieved by the Sentiment-HAR compared to the baseline HAR, the Economic-HAR to the baseline HAR and the Sentiment-HAR to the Economic-HAR, respectively, for the 0.5%, 1% and 5% value-at-risks. Panel B shows the relative number of times a model achieved the lowest loss over all analyzed companies and the stock index. For instance, the first column shows that, for the daily VaR_{1%}, the Sentiment-HAR model achieved the lowest loss for 63.16% of the analyzed assets.

improves the value-at-risk forecasts of both the baseline HAR and the Economic-HAR model for all three considered quantiles. The reduction in asymmetric VaR loss achieved by the Sentiment-HAR model compared to the Economic-HAR model increases as α becomes smaller. While for the 5%-VaR, the addition of sentiment and attention variables to a model which already accounts for economic data decreases the loss on average by only 0.23%, for the 0.5%-VaR the decrease amounts to more than 2%. Panel B in Table 1.13 reports the relative number (in %) of companies for which a model achieved the lowest loss. For all three considered quantiles we observe that, value-at-risk forecasts based on the Sentiment-HAR model clearly achieve the lowest loss for the majority of companies and the Dow Jones Index.

The detailed model confidence set results at a 10%-significance level are reported in Table 1.23 in Appendix 1.C. From the MCS results we

observe that generally the confidence set contains more than one model, and in the majority of cases all three models are kept. Except for 5%-VaR prediction of General Electrics, the Sentiment-HAR model is never eliminated from the model confidence set, and has in most cases the highest MCS p-value.¹⁹

Table 1.14: Average reduction in the asymmetric value-at-risk loss for days with returns below the historical 5%-quantile

	Sentiment/Baseline HAR	Economic/Baseline HAR	Sentiment/Economic HAR
VaR 0.5%	5.1	0.01	5
VaR 1%	3.19	0.24	2.81
VaR 5%	5.74	4.92	0.81

NOTE: The table summarises the value-at-risk results based on the asymmetric loss function introduced by González-Rivera et al. (2004) for days with an extreme negative return. The three columns show the average reduction in the asymmetric loss achieved by the Sentiment-HAR compared to the baseline HAR, the Economic-HAR to the baseline HAR and the Sentiment-HAR to the Economic-HAR, respectively, for the 0.5%, 1% and 5% value-at-risks on days where the return was below the historical 5%-quantile.

For the one-day ahead forecasts, we have found that sentiment and attention variables have higher predictive power on days characterized by high realized volatility. In Table 1.14 we report the average reductions in the asymmetric VaR loss between the three models on days where a stock's log return is extremely negative, i.e. the return is below its historical 5%-quantile. The reason for considering days with extreme negative returns, instead of days with large realized volatility, is that the risk management application presented in this study considers only the left tail of the conditional return distribution. The improvements in the value-at-risk predictions achieved by the Sentiment-HAR compared to the Economic-HAR are even larger on days with extremely negative

¹⁹The MCS p-value represents the highest significance at which a given model is still in the confidence set. For more details about the model confidence set and its p-values, we refer to Hansen et al. (2011).

returns. Furthermore, we observe again that the additional reduction in the asymmetric VaR loss achieved when adding sentiment and attention variables particularly increases as we move further into the tail, i.e. for smaller values of α .

In conclusion, this short application of the Sentiment-HAR model to a risk management task shows that enhancing the classical HAR model with sentiment and economic data can improve one-day ahead value-at-risk forecasts, even more so on days with particularly negative returns. Note that besides a direct comparison of the models, we have also evaluated the accuracy of value-at-risk forecasts individually with an unconditional coverage test (Kupiec, 1995), a conditional coverage test (Christoffersen, 1998) and the dynamic quantile test proposed by Engle and Manganelli (2004). The results (not tabulated) of these tests were, however, not able to uncover particular differences between the models.

1.5 Conclusion

We introduce a novel and extensive dataset composed of daily economic and sentiment variables covering a period of 5 years. Using a state-of-the-art sentiment classification technique and considering a wide range of sentiment sources and forms, we identify important drivers of volatility. By using a predictive regression model, we analyze in an in- and out-of-sample setting the predictive power of economic and sentiment variables. Our study shows that sentiment and attention variables clearly have predictive power for future volatility also when controlling for a large set

of economic and financial variables. Measures of investors' attention are identified to have the most significant impact on future volatility. More precisely, Google searches on financial keywords (e.g. "financial market" and "stock market") as well as the daily volume of company-specific messages posted on StockTwits are the most relevant predictors for both in-sample and out-of-sample analyses. Among the economic variables, the implied volatility index VIX and the turnover-ratio are identified as the most relevant predictors.

At the one-day horizon, we find that the inclusion of economic and financial variables in the classical HAR model increases its forecasting accuracy significantly. For the majority of stocks, the addition of sentiment variables leads to a further decrease in the mean squared prediction error, even more so on days with high realized volatility. While significant, these additional improvements are in general relatively small in magnitude. Concerning longer forecasting horizons, we find only limited evidence for the predictive power of economic and sentiment variables, indicating that these variables have a short-lasting effect on volatility. More precisely, we find that economic variables have, on average, predictive power up to the two-week horizon, and sentiment and attention variables are able to increase the predictive accuracy for the one- and two-day ahead predictions. Moreover, a simple risk management application shows that sentiment and attention variables also improve the accuracy of one-day ahead value-at-risk predictions, especially for the 0.5%-VaR.

Our results show that the informativeness of sentiment data about future volatility is generally lower for companies which either have a

small market capitalization or a high share of institutional investors. Nevertheless, in periods characterized by unexpected announcements or breaking news, micro-blogging, search engine and news article data can significantly improve the volatility forecast for these companies as well. For instance, in some preliminary investigations sentiment variables appear to be particularly useful in predicting future volatility in the following cases: (i) when it was announced in November 2016 that Kindred Healthcare was expected to purchase 36 skilled nursing facilities from Ventas Inc.; (ii) when on 5 April 2016 A. Schulman announced that it was seeking at least \$31 million compensation after a company acquired in early 2015 falsified product test results; and (iii) when at the end of January 2014 Caterpillar Inc. announced decreases in revenues and profits. A more comprehensive and detailed analysis of all such events needs a larger cross-section of companies and is left for future research.

Further research could uncover potential non-linearities in the impact of sentiment variables on volatility and investigate in more detail heterogeneities in the impact of sentiment data on volatility for different companies and over different time periods or temporal frequencies. An interesting extension of our study, might be to analyse the predictive power of sentiment and attention variables at higher frequencies, i.e. at the intraday level. Preliminary results (not tabulated) show that even after controlling for overnight returns, news article, social media and search engine data are still able to improve the prediction of future realized volatility. Moreover, the impact of sentiment and attention of different investor types is also of interest.

Acknowledgements

We are grateful to David Garcia for help in obtaining Twitter and Wikipedia data. We also thank Trevor Hastie for helping us to use the glmnet R package for estimating the recentered bootstrap adaptive lasso. Further, we are grateful for comments and suggestions by the associate editor and two anonymous reviewers which helped to improve the article. This research was partially supported by the Swiss National Science Foundation (100018-172685).

References

- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. *The Review of Economics and Statistics*, 89(4), 701–720.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43–76.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2), 579–625.
- Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbour truncation. *Journal of Econometrics*, 169(1), 75–93.
- Andrei, D., & Hasler, M. (2015). Investor Attention and Stock Market Volatility. *The Review of Financial Studies*, 28(1), 33–72.
- Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59(3), 1259–1294.
- Aouadi, A., Arouri, M., & Teulon, F. (2013). Investor attention and stock market activity: Evidence from France. *Economic Modelling*, 35, 674–681.
- Audrino, F., & Camponovo, L. (2018). Oracle Properties, Bias Correction, and Bootstrap Inference for Adaptive Lasso for Time Series M-

- Estimators. *Journal of Time Series Analysis*, 39(2), 111–128.
- Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and Bipower Variation with Stochastic Volatility and Jumps. *Journal of Financial Econometrics*, 2(1), 1–37.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2008). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3), C1–C32.
- Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2), 253–280.
- Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, 96, 355–367.
- Bergmeir, C., & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, 191, 192–213.
- Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528–543.
- Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8.
- Bollerslev, T., Litvinova, J., & Tauchen, G. (2006). Leverage and Volatility Feedback Effects in High-Frequency Data. *Journal of Financial Econometrics*, 4(3), 353–384.
- Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting.

- Journal of Econometrics*, 192(1), 1–18.
- Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web Search Queries Can Predict Stock Market Volumes. *PLOS ONE*, 7(7), 1–17.
- Borovkova, S., Garmaev, E., Lammers, P., & Rustige, J. (2017). *SenSR: A Sentiment-Based Systemic Risk Indicator* (De Nederlandsche Bank Working Paper No. 553).
- Borovkova, S., & Mahakena, D. (2015). News, volatility and jumps: the case of natural gas futures. *Quantitative Finance*, 15(7), 1217–1242.
- Bühlmann, P., & Van De Geer, S. (2011). *Statistics for high-dimensional data: methods, theory and applications*. Springer Science & Business Media.
- Buncic, D., & Gisler, K. I. (2016). Global equity market volatility spillovers: A broader role for the United States. *International Journal of Forecasting*, 32(4), 1317–1339.
- Campbell, J. Y., & Thompson, S. B. (2008). Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average? *The Review of Financial Studies*, 21(4), 1509–1531.
- Caporin, M., & Poli, F. (2017). Building News Measures from Textual Data and an Application to Volatility Forecasting. *Econometrics*, 5(3), 1–46.
- Christiansen, C., Schmeling, M., & Schrimpf, A. (2012). A comprehensive look at financial volatility prediction by economic variables. *Journal of Applied Econometrics*, 27(6), 956–977.

- Christoffersen, P. (1998). Evaluating Interval Forecasts. *International Economic Review*, 39(4), 841–862.
- Clark, T., & West, K. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics*, 138(1), 291–311.
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- Corsi, F., Audrino, F., & Renò, R. (2012). HAR Modeling for Realized Volatility Forecasting. In L. Bauwens, C. Hafner, & S. Laurent (Eds.), *Handbook of volatility models and their applications* (pp. 363–382). Hoboken (N.J.): John Wiley & Sons, Inc.
- Corsi, F., Pirino, D., & Renò, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2), 276–288.
- Corsi, F., & Renò, R. (2012). Discrete-Time Volatility Forecasting With Persistent Leverage Effect and the Link With Continuous-Time Volatility Modeling. *Journal of Business & Economic Statistics*, 30(3), 368–380.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), 1461–1499.
- Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *The Review of Financial Studies*, 28(1), 1–32.
- Daniel, K., Hirshleifer, D., & Teoh, S. H. (2002). Investor psychology in capital markets: evidence and policy implications. *Journal of*

- Monetary Economics*, 49(1), 139–209.
- Deriu, J., Gonzenbach, M., Uzdilli, F., Lucchi, A., Luca, V. D., & Jaggi, M. (2016). Swisscheese at semeval-2016 task 4: Sentiment classification using an ensemble of convolutional neural networks with distant supervision. In *Proceedings of the 10th international workshop on semantic evaluation* (pp. 1124–1128).
- Deriu, J., Lucchi, A., De Luca, V., Severyn, A., Müller, S., Cieliebak, M., ... Jaggi, M. (2017). Leveraging large amounts of weakly supervised data for multi-language sentiment classification. In *Proceedings of the 26th international conference on world wide web* (pp. 1045–1052).
- Dimpfl, T., & Jank, S. (2015). Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management*, 22(2), 171–192.
- Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics*, 22, 367–381.
- Evans, K. P. (2011). Intraday jumps and US macroeconomic news announcements. *Journal of Banking & Finance*, 35(10), 2511–2527.
- Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417.
- Fengler, M., Mammen, E., & Vogt, M. (2015). Specification and structural break tests for additive models with applications to realized variance data. *Journal of Econometrics*, 188(1), 196–218.

- Gilbert, E., & Karahalios, K. (2010). Widespread worry and the stock market. In *Proceedings of the International Conference on Weblogs and Social Media* (pp. 59–65).
- González-Rivera, G., Lee, T.-H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, 20(4), 629–645.
- Gonçalves, S., & Meddahi, N. (2011). Box–Cox transforms for realized volatility. *Journal of Econometrics*, 160(1), 129–144.
- Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y. (2016). *Deep learning* (Vol. 1). MIT press Cambridge.
- Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, 117, 62–81.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The Model Confidence Set. *Econometrica*, 79(2), 453–497.
- Ho, K.-Y., Shi, Y., & Zhang, Z. (2013). How does news sentiment impact asset volatility? Evidence from long memory and regime-switching approaches. *The North American Journal of Economics and Finance*, 26, 436–456.
- Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Hull, J., Hull, J., White, A., & White, A. (1998). Incorporating Volatility Updating Into the Historical Simulation Method for Value At Risk. *Journal of Risk*, 1, 5–19.

- Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In *Eighth international conference on weblogs and social media (icwsm-14)*. Ann Arbor, MI, June 2014.
- Johnson, E. J., & Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology*, 45(1), 20–31.
- Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27(4), 1116–1127.
- Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives*, 3(2), 73–84.
- Lahaye, J., Laurent, S., & Neely, C. J. (2011). Jumps, cojumps and macro announcements. *Journal of Applied Econometrics*, 26(6), 893–921.
- Lee, S. S., & Mykland, P. A. (2007). Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. *The Review of Financial Studies*, 21(6), 2535–2563.
- Liew, J. K.-S., & Budavári, T. (2016). *Do Tweet Sentiments Still Predict the Stock Market?* (SSRN Working Paper No. 2820269). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2820269
- Long, J. B. D., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Econ-*

- omy, 98(4), 703–738.
- Lustig, H., Roussanov, N., & Verdelhan, A. (2011). Common Risk Factors in Currency Markets. *The Review of Financial Studies*, 24(11), 3731–3777.
- Lustig, H., Roussanov, N., & Verdelhan, A. (2014). Countercyclical currency risk premia. *Journal of Financial Economics*, 111(3), 527–553.
- Mao, H., Counts, S., & Bollen, J. (2011). *Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data* (arXiv preprint arXiv:1112.1051). Retrieved from <http://cds.cern.ch/record/1404941>
- McAleer, M., & Medeiros, M. C. (2008a). A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. *Journal of Econometrics*, 147(1), 104–119.
- McAleer, M., & Medeiros, M. C. (2008b). Realized Volatility: A Review. *Econometric Reviews*, 27(1–3), 10–45.
- Mitnik, S., Robinzonov, N., & Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking & Finance*, 58(Supplement C), 1–14.
- Nakov, P., Ritter, A., Rosenthal, S., Sebastiani, F., & Stoyanov, V. (2016). SemEval-2016 task 4: Sentiment analysis in Twitter. In *Proceedings of the 10th international workshop on semantic evaluation (semeval-2016)* (pp. 1–18).
- Nofer, M., & Hinz, O. (2015, 01). Using Twitter to Predict the Stock Market. *Business & Information Systems Engineering*, 57(4), 229–

242.

- Nonejad, N. (2017). Forecasting aggregate stock market volatility using financial and macroeconomic predictors: Which models forecast best, when and why? *Journal of Empirical Finance*, 42(Supplement C), 131–154.
- Oliveira, N., Cortez, P., & Areal, N. (2013). On the Predictability of Stock Market Behavior Using StockTwits Sentiment and Posting Volume. In L. Correia, L. P. Reis, & J. Cascalho (Eds.), *Progress in artificial intelligence* (pp. 355–365). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Patton, A. J., & Sheppard, K. (2015). Good Volatility, Bad Volatility: Signed Jumps and The Persistence of Volatility. *The Review of Economics and Statistics*, 97(3), 683–697.
- Paye, B. S. (2012). ‘Déjà vol’: Predictive regressions for aggregate stock market volatility using macroeconomic variables. *Journal of Financial Economics*, 106(3), 527–546.
- Porshnev, A., Lakshina, V., & Redkin, I. (2016). *Could Emotional Markers in Twitter Posts Add Information to the Stock Market ARMAX-GARCH Model* (Higher School of Economics Research Paper No. WP BRP 54/FE/2016). Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2763583
- Preis, T., Moat, S. H., & Stanley, E. H. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, 3.
- Racine, J. (2000). Consistent cross-validators model-selection for de-

- pendent data: hv-block cross-validation. *Journal of Econometrics*, 99(1), 39–61.
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, 84, 25–40.
- Saavedra, S., Duch, J., & Uzzi, B. (2011). Tracking Traders’ Understanding of the Market Using e-Communication Data. *PLOS ONE*, 6(10), 1–7.
- Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*, 23(5), 528–543.
- See-To, E. W. K., & Yang, Y. (2017, 01). Market sentiment dispersion and its effects on stock return and volatility. *Electronic Markets*, 27(3), 283–296.
- Siganos, A., Vagenas-Nanos, E., & Verwijmeren, P. (2017). Divergence of sentiment and stock market trading. *Journal of Banking & Finance*, 78, 130–141.
- Steinberger, L. (2016). The relative effects of dimensionality and multiplicity of hypotheses on the F-test in linear regression. *Electron. J. Statist.*, 10(2), 2584–2640.
- Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: A high-frequency perspective. *Journal of Banking & Finance*, 73, 147–164.
- Tibshirani, R. (1994). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58, 267–288.

- Tseng, K. (2006). Behavioral finance, bounded rationality, neuro-finance, and traditional finance. *Investment Management and Financial Innovations*, 3(4), 7–18.
- Tumarkin, R., & Whitelaw, R. F. (2001). News or Noise? Internet Postings and Stock Prices. *Financial Analysts Journal*, 57(3), 41–51.
- Wang, S., & Cui, H. (2013). Generalized F test for high dimensional linear regression coefficients. *Journal of Multivariate Analysis*, 117, 134–149.
- Wang, Y.-H., Keswani, A., & Taylor, S. J. (2006). The relationships between sentiment, returns and volatility. *International Journal of Forecasting*, 22(1), 109–123.
- Welch, I., & Goyal, A. (2008). A Comprehensive Look at The Empirical Performance of Equity Premium Prediction. *The Review of Financial Studies*, 21(4), 1455–1508.
- Zhang, J. L., Härdle, W. K., Chen, C. Y., & Bommers, E. (2016). Distillation of News Flow Into Analysis of Stock Reactions. *Journal of Business & Economic Statistics*, 34(4), 547–563.
- Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through Twitter “I hope it is not as bad as I fear”. *Procedia - Social and Behavioral Sciences*, 26, 55–62.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.
- Zou, H., Hastie, T., & Tibshirani, R. (2007). On the “degrees of freedom” of the lasso. *The Annals of Statistics*, 35(5), 2173–2192.

Appendix

1.A Data

Table 1.15: Description of economic and financial variables

Data	Description
Equity Market variables	
Dividend-Price Ratio Dow Jones	Retrieved from Bloomberg
Dividend-Price Ratio MSCI	Retrieved from Bloomberg
Dividend-Price Ratio stocks	Retrieved from Bloomberg
Earnings-Price Ratio Dow Jones	Retrieved from Bloomberg
Earnings-Price Ratio MSCI	Retrieved from Bloomberg
Earnings-Price Ratio stocks	Retrieved from Bloomberg
Return Dow Jones	Log change in daily closing prices
Return MSCI	Log change in daily closing prices
Return stocks	Log change in daily closing prices
Equity-Market Return (Fama-French)	Retrieved from French's website, log returns.
Small-minus-Big (Fama-French)	Retrieved from French's website, log returns.
High-minus-Low (Fama-French)	Retrieved from French's website, log returns.
Short-term reversal (Fama-French)	Retrieved from French's website, log returns.
CBOE VIX - Level	Simple level (not divided by 100)
CBOE VIX - Change	Log change in the index level.
Bond Market variables	
T-Bill Rate	Annualized yield (not logarithmic).
Relative T-Bill Rate	T-Bill Rate minus 12-m moving average yield.
Long-Term Bond Return	Daily log change in index of 10+ year US gov. bonds.
Relative Bond Yield	Yield of long term gov. bonds (10y) minus 12-month MA.
Term spread	Yield difference between long term gov. Bond and T-Bill.
Bond Risk Premia	Yield difference between BAA Yield and Gov. Yield (10y).
Exchange Rates variables	
Return Spot Exchange Rate - EUR	Log-return of the spot exchange rates.
Return Spot Exchange Rate - CHF	Log-return of the spot exchange rates.
Return Spot Exchange Rate - GBP	Log-return of the spot exchange rates.
Return Spot Exchange Rate - JPY	Log-return of the spot exchange rates.
Carry Trade Factor	Return on high interest rate currencies minus return on low interest rate currencies.
Average Forward Discount	Average 1m interest rate of foreign FX minus USD interest rate.
Liquidity variables	
Default Spread	BAA minus AAA corporate bond yields.
FX Average Bid-Ask Spread	Daily average BAS for CHF, GBP, EUR, JPY exchange rates.
Turnover-Ratio Dow Jones	Daily volume divided by total market capitalization.
Turnover-Ratio Change Dow Jones	Log-change in the turnover-ratio.
Turnover-Ratio MSCI	Daily volume divided by total market capitalization.
Turnover-Ratio Change MSCI	Log-change in the turnover-ratio.
Turnover-Ratio stocks	Daily volume divided by total market capitalization.
Turnover-Ratio Change stocks	Log-change in the turnover-ratio.
TED Spread	3m USD LIBOR minus 3m T-Bill.

NOTE: The table summarizes the economic covariates included in the forecasting analysis. Variables marked with * are available only in monthly frequency. For these variables we computed the daily values by linearly interpolating the logarithmic end-of-month levels. From these approximative values we then computed month-over-month (MoM) or year-over-year (YoY) changes.

Table 1.15 continued: Description of economic and financial variables

Data	Description
Macroeconomic variables	
Inflation Rate MoM*	Monthly log change in seasonally adjusted (SA) CPI.
Inflation Rate YoY*	Yearly log change in not seasonally adjusted (NSA) CPI.
Expected Inflation	Rate at which T-Bill and TIPS achieve same yield (not log).
Expected Inflation Change	First difference in expected inflation.
Industrial Production Growth MoM*	Monthly log-difference in SA Industrial Production.
Industrial Production Growth YoY*	Yearly log-difference in NSA Industrial Production.
Housing Starts*	Monthly log-change in new private housing started.
M1 growth MoM*	Monthly log-change in SA M1 money supply.
M1 growth YoY*	Yearly log-change in NSA M1 money supply.
New Orders growth MoM*	Monthly log-change in SA New Orders.
New Orders growth YoY*	Yearly log-change in NSA New Orders.
Return CRB Spot	Log return on the CRB Index.
Capacity Utilization Level*	Level of SA capacity utilization.
Capacity Utilization change*	Monthly log change in SA capacity utilization.
Employment growth*	Monthly log-change SA employment number.
Consumer sentiment*	Monthly log-change NSA cons. sent. of the University of Michigan.
Consumer confidence*	Monthly log-change NSA cons. confidence index.
Diffusion Index*	Level of SA General Business Activity Diffusion Index.
Chicago PM Business Barometer*	Level of SA Chicago Purchasing Manager Business Barometer.
ISM PMI*	Monthly log-change in SA Purchasing Manager Index.

NOTE: The table summarizes the economic covariates included in the forecasting analysis. Variables marked with * are available only in monthly frequency. For these variables we computed the daily values by linearly interpolating the logarithmic end-of-month levels. From these approximative values we then computed month-over-month (MoM) or year-over-year (YoY) changes.

Table 1.16: Description of sentiment and attention variables

Panel A: Company specific sentiment and attention variables		
Source	Keyword/Topic	Forms
Twitter	• company name and ticker cashtag	• daily average sentiment (smoothed) • daily empirical standard deviation in sentiment (smoothed) • squared daily smoothed average sentiment (centered at zero) • logarithmic volume (total number of short messages posted) • logarithmic change in volume
StockTwits	• company name and ticker cashtag	• daily average sentiment (smoothed) • daily empirical standard deviation in sentiment (smoothed) • squared daily smoothed average sentiment (centered at zero) • logarithmic volume (total number of short messages posted) • logarithmic change in volume
Google Trends	• company name and stock ticker	• daily relative (log) search volume • logarithmic change in search volume
Wikipedia	• company's Wikipedia page	• daily (log) number of accesses to the page • logarithmic change in the daily accessing volume
RavenPack Event Sentiment Score (ESS)	• company specific news articles as identified by RavenPack*	• daily average Event Sentiment Score (smoothed) • daily standard deviation in the ESS (smoothed) • squared daily smoothed average ESS (centered at zero) • logarithmic volume (total number of news articles) • logarithmic change in volume
RavenPack Composite Sentiment Score (CSS)	• news articles mentioning the company**	• daily average Composite Sentiment Score (smoothed) • daily standard deviation in the CSS (smoothed) • squared daily smoothed average CSS (centered at zero)
RavenPack News Impact Projections (NIP)	• news articles mentioning the company**	• daily average News Impact Projections (smoothed) • daily standard deviation in the NIP (smoothed) • squared daily smoothed average NIP (centered at zero)

NOTE: The table summarises the sentiment and attention variables consider in this paper. Panel A describes the company specific variables and Panel B general measures of attention and sentiment. The first column reports the data sources, the second the keywords and/or topics for which data is collected, and the last column describes the different forms of sentiment and attention variables considered in our analysis.

*: in this category we consider only news articles which RavenPack News Analytics identified to be highly relevant for the company under consideration, i.e. the classification approach of RavenPack News Analytics assigned the news article as having a relevance score of 100 (out of 100) for a specific company.

** : in this category we consider all news articles which RavenPack News Analytics identified to be relevant for the company under consideration. Note that, sentiment scores are weighted by the relevance score assigned by the classification approach of RavenPack News Analytics.

Table 1.16 continued: Description of sentiment and attention variables

Panel B: General sentiment and attention variables		
Source	Keyword/Topic	Forms
Twitter	• “stock market”, “stocks” or “wall street” • “interest rate” • “financial markets” • “dow jones”	• daily average sentiment (smoothed) • daily empirical standard deviation in sentiment (smoothed) • squared daily smoothed average sentiment (centered at zero) • logarithmic volume (total number of short messages posted) • logarithmic change in volume
StockTwits	• “stock market”, “stocks” or “wall street” • “interest rate” • “financial markets” • “dow jones”	• daily average sentiment (smoothed) • daily empirical standard deviation in sentiment (smoothed) • squared daily smoothed average sentiment (centered at zero) • logarithmic volume (total number of short messages posted) • logarithmic change in volume
Google Trends	• “stock market”, “stocks” or “wall street” • “interest rate” • “financial markets” • “dow jones”	• daily relative (log) search volume • logarithmic change in search volume
Wikipedia	• Wikipedia page about “stock market”, “stocks” and “wall street” • Wikipedia page about “interest rates” • Wikipedia page about “financial markets” • Wikipedia page about “dow jones”	• daily (log) number of accesses to the page • logarithmic change in the daily accessing volume
RavenPack Event Sentiment Score (ESS)	• news articles about the US economy • news articles about the US dollar	• daily average Event Sentiment Score (smoothed) • daily standard deviation in the ESS (smoothed) • squared daily smoothed average ESS (centered at zero) • logarithmic volume (total number of news articles) • logarithmic change in volume

NOTE: The table summarises the sentiment and attention variables consider in this paper. Panel A describes the company specific variables and Panel B general measures of attention and sentiment. The first column reports the data sources, the second the keywords and/or topics for which data is collected, and the last column describes the different forms of sentiment and attention variables considered in our analysis.

*: in this category we consider only news articles which RavenPack News Analytics identified to be highly relevant for the company under consideration, i.e. the classification approach of RavenPack News Analytics assigned the news article as having a relevance score of 100 (out of 100) for a specific company.

**,: in this category we consider all news articles which RavenPack News Analytics identified to be relevant for the company under consideration. Note that, sentiment scores are weighted by the relevance score assigned by the classification approach of RavenPack News Analytics.

Table 1.17: Average values of stock specific sentiment and attention variables

Panel A: Average raw and smoothed sentiment scores										
	Sentiment					Sentiment smoothed				
	Twt.	StockT.	ESS	CSS	NIP	Twt.	StockT.	ESS	CSS	NIP
Intel	1.008	1.019	1.019	1.001	0.968	1.01	1.021	1.018	1.001	0.97
Microsoft	0.986	1.006	1.036	1.002	0.976	0.986	1.006	1.028	1.002	0.977
Verizon	1	1.002	1.035	1.002	0.971	1.001	1.003	1.032	1.002	0.973
Citigroup	0.998	1.007	1.012	0.999	0.973	0.995	1.009	1.009	1	0.975
BlackRock	1.005	1.01	1.005	1.001	0.947	1.014	1.008	1.005	1.001	0.95
Coca-Cola	1.031	1.011	1.011	1.001	0.977	1.034	1.013	1.013	1.001	0.978
Baxter	1.005	1.013	1.009	1.001	0.98	1.006	1.014	1.009	1.001	0.981
Exxon Mobil	0.989	1.017	1.003	1.001	0.974	0.989	1.019	1.006	1.001	0.975
Gilead Sci.	1	1.001	1	1.001	0.964	1	1.002	1.001	1.001	0.968
Hasbro	1.021	1.006	1.005	1.001	0.98	1.021	1.008	1.005	1.001	0.982
Nike	1.005	1.014	1.002	1.002	0.981	1.006	1.013	1.003	1.002	0.982
Caterpillar	0.964	0.996	1.009	1	0.968	0.966	0.998	1.011	1	0.969
General El.	0.99	1.01	1.05	1.002	0.967	0.99	1.012	1.041	1.001	0.969
Walmart Inc	0.955	1.008	1.003	1.001	0.976	0.956	1.008	1.006	1.001	0.977
Ventas	1.002	1.009	1.004	1.001	0.974	1.001	1.008	1.004	1.001	0.977
A. Schulman	1.002	1.012	1.004	1.002	0.983	1.002	1.008	1.003	1.002	0.986
Parkway	1	1.005	1.007	1.002	0.985	1	1.004	1.006	1.002	0.986
Circor	1	1.003	1.008	1.002	0.994	1	1.004	1.006	1.002	0.994
Dow Jones	0.988	1	—	—	—	0.985	1.001	—	—	—

Panel B: Average squared smoothed sentiment and average smoothed standard deviation of sentiment										
	Squared (smooth.) sent. ($\times 10^4$ /000)					Smooth. standard deviation of sent.				
	Twt.	StockT.	ESS	CSS	NIP	Twt.	StockT.	ESS	CSS	NIP
Intel	1.031	4.96	3.467	0.017	8.867	0.029	0.118	0.024	0.011	0.037
Microsoft	3.126	0.843	8.158	0.052	5.431	0.131	0.134	0.034	0.01	0.038
Verizon	7.091	0.395	13.113	0.03	7.529	0.185	0.121	0.023	0.01	0.039
Citigroup	0.825	0.781	0.91	0.015	6.371	0.059	0.122	0.039	0.014	0.041
BlackRock	2.285	1.603	0.756	0.031	29.106	0.051	0.073	0.009	0.007	0.057
Coca-Cola	14.87	1.867	1.876	0.004	5.124	0.212	0.119	0.018	0.008	0.034
Baxter	1.301	3.204	1.282	0.021	3.706	0.003	0.036	0.011	0.007	0.028
Exxon Mobil	3.229	4.997	2.035	0.008	6.265	0.043	0.122	0.024	0.01	0.038
Gilead Sci.	0.9	2.902	0.736	0.01	10.122	0.039	0.148	0.014	0.007	0.032
Hasbro	8.955	0.821	0.281	0.017	4.251	0.117	0.035	0.007	0.005	0.023
Nike	1.493	2.801	0.517	0.051	3.378	0.154	0.133	0.012	0.008	0.029
Caterpillar	12.094	2.959	2.474	0.013	10.239	0.207	0.142	0.015	0.01	0.034
General El.	2.388	1.866	16.611	0.021	9.919	0.063	0.127	0.037	0.011	0.047
Walmart Inc	25.4	1.548	0.797	0.018	5.45	0.234	0.168	0.023	0.01	0.037
Ventas	0.199	0.771	0.438	0.009	6.163	0.001	0.014	0.007	0.004	0.021
A. Schulman	0.291	0.771	0.101	0.033	2.664	0	0.007	0.003	0.002	0.009
Parkway	0.002	0.604	0.51	0.028	2.729	0	0.004	0.003	0.003	0.009
Circor	0.002	0.66	0.484	0.054	1.281	0	0.002	0.003	0.002	0.007
Dow Jones	2.579	1.688	—	—	—	0.072	0.063	—	—	—

NOTE: The table summarises company-specific sentiment and attention measures considered in this study. More precisely, we report the average over the entire sample of these variables. Panel A reports the average (raw) sentiment scores and the average smoothed sentiment scores. Panel B reports the average value of the squared smoothed sentiment and of the smoothed standard deviation in sentiment. Panel C reports the average (raw) volume and its logarithmic return. Relative search volumes obtained from Google Trends are normalized to start at one. Note that RavenPack News Analytics does not provide informations about the Dow Jones.

Table 1.17 continued: Average values of stock specific sentiment and attention variables

Panel C: Average raw volume and average log return in volume										
	Volume					Log return in volume (in %)				
	Twit.	StockT.	RavenP.	Goog.	Wiki.	Twit	StockT.	RavenP.	Goog.	Wiki.
Intel	2.82	31.591	2.386	1.003	2799.6	7.18	27.964	16.636	3.558	7.233
Microsoft	284.24	70.909	3.82	0.978	8992.9	11.025	27.262	26.187	4.676	7.492
Verizon	81.557	21.178	3.176	0.933	1116.2	6.206	26.622	20.74	0.412	9.545
Citigroup	6.363	36.798	6.236	0.88	1386.7	15.127	36.527	35.154	10.209	10.82
BlackRock	3.68	9.921	1.231	1.05	1087.8	7.093	27.979	11.554	7.521	14.38
Coca-Cola	35.765	14.987	1.564	1.053	1926.5	1.919	24.64	14.464	1.669	8.024
Baxter	0.414	3.44	1.036	0.986	328.3	1.207	18.152	13.987	7.741	16.16
Exxon Mobil	3.777	21.526	1.706	0.875	1610.6	8.034	27.311	17.141	7.248	9.051
Gilead Sci.	3.335	102.2	1.364	1.237	458.8	8.041	31.858	12.81	9.081	12.89
Hasbro	8.224	4.145	0.893	0.805	889.8	4.48	10.375	8.108	-2.114	3.901
Nike	520.19	32.975	1.287	1.115	4446.9	3.175	28.432	15.625	-2.877	9.691
Caterpillar	20.572	21.854	1.934	1.046	1147.5	4.12	26.904	15.302	3.397	9.245
General El.	4.853	19.777	3.166	0.971	3014.2	5.236	25.731	23.073	6.49	9.175
Walmart Inc	256.84	30.646	2.332	0.971	4530.6	1.268	28.545	18.852	-2.923	7.534
Ventas	0.209	1.611	0.592	0.39	2.54	0.621	11.86	6.75	8.385	3.044
A. Schulman	0.172	0.846	0.554	0.542	26.312	0.619	4.965	4.905	10.61	16.03
Parkway	0.127	0.598	0.628	0.545	6.126	0.863	6.541	6.824	8.204	7.292
Circor	0.109	0.551	0.494	0.763	0.84	0.769	6.697	5.806	16.697	3.707
Dow Jones	7.316	5.722	—	1.358	2503.8	17.739	22.583	—	19.533	13.57

NOTE: The table summarises company-specific sentiment and attention measures considered in this study. More precisely, we report the average over the entire sample of these variables. Panel A reports the average (raw) sentiment scores and the average smoothed sentiment scores. Panel B reports the average value of the squared smoothed sentiment and of the smoothed standard deviation in sentiment. Panel C reports the average (raw) volume and its logarithmic return. Relative search volumes obtained from Google Trends are normalized to start at one. Note that RavenPack News Analytics does not provide informations about the Dow Jones.

1.B Robustness checks

The forecasting analysis presented in Section 1.4 depends on some choices that have to be made such as the choice of the rolling window’s length, the adaptive lasso’s tuning parameter, the realized variance estimator, the filtering approach applied to the daily sentiment scores, and the natural language processing approach. In order to assess the sensitivity of the one-day ahead forecast with respect to these parameters and assumptions, we report in Table 1.18 the average out-of-sample R^2 for different choices of the rolling window size, adaptive lasso’s complexity, realized variance estimators, smoothing approaches, and natural language processing methodologies.

In Panel A of Table 1.18 we report the results for rolling windows of length two and a half, two, and one and a half years. Since each of these window lengths results in a different number of out-of-sample observations, we compare the forecasting accuracy only over the time period from June 2014 to December 2016. Beneath each out-of-sample R^2 , we report in parentheses the share of companies for which the more complex model has a significant superior forecasting accuracy. We observe that by changing the window size from our baseline setting (two years), the overall improvement of adding simultaneously economic and sentiment variables is between 6.5% and 7.5%. Compared to the two-year rolling window, we report a slightly lower forecasting accuracy improvement of the Sentiment-HAR compared to the Economic-HAR for the larger and smaller window lengths.

Table 1.18: Average out-of-sample R^2 and share of significant superior forecasting accuracy (in %) for different specifications of hyperparameters and data preparation approaches

Panel A: Forecasting results for different rolling window lengths			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
Window 2.5 years	7.24 (89.47%)	5.64 (89.47%)	1.7 (47.37%)
Window 2 years (baseline)	7.5 (100%)	5.52 (100%)	2.11 (63.16%)
Window 1.5 years	6.57 (100%)	4.67 (94.74%)	2 (73.68%)
Panel B: Forecasting results for different complexity parameters			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
Higher complexity ($0.8 \times \lambda_{min}$)	6.8 (100%)	5.01 (100%)	1.9 (73.68%)
Baseline complexity (λ_{min})	6.8 (100%)	5.02 (100%)	1.89 (63.16%)
Lower complexity ($1.2 \times \lambda_{min}$)	6.67 (100%)	4.95 (100%)	1.82 (57.89%)
Panel C: Forecasting results for different realized volatility estimators			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
Log median RV (baseline)	6.8 (100%)	5.02 (100%)	1.89 (63.16%)
Log bipower RV	6.88 (100%)	5.28 (100%)	1.69 (63.16%)
Log RV	6.54 (100%)	5.29 (94.74%)	1.32 (47.37%)

NOTE: The table reports the average out-of-sample R^2 and in parentheses the share of significant superior forecasting accuracy of the one-day ahead forecast for three different rolling window sizes (Panel A), three specifications of adaptive lasso's complexity (Panel B), three measures of realized volatility (Panel C), five filtering approaches (Panel D), and two natural language processing techniques (Panel E). Panel A reports the results for three different rolling window lengths, namely two and a half years, two years (our baseline setting), and one and a half years. For all three rolling window sizes, we evaluate the forecasting accuracy over the same time period (from June 2014 to December 2016). Panel B reports the results for different specifications of adaptive lasso's tuning parameter. More precisely, we report the results for λ_{min} (minimizer of the blocked cross-validation), the results with a higher complexity (decrease of λ_{min} by 20%) and those with a lower complexity (20% increase in λ_{min}). Panel C reports the average out-of-sample R^2 for three realized volatility estimators: the median realized volatility estimator, the bipower variation estimator and the plain realized volatility estimator (sum of squared 5-minute returns). Panel D reports the prediction results for five smoothing approaches applied to the sentiment variables: no filtering, Kalman filter applied to a random walk plus noise state-space model, Kalman filter applied to an autoregressive state-space model of order one, 5-day moving average and 5-day exponential moving average. Panel E shows the average R^2 values between the models for two natural language processing approaches: the Deep-MLSA technique introduced by Deriu et al. (2016) and the lexicon and rule-based sentiment analysis tool developed by Hutto and Gilbert (2014).

Table 1.18 continued: Average out-of-sample R^2 and share of significant superior forecasting accuracy (in %) for different specifications of hyperparameters and data preparation approaches

Panel D: Forecasting results for different filtering approaches of the sentiment variables			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
No filtering	6.38 (94.74%)	5.02 (100%)	1.44 (57.89%)
Kalman filter RW (baseline)	6.8 (100%)	5.02 (100%)	1.89 (63.16%)
Kalman filter AR(1)	6.53 (100%)	5.02 (100%)	1.6 (68.42%)
Moving average (5 days)	6.62 (100%)	5.02 (100%)	1.68 (68.42%)
Exponential mov. avg. (5 days)	6.63 (100%)	5.02 (100%)	1.7 (68.42%)
Panel E: Forecasting results for different natural language processing techniques			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
Deep-MLSA (baseline)	6.8 (100%)	5.02 (100%)	1.89 (63.16%)
Vader Sentiment Analysis	6.36 (100%)	5.02 (100%)	1.42 (47.37%)

NOTE: The table reports the average out-of-sample R^2 and in parentheses the share of significant superior forecasting accuracy of the one-day ahead forecast for three different rolling window sizes (Panel A), three specifications of adaptive lasso's complexity (Panel B), three measures of realized volatility (Panel C), five filtering approaches (Panel D), and two natural language processing techniques (Panel E). Panel A reports the results for three different rolling window lengths, namely two and a half years, two years (our baseline setting), and one and a half years. For all three rolling window sizes, we evaluate the forecasting accuracy over the same time period (from June 2014 to December 2016). Panel B reports the results for different specifications of adaptive lasso's tuning parameter. More precisely, we report the results for λ_{min} (minimizer of the blocked cross-validation), the results with a higher complexity (decrease of λ_{min} by 20%) and those with a lower complexity (20% increase in λ_{min}). Panel C reports the average out-of-sample R^2 for three realized volatility estimators: the median realized volatility estimator, the bipower variation estimator and the plain realized volatility estimator (sum of squared 5-minute returns). Panel D reports the prediction results for five smoothing approaches applied to the sentiment variables: no filtering, Kalman filter applied to a random walk plus noise state-space model, Kalman filter applied to an autoregressive state-space model of order one, 5-day moving average and 5-day exponential moving average. Panel E shows the average R^2 values between the models for two natural language processing approaches: the Deep-MLSA technique introduced by Deriu et al. (2016) and the lexicon and rule-based sentiment analysis tool developed by Hutto and Gilbert (2014).

Panel B in Table 1.18 reports the forecasting results for small changes to adaptive lasso's tuning parameter. For our analysis, we chose λ to minimize the mean squared cross-validation error (denoted by λ_{min}). We repeat the forecasting exercise for a tuning parameter increased and

decreased by 20%. The average out-of-sample R^2 shows that while a decrease of λ_{min} marginally increases the forecasting accuracy of the Sentiment-HAR model, an increase of adaptive lasso's tuning parameter decreases slightly the average forecasting improvements achieved by the sentiment enhanced HAR model. The changes in average out-of-sample R^2 are, however, small in magnitude. We therefore do not regard the choice of λ_{min} as critical for our results.

Panel C in Table 1.18 compares the forecasting results reported in our paper with those obtained when predicting the log realized volatility estimated by the sum of squared 5-minute returns and by the bipower variation estimator (Barndorff-Nielsen & Shephard, 2004). We observe qualitatively the same forecasting power of social media, news and search engine data when the realized volatility is estimated with the bipower variation estimator. A slightly lower improvement in forecasting accuracy of the Sentiment-HAR model compared to the Economic-HAR model is observed when we estimate realized volatility with the sum of squared 5-minute returns. The overall prediction improvement achieved by social media, news and search engine data in comparison to the baseline HAR model is instead qualitatively the same for all three volatility estimators.

Panel D in Table 1.18 assesses the robustness of our results with respect to the smoothing approach applied to the sentiment measures in our dataset. We observe that, when no smoothing is applied to the sentiment variables, micro-blogging, news and search engine data still have predictive power for tomorrow's realized volatility: on average the Sentiment-HAR model reduces the Economic-HAR model's MSPE by

1.44%. When smoothing sentiment predictors with a simple 5-day moving average or exponential moving average filter, the Sentiment-HAR model achieves an out-of-sample R^2 compared to the Economic-HAR model of 1.68% and 1.7%, respectively. When applying the Kalman filter to an autoregressive state-space model of order one, sentiment and attention variable decrease on average the MSPE by additional 1.6%. These results underline the importance of removing noise from sentiment measures as the smallest prediction improvements are achieved when the sentiment variables are not smoothed. Even when a trivial filter is used (e.g. moving average) we observe an improvement in the predictive accuracy of the Sentiment-HAR model. The difference between the simple 5-day filters and the Kalman filter applied to a random walk plus noise state-space model (baseline) might be attributed to the fact that for the latter, the degree of smoothing is determined by the data.

Finally, Panel E in Table 1.18 reports the average R^2 values between the analysed models for two natural language processing approaches: the Deep-MLSA technique introduced by Deriu et al. (2016) and the lexicon and rule-based sentiment analysis tool developed by Hutto and Gilbert (2014). Even though our results showed that attention measures are the most relevant predictors for realized volatility, the natural language processing methodology applied to the social media data has an impact on the results. The average decreases in MSPE achieved by the Sentiment-HAR compared to the Economic-HAR model amounts to 1.42% when the sentiment scores of social media messages are computed with a lexicon and rule-based approach. In contrast, when processing micro-

blogging data with the Deep-MLSA technique (baseline), the R^2 between Sentiment-HAR and Economic-HAR amounts to 1.89%.

1.C Supplementary tables

1.C.1 In-sample

Table 1.19: Goodness-of-fit-measures for the Sentiment-HAR, Economic-HAR and baseline HAR models estimated over the entire sample

	Sentiment-HAR			Economic-HAR			Baseline HAR		
	R^2	BIC	AIC	R^2	BIC	AIC	R^2	BIC	AIC
Intel	45.76	1.3	1.25	44.75	1.35	1.28	39.24	1.4	1.38
Microsoft	44.82	1.34	1.25	40.26	1.38	1.33	35.66	1.44	1.42
Verizon	37.22	1.37	1.24	29.55	1.37	1.34	26.07	1.42	1.4
Citigroup	54.85	1.29	1.24	53.73	1.31	1.27	49.56	1.39	1.38
BlackRock	49.76	1.35	1.2	43.24	1.32	1.3	39.39	1.41	1.39
Coca-Cola	40.53	1.34	1.29	40.01	1.33	1.29	36.61	1.37	1.36
Baxter	44.78	1.33	1.25	45.09	1.34	1.25	39.33	1.36	1.35
Exxon Mobil	60.41	1.35	1.29	59.74	1.37	1.31	55.36	1.45	1.44
Gilead Sci.	46.28	1.34	1.24	43.03	1.33	1.29	40.81	1.34	1.33
Hasbro	35.87	1.25	1.19	35.3	1.26	1.2	28.36	1.33	1.31
Nike	46.31	1.37	1.29	43.19	1.4	1.35	37.64	1.49	1.47
Caterpillar	52.51	1.33	1.28	51.82	1.34	1.3	48.43	1.4	1.38
General El.	50.4	1.27	1.2	48.88	1.29	1.23	43.32	1.36	1.35
Walmart Inc	37.59	1.27	1.23	35.83	1.29	1.26	33.7	1.31	1.29
Ventas	44.08	1.27	1.23	42.12	1.29	1.26	40.76	1.31	1.29
A. Schulman	38.48	1.26	1.18	34.24	1.29	1.25	30.22	1.33	1.32
Parkway	29.12	1.3	1.21	27.91	1.3	1.23	22.19	1.32	1.3
Circor	29.14	1.27	1.2	26.41	1.27	1.23	22.36	1.3	1.28
Dow Jones	61.52	1.28	1.21	61.36	1.29	1.22	55.55	1.39	1.38

NOTE: The table reports the regression R^2 , the Bayes Information Criterion (BIC) and the Akaike Information Criterion for the Sentiment-HAR, Economic-HAR (both estimated with adaptive lasso) and the baseline HAR (estimated with least squares). Note that the degrees of freedom for Lasso correspond to the expected number of non-zero estimates (Zou et al., 2007); a simple unbiased estimator of the degrees of freedom is the number of estimated non-zero coefficients (Bühlmann & Van De Geer, 2011). We follow the definition of Zou et al. (2007) for the information criteria of lasso.

1.C.2 Forecasting

Table 1.20: Mean squared prediction error and model confidence set p-values for the forecasting results at three different horizons

Panel A: One-day ahead forecasts						
	Sentiment- HAR	MSPE		MCS p-values (in %)		
		Economic- HAR	Baseline HAR	Sentiment- HAR	Economic- HAR	Baseline HAR
Intel	0.064	0.065	0.069	100	9.8	7.8
Microsoft	0.066	0.07	0.074	100	0	0
Verizon	0.066	0.069	0.073	100	0.4	0
Citigroup	0.068	0.07	0.076	100	4.8	0
BlackRock	0.078	0.079	0.085	100	25.4	7.6
Coca-Cola	0.058	0.058	0.062	100	85.8	20.4
Baxter	0.067	0.067	0.071	100	70.8	1.4
Exxon Mobil	0.054	0.056	0.058	100	3.4	3.4
Gilead Sci.	0.076	0.077	0.078	100	43	43
Hasbro	0.083	0.083	0.088	100	30.6	5
Nike	0.059	0.062	0.067	100	2.2	0.2
Caterpillar	0.067	0.067	0.069	100	76.4	11.8
General El.	0.068	0.069	0.072	100	9.6	9.6
Walmart Inc	0.07	0.07	0.074	57.4	100	10.2
Ventas	0.06	0.062	0.063	100	4.6	4.6
A. Schulman	0.182	0.191	0.197	100	3.4	1.6
Parkway	0.082	0.082	0.084	19.8	100	19.8
Circor	0.206	0.207	0.217	100	74.4	0.8
Dow Jones	0.061	0.063	0.07	100	0.4	0.4

NOTE: The table reports the mean squared prediction error and the corresponding model confidence set p-values (Hansen et al., 2011) for the three analysed models at three different forecasting horizons (one day, one week and one month). More precisely, the last three columns in each panel report the MCS p -values as defined by Hansen et al. (2011) for a model confidence set constructed with the test statistic $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{ij}|$, where t_{ij} is the relative (scaled) sample loss between the i th and the j th model and $\mathcal{M}$ is a set of candidate models. Models included in the 90%-confidence set are highlighted in bold.

Table 1.20 continued: Mean squared prediction error and model confidence set p-values for the forecasting results at three different horizons

Panel B: One-week ahead forecasts						
	MSPE			MCS p-values (in %)		
	Sentiment- HAR	Economic- HAR	Baseline HAR	Sentiment- HAR	Economic- HAR	Baseline HAR
Intel	0.044	0.045	0.047	100	71.6	19.6
Microsoft	0.046	0.043	0.049	1	100	1.4
Verizon	0.036	0.035	0.039	19.8	100	0
Citigroup	0.048	0.047	0.051	26.6	100	2.2
BlackRock	0.055	0.053	0.057	3.6	100	1.8
Coca-Cola	0.036	0.037	0.036	100	69.2	96.6
Baxter	0.044	0.045	0.046	100	82.6	63.8
Exxon Mobil	0.047	0.047	0.046	35.4	44	100
Gilead Sci.	0.053	0.053	0.053	99.4	99.4	100
Hasbro	0.045	0.046	0.049	100	45.6	45.6
Nike	0.041	0.04	0.041	25	100	47.8
Caterpillar	0.042	0.04	0.041	11.8	100	24.2
General El.	0.05	0.051	0.054	100	54.2	23
Walmart Inc	0.04	0.041	0.041	100	65.2	65.2
Ventas	0.037	0.036	0.035	16.6	37.2	100
A. Schulman	0.094	0.094	0.099	100	86.8	17.2
Parkway	0.039	0.039	0.04	74.2	100	74.2
Circor	0.095	0.092	0.097	0	100	0
Dow Jones	0.061	0.06	0.069	63.6	100	2.2
Panel C: One-month ahead forecasts						
	MSPE			MCS p-values (in %)		
	Sentiment- HAR	Economic- HAR	Baseline HAR	Sentiment- HAR	Economic- HAR	Baseline HAR
Intel	0.043	0.046	0.041	65.2	59.2	100
Microsoft	0.039	0.038	0.04	78.6	100	78.6
Verizon	0.034	0.034	0.025	1	0.4	100
Citigroup	0.051	0.051	0.045	15.2	3	100
BlackRock	0.065	0.065	0.05	3.4	3.4	100
Coca-Cola	0.037	0.034	0.027	12.8	28.2	100
Baxter	0.048	0.035	0.039	0.4	100	28
Exxon Mobil	0.059	0.052	0.049	3.8	11	100
Gilead Sci.	0.05	0.06	0.051	100	5.8	65.2
Hasbro	0.055	0.065	0.042	26.2	23.8	100
Nike	0.045	0.039	0.037	9.8	33.6	100
Caterpillar	0.04	0.034	0.041	15	100	15
General El.	0.045	0.048	0.051	100	59.2	42
Walmart Inc	0.035	0.033	0.032	54.8	86.4	100
Ventas	0.031	0.044	0.025	10	10	100
A. Schulman	0.058	0.058	0.071	81	100	9.4
Parkway	0.046	0.032	0.029	22.6	46.4	100
Circor	0.054	0.05	0.054	13.8	100	14.6
Dow Jones	0.091	0.068	0.074	30.2	100	49.2

NOTE: The table reports the mean squared prediction error and the corresponding model confidence set p-values (Hansen et al., 2011) for the three analysed models at three different forecasting horizons (one day, one week and one month). More precisely, the last three columns in each panel report the MCS p -values as defined by Hansen et al. (2011) for a model confidence set constructed with the test statistic $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{ij}|$, where t_{ij} is the relative (scaled) sample loss between the i th and the j th model and $\mathcal{M}$ is a set of candidate models. Models included in the 90%-confidence set are highlighted in bold.

Table 1.21: Out-of-sample R^2 for comparing the MSPE of the three different HAR models for days with returns below the historical 2.5%-quantile or above the 97.5%-quantile (1-day ahead forecasts based on a rolling window over the time period 01.01.2014 - 31.12.2016)

	Sentiment/Baseline HAR	Economic/Baseline HAR	Sentiment/Economic HAR
Intel	6.76	-1.33	7.99
Microsoft	18.33	6.18	12.94
Verizon	8.15	7.63	0.56
Citigroup	29.6	29.92	-0.45
BlackRock	5.88	10.84	-5.56
Coca-Cola	19.48	14.49	5.84
Baxter	12.39	16.67	-5.14
Exxon Mobil	20.72	16.09	5.52
Gilead Sciences	10.8	9.42	1.53
Hasbro	13.28	14.38	-1.29
Nike	31.93	23.21	11.35
Caterpillar	-2.83	1.44	-4.32
General Electric	3.64	2.06	1.62
Walmart Inc	12.26	12.59	-0.37
Ventas	5.37	2.4	3.04
A. Schulman	22.77	3.46	20
Parkway	-3.3	-1.61	-1.66
Circor	12.17	9.07	3.41
Dow Jones	20.39	20.46	-0.09
Average R^2	13.04	10.39	2.89

NOTE: The table reports the out-of-sample R^2 as reported in Campbell and Thompson (2008), i.e. $R^2_{i,j} = 1 - MSPE_i/MSPE_j$, for days where a company's return was below its historical 2.5%-quantile or above its 97.5%-quantile. The results are all reported in percent and can be interpreted as the increase (or decrease) in forecasting accuracy of a model compared to a benchmark model on days with particularly negative or positive returns. In the first and second columns we report the decrease in MSPE achieved by the Sentiment-HAR and Economic-HAR models compared to the baseline HAR model, respectively. The last column reports the decrease (increase) in MSPE from the Economic- to the Sentiment-HAR model.

Table 1.22: Average one-day ahead out-of-sample R^2 for sub-groups of companies

Panel A: Avg. R^2 (in %) of the five companies with the lowest/highest insti. ownership			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
Lowest insti. ownership	7	4.82	2.29
Highest insti. ownership	4.99	3.64	1.39
Panel B: Avg. R^2 (in %) of the five companies with the lowest/highest market cap.			
	Sentiment/ Baseline HAR	Economic/ Baseline HAR	Sentiment/ Economic HAR
Lowest market cap.	4.99	3.64	1.39
Highest market cap.	6.73	4.76	2.07

NOTE: The table reports the average out-of-sample R^2 for one-day ahead predictions grouped by the size of the companies and their share of institutional investors. Panel A reports the average reductions in MSPE for the five companies with the lowest and highest share of institutional investors. Panel B shows the average out-of-sample R^2 of the five companies with the lowest and highest market capitalization.

1.C.3 Risk management application

Table 1.23: Average asymmetric value-at-risk losses and model confidence set p-values

Panel A: 0.5%-Value-at-risk						
	Average loss ($\times 10^4$ 000)			MCS p-values (in %)		
	Sentiment- HAR	Economic- HAR	Baseline HAR	Sentiment- HAR	Economic- HAR	Baseline HAR
Intel	3.22	3.39	3.3	100	13	38.6
Microsoft	3.27	3.52	3.5	100	9.8	9.8
Verizon	1.96	2.02	1.94	44.8	5.4	100
Citigroup	3.32	3.36	3.56	100	32.8	33.2
BlackRock	2.87	2.88	2.85	55.8	48.6	100
Coca-Cola	2.77	2.81	2.73	37.6	7.8	100
Baxter	2.49	2.52	2.49	97.2	75.4	100
Exxon Mobil	2.68	2.7	2.72	100	51.2	57.4
Gilead Sci.	5.6	5.7	5.78	100	49.6	49.6
Hasbro	3.53	3.59	3.7	100	1.8	23.8
Nike	2.15	2.31	2.31	100	27.4	27.4
Caterpillar	3.65	3.67	3.74	100	71.4	13.6
General El.	2.06	2.07	2.16	100	79.2	32.4
Walmart Inc	3.09	3.12	3.04	52.4	43.8	100
Ventas	3.2	3.27	3.27	100	23.8	48.8
A. Schulman	10.44	10.66	10.63	100	73.2	73.2
Parkway	2.7	2.74	2.93	100	51.2	41.2
Circor	6.29	6.41	6.57	100	21.2	12.2
Dow Jones	1.52	1.53	1.51	99.8	91.8	100

NOTE: The table reports the average asymmetric losses as defined by González-Rivera et al. (2004) and the corresponding model confidence set p-values (Hansen et al., 2011) for the three analysed models at three different value-at-risk levels (0.5%, 1% and 5%). More precisely, the last three columns in each panel report the MCS p -values as defined by Hansen et al. (2011) for a model confidence set constructed with the test statistic $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{ij}|$, where t_{ij} is the relative (scaled) sample loss between the i th and the j th model and $\mathcal{M}$ is a set of candidate models. Models included in the 90%-confidence set are highlighted in bold.

Table 1.23 continued: Average asymmetric value-at-risk losses and model confidence set p-values

Panel B: 1%-Value-at-risk						
	Average loss ($\times 10^4$)			MCS p-values (in %)		
	Sentiment- HAR	Economic- HAR	Baseline HAR	Sentiment- HAR	Economic- HAR	Baseline HAR
Intel	5.5	5.64	5.48	88.8	26.4	100
Microsoft	5.33	5.63	5.6	100	5.8	5.8
Verizon	3.54	3.57	3.44	24.4	24.4	100
Citigroup	5.39	5.39	5.67	100	98.6	49.8
BlackRock	4.92	4.95	4.98	100	75	75
Coca-Cola	4.14	4.22	4.16	100	4.8	71
Baxter	4.13	4.14	4.11	97	97	100
Exxon Mobil	4.12	4.2	4.26	100	32.6	48.4
Gilead Sci.	8.52	8.67	8.81	100	37.8	22.4
Hasbro	5.56	5.63	6	100	23.4	12.2
Nike	3.91	4.02	3.88	80	36.8	100
Caterpillar	6.13	6.13	6.12	98.2	98.2	100
General El.	3.36	3.34	3.41	49.8	100	49.8
Walmart Inc	4.76	4.84	4.85	100	56	56
Ventas	5.6	5.59	5.59	98.4	99.8	100
A. Schulman	13.64	13.91	13.97	100	48.4	48.4
Parkway	4.7	4.83	5.05	100	12.6	12.6
Circor	9.8	10.03	10.05	100	21.8	21.8
Dow Jones	2.61	2.64	2.66	100	65.2	65.2
Panel C: 5%-Value-at-risk						
	Average loss ($\times 10^4$)			MCS p-values (in %)		
	Sentiment- HAR	Economic- HAR	Baseline HAR	Sentiment- HAR	Economic- HAR	Baseline HAR
Intel	16.45	16.6	16.76	100	11.4	8.8
Microsoft	15.73	16	16.24	100	7.6	7.6
Verizon	11.29	11.31	11.29	94.4	92	100
Citigroup	17.93	17.79	18.52	11.8	100	8
BlackRock	16.08	16.02	16.07	68.6	100	81.8
Coca-Cola	11.57	11.62	11.65	100	39.6	44
Baxter	13.26	13.27	13.64	100	89.8	12.8
Exxon Mobil	14.13	14.16	14.18	100	90.4	90.4
Gilead Sci.	22.92	22.98	23.05	100	83	83
Hasbro	16.57	16.72	16.98	100	13.6	13.6
Nike	13.65	13.81	13.86	100	41	41
Caterpillar	17.66	17.61	17.61	57.6	99	100
General El.	11.51	11.41	11.75	6.2	100	6.2
Walmart Inc	13.48	13.45	13.55	78.6	100	78.6
Ventas	17.49	17.53	17.76	100	68.4	22
A. Schulman	30.48	30.63	30.99	100	63	46.8
Parkway	16.08	16.13	16.15	100	82.6	82.6
Circor	27.86	27.94	28.04	100	54.6	54.6
Dow Jones	9.19	9.19	9.31	94.2	100	85

NOTE: The table reports the average asymmetric losses as defined by González-Rivera et al. (2004) and the corresponding model confidence set p-values (Hansen et al., 2011) for the three analysed models at three different value-at-risk levels (0.5%, 1% and 5%). More precisely, the last three columns in each panel report the MCS p -values as defined by Hansen et al. (2011) for a model confidence set constructed with the test statistic $T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} |t_{ij}|$, where t_{ij} is the relative (scaled) sample loss between the i th and the j th model and $\mathcal{M}$ is a set of candidate models. Models included in the 90%-confidence set are highlighted in bold.

Chapter 2

When does attention matter? The effect of investor attention on stock market volatility around news releases

Co-authors: Francesco Audrino, Fabio Sigris

Abstract

We empirically investigate how retail and institutional investor attention is related to the way stock markets process information. With a focus on 360 US stocks in the S&P 500 universe, our results show that higher investors' attention around news releases is related to higher contemporaneous volatility. Further, retail investor attention increases the post-announcement volatility, whereas institutional investor attention has a small but negative impact on volatility on days following news releases on average over the cross-section of companies. These findings are in line with the hypotheses that attention of retail investors slows price-adjustments to new information and attention of institutional investors results in the opposite reaction. We show that these effects are heterogeneous in the type of news and the topic of the information being released. A portfolio allocation application highlights that these results are not only statistically significant but also sizeable in economic terms and can lead to an overperformance as large as dozens of basis points.

Keywords: limited attention; realized volatility; investor attention; retail investors; institutional investors; StockTwits; Twitter; news releases

2.1 Introduction

In today's world, we are flooded with information of all kinds but have only limited time and resources to properly digest it. In particular, investors have to decide where to allocate their limited attention and, thus, not all information receives the same amount of dedication (Kahneman, 1973). Empirical results presented in the prior literature show that high levels of attention towards a company increase its stock price volatility and trading volume (Antweiler & Frank, 2004; Audrino, Sigrist, & Ballinari, 2020; Dimpfl & Jank, 2016; Goddard, Kita, & Wang, 2015; Hamid & Heiden, 2015; Rakowski, Shirley, & Stark, 2018; Vlastakis & Markellos, 2012, among others).

A possible explanation for these empirical findings is the *limited attention hypothesis*. Theoretical models along the lines of Peng (2005), Peng and Xiong (2006), DellaVigna and Pollet (2009), Hirshleifer, Lim, and Teoh (2009), Andrei and Hasler (2014) postulate that limited attention towards specific events reduces the speed at which the new information is properly interpreted and incorporated in stock prices. While standard asset-pricing models assume that new information is instantaneously reflected in stock prices, these theoretical models assume that attention is needed for the information to be distilled and incorporated in the market value of a company. In the literature, this mechanism is often referred to as the limited attention hypothesis. Based on the implication of these theoretical models, we first investigate whether higher levels of attention around news releases are associated with stronger price reactions:

Hypothesis 1 *Higher attention around firm-specific news releases is related to higher contemporaneous volatility.*

Along the lines of the model introduced by Andrei and Hasler (2014), we investigate this hypothesis by studying the contemporaneous relation between daily volatility and investors' attention around news releases.

Theoretical behavioral market models broadly distinguish between rational well informed institutional investors and retail investors (Barberis, Shleifer, & Vishny, 1998; Black, 1986; De Long, Shleifer, Summers, & Waldmann, 1990; Shleifer & Vishny, 1997). Further, prior empirical research suggests that stock returns react differently to institutional and retail investor attention (Ben-Rephael, Carlin, Da, & Israelsen, in press; Ben-Rephael, Da, & Israelsen, 2017; Da, Engelberg, & Gao, 2011) and also that institutional and retail investors allocate attention around earnings and macroeconomic news announcements differently (Hirshleifer & Sheng, 2019; Liu, Peng, & Tang, 2019). Concerning the impact of attention on volatility, it is thus likely that it makes a difference whether retail or institutional investors are paying attention to information releases. In particular, we expect that higher retail investor attention around news releases also increases the volatility on days following the publication of news. The motivation for this hypothesis is twofold: First, retail investors paying attention to a news release are more likely to wrongly interpret the new information and, therefore, create disagreement in the market and reduce the speed of price-adjustment. Second, their

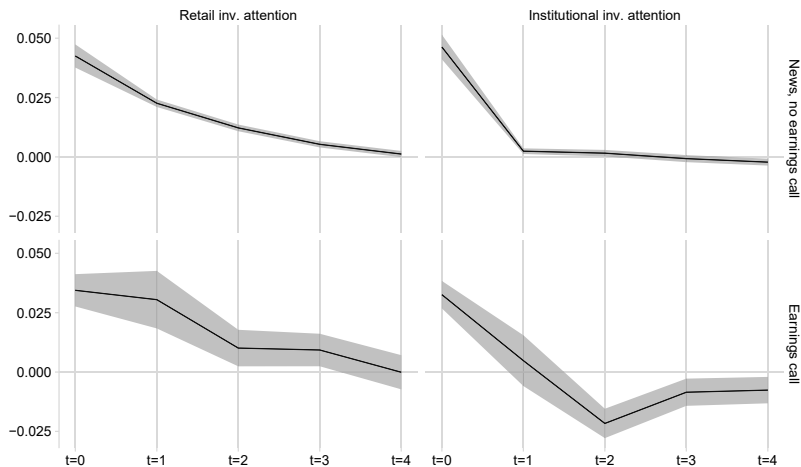


Figure 2.1: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around news releases

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news event day and the four post-release days. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

noise-based trading activity adds additional risk to the market.¹ Further, we expect institutional investors' attention to have the opposite effect, i.e. increase the speed at which stock prices adjust to the new information and therefore reduce (or at least not increase) the return volatility on days following news releases. In summary, the second hypothesis that we analyze is given by:

Hypothesis 2 *Retail investors' attention around firm-specific news releases increases the return volatility on days following the publication of news. Institutional investors' attention has the opposite effect.*

¹Models along the lines of De Long et al. (1990) postulate that noise traders add additional risk to the market since sophisticated investors cannot anticipate how the irrational beliefs of these traders change in the future.

Figure 2.1 summarizes our main findings concerning the first two hypotheses. The figure depicts the change in volatility on the day of a news event ($t = 0$) and on the following four days ($t = 1, \dots, 4$) associated with a one standard deviation increase in retail and institutional attention around the firm-specific news release.² Further, we distinguish between earnings announcements and other types of firm-specific news. Both retail and institutional attention are significantly positively related to the stock return volatility observed on days with news releases. In other words, when investors are paying attention to a news event, immediate price reactions are stronger as postulated by the limited attention hypothesis. Higher levels of retail attention around news events are also associated with increased volatility on days after news releases. In contrast, we find no evidence of a positive relation between the attention of institutional investors to news releases and the return volatility in the post-announcement period. In fact, our results show that higher attention levels of institutional investors around earnings calls are associated with lower volatility levels two days after the event. These results confirm our second hypothesis and show that when retail investors pay attention to news releases, price-adjustments are slower and the risk in the stock market is higher. The attention of institutional investors, on the other hand, increases the speed at which prices adjust to the new information and dampens the volatility on days following news releases.

²We use realized volatility measures based on five minute intraday returns and measure retail investors' attention through traders' social media activity on StockTwits and institutional attention with Bloomberg's *News Heat* measure, an indicator of firm-specific news articles' retrievals from Bloomberg-terminals. More details about the attention measures and the regression framework are given in Section 2.4.

The results depicted in Figure 2.1 roughly distinguish between earnings announcements and all other news releases. However, it is intuitively plausible that the reaction of investors to different news types and topics is not homogeneous. As an illustrative example, investors' attention to a news article about corporate responsibility and to one about a potential merger is most likely not associated with the same post-announcement price reactions. For instance, when McDonald's Corp. announced on July 19, 2016 that they were going to sponsor the launch of the popular mobile game Pokémon Go in Japan, the post-event volatility was low even though retail investors' attention to this news release was very high (social media activity was above its historical 97% quantile) and the contemporaneous volatility spiked (volatility level was above its historical 86% quantile). On the other hand, when the retail company Lowe's Cos. Inc. announced on February 3, 2016 the acquisition of the Canadian home improvement chain Rona Inc. for 2.3 billion US dollars, retail investors' attention was high (the volume of messages shared about the company was above its historical 99% quantile), and the stock price adjusted not only on the announcement day (volatility above its historical 99% quantile), but also over the post-release period (on the following two trading days the volatility was above its historical 96% quantile). Our third hypothesis of interest is therefore:

Hypothesis 3 *The relation between retail investors' attention to news releases and the post-release volatility is heterogeneous in the type and topic of the firm-specific news release.*

Our findings show that higher levels of retail investors' attention are more strongly associated with higher post-release volatility when the news release is unexpected (e.g. "flash news") or when the new information potentially affects a company's market value (e.g. news about mergers and acquisitions, revenues or earnings). The novelty and complexity of new information increase the difficulty for retail investors to correctly interpret the implications for a company's market value and hence reduce the speed at which prices adjust. In contrast, we do not find such a heterogeneous effect for institutional investors' attention, i.e. higher institutional investors' attention to news releases leads to lower volatility on days following news releases in a homogeneous way for different news types and topics. Figure 2.2 depicts in exemplary fashion our findings concerning Hypothesis 3. The figure shows the volatility change associated with a one standard deviation increase in retail and institutional investors' attention around news releases about mergers and acquisitions, corporate responsibility and news flashes. When retail investors are paying attention to news flashes or information releases about mergers and acquisitions, volatility increases in the post-release period. Instead, institutional investors' attention to these news releases is associated with lower volatility on the days after the announcement. For news releases about corporate responsibility, neither retail nor institutional investors' attention is significantly associated with changes in volatility.

Finally, we apply our findings to generate out-of-sample predictions and measure the economic value gained from this. More precisely, we take advantage of the relation between retail investors' attention and

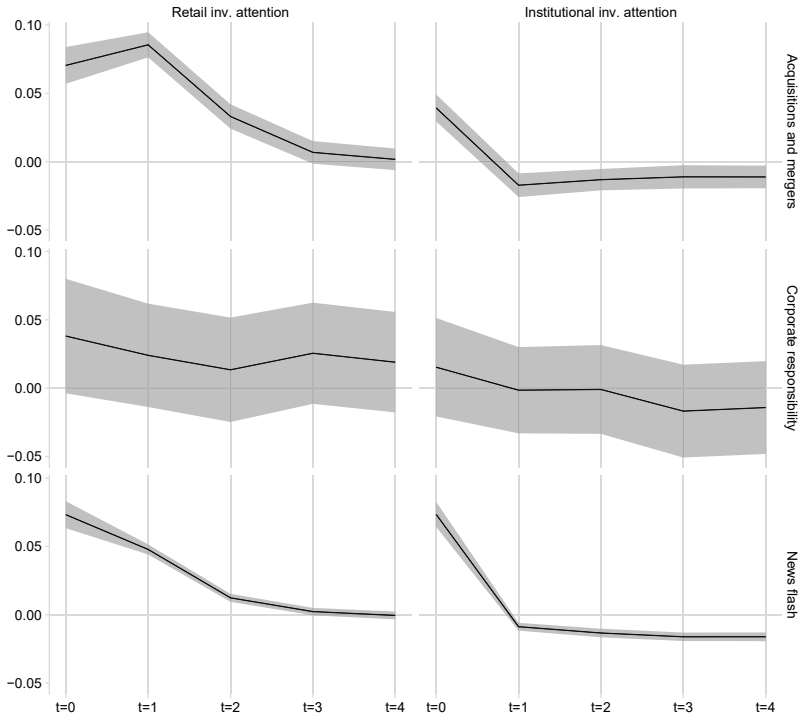


Figure 2.2: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around news releases about mergers and acquisitions, corporate responsibility and news flashes

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news event day and the four post-release days. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

the increased post-announcement volatility to improve one-day ahead volatility forecasts. The improvements thus obtained are not only statistically significant but also economically meaningful. For example, in order to obtain volatility predictions which take into account retail investors' attention to news flashes and news about acquisitions and mergers, an investor with mean-variance preferences is willing to give up 57 and 45 basis points per year of her wealth, respectively.

We contribute to the existing literature in several ways. First, we contribute to the literature concerned with the limited attention hypothesis by studying its implications for a wider range of news releases and using a novel direct measure of investors' attention constructed from social media data. Moreover, we investigate the limited attention hypothesis from a new perspective by analyzing the relation between attention to news and stock market volatility rather than returns. Furthermore, we extend the analysis of the limited attention hypothesis by taking into account the type of investors. In particular, to the best of our knowledge, this is the first study that empirically untangles the role played by retail and institutional investors' attention in digesting a wide and heterogeneous range of information releases. In this respect, our results contribute to the existing empirical and theoretical behavioral literature by providing evidence that while the attention level of sophisticated investors fosters the price-adjustment, noise traders' attention slows this process down and increases post-announcement market risk. Moreover, we contribute to the literature concerned with predicting volatility by providing evidence that volatility forecasts can be significantly improved based on our findings. Finally, we extend the approach to economically evaluating volatility predictions introduced by Bollerslev, Hood, Huss, and Pedersen (2018) to a multivariate setting and by optimally including transaction costs.

The remainder of this paper is organized as follows. In Section 2.2, we review the related literature. Section 2.3 describes our methodology. In Section 2.4 we introduce the dataset and in particular our measures of attention and the identification of news articles. In Section 2.5 we

present the empirical analysis of our hypotheses, and in Section 2.6 we use our findings to predict volatility and improve the asset allocation of a representative investor. Lastly, Section 2.7 offers some concluding remarks.

2.2 Related literature

Building upon the implications of the limited attention hypothesis developed in theoretical models (e.g. Andrei & Hasler, 2014; DellaVigna & Pollet, 2009; Hirshleifer et al., 2009; Peng, 2005; Peng & Xiong, 2006), several empirical studies have examined the impact of attention on stock price reactions to news releases. Hirshleifer et al. (2009) show that price and volume reactions to earnings announcements are weaker when investors are distracted, i.e. attention is low, where distraction is measured as the number of concurrent earnings releases. Similarly, DellaVigna and Pollet (2009) compare the price reactions observed during and after earnings announcements made on Fridays with those observed when the announcement is made on other week days. Their results show that the immediate price reaction to the announcement is weaker when the earnings call happens on a Friday, since investors are generally more distracted on the last working day of the week. Following the implications of their theoretical model, Andrei and Hasler (2014) provide empirical evidence that stock market volatility and a measure of investors' attention constructed from Google web search queries are contemporaneously positively related. Ben-Rephael et al. (2017) analyze the effect of in-

stitutional and retail investor attention on price adjustments observed around earnings announcements and analyst recommendation changes. The study of Ben-Rephael et al. (2017) is closely related to our paper, in that they investigate the implications of attention on stock price reactions to earnings announcements and analyst recommendation changes, distinguishing between retail and institutional investors. They find that institutional attention increases the immediate reaction to the news and reduces the post-announcement drift in the stock price. While their results provide limited evidence that retail investors' attention reduces the price-adjustment speed in the post-announcement period after earnings calls, they find no such evidence for analyst recommendation changes. Our study differs in several aspects. First, we analyze reactions in volatility rather than returns. This approach has several advantages; e.g. our measure of volatility better reflects the trading activity of both retail and institutional investors (see Section 2.4.3). Moreover, by analyzing the reaction in volatility we are able to investigate the implications of limited attention on a wider range of news, since we do not need to know whether the news is positive or negative. Finally, we consider a different measure of retail investors' attention (see Section 2.4.1). For a recent survey of behavioral inattention in the field of economics and finance, we refer to Gabaix (2019).

Besides the measures of attention used in the above-mentioned studies, several other proxies have been proposed in the empirical literature. In their pioneering work, Antweiler and Frank (2004) proxy attention with the posting volume on Yahoo! Finance message boards and report the

predictive power of their measure for the volatility of 45 US stocks. Barber and Odean (2008) use abnormal trading volumes, extreme returns and the release of news as proxies for investors' awareness of a company. Studying the effect of attention on a company's stock price around earnings calls, Aboody, Lehavy, and Trueman (2010) use the 12-month return preceding the announcement. Rubin and Rubin (2010) use the frequency with which a company's Wikipedia article is edited as a proxy of attention towards changes in firm-specific information. More recently, a series of studies used the search-traffic on the Electronic Data Gathering, Analysis, and Retrieval (EDGAR) system to measure investors' attention (see, among others, Drake, Roulstone, & Thornock, 2015; Lee, Ma, & Wang, 2015).

Several past studies have used direct attention measures for volatility prediction. In particular, following Da et al. (2011), several studies have used Google web search queries to proxy for investors' attention. Using web search queries for company names and the S&P 500 Index, Vlastakis and Markellos (2012) analyze the effect of information demand on stock market volatility. Their results show that Google searches at the market level have predictive power for future volatility and trading volume. Hamid and Heiden (2015) report predictive power for attention measures constructed from Google search queries for the keyword "dow" for the weekly volatility of the Dow Jones Industrial Average Index. Using a similar proxy, Dimpfl and Jank (2016) augment the classical HAR model for realized volatility with an attention measure and also report the predictive power for the Dow Jones Industrial Average Index at the daily frequency. Furthermore, several studies have proposed

using Google searches for stock tickers as a proxy of retail investors' attention (e.g. Ben-Rephael et al., in press, 2017; Da et al., 2011; Drake, Roulstone, & Thornock, 2012). In our opinion, it is questionable whether retail investors gather information about stocks by searching for their ticker. For stock market volatility, most empirical studies document an effect of general market-related web search queries (see, among others, Dimpfl & Jank, 2016), whereas the impact of company-specific Google searches is limited (see e.g. Vlastakis & Markellos, 2012). Comparing the forecasting power of different sentiment and attention measures for return volatility, the findings of Audrino et al. (2020) show that the most relevant predictors are indeed company-specific attention measures constructed from social media data and general market attention proxies obtained from web search query data. Moreover, Google does not provide the absolute number of web searches, but only a relative measure ranging from 0 to 100 based on a random sub-sample of all searches. This might be especially problematic for stock tickers with a low number of daily web search queries. Additionally, the daily trend measure provided by Google is aggregated in terms of UTC time, and not Eastern Time. As such, we regard a measure based on social media data as being more transparent and representative for retail investors' attention.

Our analysis is also related to the accounting literature concerned with firm disclosures. Since the pioneering work of Morse (1982), the relation between news releases and stock prices has been widely studied over the past decades. Among others, the empirical results of Ryan and Taffler (2004) show that major price movements and trading volumes are driven

by firm-specific news releases: Roughly 65% of significant price changes and trading volumes are explained by publicly available information. Analyzing the role of press coverage around earnings announcements, Bushee, Core, Guay, and Hamm (2010) report a reduced information asymmetry for companies with a broader news coverage. Focusing on a wide range of news announcements, Boudoukh, Feldman, Kogan, and Richardson (2019) show that news releases provide a significant economic contribution to stock market volatility. In a recent study, Lerman (2020) analyzes the discussion of accounting information on the social media platform StockTwits and Yahoo! message boards. Her results show that online financial communities pay attention to the release of accounting information, in particular earnings. However, she finds no evidence that higher attention is associated with an improvement in the processing of information.

2.3 Methodology

In the following, we introduce the econometric framework used to test the hypotheses introduced in the introduction. In our first hypothesis, we investigate the contemporaneous relation between volatility and investors' attention to a news article. We test Hypothesis 1 within a fixed effects panel data framework. More precisely, using a within-transformation we estimate the following regression model:

$$\begin{aligned}
 \log RV_{i,t} = & c_i + X'_{i,t} \beta \\
 & + \left[(1 - N_{i,t})(1 - E_{i,t}) \gamma_1 + N_{i,t}(1 - E_{i,t}) \gamma_2 + E_{i,t} \gamma_3 \right] A^R_{i,t} \\
 & + \left[(1 - N_{i,t})(1 - E_{i,t}) \phi_1 + N_{i,t}(1 - E_{i,t}) \phi_2 + E_{i,t} \phi_3 \right] A^I_{i,t} + \epsilon_{i,t},
 \end{aligned} \tag{2.1}$$

where $\log(\cdot)$ denotes the natural logarithm, $RV_{i,t}$ is the whole day realized volatility of company i on day t , c_i is a company-specific fixed effect, $X_{i,t}$ is a vector of control variables, $A^R_{i,t}$ and $A^I_{i,t}$ are the measures of retail and institutional investors' attention, respectively, $N_{i,t}$ and $E_{i,t}$ are dummy variables being equal to one if news and earnings are released on day t about company i and zero otherwise, and $\{\epsilon_{i,t}\}$ is a zero-mean innovation process. Note that the attention measures $A^R_{i,t}$ and $A^I_{i,t}$ are standardized and, therefore, we can interpret the estimated coefficients as the percentual change in volatility associated with a one standard deviation increase in attention.

In the vector of control variables, we include dummy variables for news and earnings releases, lagged values of (log) realized volatility, the turnover-ratio defined as the daily traded (dollar) volume divided by the market capitalization, and the log CBOE Market Volatility Index (VIX).³ The inclusion of lagged volatilities is motivated by the high persistence observed in the stock return's second moment. More precisely, building upon the model introduced by Corsi (2009) we include the lagged log realized volatility and the lagged average log realized volatility observed over the previous week and month. The inclusion of the turnover-ratio and the VIX are motivated by their predictive power for volatility documented in several empirical studies (e.g. Audrino et al., 2020; Buncic & Gisler, 2016; Mitnik, Robinsonov, & Spindler, 2015). The results presented in the remainder of this paper are robust to the inclusion of other control variables, e.g. weekday dummies and measures of earnings surprises.⁴

The coefficients of interest from the perspective of our first hypothesis, are γ_2 , γ_3 , ϕ_2 and ϕ_3 . More precisely, under Hypothesis 1 we expect them to be positive, i.e. higher levels of retail and institutional investors' attention on the day a news article is released (γ_2 and ϕ_2) or an earnings announcement is made (γ_3 and ϕ_3) are associated with higher volatility.

Concerning Hypothesis 2, we investigate if retail investors' attention

³We note that, in the contemporaneous setting, there could be several forms of potential endogeneity such as reverse causality and hidden confounders. In the predictive setting, this is mitigated as there is obviously no reverse causality and the inclusion of the lagged volatility blocks many potential confounders. Nonetheless, it is technically possible that there are hidden confounders that directly impact both the response $\log RV_{t+h}$ and A_t .

⁴Following Livnat and Mendenhall (2006), we define earnings surprises as the difference between the current earnings per share (EPS) and the previous year's EPS, divided by the stock price (i.e. EPS are assumed to follow a seasonal random walk). The results are available from the authors upon request.

decreases and institutional investors' attention increases the speed of price-adjustment to new information. In particular, we conjecture that retail investors' attention to news is associated with an increase in post-release volatility, i.e. the volatility observed the days after the news release.⁵ On the other hand, we conjecture that institutional investors' attention to news is instead associated with a decrease in post-release volatility. We investigate our second hypothesis by estimating the following fixed effects panel predictive regression model:

$$\begin{aligned} \log RV_{i,t+h} = & c_i + X'_{i,t} \beta \\ & + \left[(1 - N_{i,t})(1 - E_{i,t}) \gamma_1 + N_{i,t}(1 - E_{i,t}) \gamma_2 + E_{i,t} \gamma_3 \right] A_{i,t}^R \\ & + \left[(1 - N_{i,t})(1 - E_{i,t}) \phi_1 + N_{i,t}(1 - E_{i,t}) \phi_2 + E_{i,t} \phi_3 \right] A_{i,t}^I + \epsilon_{i,t+1}, \end{aligned} \tag{2.2}$$

where we consider different post-event horizons, i.e. $h = 1, 2, 3, 4$. The main differences to the regression model defined in Equation (2.1) are that the dependent variable is the realized volatility on subsequent days after news has been released, and that the vector of control variables contains the realized volatility and average weekly and monthly realized volatilities observed on day t , rather than $t - 1$. Under our second hypothesis, we expect the coefficients γ_2 and γ_3 to be positive, and ϕ_2 and ϕ_3 to be negative. In other words, higher attention of retail (institutional) investors to a news release is expected to be associated

⁵Throughout the paper, we will refer to the trading days after the day news has been released as post-event, post-release or post-announcement period.

with higher (lower) volatility in the days after the publication.

In Hypothesis 3, we hypothesize that the relation between retail investors' attention to news releases and the post-release volatility is heterogeneous in the type and topic of the firm-specific news. From an econometric perspective, we add to the regression defined in Equation (2.2) additional news dummy interactions which distinguish between different types and topics of news articles. More precisely, we use the following predictive regression model:

$$\begin{aligned}
 \log RV_{i,t+h} = & c_i + X'_{i,t} \beta \\
 & + \left[(1 - N_{i,t})(1 - E_{i,t}) \gamma_1 + (1 - E_{i,t}) \sum_j D_{i,t}(j) \gamma_2^{(j)} + E_{i,t} \gamma_3 \right] A_{i,t}^R \\
 & + \left[(1 - N_{i,t})(1 - E_{i,t}) \phi_1 + N_{i,t}(1 - E_{i,t}) \phi_2 + E_{i,t} \phi_3 \right] A_{i,t}^I + \epsilon_{i,t+1},
 \end{aligned} \tag{2.3}$$

where $D_{i,t}(j)$ is a dummy variable being equal to one when a news article of type or topic j is released. For example, when comparing different types of news articles, we include a dummy variable for “tabular material,” “full article,” “press release,” “news flash” and “hot news flash.” Under Hypothesis 3 we expect $\gamma_2^{(j)} \neq \gamma_2^{(s)}$ for at least one $j \neq s$. Note that while formally our third assumption solely focuses on retail investors' attention, we also test for potential heterogeneities in institutional investors' attention.

2.4 Data

The time period for our analysis ranges from January 2011 to December 2017. We consider 360 companies which were part of the S&P 500 throughout the time period of our analysis.⁶ Our basic frequency for the analysis is daily data. We define a day to start at 9:30 AM and end at 9:30 AM Eastern Time of the next trading day. Accordingly, a day consists of a trading time period followed by a period where the stock market is closed. For example, the last day of each week starts on Friday at 9:30 AM and ends on Monday morning at 9:30 AM Eastern Time when markets open for Monday.

2.4.1 Attention measures

In this study, we consider two different measures of attention: One for institutional investors' and one for retail investors' attention. Following the approach of Audrino et al. (2020), we measure retail investors' attention by counting the daily number of messages shared on social media platforms about a company. The choice of using social media activity to proxy attention is mainly motivated by four aspects. First, in contrast to many measures provided by commercial vendors, our attention proxy is transparent in terms of its construction since we have full control over the collection and filtering approach (more details are provided in the next paragraph). Second, financial measures which are typically used as

⁶Note that from 2011 to 2017, 368 companies were constantly part of the S&P 500. We excluded eight companies due to data issues (missing intraday trade information).

proxies for investors' attention, such as extreme returns and abnormal trading volumes, not only reflect the attention and awareness of investors towards a company, but are the outcome of different economic forces (Da, Engelberg, & Gao, 2014). Third, the use of social media data allows us to capture different awareness dimensions, such as attention towards the stock price, the firm's fundamentals, its products or the company's social value. Finally, as discussed in Section 2.2, we regard our measure based on social media data as being more transparent and representative for retail investors' attention compared to measures constructed from web search queries. However, when measuring retail investors' attention through social media activity, we need to make the assumption that traders are in fact sharing their thoughts, ideas and fears about a company on the online platforms considered.⁷

The primary social media platform considered in this study is StockTwits.⁸ Similarly to Twitter, StockTwits is a social media platform where users can share short messages with the online community, the difference being that it is specifically designed for sharing thoughts, information and ideas between investors. We collect all messages shared on these platform, i.e. *tweets*, which either mention a company's name or its *cashtag*, i.e. the company's ticker symbol preceded by the dollar sign which is a common convention adopted by stock market participants

⁷See Cookson and Niessner (2020) for a discussion about the plausibility of the assumption that investors share their "true" thoughts about companies and investment opportunities on social media.

⁸See <https://stocktwits.com>.

on social media platforms⁹ (for instance, the *cashtag* of Apple Inc. is \$AAPL). We include in our dataset both original and re-shared messages. When collecting data from the social media platform, we account for changes of a company’s name or ticker.¹⁰ As a further robustness check, we also consider an alternative social media platform, namely Twitter.¹¹ We report in the main part of this paper only the results obtained with StockTwits; the results obtained with Twitter are summarized in Appendix 2.B. The data for Twitter is collected using a scraper, and for StockTwits we have obtained the full historical data containing all short messages for the period considered in this study. For the 360 analyzed companies, we identify and collect a total of 9,890,132 and 30,520,617 messages shared on StockTwits and Twitter, respectively. Summary statistics of the raw total volume of social media messages collected for the analyzed companies can be found in Panel A of Table 2.1. Summary statistics of the correlation between the attention measures constructed from the two social media platforms are reported in Panel A of Table 2.2. The average correlation amounts to 0.44 on days without any information release, 0.57 on days with news and 0.56 on earnings call days suggesting that the two data sources provide to some extent different information.

Raw daily retail investors’ attention variables are defined as the logarithm of one plus the number of messages posted on a social media platform during a day. We follow the approach adopted by Garcia (2013)

⁹We also investigated the effect on our results when using a more restrictive filtering approach, e.g. considering only messages mentioning a unique cashtag. Our results are, however, not affected by the filtering approach.

¹⁰Between 2011 and 2017 we identified 41 out of the 360 analyzed companies that changed their name or ticker at least once.

¹¹See <https://twitter.com>.

Table 2.1: Summary statistics of news and social media coverage across 360 large cap US stocks

Panel A: Number of news articles and social media posts from 2011 to 2017				
	$Q_{25\%}$	Average	Median	$Q_{75\%}$
StockTwits messages	4,709.00	27,471.82	7,900.50	16,609.00
Twitter messages	20,000.00	84,772.67	32,815.00	66,512.75
News articles	1,280.75	1,689.53	1,480.00	1,849.00
Panel B: Number of days without news, news and earnings calls from 2011 to 2017				
	$Q_{25\%}$	Average	Median	$Q_{75\%}$
No news	861.75	933.51	972.00	1052.00
News, no earnings call	680.00	798.49	760.00	870.25
Earnings calls	28.00	28.00	28.00	28.00

NOTE: In Panel A the table depicts the average, median, 25%- and 75%-quantiles across the 360 analyzed companies of the total number of news articles and social media messages published between 2011 and 2017. Panel B summarizes the average, median, 25%- and 75%-quantiles of the total number of days a company has had no news releases, at least one news article (other than earnings calls), and an earnings call. Not surprisingly, since each company has four earnings calls per year, all firms in our sample have 28 earnings announcements over the eight analyzed years.

and remove trends and seasonal components from the raw daily attention measure by running the following regression at the company level:

$$\tilde{A}_{i,t}^R = \beta L_5(\tilde{A}_{i,t}^R) + \gamma Z_t + \epsilon_{i,t},$$

where $\tilde{A}_{i,t}^R$ is the raw (log) posting volume about company i , L_p is a p -lag operator which creates a vector of the past five lags, and Z_t is a vector including a constant, dummy variables for weekdays and months. The attention measure is then defined as $A_{i,t}^R = \hat{\epsilon}_{i,t}/s_{i,\epsilon}$ where $\hat{\epsilon}_{i,t}$ are the estimated residuals and $s_{i,\epsilon}$ are their standard deviations. It follows that the attention measures have zero mean and unit standard deviation. Note that the results presented in the following are qualitatively unchanged when defining investors' attention as the raw (log) daily posting volume.¹²

¹²The results obtained with the raw (log) attention measures are available from the authors upon request.

In order to measure institutional investors' attention, we follow the approach of Ben-Rephael et al. (2017) and use the news searching and reading activity for specific stocks on Bloomberg terminals. More precisely, Bloomberg records the number of times firm-specific news articles are read and searched for by terminal users. Comparing these counts with those observed in the past 30 days, Bloomberg constructs a daily index of attention which attains the values 0 (lowest attention), 1, 2, 3 or 4 (highest attention). Following the approach proposed by Ben-Rephael et al. (2017), we transform these discrete scores into continuous values and standardize the variable at the company level to have zero mean and unit standard deviation.¹³ Historical data provided by Bloomberg are missing for the first week of 2011 and from August 17 to November 2, 2011. We refer to Ben-Rephael et al. (2017) for a more detailed description of Bloomberg's news reading and searching index, and for a discussion of why this index captures the attention of institutional investors.

Panel B and Panel C of Table 2.2 summarize the correlations between Bloomberg's news attention measure and the volume of social media posts shared on StockTwits and Twitter, respectively. The correlation coefficients reported in Panel B and Panel C are small, in particular on days with earnings calls, suggesting that the information provided by social media platforms differs from the one that is provided by the Bloomberg news retrieval measure. As expected, the social media platforms reflect a different type of investors' attention compared to the

¹³Based on Bloomberg's definition of the index, to construct a continuous attention variable, Ben-Rephael et al. (2017) assign to the discrete scores 0, 1, 2, 3 and 4 the values -0.350, 1.045, 1.409, 1.647 and 2.154.

measure of institutional investors' attention introduced by Ben-Rephael et al. (2017).

Table 2.2: Summary statistics of the correlation between the attention measures

Panel A: Correlation between StockTwits and Twitter attention				
	$Q_{25\%}$	Average	Median	$Q_{75\%}$
No news	0.32	0.44	0.44	0.55
News, no earnings call	0.45	0.57	0.58	0.70
Earnings calls	0.42	0.56	0.58	0.75
Panel B: Correlation between StockTwits and Bloomberg attention				
	$Q_{25\%}$	Average	Median	$Q_{75\%}$
No news	0.08	0.13	0.14	0.18
News, no earnings call	0.15	0.21	0.21	0.27
Earnings calls	-0.15	0.00	0.03	0.19
Panel C: Correlation between Twitter and Bloomberg attention				
	$Q_{25\%}$	Average	Median	$Q_{75\%}$
No news	-0.01	0.05	0.05	0.11
News, no earnings call	0.11	0.17	0.17	0.23
Earnings calls	-0.18	0.01	0.03	0.19

NOTE: In Panel A the table depicts the average, median, 25%- and 75%-quantiles across the 360 analyzed companies of the correlation between the daily volume of posts shared on StockTwits and Twitter. Similarly, Panel B and Panel C summarize the correlations between Bloomberg's news attention measure and the volume of social media posts shared on StockTwits and Twitter, respectively.

2.4.2 News releases

As we are interested in investors' attention around firm-specific news releases, we need to identify all such events. The majority of empirical studies analyzing the impact of investors' attention on stock markets have focused on price reactions around earnings releases (e.g. Ben-Rephael et al., 2017; DellaVigna & Pollet, 2009; Hirshleifer et al., 2009). In the first part of this paper, we therefore distinguish between earnings announcements and other news releases.

For the identification of scheduled earnings releases, we use the Thom-

son Reuters' I/B/E/S database, which provides dates and time-stamps of quarterly earnings calls. The identification of news releases is done using the news article database obtained from RavenPack News Analytics. The database covers information from Dow Jones Newswire, regional editions of the Wall Street Journal, Barron's and MarketWatch as well as press releases from leading global media organizations. Using textual analysis techniques, RavenPack News Analytics detects the entities in each news release and determines how novel and relevant the story is for each mentioned company. Moreover, each news release is automatically categorized based on its topic (e.g. earnings, labor issues, legal). Finally, the database also provides information about the type of the news release (obtained directly from the news providers), i.e. tabular material, full article, press release, news flash and hot news flash.¹⁴ We collect all news releases with the highest novelty and relevance for the analyzed companies.¹⁵ This is motivated by the findings of Groß-Klußmann and Hautsch (2011), which show that the omission of news articles containing irrelevant and old information is crucial for filtering out noise. A similar filtering approach is also adopted by Boudoukh et al. (2019). For the 360 analyzed companies, we identify a total of 608,298 news releases from 2011 to 2017. As for the social media data, the news coverage varies greatly among the analyzed companies. Summary statistics of the total number of news articles published between 2011 and 2017 for the 360

¹⁴The database provided by RavenPack News Analytics offers a much wider range of indicators than the ones mentioned in this study; for more information we refer directly to <https://www.ravenpack.com>.

¹⁵RavenPack News Analytics assigns to each news release and for each mentioned entity a novelty and relevance score ranging from 0 to 100. In this study only news articles with a relevance and novelty score of 100 are considered.

analyzed US companies can be found in Panel A of Table 2.1. Summary statistics of the total number of days a company has had no news, at least one news release (other than an earnings announcement), and earnings calls are depicted in Panel B of Table 2.1. While the news coverage (number of articles) varies greatly across the companies, the number of events (days with at least one news article) is more homogeneous.

Relevance of news releases to retail investors

Being particularly interested in the relation between retail investors' attention and price reactions to news releases, we introduce the concept of relevance of a news event. More precisely, we compare the textual data of social media posts and news article headlines to identify news that is noticed and discussed by retail investors. The fact that RavenPack News Analytics classifies an article as being highly relevant for a specific company does not necessarily imply that the new information is noticed by investors, in particular by retail investors. To this end, we use a standard approach taken from the natural language and information science literature and compute for each news article the cosine-similarity between its headline and social media messages published on the same day.¹⁶ The main idea of cosine-similarity is to first cast documents in a (high-dimensional) vector space, and then measure the similarity among textual data by computing the angle between the corresponding vectors:

$$\text{sim}(h_i, m_j) = \frac{\langle h_i, m_j \rangle}{\|h_i\| \|m_j\|}, \quad (2.4)$$

¹⁶As described previously, we consider a day to start at 9:30 AM and end at 9:30 AM Eastern Time of the next trading day.

where h_i and m_j are the vectorial representation of a news headline and a social media message, respectively, $\langle \cdot, \cdot \rangle$ stands for the scalar product and $\| \cdot \|$ for the Euclidean norm. To obtain a vectorial representation of the textual data, we first clean and tokenize all social media messages and news headlines, i.e. (i) we remove stop-words, links, emoticons and references to other users, (ii) convert all letters to lower case and reduce inflected words to their word stem by applying a Porter stemmer, and (iii) break up all headlines and messages into single words (uni-grams). A term-document matrix is then created, where each column represents a word and each row corresponds to a document, i.e. a news headline or a message shared on Twitter or StockTwits. In other words, each row of the matrix corresponds to the vectorial representation of a document. In its most simple form, the matrix is filled with the relative number of times a term occurs in a given document, often referred to as term-frequency (TF). In this study we follow, among others, Brown and Tucker (2011) and apply a common weighting refinement which multiplies the term-frequency with the inverse document-frequency (IDF), resulting in the so-called term-frequency inverse document-frequency (TF-IDF).¹⁷ This approach has the advantage that common words, such as “company,” are down-weighted and contribute only marginally to the cosine-similarity between two documents. Note that TF-IDF is non-negative and as such the cosine-similarity between a news headline and a social media message is bounded between 0 and +1. Given the similarity between a news

¹⁷More precisely, the inverse document-frequency is defined as the logarithm of the total number of documents divided by the number of documents that contain a particular term, i.e. $IDF(w) = \log \frac{N}{n_w}$ where N is the number of documents and n_w the number of documents that contain word w .

headline and Twitter or StockTwits messages, we define the relevance score of a news article as:

$$\text{rel}(h_{i,t}) = \sum_{t_j \in t} \text{sim}(h_{i,t}, m_{t_j}), \quad (2.5)$$

where $\text{sim}(h_{i,t}, m_{t_j})$ is the cosine-similarity between headline i published on day t and a social media message published at time t_j , and the sum is taken over all messages published on day t . In this study we define a news article to be relevant for retail investors if $\text{rel}(h_{i,t}) > 0$. We also considered alternative thresholds. In particular, choosing a higher thresholds leads to even more favorable (unreported) results for Hypothesis 3. In order to avoid an arbitrary choice of the threshold value, for the remainder of this paper we adopt a conservative approach and define an article to be relevant for retail investors if its relevance score is larger than zero. We conducted several other robustness checks for the relevance score measure. In particular, we assessed the effect of removing company names and tickers from the text data before calculating the cosine similarity, as well as computing the relevance score of an article using only social media messages posted after the publication time of the news article. Our results are unaffected by these changes, as depicted in Appendix 2.C.

Figure 2.3 shows the distribution of social media messages and news articles over one-hour intervals of a day. More precisely, the plots in Figure 2.3 depict the hourly average number (per company) of StockTwits messages (a), Twitter microblogs (b) and news articles (c, d) published

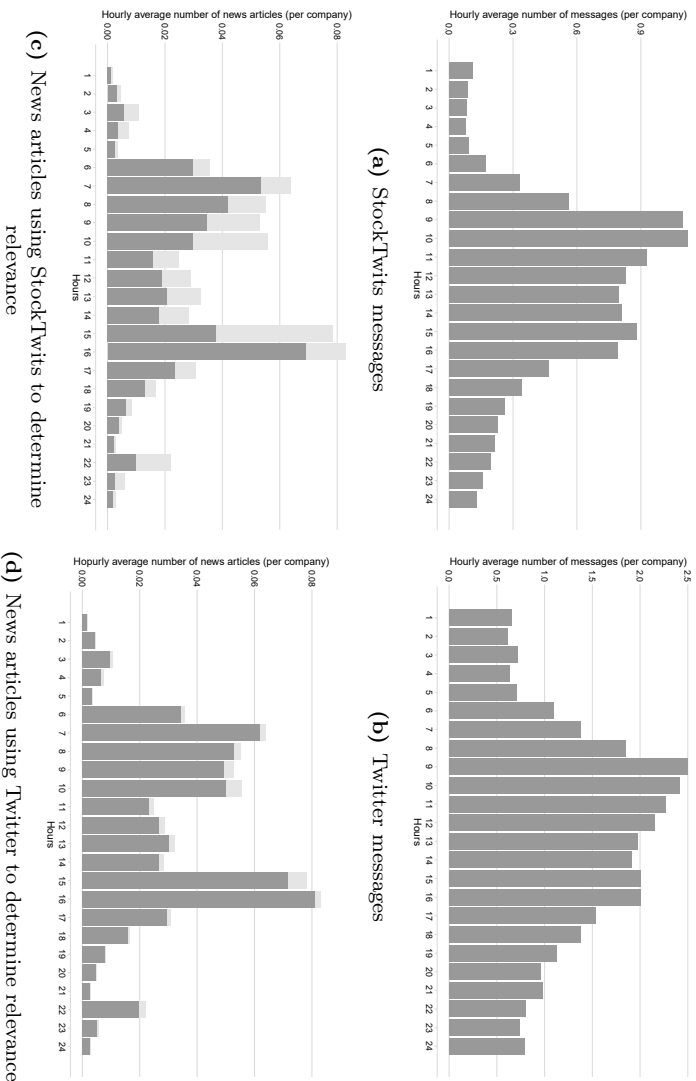


Figure 2.3: Number of social media messages and news articles published in one-hour intervals over a day

NOTE: The figure reports the average number (across companies) of StockTwits messages (a), Twitter messages (b), and news articles, with the relevance determined by StockTwits (c) or Twitter messages (d), published in one-hour intervals over a day. More precisely, each plot shows the average number of social media messages published within a certain hour of the day across all companies over the entire analysis period, i.e., from 2011 to 2017. The average number of relevant news articles published within a one-hour interval is represented in dark-grey, and the irrelevant articles in light-grey. A news article is classified as being relevant if the cosine-similarity between its headline and social media messages is positive.

within a one-hour interval. For the volume of news releases, we distinguish between relevant (dark-grey) and irrelevant (light-grey) articles as defined by the cosine-similarity between the headlines and messages posted on StockTwits (c) and Twitter (d). The social media activity is particularly concentrated over the trading hours, with the highest number of messages shared between 9:00 AM and 10:00 AM (Eastern Time), i.e. during market opening. Most news articles are instead published between 7:00 AM and 8:00 AM and between 3:00 PM and 4:00 PM (Eastern Time), i.e. before the market closes. Interestingly, when StockTwits data are used, fewer news articles have a positive relevance score, compared to when Twitter messages are used. In particular, a considerable share of articles published between 3:00 PM and 4:00 PM are not discussed on the social media platform StockTwits.

Categorization of days based on type and topic of news

The computation of relevance scores for each news article also allows us to identify, from the perspective of retail investors, the most relevant articles published on a given day per company. For every company, we create two categorizations of days by identifying the category with the highest relevance scores using the following grouping factors: (i) news type, and (ii) topic of the articles. In other words, we first sum the relevance scores of all articles published on a given day per category of the chosen grouping. We then categorize the day based on the category with the highest total relevance score. This approach allows us to study potential heterogeneities in the effect of investors' attention on stock

market volatility. In particular, we are not only able to study the effect of investors' attention around news releases in general, but we can also shed light on potential heterogeneities due to different news types and topics.

2.4.3 Volatility measure

We measure daily volatility using the concept of realized volatility. I.e., instead of treating the returns' conditional second moment as latent, we directly estimate it using high-frequency intraday data. In detail, we construct realized volatility measures for the time period between 2011 and 2017 using intraday transaction prices obtained from the NYSE Trade and Quote (TAQ) database.¹⁸ The high-frequency data are cleaned following the procedure suggested by Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). Realized volatility is constructed as the square root of the sum of squared intraday log returns. Following the existing realized volatility literature, we use a five-minute sampling frequency for calculating intraday returns.

The volatility measure thus constructed uses only data covering part of the day, i.e. from 9:30 AM to 4:00 PM. Since we define a day to start at 9:30 AM and end at 9:30 AM of the following trading day, we adopt an estimator suggested by Hansen and Lunde (2005) and use overnight returns to construct a volatility measure for the whole day. More precisely, a variance measure for the whole day is obtained by a linear combination

¹⁸For August 1, 2012, trade data of 35 companies in our sample are missing. For this day, realized volatility measures are constructed using quote data.

of the realized variance estimated from high-frequency transaction data and the squared overnight return. The optimal combination is determined by minimizing a mean-squared error (MSE) criterion.¹⁹ In order to obtain more robust optimal weights, we compute the optimal linear combination cross-sectionally.²⁰ Our main findings remain qualitatively the same when the classical realized volatility estimator (i.e no use of overnight returns) is used, when the optimal combination is computed at the company-level, or when one of the other two estimators suggested by Hansen and Lunde (2005) is used.

Analyzing the volatility reaction to news rather than the stock return reaction has two main advantages. First, we do not need to know whether a news release is positive or negative for the considered company, i.e. regardless of whether the return is positive or negative, if prices react to news, volatility will increase. Secondly, daily returns reflect the aggregate opinion of market participants and are highly influenced by the most dominant market participants. If, for example, the opinions between institutional and retail investors about a news release diverge, and the former push stock prices up while the latter start selling their shares, the daily return will most likely be positive since institutional investors' trading volume is usually considerable larger. In other words, when news is relevant to both retail and institutional investors, daily stock returns

¹⁹Since the true daily integrated variance is not observed, Hansen and Lunde (2005) obtain the optimal combination of realized variance and squared overnight returns that minimizes the MSE criterion by restricting their attention to the class of conditionally unbiased estimators. For a more detailed description of the estimation procedure, we refer to Hansen and Lunde (2005).

²⁰In order to obtain true ex-ante predictions, in the out-of-sample analysis the optimal weights are computed using only data from the rolling estimation window.

do not always reflect the reaction of all market participants. Realized volatility estimated from intraday trade data accounts by construction for the trades of all investors; i.e. the trades of retail investors also increase the daily volatility.²¹

Table 2.3 reports summary statistics for the three main variables of interest, i.e. volatility, retail and institutional investors' attention. More precisely, Panel A reports the average volatility and attention on days with news releases relative to days without news releases. A value above one indicates that volatility or attention is on average higher compared to a day without any new information release. Not surprisingly, volatility is on average higher when news is released, in particular when earnings are announced. Attention of both retail and institutional investors follows

Table 2.3: Summary statistics of volatility and attention measures

Panel A: Volatility and attention on days with relative to days without news releases				
	RV_t	RV_{t+1}	A_t^R	A_t^I
No news in t	1.00	1.00	1.00	1.00
News, no earnings call in t	1.07	1.05	1.32	1.44
Earnings calls in t	1.31	1.92	6.41	3.73
Panel B: Correlation between attention and volatility				
	Contemp. corr.		Lagged corr.	
	A_t^R, RV_t	A_t^I, RV_t	A_t^R, RV_{t+1}	A_t^I, RV_{t+1}
No news in t	0.16	0.25	0.05	0.13
News, no earnings call in t	0.30	0.30	0.12	0.16
Earnings calls in t	0.13	0.22	0.13	-0.02

NOTE: Panel A reports the average volatility and attention on days with news releases relative to days without news releases. A value above one indicates that volatility or attention is on average higher compared to a day without any new information release. Panel B reports average (across all companies) contemporaneous and predictive correlation coefficients between volatility and attention. Retail investors' attention is constructed from StockTwits data.

²¹On days with news releases, the mean squared difference between the open-to-open daily absolute return and the whole day realized volatility is four times larger when both retail and institutional attention are above their historical 90% quantile, compared to when only institutional attention is above its 90% quantile.

a similar pattern. Interestingly, volatility is also on average higher the day after the news has been released. Panel B of Table 2.3 reports average (across all companies) contemporaneous and predictive correlation coefficients between volatility and attention. Consistent with the widely reported high persistence in return volatility, the auto-correlation of realized volatility is considerably high (e.g. Andersen, Bollerslev, Diebold, & Ebens, 2001). The first two columns in Panel B show that the contemporaneous correlation between volatility and attention increases when news articles are released (other than earnings calls), in particular for retail investors. A similar pattern is also observed for the predictive correlation between attention and volatility.

2.5 The relation between investor attention, news, and volatility

2.5.1 News releases and earnings announcements

In this section, we analyze our first two hypotheses. Concerning Hypothesis 1, Panel A of Table 2.4 reports the estimates of the attention coefficients for the fixed effect panel regression defined in Equation (2.1).²² More precisely, Panel A reports the increases in volatility associated with a one standard deviation increase of retail investors' and institutional investors' attention on days with earnings calls and days with other news releases. For completeness, we also report the coefficients for days without

²²The results obtained when measuring retail investors' attention with Twitter data are reported in Table 2.11 in Appendix 2.B.

any news releases. For both retail investors and institutional investors, increases in attention on days with news releases are associated with contemporaneous increases in volatility. The estimated coefficients are positive and highly significant. On days with an earnings call, a one standard deviation increase in retail and institutional attention is associated with a 3.4% and 3.3% increase in volatility, respectively. This finding confirms the results reported in other empirical studies (e.g. Aboody et al., 2010; Ben-Rephael et al., 2017; DellaVigna & Pollet, 2009; Hirshleifer et al., 2009; Hirshleifer & Sheng, 2019). Note that the earnings day dummy variable coefficients are large and highly significant (results not tabulated). I.e., a large part of the volatility spike observed around earnings announcements is explained by this dummy variable. The estimated attention coefficients are even larger on days with other news releases: the increase in contemporaneous volatility associated with a one standard deviation increase in attention amounts to 4.2% for retail investors and 4.6% for institutional investors. Our findings therefore extend the results reported in the literature about earnings announcements to a wider range of news releases. In summary, the results reported in Panel A of Table 2.4 clearly confirm our first hypothesis. When investors are paying attention to news releases, price reactions are stronger and volatility increases.

Panel B of Table 2.4 reports the estimated attention coefficients of the predictive regression model defined in Equation (2.2). In detail, the panel reports the increase in volatility observed the day after a news event for a one standard deviation increase in investors' attention. Further, changes in volatility observed up to four days after the news announcement

Table 2.4: Investors' attention and volatility reaction

Panel A: Contemporaneous relation between attention and volatility		
	Estimate	t-statistic
Retail investor attention		
No news	0.028	17.99
News, no earnings call	0.042	17.10
Earnings calls	0.034	9.91
Institutional investor attention		
No news	0.037	19.48
News, no earnings call	0.046	17.48
Earnings calls	0.033	11.05
Adjusted R^2	58.48%	
Number observations	587,762	
Panel B: Relation between attention and next day's volatility		
	Estimate	t-statistic
Retail investor attention		
No news	0.012	21.36
News, no earnings call	0.023	28.34
Earnings calls	0.030	4.93
Institutional investor attention		
No news	0.002	3.31
News, no earnings call	0.002	3.99
Earnings calls	0.005	0.88
Adjusted R^2	53.78%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase on the news release day (Panel A) and the day after (Panel B) associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, Panel A summarizes the attention estimates of the regression defined in Equation (2.1), and Panel B those of the regression defined in Equation (2.2) with $h = 1$. For completeness, the table also reports the estimated increase in contemporaneous and future volatility associated with a one standard deviation increase in attention when no news article is released. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

associated with a one standard deviation increase in investors' attention are summarized in Figure 2.1. Note that we measure investors' attention only on the day the news is released. Retail investors' attention to news releases is associated with an increase of next day's volatility of 3% for earnings announcements and 2.3% for other news. These estimates are highly significant. Considering up to four days after the news release, we observe that this relation is still significantly positive and decreasing over time for retail investors. In contrast, we find only limited evidence

that institutional investors' attention to news releases is associated with higher volatility in the post-event trading days. The estimated coefficient for earnings announcements is 0.005 and not significant, and for other news releases 0.002 and marginally significant. In fact, Figure 2.1 shows that the relation between institutional investors' attention to earnings announcements and volatility becomes significantly negative two days after the event. Further, our findings show that for other news releases, the attention of institutional investors is not associated with changes in volatility in the post-event period. The results presented in Panel B of Table 2.4 and in Figure 2.1 support Hypothesis 2. In summary, institutional investors' attention to earnings calls reduces the volatility in the days after the announcement, an indication that the new information is incorporated more quickly into prices. This finding confirms the results reported by Ben-Rephael et al. (2017). There is instead only limited evidence of a relation between institutional investors' attention and stock market volatility after the release of other information. In addition, we find clear evidence that a higher attention of retail investors to news releases is associated with a higher volatility over the next trading day. This result is consistent with the idea that retail investors are reducing the speed at which prices adjust to new information and increasing the risk in the stock market.

The relation between attention and trading volume

The results presented thus far show that retail investors' attention to news releases is associated with increases in volatility up to four days

after the event. If the higher volatility is in fact related to increased trading activities of investors, we would expect attention also to be positively related to the trading volume around the news release. Ideally, we would analyze the relation between retail investors' attention and trading activity. However, the trading volume of retail investors is not directly observable and we therefore follow Dimpfl and Jank (2016) and use the total trading volume. We adopt the approach by Garcia (2013) and use the same procedure applied previously to the attention measures to remove trend and seasonal components from the dollar trading volume. We regress the log trading volume at the company level on five lagged values and dummy variables for weekdays and months (for more details see Section 2.4). We then analyze the effect of investors' attention on the normalized trading volume $\hat{V}_{i,t} = \hat{\epsilon}_{i,t}/s_{i,\epsilon}$, where $\hat{\epsilon}_{i,t}$ are the estimated residuals from the aforementioned regression and $s_{i,\epsilon}$ its standard deviation, within a fixed effects panel regression framework. More precisely, we estimate the same regressions defined in Equation (2.1) and Equation (2.2) using as a dependent variable the normalized trading volume and additionally control for daily returns and squared returns. The results are depicted in Figure 2.4.

For retail investors, we observe a similar pattern as for the realized volatility. In particular, when retail investors are paying attention to a news event, the trading volume not only increases on the day of the news release, but also in the post-release period. This finding indicates that higher attention of retail investors increases the market risk and reduces the price-adjustment speed through excessive trading. In contrast, when

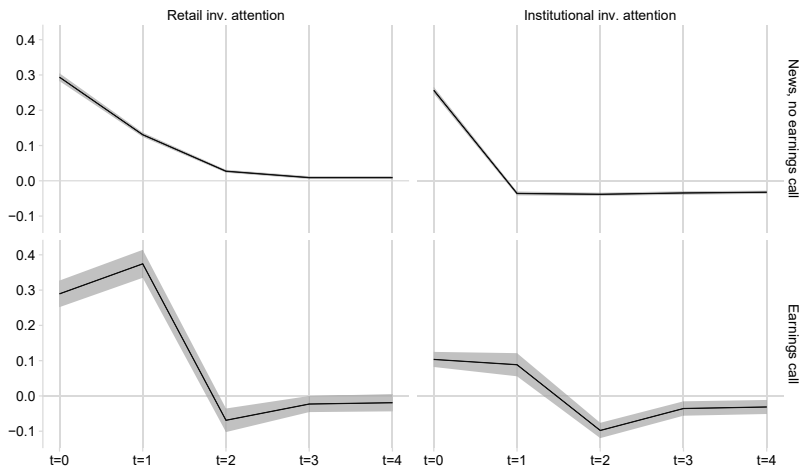


Figure 2.4: Trading volume increase associated with a one standard deviation increase in retail and institutional investors' attention around news releases

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the normalized trading volume on the news event day and the four post-release days. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

institutional investors pay attention to a news release we even observe a reduction in the trading activity, i.e. a one standard deviation increase in institutional investors' attention on the news release day is associated with a reduction in trading volume over the post-event period. When paying attention to a news release, sophisticated investors quickly interpret the new information and rapidly react to it, leading to a higher trading activity and stronger price adjustments on the event-day, and to calmer markets in the following trading days. Around earnings calls, however, institutional investors' attention is associated with higher trading activities even on the first day after the announcement. This finding can be explained by our definition of a trading day: We define a day to start at 9:30 AM and

end at 9:30 AM of the next trading day. As such, when earnings are released in the morning, the trades of institutional investors are reflected in the trading volume of the day after the announcement. This effect is not observed for the volatility, since we take into account overnight returns when measuring the second moment of stock returns.²³ Given that the focus of this study is on volatility, a more detailed analysis of the relation between investors' attention and their trading activity is left for future research.

Weekday effects and the reaction to news

Several studies have analyzed a similar hypothesis focusing primarily on earnings announcements and using indirect measures of investors' attention (or distraction). DellaVigna and Pollet (2009), for example, analyze price-adjustments to earnings calls in dependence on the release day, arguing that on Fridays investors are more distracted and as such, attention to the newly released information is lower. Their results show that the post-announcement price-drift is stronger when the earnings call happens on a Friday. In order to ensure that our results are not driven by potential weekday effects, we include weekday-dummy variables and interact them with the dummies for news releases and earnings calls as a robustness check. In this way, we control for potential spikes in volatility associated with the publication-day of the news article and earnings.

²³Because of the potential issues with the timing of the news releases, we estimate the regression defined in Equation (2.2) for $h = 2$ by defining the attention measures over a two-day period. In other words, the attention to a news article is defined by the attention on the release day and the next trading day. The estimation results (not reported) are qualitatively unaffected.

The coefficients of interest for our first two hypotheses estimated when controlling for weekday-effects are summarized in Table 2.8 in Appendix 2.A. The contemporaneous relation between attention and volatility is not affected by the additional control variables, confirming our first hypothesis. The results for retail investors' attention reported in Panel B of Table 2.8 are qualitatively unaffected when controlling for weekday-effects. The coefficients of institutional investors' attention are instead now even smaller and none of them are significant at any conventional level. Even when accounting for the Friday-induced distraction, retail investors' attention is still associated with increases in volatility over the post-event period.

2.5.2 Heterogeneities in retail investors' attention

In the following, we focus on our third hypothesis and analyze whether there are heterogeneities in retail investors' reaction to information depending on the type and topic of the news release. Note that in Hypothesis 3 we focus on the relation between retail investors' attention to news and post-announcement volatility. As such we do not report detailed results for the contemporaneous regression.²⁴

Table 2.5 reports the estimated coefficients of attention variables for the regression defined in Equation (2.3) with $h = 1$ distinguishing between the most relevant news release types. In other words, the results

²⁴The contemporaneous regression results which account for potential heterogeneities are available upon request from the authors. In general the contemporaneous results are in line with those obtained from the regression defined in Equation (2.3), in the sense that we find in both cases similar heterogeneities.

summarized in Table 2.5 can be interpreted as the change in volatility observed the day after a news release associated with a one standard deviation increase in attention on the news-event day. Instead of simply distinguishing between earnings calls and other types of news releases, we investigate whether the relation between volatility and attention is heterogeneous in the type of news. Recall that we identify the most relevant news article type from the perspective of retail investors by identifying the articles most discussed on social media (see Section 2.4.2). The largest increase in volatility associated with an increase in retail attention is observed when the most relevant released information is unexpected: The news formats “news flash” and “hot news flash” are associated with a 4.8% and 7.5% increase in volatility, respectively. The estimated coefficients for institutional investors’ attention are smaller and often not significant. Only when the most relevant released news articles are in the format of a “full article” or a “news flash” are the estimates significant, but the volatility increases associated with a one standard deviation increase in institutional attention amount to only 0.9% and -0.9%, respectively.

Figure 2.5 depicts the changes in volatility associated with a one standard deviation increase in retail and institutional investors’ attention to press releases, news flashes and hot news flashes, over the four days after the news release. In particular, for “press release” and “news flash,” retail investors’ attention is associated with higher volatility still two days after the release of the information. When institutional investors are paying attention to these releases, volatility decreases the days after the

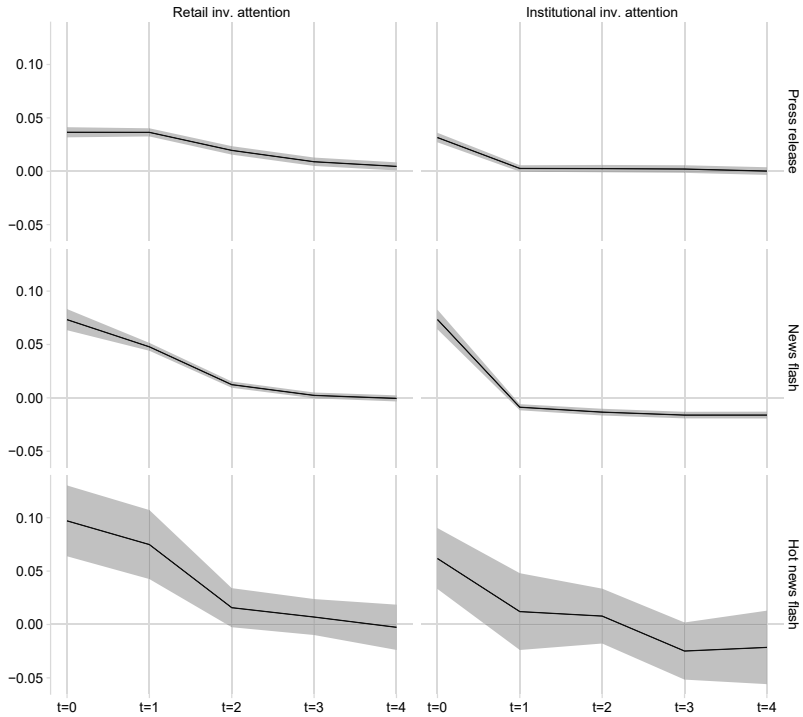


Figure 2.5: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around different news release types

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news event day and the four post-release days for different news types. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

announcement. For news articles in the format of “hot news flash,” retail investors' attention is associated with significant increases in volatility only up to one day after the news release.

Table 2.6 reports similar results but distinguishes between the most relevant topic for retail investors. Given the large number of different news topics in our dataset, the table reports only the most important results,

Table 2.5: Investors' attention to news types and next day's volatility reaction

	Estimate	t-statistic
Retail investor attention		
<i>No news</i>	<i>0.012</i>	<i>21.61</i>
<i>Not relevant news</i>	<i>0.010</i>	<i>13.12</i>
Tabular material	0.027	17.32
Full article	0.017	9.33
Press release	0.036	18.71
News flash	0.048	25.39
Hot news flash	0.075	4.54
<i>Earnings calls</i>	<i>0.030</i>	<i>4.90</i>
Institutional investor attention		
<i>No news</i>	<i>0.002</i>	<i>3.62</i>
<i>Not relevant news</i>	<i>0.002</i>	<i>3.22</i>
Tabular material	0.001	1.03
Full article	0.009	7.52
Press release	0.003	1.70
News flash	-0.009	-5.98
Hot news flash	0.012	0.66
<i>Earnings calls</i>	<i>0.005</i>	<i>0.90</i>
Adjusted R^2	53.86%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase the day after a news release associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, the table reports the coefficients for attention measures obtained from the estimation of the regression defined in Equation (2.3) with $h = 1$. In particular, we distinguish between different news types defined as the most discussed articles by retail investors. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

i.e. those news topics for which we find the smallest and largest estimates for changes in volatility associated with a one standard deviation increase in retail investors' attention. For retail investors' attention, all estimates are positive and the majority of them also significant. On the day after a news release, the largest percentual changes in volatility associated with a one standard deviation increase in retail investors' attention are observed when the most relevant published articles are about "stock prices" (5.5%), "analyst ratings" (6.6%), "acquisition and merger" (8.5%) or "earnings" (8.8%) and "revenues" (10.4%).²⁵ Concerning institutional

²⁵Note that articles categorized as being about "earnings" or "revenues" refer to news articles about a company's earnings or revenues which have not been published on an earnings call day.

investors' attention, most estimates are small in absolute terms and often not significant at any conventional level. However, the results for the news topics "stock prices," "analyst ratings" and "acquisition and merger" are of particular interest: A one standard deviation increase in institutional investors' attention is associated with a reduction in next day's volatility of 2.2%, 1.6% and 1.7%, respectively.

Figure 2.6 summarizes the results for the most interesting news topics for different time horizons. In general, the results show that when retail investors are paying attention to these news events, volatility is not only higher on the day the news is released but also during the subsequent four trading days. Institutional investors' attention to a news release is only associated with a contemporaneous increase in volatility. Further, when the most relevant articles are about stock prices, mergers and acquisitions, or analyst ratings, institutional investors' attention is associated with a decrease in volatility over the four days after the release of the news. In other words, when the most relevant published articles are about topics directly related to a company's market value, high attention of retail investors is associated with an increase in volatility over the following trading days, whereas institutional investors' attention appears to increase the speed at which prices adjust to the new information reducing the post-event volatility. Indeed, Ben-Rephael et al. (2017) find a similar result for news about changes in analyst recommendations, showing that higher institutional attention is associated with a faster price-adjustment to the new information. Our results extend this finding by showing that this relation is also valid for other news topics, and that retail investors'

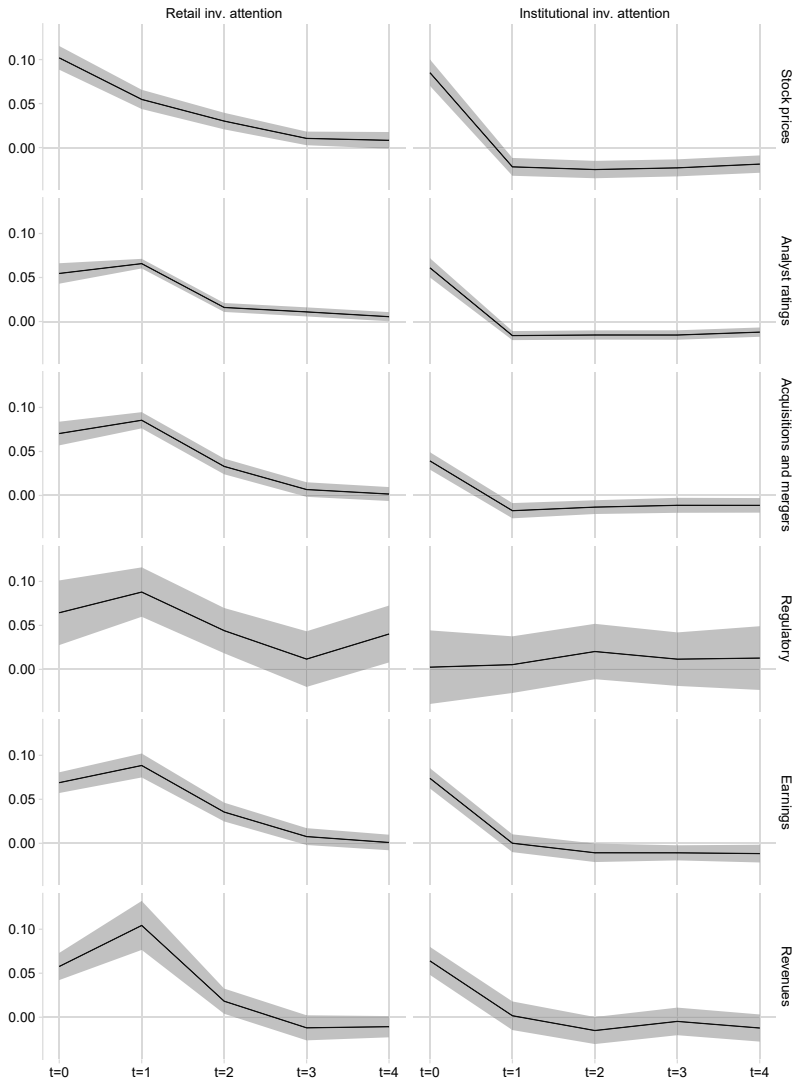


Figure 2.6: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around different news release topics

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news-event day and the four post-release days for different news topics. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

Table 2.6: Investors' attention to news topics and next day's volatility reaction

	Estimate	t-statistic
Retail investor attention		
<i>No news</i>	0.012	21.57
<i>Not relevant news</i>	0.010	13.05
Insider trading	0.010	5.95
Order imbalances	0.013	5.80
Credit	0.016	1.88
Credit ratings	0.020	4.29
Corporate responsibility	0.024	1.24
Technical analysis	0.026	16.57
	⋮	⋮
Stock prices	0.055	9.81
Analyst ratings	0.066	23.65
Acquisitions and mergers	0.085	18.19
Regulatory	0.088	6.13
Earnings	0.088	12.75
Revenues	0.104	7.35
<i>Earnings call</i>	0.031	4.93
Institutional investor attention		
<i>No news</i>	0.002	3.68
<i>Not relevant news</i>	0.003	3.26
Insider trading	0.011	8.21
Order imbalances	-0.007	-3.02
Credit	-0.009	-1.45
Credit ratings	-0.004	-1.10
Corporate responsibility	-0.002	-0.09
Technical analysis	0.002	1.20
	⋮	⋮
Stock prices	-0.022	-4.24
Analyst ratings	-0.016	-6.11
Acquisitions and mergers	-0.017	-3.93
Regulatory	0.005	0.33
Earnings	0.000	0.02
Revenues	0.002	0.21
<i>Earnings call</i>	0.005	0.90
Adjusted R^2	53.94%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase the day after a news release associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, the table reports the coefficients for attention measures obtained from the estimation of the regression defined in Equation (2.3) with $h = 1$. In particular, we distinguish between different news topics defined as the most discussed articles by retail investors. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). Note that the table only reports partial results (the news articles with the largest absolute estimates for retail investors); results for the omitted events are reported in Table 2.9 in Appendix 2.A.

attention has the opposite relation, i.e. it reduces the price-adjustment speed.

In summary, we find clear evidence that the volatility increase asso-

ciated with investors' attention is not homogeneous depending on the type or the topic of the most relevant articles for retail investors. In particular, we find that when retail investors are paying attention to an unexpected news event or to a news article about topics directly linked to a company's market value (e.g. mergers and acquisitions, analyst ratings), the relation is stronger. Due to their complexity, the implications of these topics for a company's market value are generally difficult to understand, in particular for unsophisticated investors. This finding confirms our third hypothesis. Moreover, while for most news topics institutional investors' attention has no significant relation to the post-event volatility, we find that when sophisticated investors are paying attention to news releases about analyst ratings, mergers and acquisitions, or stock prices, the market risk decreases in the four days after the publication of the new information.

2.6 The economic value of accounting for attention around news releases

In this section, we show that the empirical findings reported in the previous section can be used to generate an economic value for investors. More precisely, we propose two prediction models and use out-of-sample volatility forecasts in a portfolio selection problem for a representative investor. This approach allows us to compare how much of her wealth the investor is willing to give up in order to have access to the volatility predictions of the more sophisticated model.

We follow the volatility prediction literature and define the benchmark model as the heterogeneous autoregressive realized volatility (HAR) model introduced by Corsi (2009) augmented by the (log) VIX level, the turnover-ratio and dummy variables for news releases and earnings calls. Specifically, building upon the regression framework used in this study, we obtain predictions from the following fixed effects panel HAR model:

$$\log RV_{i,t+1} = c_i + \theta^{(d)} \log RV_{i,t} + \theta^{(w)} \log RV_{i,t}^{(w)} + \theta^{(m)} \log RV_{i,t}^{(m)} + X'_{i,t} \theta' + \epsilon_{i,t+1}, \quad (2.6)$$

where $X_{i,t}$ contains the (log) VIX, the turnover-ratio and dummy variables for news releases. More precisely, we include dummies for earnings calls, days without any news articles, days with relevant news releases and not relevant ones.²⁶ The second model additionally includes the measure of retail investors' attention interacted with the news dummies. We estimate the fixed effects panel models using the within-transformation over a rolling window of four years and produce one-day ahead forecasts of (log) realized volatility over the following year. The starting estimation window ranges from January 2011 to December 2014. In order to obtain true ex-ante predictions, all the involved variables are constructed using only data from the rolling window. For example, when removing trend and seasonal components from the (log) posting volumes, we estimate

²⁶We categorized a day to only have not relevant news releases when the cumulative relevance score across all articles published for a specific company is zero. Recall that the relevance score is defined as the total cosine-similarity between news headlines and messages shared on social media (see Section 2.4.2).

the regression coefficients using only data from the estimation window.

Results regarding the improvements in prediction accuracy achieved by accounting for retail investors' attention when predicting one-day ahead (log) realized volatility are relegated to Appendix 2.D. On average, retail investors' attention reduces the mean-squared prediction error by 0.79%. Moreover, the inclusion of the attention measure based on StockTwits data improves the predictive accuracy for 316 stocks, and for 196 these improvements are statistically significant at the 5% level.

2.6.1 Evaluating the economic value of volatility predictions

The evaluation of the economic value of the volatility predictions closely follows the approach introduced by Bollerslev et al. (2018). In particular, we extend their approach to a multivariate setting, i.e. for multiple risky assets. We assume that a representative investor has mean-variance preferences:

$$\mathbb{E}_t[u(W_{t+1})] = \mathbb{E}_t[W_{t+1}] - \frac{k}{2} \text{Var}_t[W_{t+1}],$$

where $\mathbb{E}_t[\cdot]$ and $\text{Var}_t[\cdot]$ denote the conditional expectation and variance, W_{t+1} is the investor's wealth and k denotes the absolute risk aversion. In each trading period, the investor can allocate her wealth in a risk-free asset and n risky stocks. Her wealth therefore evolves as $W_{t+1} = W_t(1 + (1 - \boldsymbol{\omega}_t^\top \mathbf{1})R_{f,t+1} + \boldsymbol{\omega}_t^\top R_{t+1})$ where $\boldsymbol{\omega}_t$ is a n -dimensional vector of the shares of wealth invested in the risky assets, R_t is a random vector

of the risky assets' returns, and $R_{f,t}$ is the risk-free rate. Assuming that the relative risk aversion $\gamma = kW_t$ is constant, the expected utility the investor wishes to maximize is:

$$U(\boldsymbol{\omega}_t) = W_t \left(1 + R_{f,t+1} + \boldsymbol{\omega}_t^\top \mathbb{E}_t[R_{t+1}^e] - \frac{\gamma}{2} \boldsymbol{\omega}_t^\top \text{Cov}_t[R_{t+1}^e] \boldsymbol{\omega}_t \right),$$

where $R_{t+1}^e = R_{t+1} - R_{f,t+1}$ is the vector of excess returns of the risky assets and $\text{Cov}_t[R_{t+1}^e]$ is their conditional covariance matrix. As in Bollerslev et al. (2018), we assume that the Sharpe ratio of each asset is constant over time, but we allow for different sharp ratios across different sectors.²⁷ The conditional expected excess returns are $\mathbb{E}_t[R_{t+1}^e] = SR \odot \mathbb{E}_t[RV_{t+1}]$ where RV_{t+1} is a n -dimensional vector of realized volatilities, SR a n -dimensional vector of Sharpe ratios, and $\odot$ denotes the element-wise product of the two vectors. In this way, we avoid influencing the economic evaluation of volatility predictions by the modeling of conditional expected excess returns. Moreover, we assume that the risky assets have a constant correlation matrix Γ , and that the conditional covariance matrix is therefore $\text{Cov}_t[R_{t+1}^e] = \text{diag}(\mathbb{E}_t[RV_{t+1}]) \Gamma \text{diag}(\mathbb{E}_t[RV_{t+1}])$ where $\text{diag}(x)$ stands for the diagonal matrix with diagonal entries given by the vector x .

The shares of wealth invested in the risky assets that maximize the

²⁷In the empirical application, we estimate sector-specific Sharpe ratios. Sectors are determined based on Standard & Poor's Global Industry Classification Standard (GICS).

expected utility are:

$$\omega_t^* = \frac{1}{\gamma} \text{diag}(\mathbb{E}_t[RV_{t+1}]^{-1}) \Gamma^{-1} SR$$

and the optimal allocation implied by the predictions of a model is then:

$$\hat{\omega}_t = \frac{1}{\gamma} \text{diag}(\hat{\mathbb{E}}_t[RV_{t+1}]^{-1}) \Gamma^{-1} SR,$$

where $\hat{\mathbb{E}}_t[\cdot]$ denotes the conditional expectation from a model, i.e. the one-day ahead prediction.²⁸ The conditional expected utility obtained with the asset allocation $\hat{\omega}_t$ is:

$$\begin{aligned} U(\hat{\omega}_t) = & W_t \left(1 + R_{f,t+1} + \frac{1}{\gamma} SR^\top \Gamma^{-1} \text{diag} \left(\frac{\mathbb{E}_t[RV_{t+1}]}{\hat{\mathbb{E}}_t[RV_{t+1}]} \right) SR \right. \\ & \left. - \frac{1}{2\gamma} SR^\top \Gamma^{-1} \text{diag} \left(\frac{\mathbb{E}_t[RV_{t+1}]}{\hat{\mathbb{E}}_t[RV_{t+1}]} \right) \Gamma \text{diag} \left(\frac{\mathbb{E}_t[RV_{t+1}]}{\hat{\mathbb{E}}_t[RV_{t+1}]} \right) \Gamma^{-1} SR \right) \end{aligned}$$

and the expected utility achieved with the asset allocation ω_t^* is:

$$U(\omega_t^*) = W_t \left(1 + R_{f,t+1} + \frac{1}{2\gamma} SR^\top \Gamma^{-1} SR \right).$$

It is easily seen that when only one risky asset is available, we obtain the same results as in Bollerslev et al. (2018). In order to economically evaluate the utility achieved with the predictions made by a model, Bollerslev et al. (2018) propose computing the empirical expected utility

²⁸Note that we predict the log realized volatility, i.e. $\hat{\mathbb{E}}_t[\log RV_{t+1}]$. We approximate the prediction of the realized volatility by $\hat{\mathbb{E}}_t[RV_{t+1}] \approx \exp \hat{\mathbb{E}}_t[\log RV_{t+1}]$.

per unit of wealth by its realized average:

$$\begin{aligned} \overline{UoW}(\{\hat{\omega}_t\}_{t=0}^{T-1}) = \\ T^{-1} \sum_{t=0}^{T-1} \left(1 + R_{f,t+1} + \frac{1}{\gamma} SR^\top \Gamma^{-1} \text{diag}\left(\frac{RV_{t+1}}{\hat{E}_t[RV_{t+1}]}\right) SR \right. \\ \left. - \frac{1}{2\gamma} SR^\top \Gamma^{-1} \text{diag}\left(\frac{RV_{t+1}}{\hat{E}_t[RV_{t+1}]}\right) \Gamma \text{diag}\left(\frac{RV_{t+1}}{\hat{E}_t[RV_{t+1}]}\right) \Gamma^{-1} SR \right). \end{aligned}$$

The economic value of the predictions $\hat{E}_t[RV_{t+1}]$ can then be quantified by comparing the above realized average utility with that obtained when using the volatility predictions of a competing model. The value thus obtained can be interpreted as the fraction of wealth an investor with mean-variance preference is willing to give up for the volatility predictions produced by a more sophisticated model.

This approach can be further extended by taking into account transaction costs. Specifically, we assume that the costs for trading in the assets are linear in the absolute magnitude of the change in the respective position, i.e. $|\hat{\omega}_{t,i} - \hat{\omega}_{t-1|t,i}|$, where $\hat{\omega}_{t-1|t,i}$ is the share of wealth invested in asset i just before changing the asset allocation.²⁹ The expected utility a mean-variance investor is maximizing becomes:

$$\begin{aligned} U(\omega_t) = & W_t \left(1 + R_{f,t+1} + \omega_t^\top E_t[R_{t+1}^e] - \frac{\gamma}{2} \omega_t^\top \text{Cov}_t[R_{t+1}^e] \omega_t \right. \\ & \left. - |\omega_t - \omega_{t-1|t}|^\top TC_t \right), \end{aligned}$$

²⁹At time t , the share of wealth invested at time $t-1$ in asset i is $\omega_{t-1|t,i} = \omega_{t-1,i}(1 + R_{t,i})/(1 + R_t)$, where $R_{t,i}$ is the return of asset i from $t-1$ to t , and R_t is the portfolio return over the same time period.

where TC_t is a n -dimensional vector of transaction costs. Since the objective function is no longer differentiable, we obtain the optimal allocation by maximizing the expected utility with a cyclic coordinate descent algorithm (see Appendix 2.E).

2.6.2 The value of attention for news-event portfolios

In the following, we analyze the value of retail investors' attention for volatility predictions in a multivariate setting, i.e. a mean-variance investor can allocate her wealth in several risky assets and a risk-free asset. We build specific news-event portfolios that invest exclusively in stocks for which a particular news topic or type has been published on the day the asset allocation is determined. More precisely, on each trading day, the investor only considers those companies for which a news article about a specific topic (e.g. "analyst ratings") or of a particular type (e.g. "press release") has been released on that day. When no stock is subject to a specific news event, the investor stays out of the markets and invests her entire wealth in the risk-free asset. We determine the economic value of accounting for retail investors' attention by comparing the average realized utility achieved by the two volatility prediction models.

The risk-free rate is the effective federal fund rate. The Sharpe-ratio is estimated over the rolling estimation window at the industry level and kept constant over the following year. Similarly, the correlation matrix is estimated using daily returns in the rolling estimation window and kept fixed for the following year. Following Bollerslev et al. (2018),

Table 2.7: The economic value of volatility predictions for event-based portfolios

	Num. events	Transaction costs	Performance fee in bps	t-statistic
Not relevant news	528	No	13.67	2.47
		Yes	4.30	1.55
News flash	754	No	118.36	8.53
		Yes	56.89	8.29
Press release	753	No	21.96	2.03
		Yes	14.10	4.17
Acquisitions and mergers	680	No	71.28	5.73
		Yes	45.30	6.08
Analyst ratings	702	No	54.52	7.11
		Yes	37.82	7.07
Revenues	335	No	31.44	5.11
		Yes	17.14	4.88
Earnings call	528	No	44.53	6.24
		Yes	35.53	6.69

NOTE: The table summarizes the fraction of wealth an investor with mean-variance preferences is willing to give up for the volatility predictions obtained from the attention-based model compared to those obtained from the benchmark model (annualized in basis points). The investor only trades in stocks subject to specific news types and topics. The first column reports for how many days at least one company is subject to that event. The second column indicates whether transaction costs have been taken into account or not. The last column reports t-statistics of the performance fees, using Newey-West standard errors.

we estimate transaction costs with the median bid-ask spread over the previous nine months. In line with the literature, we set the relative risk aversion to be 2.

Table 2.7 summarizes the annual performance fee (in basis points) the investor is willing to pay for attention-based volatility predictions for different news-event portfolios. The first column reports for how many days at least one company is subject to a specific news event. The second column indicates whether transaction costs have been taken into account, the third reports the performance fee and the last one t-statistics of the difference in average realized utility using Newey-West standard errors.

When trading only in companies with irrelevant news releases to which retail investors do not pay attention, the performance fee the investor is willing to pay is rather small and amounts to 13.67 basis points. In fact,

when also considering transaction costs, the fee becomes even smaller and is no longer significant at any conventional level. For the other news-events, the fees are large and highly significant. When investing in stocks for which a news article about acquisition and mergers has been released, the investor is willing to give up 71.28 basis points per year of her wealth, and 45.30 when considering transaction costs for predictions based on attention measures.

These findings are confirmed by cross-sectional results obtained at the company level. More precisely, we also analyze the economic value of retail investors' attention for a mean-variance investor who can allocate her wealth in either the risk-free or a single risky asset, the latter being one of the 360 companies considered in our study. The results are reported in Table 2.10 in Appendix 2.A. In particular, we find that attention-based volatility forecasts lead to a higher average realized utility for 345 companies, and for 237 the improvement is significant at the 5% significance level of a two-sided test. Moreover, the cross-sectional results reflect the same heterogeneities presented previously. When no news or only irrelevant news is released, the economic value of attention is low. Similarly, when articles about "credit" or "corporate responsibility" are released, the value of taking into account retail investors' attention is low. In contrast, the mean-variance investor is willing to give up more of her wealth for attention-based forecasts when, for example, a "news flash" or an article about "acquisition and mergers" is released.

In summary, the news-event portfolios clearly show that taking into account retail investors' attention around news releases for predicting

volatility is valuable in economic terms. These results strengthen the findings presented in this paper showing that our hypotheses hold not only in a classical fixed effect panel regression framework but also translate into economically meaningful volatility predictions.

2.7 Conclusion

In this study, we introduce three hypotheses related to the behavioral finding of people's limited cognitive ability. More precisely, we investigated the hypotheses that (i) when investors pay attention to news releases, the contemporaneous volatility increases, (ii) the volatility observed in the subsequent trading days is positively associated with higher attention of retail investors and negatively with that of institutional investors, and finally (iii) that this relation is heterogeneous in the type of the information release and the topic covered by the news. Using a novel measure of retail investors' attention obtained from the daily posting volume on social media platforms, we find significant evidence in favor of all three hypotheses.

Our findings show that, consistent with the noise trader models, when retail investors are paying attention to news releases, the price-adjustment to the new information is slowed down and the market risk increases. In contrast, institutional investors' attention to the release of news fosters price-adjustments and reduces the volatility in the post-announcement period.

Building upon these findings, we propose a volatility prediction model

that accounts for retail investors' attention. To quantify the economic value of these predictions, we extend the approach introduced by Bollerslev et al. (2018) to a multivariate setting. In particular, our findings show that around the release of unexpected news or news about specific topics (e.g. analyst ratings, acquisition and mergers), volatility predictions that take into account retail investors' attention have significant economic value for an investor with mean-variance preferences.

In summary, our findings have several implications for the academic community and practitioners. First, we show that the attention of sophisticated institutional investors is of central importance for the price-adjustment speed to new information. This finding is also of relevance for the asset pricing literature in that prices do not instantaneously reflect all available information. Moreover, the type of news release and the topic of the new information have an influence on how prices react when retail investors are paying attention. These findings might be incorporated in models of investors' inattention by allowing for a heterogeneous reaction to different news releases. Finally, the findings presented in this paper contribute to the volatility prediction literature by showing that measures of retail investors' attention around the release of news significantly improve forecasts.

Acknowledgements

The research project is funded by the Swiss National Science Foundation (Grant 100018-172685).

References

- Aboudy, D., Lehavy, R., & Trueman, B. (2010). Limited attention and the earnings announcement returns of past stock market winners. *Review of Accounting Studies*, 15(2), 317–344.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43–76.
- Andrei, D., & Hasler, M. (2014). Investor Attention and Stock Market Volatility. *The Review of Financial Studies*, 28(1), 33–72.
- Antweiler, W., & Frank, M. Z. (2004). Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59(3), 1259–1294.
- Arellano, M. (1987). Computing Robust Standard Errors for Within-groups Estimators. *Oxford Bulletin of Economics and Statistics*, 49(4), 431–434.
- Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36(2), 334–357.
- Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 785–818.
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307–343.

- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3), C1–C32.
- Ben-Rephael, A., Carlin, B. I., Da, Z., & Israelsen, R. D. (in press). Information consumption and asset pricing. *The Journal of Finance*.
- Ben-Rephael, A., Da, Z., & Israelsen, R. D. (2017). It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *The Review of Financial Studies*, 30(9), 3009–3047.
- Black, F. (1986). Noise. *The Journal of Finance*, 41(3), 528–543.
- Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk Everywhere: Modeling and Managing Volatility. *The Review of Financial Studies*, 31(7), 2729–2773.
- Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2019). Information, Trading, and Volatility: Evidence from Firm-Specific News. *The Review of Financial Studies*, 32(3), 992–1033.
- Brown, S. V., & Tucker, J. W. (2011). Large-Sample Evidence on Firms’ Year-over-Year MD&A Modifications. *Journal of Accounting Research*, 49(2), 309–346.
- Buncic, D., & Gisler, K. I. (2016). Global equity market volatility spillovers: A broader role for the United States. *International Journal of Forecasting*, 32(4), 1317–1339.
- Bushee, B. J., Core, J. E., Guay, W., & Hamm, S. J. (2010). The Role of the Business Press as an Information Intermediary. *Journal of Accounting Research*, 48(1), 1–19.
- Cookson, J. A., & Niessner, M. (2020). Why Don’t We Agree? Evidence

- from a Social Network of Investors. *Journal of Finance*, 75(1), 173–228.
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), 1461–1499.
- Da, Z., Engelberg, J., & Gao, P. (2014). The Sum of All FEARS Investor Sentiment and Asset Prices. *The Review of Financial Studies*, 28(1), 1–32.
- DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *The Journal of Finance*, 64(2), 709–749.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4), 703–738.
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263.
- Dimpfl, T., & Jank, S. (2016). Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management*, 22(2), 171–192.
- Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor Information Demand: Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research*, 50(4), 1001–1040.
- Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2015). The Determinants and Consequences of Information Acquisition via EDGAR.

- Contemporary Accounting Research*, 32(3), 1128–1161.
- Gabaix, X. (2019). Chapter 4 - Behavioral inattention. In *Handbook of behavioral economics - foundations and applications 2* (Vol. 2, pp. 261–343). North-Holland.
- Garcia, D. (2013). Sentiment during Recessions. *The Journal of Finance*, 68(3), 1267–1300.
- Goddard, J., Kita, A., & Wang, Q. (2015). Investor attention and fx market volatility. *Journal of International Financial Markets, Institutions and Money*, 38, 79–96.
- Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, 18(2), 321–340.
- Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and google trends. *Journal of Economic Behavior & Organization*, 117, 62–81.
- Hansen, P. R., & Lunde, A. (2005). A Realized Variance for the Whole Day Based on Intermittent High-Frequency Data. *Journal of Financial Econometrics*, 3(4), 525–554.
- Hirshleifer, D., Lim, S. S., & Teoh, S. H. (2009). Driven to Distraction: Extraneous Events and Underreaction to Earnings News. *The Journal of Finance*, 64(5), 2289–2325.
- Hirshleifer, D., & Sheng, J. (2019). *The Attention Trigger Effect: Macro News and Efficient Processing of Micro News*. Retrieved from <https://ssrn.com/abstract=2890864> (Working paper)

- Kahneman, D. (1973). *Attention and effort* (Vol. 1063). Englewood Cliffs, NJ: Prentice-Hall.
- Lee, C. M., Ma, P., & Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, 116(2), 410–431.
- Lerman, A. (2020). Individual investors’ attention to accounting information: Evidence from online financial communities. *Contemporary Accounting Research*, *Forthcoming*.
- Liu, H., Peng, L., & Tang, Y. (2019). *Investor Attention: Endogenous Allocations, Clientele Effects, and Asset Pricing Implications*. (Working paper)
- Livnat, J., & Mendenhall, R. R. (2006). Comparing the Post-Earnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts. *Journal of Accounting Research*, 44(1), 177–205.
- Mittnik, S., Robinsonov, N., & Spindler, M. (2015). Stock market volatility: Identifying major drivers and the nature of their impact. *Journal of Banking & Finance*, 58(Supplement C), 1–14.
- Morse, D. (1982). Wall Street Journal Announcements and the Securities Markets. *Financial Analysts Journal*, 38(2), 69–76.
- Peng, L. (2005). Learning with Information Capacity Constraints. *The Journal of Financial and Quantitative Analysis*, 40(2), 307–329.
- Peng, L., & Xiong, W. (2006). Investor attention, overconfidence and category learning. *Journal of Financial Economics*, 80(3), 563–602.
- Rakowski, D., Shirley, S. E., & Stark, J. R. (2018). Twitter activity,

investor attention, and the diffusion of information. *Financial Management*.

Rubin, A., & Rubin, E. (2010). Informed investors and the internet.

Journal of Business Finance & Accounting, 37(7–8), 841–865.

Ryan, P., & Taffler, R. J. (2004). Are Economically Significant Stock Returns and Trading Volumes Driven by Firm-specific News Releases?

Journal of Business Finance & Accounting, 31(1–2), 49–82.

Shleifer, A., & Vishny, R. W. (1997). The Limits of Arbitrage. *The*

Journal of Finance, 52(1), 35–55.

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6),

1808–1821.

Appendix

2.A Tables

Table 2.8: Investors' attention and volatility reaction to news controlling for weekday-effects

Panel A: Contemporaneous relation between attention and volatility		
	Estimate	t-statistic
Retail investor attention		
No news	0.028	18.58
News, no earnings call	0.042	17.36
Earnings calls	0.036	10.49
Institutional investor attention		
No news	0.038	19.95
News, no earnings call	0.048	17.59
Earnings calls	0.033	11.11
Adjusted R^2	59.07%	
Number observations	587,762	
Panel B: Relation between attention and next day's volatility		
	Estimate	t-statistic
Retail investor attention		
No news	0.013	22.62
News, no earnings call	0.022	27.84
Earnings calls	0.033	5.46
Institutional investor attention		
No news	0.000	-0.40
News, no earnings call	0.000	0.11
Earnings calls	0.005	0.96
Adjusted R^2	54.23%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase on the news release day (Panel A) and the day after (Panel B) associated with a one standard deviation increase in retail and institutional investors' attention, controlling for potential weekday-effects. More precisely, Panel A summarizes the attention estimates of the regression defined in Equation (2.1), and Panel B those of the regression defined in Equation (2.2) with $h = 1$. For completeness, the table also reports the estimated increase in contemporaneous and future volatility associated with a one standard deviation increase in attention when no news article is released. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

Table 2.9: Investors' attention to news topics and next day's volatility reaction

	Estimate	t-statistic
Retail investor attention		
<i>No news</i>	<i>0.012</i>	<i>21.57</i>
<i>Not relevant news</i>	<i>0.010</i>	<i>13.05</i>
Insider trading	0.010	5.95
Order imbalances	0.013	5.80
Credit	0.016	1.88
Credit ratings	0.020	4.29
Corporate responsibility	0.024	1.24
Technical analysis	0.026	16.57
Investor relations	0.027	8.02
Partnerships	0.028	5.67
Dividends	0.029	7.01
Equity actions	0.029	5.48
Marketing	0.030	5.29
Products and services	0.030	9.87
Legal	0.033	3.83
Assets	0.035	4.62
Price targets	0.041	6.05
Labor issues	0.041	10.36
Stock prices	0.055	9.81
Analyst ratings	0.066	23.65
Acquisitions and mergers	0.085	18.19
Regulatory	0.088	6.13
Earnings	0.088	12.75
Revenues	0.104	7.35
<i>Earnings call</i>	<i>0.031</i>	<i>4.93</i>

NOTE: The table reports the estimated volatility increase the day after a news release associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, the table reports the coefficients for attention measures obtained from the estimation of the regression defined in Equation (2.3) with $h = 1$. In particular, we distinguish between different news topics defined as the most discussed articles by retail investors. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

Table 2.9 continued: Investors' attention to news topics and next day's volatility reaction

	Estimate	t-statistic
Institutional investor attention		
<i>No news</i>	<i>0.002</i>	<i>3.68</i>
<i>Not relevant news</i>	<i>0.003</i>	<i>3.26</i>
Insider trading	0.011	8.21
Order imbalances	-0.007	-3.02
Credit	-0.009	-1.45
Credit ratings	-0.004	-1.10
Corporate responsibility	-0.002	-0.09
Technical analysis	0.002	1.20
Investor relations	-0.004	-1.62
Partnerships	-0.007	-1.44
Dividends	0.005	1.42
Equity actions	0.002	0.32
Marketing	0.006	1.44
Products and services	0.002	0.85
Legal	0.001	0.20
Assets	0.002	0.29
Price targets	-0.004	-0.70
Labor issues	0.005	1.58
Stock prices	-0.022	-4.24
Analyst ratings	-0.016	-6.11
Acquisitions and mergers	-0.017	-3.93
Regulatory	0.005	0.33
Earnings	0.000	0.02
Revenues	0.002	0.21
<i>Earnings call</i>	<i>0.005</i>	<i>0.90</i>
Adjusted R^2	53.94%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase the day after a news release associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, the table reports the coefficients for attention measures obtained from the estimation of the regression defined in Equation (2.3) with $h = 1$. In particular, we distinguish between different news topics defined as the most discussed articles by retail investors. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

Table 2.10: The economic value of retail investors' attention in predicting volatility (effective annual performance fee in basis points)

	N	TC	$Q_{25\%}$	Avg.	Med.	$Q_{75\%}$
Overall	754	No	1.43	5.17	3.44	7.41
		Yes	-0.13	2.89	0.75	3.49
No news and earnings calls						
Not relevant news	85	No	-0.02	0.17	0.04	0.19
		Yes	-0.10	0.08	0.00	0.13
Earnings call	12	No	-0.02	0.28	0.09	0.42
		Yes	-0.02	0.43	0.15	0.56
News types						
Tabular material	77	No	-0.01	0.44	0.25	0.78
		Yes	-0.07	0.38	0.01	0.31
Full article	71	No	0.02	0.66	0.32	0.95
		Yes	-0.05	0.28	0.03	0.28
Press release	36	No	-0.01	0.33	0.12	0.47
		Yes	-0.04	0.13	0.00	0.09
News flash	41	No	0.22	1.49	0.64	1.74
		Yes	-0.03	0.69	0.04	0.53
Hot news flash	6	No	0.24	0.56	0.55	0.62
		Yes	-0.03	0.10	0.01	0.06
News topics						
Insider trading	60	No	-0.01	0.36	0.17	0.52
		Yes	-0.05	0.10	0.01	0.16
Credit	8	No	-0.02	0.01	-0.01	0.02
		Yes	-0.01	0.01	0.00	0.01
Credit ratings	9	No	-0.05	0.04	0.01	0.07
		Yes	-0.03	0.04	0.00	0.02
Corporate responsibility	7	No	-0.02	-0.01	-0.01	0.01
		Yes	-0.04	-0.03	-0.03	-0.03
Technical analysis	72	No	-0.01	0.40	0.21	0.71
		Yes	-0.07	0.37	0.01	0.28
			:	:	:	:
			:	:	:	:
Stock prices	14	No	0.02	0.91	0.28	1.20
		Yes	-0.01	0.64	0.07	0.55
Analyst ratings	12	No	0.05	0.42	0.20	0.51
		Yes	-0.02	0.14	0.01	0.10
Acquisitions and mergers	11	No	0.04	0.78	0.21	0.83
		Yes	-0.02	0.42	0.01	0.18
Regulatory	6	No	0.29	0.46	0.50	0.67
		Yes	0.05	0.10	0.11	0.16
Earnings	8	No	0.00	0.11	0.05	0.17
		Yes	-0.04	0.15	0.00	0.03
Revenues	18	No	0.03	1.21	0.11	1.54
		Yes	0.00	0.83	0.03	1.00

NOTE: The table shows summary statistics of the fraction of wealth an investor is willing to give up in order to obtain volatility predictions from the attention-based model compared to the benchmark model. The second column summarizes the average number of observations, and the column TC indicates whether transaction costs have been taken into account. For each grouping criterion, we only consider companies which have at least five observations in that category.

2.B Results using Twitter attention

Table 2.11: Investors' attention and volatility reaction using Twitter data

Panel A: Contemporaneous relation between attention and volatility		
	Estimate	t-statistic
Retail investor attention		
No news	0.020	17.12
News, no earnings call	0.030	15.45
Earnings calls	0.015	4.15
Institutional investor attention		
No news	0.038	18.78
News, no earnings call	0.048	16.92
Earnings calls	0.032	10.49
Adjusted R^2	58.03%	
Number observations	587,762	
Panel B: Relation between attention and next day's volatility		
	Estimate	t-statistic
Retail investor attention		
No news	0.009	14.31
News, no earnings call	0.018	23.48
Earnings calls	0.003	0.48
Institutional investor attention		
No news	0.002	3.60
News, no earnings call	0.003	4.19
Earnings calls	0.005	0.99
Adjusted R^2	53.69%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase on the news release day (Panel A) and the day after (Panel B) associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, Panel A summarizes the attention estimates of the regression defined in Equation (2.1), and Panel B those of the regression defined in Equation (2.2) with $h = 1$. For completeness, the table also reports the estimated increase in contemporaneous and future volatility associated with a one standard deviation increase in attention when no news article is released. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). Retail investors' attention is measured with Twitter data.

Table 2.12: Investors' attention to news types and next day's volatility reaction using Twitter data

	Estimate	t-statistic
Retail investor attention		
<i>No news</i>	<i>0.009</i>	<i>14.35</i>
<i>Not relevant news</i>	<i>0.007</i>	<i>5.16</i>
Tabular material	0.013	12.70
Full article	0.012	8.49
Press release	0.022	15.42
News flash	0.030	20.05
Hot news flash	0.060	4.57
<i>Earnings calls</i>	<i>0.003</i>	<i>0.46</i>
Institutional investor attention		
<i>No news</i>	<i>0.002</i>	<i>3.68</i>
<i>Not relevant news</i>	<i>0.000</i>	<i>0.10</i>
Tabular material	0.004	4.82
Full article	0.007	6.08
Press release	0.003	2.57
News flash	-0.004	-3.85
Hot news flash	0.009	0.53
<i>Earnings calls</i>	<i>0.005</i>	<i>0.99</i>
Adjusted R^2	53.72%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase the day after a news release associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, the table reports the coefficients for attention measures obtained from the estimation of the regression defined in Equation (2.3) with $h = 1$. In particular, we distinguish between different news types defined as the most discussed articles by retail investors. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987).

Table 2.13: Investors' attention to news topics and next day's volatility reaction using Twitter data

	Estimate	t-statistic
Retail investor attention		
<i>No news</i>	0.009	14.36
<i>Not relevant news</i>	0.006	4.72
Credit ratings	0.003	0.78
Insider trading	0.007	5.18
Order imbalances	0.009	6.27
Credit	0.010	1.47
Legal	0.012	2.02
Investor relations	0.012	4.59
	:	:
Stock prices	0.040	6.93
Regulatory	0.056	3.83
Analyst ratings	0.056	18.29
Earnings	0.057	10.22
Acquisitions and mergers	0.065	17.30
Revenues	0.084	7.25
<i>Earnings call</i>	0.003	0.49
Institutional investor attention		
<i>No news</i>	0.002	3.82
<i>Not relevant news</i>	0.000	0.11
Credit ratings	-0.003	-1.12
Insider trading	0.008	6.72
Order imbalances	-0.004	-2.55
Credit	-0.005	-0.81
Legal	0.007	1.18
Investor relations	0.003	1.63
	:	:
Stock prices	-0.018	-3.15
Regulatory	-0.005	-0.34
Analyst ratings	-0.013	-5.84
Earnings	0.005	1.23
Acquisitions and mergers	-0.019	-5.01
Revenues	0.011	1.30
<i>Earnings call</i>	0.005	1.00
Adjusted R^2	53.80%	
Number observations	588,104	

NOTE: The table reports the estimated volatility increase the day after a news release associated with a one standard deviation increase in retail and institutional investors' attention. More precisely, the table reports the coefficients for attention measures obtained from the estimation of the regression defined in Equation (2.3) with $h = 1$. In particular, we distinguish between different news topics defined as the most discussed articles by retail investors. The t-statistics are computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). Note that the table only reports partial results (the news articles with the largest absolute estimates for retail investors).

2.C Results with alternative relevance scores

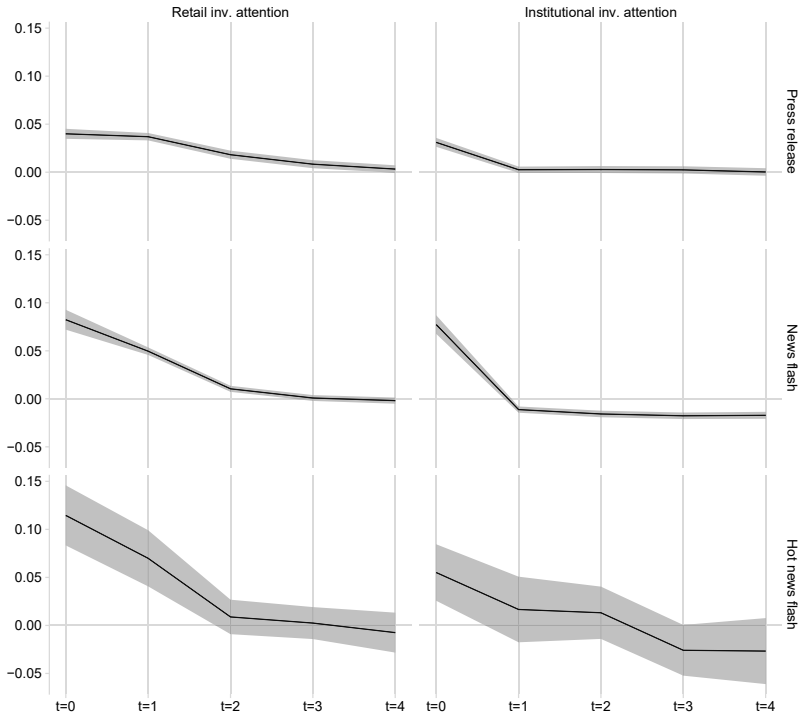


Figure 2.7: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around different news release types – Relevance score based on text data without company names and tickers

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news event day and the four post-release days for different news types. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). The relevance score of an article is defined as the sum of cosine similarities between the article's headline and StockTwits messages, removing from both texts company names and tickers.

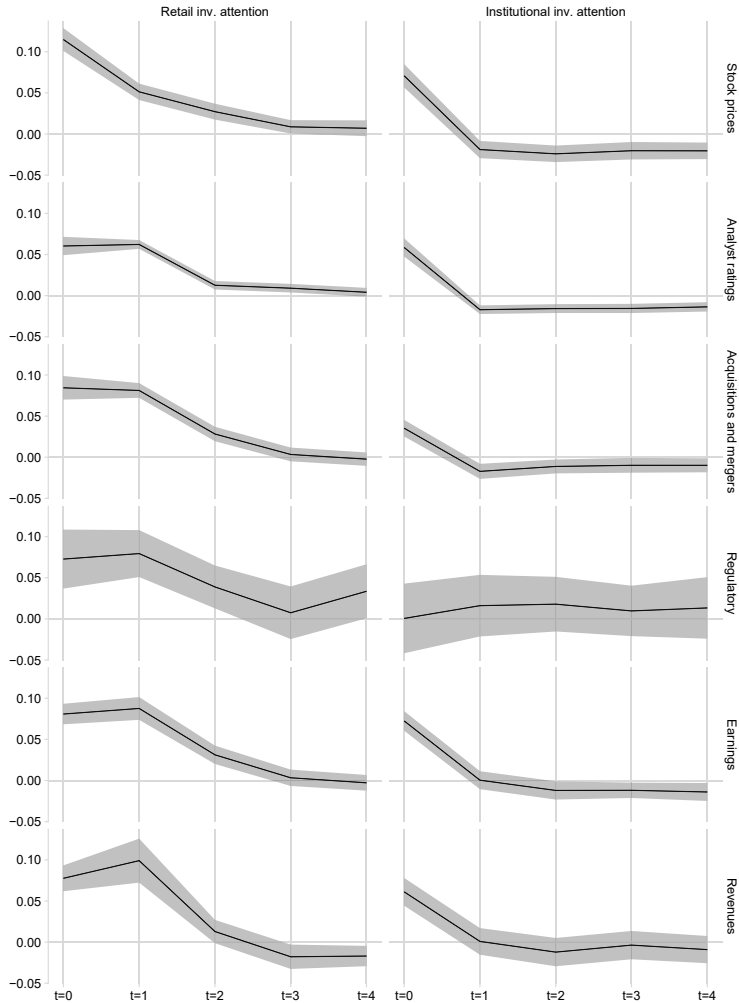


Figure 2.8: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around different news release topics – Relevance score based on text data without company names and tickers

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news-event day and the four post-release days for different news topics. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). The relevance score of an article is defined as the sum of cosine similarities between the article's headline and StockTwits messages, removing from both texts company names and tickers.

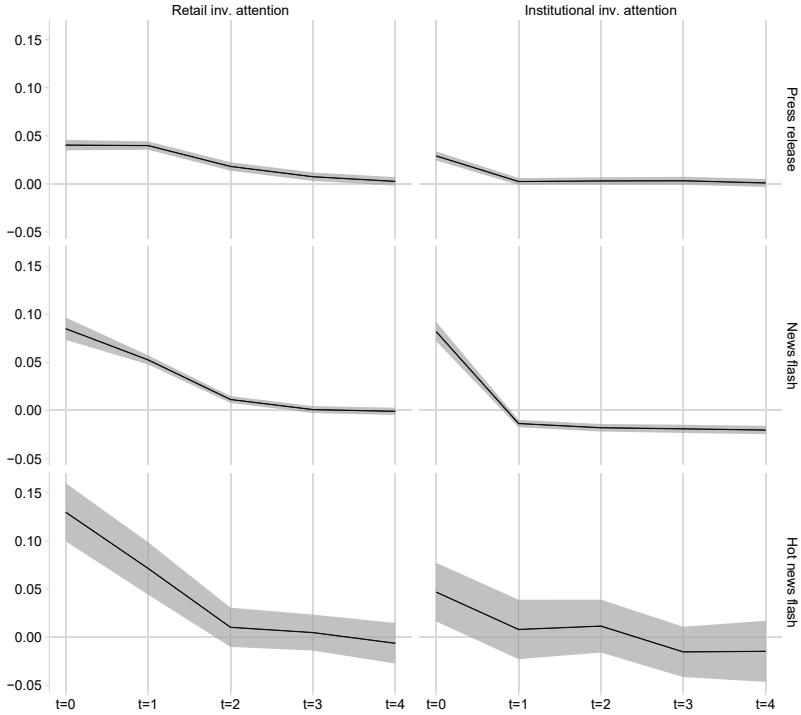


Figure 2.9: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around different news release types – Relevance score based on social media messages published after the news article and text data without company names and tickers

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news event day and the four post-release days for different news types. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). The relevance score of an article is defined as the sum of cosine similarities between the article's headline and StockTwits messages, removing from both texts company names and tickers. Moreover, only social media messages published after the news article are considered for the computation of relevance scores.

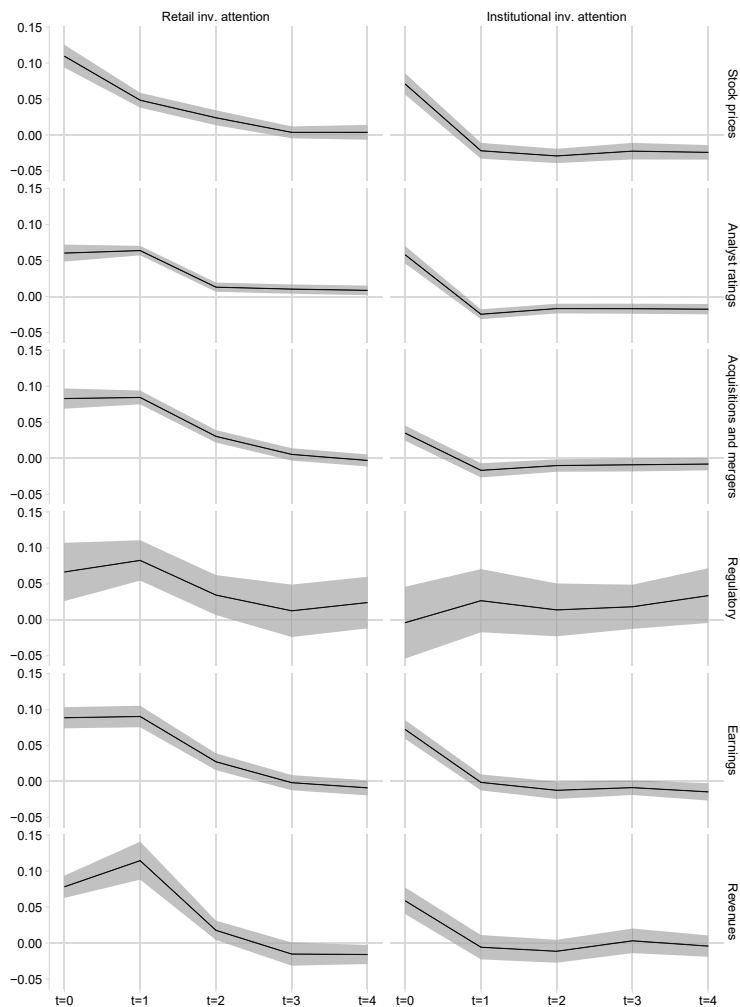


Figure 2.10: Change in volatility associated with a one standard deviation increase in retail and institutional investors' attention around different news release topics – Relevance score based on social media messages published after the news article and text data without company names and tickers

NOTE: The figure depicts the fixed effects panel regression coefficients of attention measures where the dependent variable is the realized volatility on the news-event day and the four post-release days for different news topics. The gray-shaded areas represent 95% confidence intervals computed using clustered standard errors obtained with the estimator introduced by Arellano (1987). The relevance score of an article is defined as the sum of cosine similarities between the article's headline and StockTwits messages, removing from both texts company names and tickers. Moreover, only social media messages published after the news article are considered for the computation of relevance scores.

2.D Predictive accuracy of attention-based volatility forecasts

In the following, we briefly present the results when analyzing the predictive power of retail investors' attention in forecasting volatility. The models and estimation techniques used are described at the beginning of Section 2.6. The accuracy of the forecasts is measured in terms of mean-squared prediction error (MSE). We assess the significance of the differences in prediction accuracy with the Diebold-Mariano test statistic (Diebold & Mariano, 1995). Table 2.14 reports summary statistics (quartiles, median and average) of reductions in MSE as well as corresponding Diebold-Mariano test statistics when comparing predictions from a model including attention variables with those obtained without attention variables for the cross-section of 360 stocks. Figure 2.11 depicts these results graphically.

Table 2.14: Predictive accuracy of retail investors' attention for one-day ahead realized volatility

	$Q_{25\%}$	Avg.	Med.	$Q_{75\%}$
MSE reduction (in %)	0.33	0.79	0.80	1.28
DM statistic	0.75	1.75	1.81	2.71

NOTE: The first summarizes summary statistics of the reduction in MSE (in %) of one-day ahead prediction of (log) realized volatility achieved by augmenting a benchmark model with retail investors' attention. The second row reports instead summary statistics of Diebold-Mariano test statistics for the difference in predictive accuracy.

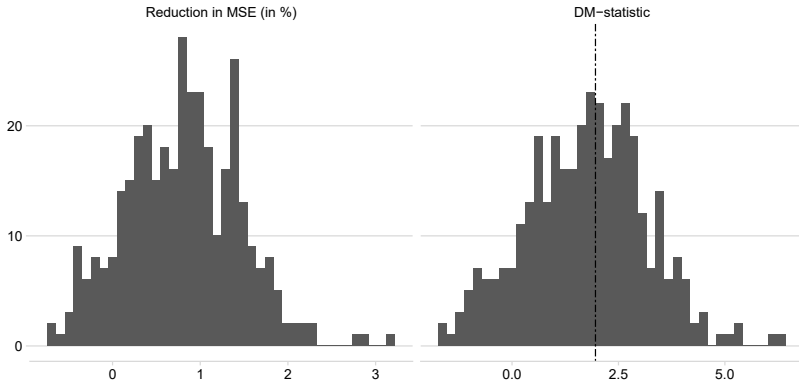


Figure 2.11: Improvements in volatility predictions achieved by accounting for retail investors' attention

NOTE: The figure depicts the percentual reduction in MSE (left) and the respective Diebold-Mariano statistics (right) achieved by the addition of retail investors' attention in a predictive volatility model. The vertical line indicates the threshold for the 5% significance level of a two-sided Diebold-Mariano test.

Both Table 2.14 and Figure 2.11 clearly show that for the great majority of companies, we improve the one-day ahead volatility forecasts by accounting for retail investors' attention. In fact, for 316 of the stocks, the MSE is reduced when including the attention measure based on StockTwits data, and for 196 the improvement is significant at the 5% level.

2.E Maximization of expected mean-variance utility with transaction costs

The objective function to be maximized is the expected mean-variance utility:

$$U(\boldsymbol{\omega}_t) = W_t \left(1 + R_{f,t+1} + \boldsymbol{\omega}_t^\top \mathbb{E}_t[R_{t+1}^e] - \frac{\gamma}{2} \boldsymbol{\omega}_t^\top \text{Cov}_t[R_{t+1}^e] \boldsymbol{\omega}_t - |\boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1|t}|^\top TC_t \right).$$

Dropping terms which do not affect the optimization and rewriting the objective function in terms of $\Delta\boldsymbol{\omega}_t = \boldsymbol{\omega}_t - \boldsymbol{\omega}_{t-1|t}$:

$$\begin{aligned} U(\Delta\boldsymbol{\omega}_t) &\propto (\boldsymbol{\omega}_{t-1|t} + \Delta\boldsymbol{\omega}_t)^\top \mathbb{E}_t[R_{t+1}^e] \\ &\quad - \frac{\gamma}{2} (\boldsymbol{\omega}_{t-1|t} + \Delta\boldsymbol{\omega}_t)^\top \text{Cov}_t[R_{t+1}^e] (\boldsymbol{\omega}_{t-1|t} + \Delta\boldsymbol{\omega}_t) \\ &\quad - |\Delta\boldsymbol{\omega}_t|^\top TC_t, \end{aligned}$$

where $\boldsymbol{\omega}_{t-1|t}$ are the shares invested in the risky assets just before determining the new allocation and therefore can be treated as being known and fixed.

We maximize $U(\Delta\boldsymbol{\omega}_t)$ with respect to $\Delta\boldsymbol{\omega}_t$ using a cyclic coordinate descent algorithm. Even though $U(\Delta\boldsymbol{\omega}_t)$ is non-differentiable, it has directional derivatives along each forward and backward coordinate. Let where v_k be the coordinate direction along which $\Delta\omega_{t,k}$ varies. The directional derivative of the expected utility along its forward coordinate

is:

$$\begin{aligned}
 \nabla_{v_k} U &= \lim_{\tau \downarrow 0} \frac{U(\Delta \omega_t + \tau v_k) - U(\Delta \omega_t)}{\tau} \\
 &= E_t[R_{t+1}^e] - \gamma \sum_{j=1}^n (w_{t-1|t,j} + \Delta \omega_{t,j}) \text{Cov}_t[R_{t+1,k}^e, R_{t+1,j}^e] \\
 &\quad + \begin{cases} -TC_{t,k} & \Delta \omega_t \geq 0 \\ TC_{t,k} & \Delta \omega_t < 0 \end{cases}
 \end{aligned}$$

Analogously, the directional derivative of the expected utility along its backward coordinate is:

$$\begin{aligned}
 \nabla_{-v_k} U &= \lim_{\tau \downarrow 0} \frac{U(\Delta \omega_t - \tau v_k) - U(\Delta \omega_t)}{\tau} \\
 &= -E_t[R_{t+1}^e] + \gamma \sum_{j=1}^n (w_{t-1|t,j} + \Delta \omega_{t,j}) \text{Cov}_t[R_{t+1,k}^e, R_{t+1,j}^e] \\
 &\quad + \begin{cases} TC_{t,k} & \Delta \omega_t > 0 \\ -TC_{t,k} & \Delta \omega_t \leq 0 \end{cases}
 \end{aligned}$$

For $k = 1, \dots, n$, we then iteratively update $\Delta \omega_{t,k}$ with:

$$\Delta \omega_{t,k} = \Delta \tilde{\omega}_{t,k}^+ + \Delta \tilde{\omega}_{t,k}^-,$$

where $\Delta \tilde{\omega}_{t,k}^+$ and $\Delta \tilde{\omega}_{t,k}^-$ are the candidate solutions to the right and left

of zero (only one of them is non-zero). They are defined as:

$$\Delta\tilde{\omega}_{t,k}^+ = \max \left(0, \Delta\omega_{t,k} + \frac{E_t[R_{t+1}^e] - \gamma \sum_{j=1}^n (w_{t-1|t,j} + \Delta\omega_{t,j}) \text{Cov}_t[R_{t+1,k}^e, R_{t+1,j}^e] - TC_{t,k}}{\gamma \text{Var}_t[R_{t+1,k}^e]} \right),$$

$$\Delta\tilde{\omega}_{t,k}^- = \min \left(0, \Delta\omega_{t,k} + \frac{E_t[R_{t+1}^e] - \gamma \sum_{j=1}^n (w_{t-1|t,j} + \Delta\omega_{t,j}) \text{Cov}_t[R_{t+1,k}^e, R_{t+1,j}^e] + TC_{t,k}}{\gamma \text{Var}_t[R_{t+1,k}^e]} \right).$$

If, for a given k , neither of the directional derivatives is positive, we skip the update for asset k . We repeat the iterative update of $\Delta\omega_t$ until the improvements in utility no longer exceed a tolerance level η .

Chapter 3

Retail investors' trading activity and
the predictability of stock return cor-
relations

Abstract

Considerable theoretical and empirical evidence links price comovements with the behavior of retail investors. Nevertheless, when predicting stock return correlations, research has focused on the leverage effect. We propose a new model of realized covariances that allows exogenous predictors to influence the correlation dynamics while ensuring the predicted matrices' positive definiteness. Using this model, the predictive power of retail investors' sentiment and attention for the correlations of 35 Dow Jones stocks is analyzed. We find retail investors' attention to have predictive power for return correlations, especially for longer forecasting horizons and during the COVID-19 pandemic. Value-at-risk forecasts confirm these results.

Keywords: realized correlation; realized covariance; retail investors; investor sentiment; investor attention

3.1 Introduction

In predicting stock return comovements, research has focused on the leverage effect, being the empirical evidence that stock prices correlate more during market downturns than upturns (Ang & Chen, 2002). In addition to these asymmetries, the academic literature has documented several other stylized facts about the correlation between stock returns. Two broad theories regarding return comovements have been analyzed (Barberis, Shleifer, & Wurgler, 2005). The traditional theory assumes that the market is frictionless and states that stock price correlations reflect the correlations in fundamental values. In contrast, sentiment-based theories assumes that market frictions exist and that stock price comovements are attributed to investors' behavior rather than to the comovement in fundamental values. For example, investors have been found to simplify investment decisions by grouping stocks into categories (Barberis & Shleifer, 2003). Changes in their aggregated sentiment induce comovements in the return of stocks belonging to the same category (Kumar & Lee, 2006). Similarly, when investors' attention to the market or an industry is higher than the focus on single stocks, stock returns are more correlated (Drake, Jennings, Roulstone, & Thornock, 2017; Huang, Huang, & Lin, 2019).

While there is considerable empirical evidence on the factors influencing stock price comovements, their (out-of-sample) predictive power for return covariances, excluding the leverage effect, has received little attention. Although the predictive content of retail investors' trading

behavior has been documented for stock return volatility (see, e.g., Audrino, Sigrist, & Ballinari, 2020), its forecasting power for correlations has not yet been analyzed. Numerous studies have proposed extensions of classical multivariate volatility models to account for asymmetries in volatilities and correlations (see, among others, Bollerslev, Patton, & Quaedvlieg, 2020; Cappiello, Engle, & Sheppard, 2006; de Goeij & Marquering, 2004; Francq & Zakoïan, 2012; McAleer, Hoti, & Chan, 2009). However, only a few models have been developed that allow exogenous variables to influence the prediction of the returns' covariance matrix (e.g. Audrino & Corsi, 2010; Audrino & Trojani, 2011; Bauer & Vorkink, 2011). There are two potential reasons for the lack of studies that analyze the predictive power of sentiment-based theories of price comovement: (i) these stylized facts do not have out-of-sample predictive power for the comovement of stock returns, or (ii) the literature lacks a suitable model to analyze their forecasting ability.

This paper focuses on the correlation between stock returns. First, we introduce a new model of realized correlations, allowing exogenous predictors to have a non-linear influence on the model's forecasts. Building on the model of Audrino and Trojani (2011), we propose a tree-structured version of the realized covariance HAR-DRD model of Oh and Patton (2016). Their model defines different dynamics for the realized volatilities and correlation matrix. Through a binary tree, we determine a partition of the predictor space and fit local versions of the HAR-DRD model's correlation dynamics over each partition cell. This approach allows exogenous predictors to directly influence the model's parameters by ensuring

the predicted covariance matrix's positive definiteness. Second, we use the new model to analyze the predictive power of retail investors' trading activity. Using high-frequency transaction data, we measure retail traders' daily sentiment and attention through their buy-sell imbalance and total trading volume, respectively. The findings indicate that for 35 U.S. stocks that are or have been part of the Dow Jones Industrial Average Index from January 2010 to June 2020, retail investors' aggregated attention has significant forecasting power for the realized correlation matrix. Even after controlling for the classical leverage effect, retail investors' attention still has incremental predictive power. Moreover, we find no evidence that retail investors' aggregated sentiment has predictive power for stock return correlations. Value-at-risks forecasts confirm these findings. Finally, the results show that retail traders' behavior is particularly valuable during the COVID-19 pandemic, a period characterized by a considerable increase in the number of retail investors.

This paper contributes to the existing literature in several ways. First, a new predictive model for realized covariance matrices that allows the inclusion of exogenous predictors is introduced. Our model is most closely related to the tree-structured multivariate DCC-GARCH model introduced by Audrino and Trojani (2011) and the tree-structure HAR model used by Audrino and Corsi (2010) to forecast the realized correlation between the S&P 500 futures and the 30-year Treasury Bond futures. However, the latter model is limited in the number of stocks because its forecasts are not guaranteed to provide positive definite correlation matrices for more than two assets. Second, to the best of our knowledge,

this is the first paper that analyzes the out-of-sample predictive power of sentiment-based theories of return comovements. It is the first study that considers whether retail investors' trading activity can improve the forecasting accuracy of the realized correlation matrix in statistical and economic terms. Finally, this is among the first studies to analyze the predictive power of multivariate volatility models during the COVID-19 pandemic.

The remainder of this paper is organized as follows. In Section 3.2, we specify the realized covariance model with exogenous predictors. Section 3.3 outlines the empirical methodology and describes the dataset. In Section 3.4, a summary of the estimation and prediction results is presented. Lastly, the paper concludes in Section 3.5.

3.2 A realized covariance model with exogenous predictors

3.2.1 Model specification

Inspired by the dynamic correlation multivariate GARCH model (Engle, 2002), Oh and Patton (2016) suggest to model the realized variances and correlations separately. Let RCov_t be the $N \times N$ realized covariance matrix computed as (see, e.g., Barndorff-Nielsen & Shephard, 2004):

$$\text{RCov}_t = \sum_{s=1}^m \mathbf{r}_{t-1+s \cdot \Delta} \mathbf{r}_{t-1+s \cdot \Delta}', \quad (3.1)$$

where $\mathbf{r}_{t-1+s\cdot\Delta} \in \mathbb{R}^N$ is a vector of high-frequency (log) returns computed at the frequency Δ and m is the number of intra-day observations. Let $\text{RVar}_t = \text{diag}(\text{RCov}_t)$ where $\text{diag}(\cdot)$ is the operator that sets the off-diagonal elements of a matrix to zero. The realized correlation matrix is defined as (see, e.g., Barndorff-Nielsen & Shephard, 2004):

$$\text{RCorr}_t = \text{RVar}_t^{-1/2} \text{RCov}_t \text{RVar}_t^{-1/2}. \quad (3.2)$$

Oh and Patton (2016) model the N (log) realized variances, being the diagonal elements of RVar_t , as univariate heterogeneous autoregressive (HAR) models (Corsi, 2009):

$$\begin{aligned} \log \text{RVar}_{j,t+1} = & \alpha_j + \beta_j^{(d)} \cdot \log \text{RVar}_{j,t} \\ & + \beta_j^{(w)} \cdot \frac{1}{5} \sum_{s=0}^4 \log \text{RVar}_{j,t-s} \\ & + \beta_j^{(m)} \cdot \frac{1}{22} \sum_{s=0}^{21} \log \text{RVar}_{j,t-s} + u_{j,t+1}, \end{aligned} \quad (3.3)$$

where $\text{RVar}_{j,t}$ denotes the j -th diagonal element of RVar_t , $\log$ is the natural logarithm, and the parameter vector $\theta'_j = (\alpha_j, \beta_j^{(d)}, \beta_j^{(w)}, \beta_j^{(m)})$ is estimated by ordinary least squares (OLS) for each variance process.

The dynamics of the realized correlations are defined as:

$$\begin{aligned}
\text{vech}(\text{RCorr}_{t+1}) = & \text{vech}(\overline{\text{RCorr}})(1 - a - b - c) \\
& + a \cdot \text{vech}(\text{RCorr}_t) \\
& + b \cdot \frac{1}{5} \sum_{s=0}^4 \text{vech}(\text{RCorr}_{t-s}) \\
& + c \cdot \frac{1}{22} \sum_{s=0}^{21} \text{vech}(\text{RCorr}_{t-s}) + \epsilon_{t+1}, \tag{3.4}
\end{aligned}$$

where $\overline{\text{RCorr}} = T^{-1} \sum_{t=1}^T \text{RCorr}_t$, $\text{vech}(\cdot)$ represents the half-vectorization operator,¹ and a , b , and c are scalars estimated with OLS. For $a, b, c > 0$ and $a + b + c < 1$, Theorem 2 in Oh and Patton (2016) ensures that the forecasts obtained from the model in Equation (3.4) are positive definite correlation matrices. For the remainder of this paper, the above model is referred to as HAR-DRD.

The motivation for the restrictive parametrization of the correlation dynamics is twofold. First, imposing the parameters a , b , and c as scalars allows us to overcome the curse of dimensionality. Oh and Patton (2016) argue that a more flexible version of their model, which replaces the scalar parameters with matrices, is not feasible in higher dimensions. Related models, such as the DCC-GARCH, use a similar approach to overcome the curse of dimensionality (see, e.g., Engle, 2002; Tse & Tsui, 2002). Second, while formally not belonging to the class of long-memory models, the HAR-structure used in Equations (3.3) and (3.4) allows the slowly

¹The vectorization of a matrix is obtained by stacking its columns, and the half-vectorization of a symmetric matrix is obtained by vectorizing only the lower triangular part of the matrix.

decaying autocorrelation of volatilities and correlations to be captured in a parsimonious way (see, e.g., Andersen, Bollerslev, Diebold, & Ebens, 2001; Corsi, 2009).

The 1-day ahead predictions based on the models defined in Equations (3.3) and (3.4), as well as the estimated parameters, are denoted as $\widehat{\log \text{RVar}}_{j,t+1|t}$ for $j = 1, \dots, N$ and $\widehat{\text{RCorr}}_{t+1|t}$, respectively. Volatility and correlation forecasts for horizons longer than one day are obtained by direct forecasts.² The h -day ahead prediction for the entire realized covariance matrix is then:

$$\widehat{\text{RCov}}_{t+h|t} = \widehat{\text{RVar}}_{t+h|t}^{\frac{1}{2}} \widehat{\text{RCorr}}_{t+h|t} \widehat{\text{RVar}}_{t+h|t}^{\frac{1}{2}}, \quad (3.5)$$

where:

$$\widehat{\text{RVar}}_{t+h|t} = \text{diag} \left(\exp \left(\widehat{\log \text{RVar}}_{1,t+h|t} \right), \dots, \exp \left(\widehat{\log \text{RVar}}_{N,t+h|t} \right) \right). \quad (3.6)$$

Due to the non-linear nature of the aggregation of variance and correlation predictions, the realized covariance matrix forecasts are biased. However, other approaches to ensure positive definiteness, such as the matrix logarithm or the Cholesky decomposition, face similar issues. We assess the severity of the bias induced by the transformation outlined in Equation (3.5) in Appendix 3.A. For the dataset at hand, the (empirical) bias of Oh and Patton's (2016) model is small and of a similar magnitude to

²We obtain direct forecasts for longer prediction horizons by replacing the 1-day ahead realized variance in Equation (3.3) and the 1-day ahead realized correlation matrix in Equation (3.4) by their respective measures over the relevant horizon of interest. The models' parameters are estimated separately for each prediction horizon.

the one caused by the Cholesky decomposition (see Chiriac & Voev, 2011). Moreover, an empirical correction of the bias does not improve the forecasting results.³ In this study, the bias induced by the non-linearity of Equation (3.5) is not corrected and the derivation of an analytical bias correction remains a topic for future research.

While the model defined in Equations (3.3) and (3.4) is more flexible compared to realized covariance models suggested in other studies, it does not allow exogenous variables to influence the correlation forecasts. Focusing on the dynamics of the realized correlations, we introduce an extension of the above model that allows exogenous predictors to have a non-linear influence on the predicted realized correlation matrices while still ensuring its positive definiteness. Building on the modeling framework introduced by Audrino and Bühlmann (2001), Audrino and Corsi (2010), and Audrino and Trojani (2006, 2011), we propose to model the realized correlations by local versions of the model defined in Equation (3.4). In other words, while the structure of the model introduced by Oh and Patton (2016) is preserved locally, the parameters of the model are allowed to change over time. The changes in the parameters are controlled by exogenous predictors. More precisely, the realized correlation matrix is modeled as:

$$\text{vech}(\text{RCorr}_{t+1}) = \mathbb{E}[\text{vech}(\text{RCorr}_{t+1})|\mathcal{F}_t] + \sigma_{t+1} \cdot z_{t+1}, \quad (3.7)$$

where $\mathcal{F}_t$ denotes the information set at time t , and $\{z_t\}_{t \geq 1}$ is a sequence

³See the results in Panel B of Table 3.5 in Appendix 3.A.

of i.i.d. innovation vectors with expected value $\mathbf{0}$ and covariance $\mathbf{I}_{\frac{N(N-1)}{2}}$ (i.e., the identity matrix). The dynamics of the conditional expected realized correlations are given by:

$$\begin{aligned} \mathbb{E}[\text{vech}(\text{RCorr}_{t+1})|\mathcal{F}_t] = & \sum_{i=1}^k \left\{ \text{vech}(\overline{\text{RCorr}}_i)(1 - a_i - b_i - c_i) \right. \\ & + a_i \cdot \text{vech}(\text{RCorr}_t) \\ & + b_i \cdot \frac{1}{5} \sum_{s=0}^4 \text{vech}(\text{RCorr}_{t-s}) \\ & \left. + c_i \cdot \frac{1}{22} \sum_{s=0}^{21} \text{vech}(\text{RCorr}_{t-s}) \right\} \mathbb{1}_{[\mathbf{x}_t \in \mathcal{R}_i]}, \quad (3.8) \end{aligned}$$

and the dynamics of $\{\sigma_t^2\}_{t \geq 1}$ are:

$$\sigma_{t+1}^2 = \sum_{i=1}^k \sigma_i^2 \mathbb{1}_{[\mathbf{x}_t \in \mathcal{R}_i]}, \quad \text{with } \sigma_i^2 > 0, \quad i = 1, \dots, k, \quad (3.9)$$

where the parameters $\{a_i, b_i, c_i, \sigma_i^2\}_{i=1}^k$ define the local dynamics of the model, $\mathbf{X}_t \in \mathbb{R}^p$ is a vector of exogenous predictors known at time t , $\overline{\text{RCorr}}_i = T_i^{-1} \sum_{t=1}^T \text{RCorr}_t \cdot \mathbb{1}_{[\mathbf{x}_t \in \mathcal{R}_i]}$ with $T_i = \sum_{t=1}^T \mathbb{1}_{[\mathbf{x}_t \in \mathcal{R}_i]}$, and $\mathbb{1}_{[\cdot]}$ being an indicator function. Following Audrino and Corsi (2010), the local models are characterized by the partition $\mathcal{P} = \{\mathcal{R}_1, \dots, \mathcal{R}_k\}$ of the predictor space, defined as:

$$\bigcup_{i=1}^k \mathcal{R}_i = \mathbb{R}^p, \quad \text{with } \mathcal{R}_i \cap \mathcal{R}_j = \emptyset \quad \forall i \neq j. \quad (3.10)$$

Note that the model introduced by Oh and Patton (2016) is nested in this model, meaning that for $k = 1$ the model defined in Equations (3.7) to (3.9) is simply the HAR-DRD model. For $k > 1$, the model becomes more flexible and allows for changes in the average level and persistence of the correlations. For example, two partition cells $\mathcal{R}_1$ and $\mathcal{R}_2$ could represent market downturns and upturns. The model allows the on average higher correlation observed during market downturns, known as the leverage effect, to be captured through the separate estimation of $\overline{\text{RCorr}}_1$ and $\overline{\text{RCorr}}_2$.

The following theorem extends Theorem 2 in Oh and Patton (2016) and provides conditions under which the predicted correlation matrix obtained from the model defined in Equation (3.8) is positive definite. Let $\mathbf{r}_t$ be a vector of (log) returns and m the number of intraday observations used for the estimation of the realized covariance matrix.

Theorem 1 *Assume that (i) $\mathbb{P}[\mathbf{w}' \mathbf{r}_t = 0] = 0$ for any nonzero vector $\mathbf{w} \in \mathbb{R}^N$, (ii) $a_i, b_i, c_i \geq 0$ for $i = 1, \dots, k$, (iii) $a_i + b_i + c_i < 1$ for $i = 1, \dots, k$, (iv) $T_i \cdot m > N$ for $i = 1, \dots, k$, and (v) $\mathcal{R}_i \cap \mathcal{R}_j = \emptyset$ for $i \neq j$. Then, $\mathbb{E}[\text{RCorr}_{t+1} | \mathcal{F}_t]$ is a positive definite correlation matrix.*

For a specific partition cell $\mathcal{R}_i$, under assumptions (i) to (iv), the proof of Theorem 1 is analogous to that of Theorem 2 in Oh and Patton (2016). Assumption (v) ensures that the prediction is a positive definite correlation matrix over the entire predictor space.

3.2.2 Model estimation

Oh and Patton (2016) estimate the HAR-DRD model with OLS.⁴ For a given partition of the predictor space, the model defined in Equations (3.7) to (3.9) can also be estimated using OLS. To define the partition $\mathcal{P}$ of the predictor space, we follow Audrino and Bühlmann (2001) and impose the partition cells to be rectangular, meaning that the edges of a partition cell are determined by thresholds on the predictor variables. In the spirit of the standard classification and regression tree (CART) procedure (Breiman, Friedman, Stone, & Olshen, 1984), the construction of the partition is based on a binary tree. For a given partition cell $\mathcal{R}_i$, we compare the fit of a standard HAR-DRD model to two local HAR-DRD models estimated over different binary partitions of the partition cell $\mathcal{R}_i$. The fit of the models is evaluated in terms of the sum of squared residuals.⁵ The data-driven construction of the partition is summarized by the following algorithm:

Step 1: Set $\mathcal{P}^{(0)} = \mathbb{R}^p$ and $q = 0$.

Step 2: While $q < q_{\max}$:

Step 2.1: Set $q = q + 1$.

Step 2.2: Find the optimal binary partition of all partition cells in $\mathcal{P}^{(q-1)}$. In other words, let $\mathcal{R}_i$ be the i -th partition

⁴For the data and sample period considered in this study, the estimates of the HAR-DRD model always fulfill the assumptions of Theorem 2 in Oh and Patton (2016), without the need to impose any additional restrictions in the estimation procedure.

⁵Since we consider the half-vectorized correlation matrix in the estimation procedure, minimizing the sum of squared residuals is equivalent to using the Euclidean distance as an optimization criterion, which is often used as a loss function for multivariate volatility models (see Section 3.3.2).

cell of $\mathcal{P}^{(q-1)}$. Determine the index ι_i and threshold value d_{ι_i} such that the sum of squared residuals of the model defined in Equations (3.7) to (3.9) estimated over the partition

$$\mathcal{R}_i = \mathcal{R}_{i,\text{left}} \cup \mathcal{R}_{i,\text{right}}$$

with $\mathcal{R}_{i,\text{left}} = \{\mathbf{x}_t \in \mathbb{R}^p \mid x_{\iota_i} \leq d_{\iota_i}\}$ and $\mathcal{R}_{i,\text{right}} = \{\mathbf{x}_t \in \mathbb{R}^p \mid x_{\iota_i} > d_{\iota_i}\}$ is minimized.

Step 2.3: Let i^* be the index of the partition cell whose binary partition leads to the smallest sum of squared residuals among all other partition cells. Define $\mathcal{P}^{(q)} = \left(\bigcup_{i \neq i^*} \mathcal{R}_i\right) \cup (\mathcal{R}_{i^*,\text{left}} \cup \mathcal{R}_{i^*,\text{right}})$, where the i^* -th partition cell is replaced with its optimal binary split.

The total number of partition cells corresponds to $q_{\max} + 1$. To find the optimal binary split in *Step 2.2*, we follow Audrino and Bühlmann (2001) and use a grid search. For each predictor variable, we construct a binary split of the partition cell for a grid of threshold values, where the grid is defined as empirical α -quantiles of the predictor variable with $\alpha = v/V$ for $v = 1, \dots, V - 1$. Following Audrino and Bühlmann (2001), for the empirical analysis we set $q_{\max} = 7$ and $V = 16$. Moreover, for the practical implementation of the estimation procedure, each partition cell is required to consists of at least 50 observations of each stock-pair and

the estimated parameters must satisfy the assumptions of Theorem 1.⁶

For the purpose of out-of-sample forecasting, the constructed partition $\mathcal{P}^{(q_{\max})}$ could be too fine, resulting in the model overfitting the data. Exploiting the fact that the constructed partition is a binary tree, we adopt classical tree-pruning approaches and search for the best sub-tree in terms of the Akaike information criteria. This approach is also used in other related studies (see, e.g., Audrino & Bühlmann, 2001; Audrino & Corsi, 2010; Audrino & Trojani, 2006, 2011). An alternative approach to pruning is the idea of “bagging” and random predictor selection to reduce the variance of the model’s forecasts (see, e.g., Breiman, 2001). Since the average of positive definite correlation matrices is a positive definite correlation matrix, predictions obtained from a “bagging” approach are also positive definite correlation matrices. Being computationally very intensive, the analysis of this approach remains a topic for future research. For the remainder of this paper, this model is referred to as the tree-structured HAR-DRD model.

Let $\widehat{\text{RCorr}}_{t+1|t}^{\text{tree}}$ denote the 1-day ahead forecast of the correlation matrix based on Equations (3.7) to (3.9), as well as the estimated parameters and partition cells. Multi-period predictions are obtained by direct forecasts. Realized variance forecasts are obtained as in Oh and Patton (2016) by using the model defined in Equation (3.3).⁷ Multi-period

⁶In practice, if a candidate index-threshold pair (ι_i, d_{ι_i}) defines a partition cell over which the parameter estimates do not satisfy the assumptions of Theorem 1, the respective sum of squared residuals is set to $+\infty$.

⁷In practice, more sophisticated realized variance models could be used (see, e.g., Bollerslev, Patton, & Quaedvlieg, 2018; Ghysels, Santa-Clara, & Valkanov, 2006). Since the focus of this study is on the prediction of realized correlations, we mirror the approach of Oh and Patton (2016) and use the HAR model of Corsi (2009).

forecasts of the realized covariance matrix are computed analogously to Equation (3.5).

3.3 Data and forecasting procedure

3.3.1 Data

The dataset consists of 35 U.S. stocks that are or have been part of the Dow Jones Industrial Average Index (DJIA) from January 2010 to June 2020, with a total of 2,527 daily observations.⁸ High-frequency returns are computed using transaction data from 9:30 AM to 4:00 PM Eastern Time. They are obtained from the NYSE Trade and Quote database (TAQ) and cleaned following the guidance of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009). We follow, among others, Bollerslev et al. (2018) along with Oh and Patton (2016) and use a 5-minute sampling frequency. Panel A in Table 3.1 reports the summary statistics of the annualized realized volatilities and realized correlations across the 35 stocks in the sample.

The measures of sentiment and attention used in this study are based

⁸The stocks in our dataset are American Express Co. (AXP), Apple Inc. (AAPL), Verizon Communications Inc. (VZ), Boeing Co. (BA), Caterpillar Inc. (CAT), JPMorgan Chase & Co. (JPM), Chevron Corp. (CVX), Coca-Cola Co. (KO), Walt Disney Co. (DIS), DuPont de Nemours Inc. (DD), Exxon Mobil Corp. (XOM), General Electric Co. (GE), Hewlett-Packard Inc. (HPQ), Home Depot Inc. (HD), Intel Corp. (INTC), IBM Corp. (IBM), Johnson & Johnson (JNJ), McDonald's Corp. (MCD), Merck & Co. Inc. (MRK), 3M Co. (MMM), Bank of America Corp. (BAC), Nike Inc. (NKE), Pfizer Inc. (PFE), Procter & Gamble Co. (PG), AT&T Inc. (T), UnitedHealth Group Inc. (UNH), Walmart Inc. (WMT), Walgreens Boots Alliance Inc. (WBA), Microsoft Corpor. (MSFT), Cisco Systems Inc. (CSCO), Travelers Companies Inc. (TRV), Goldman Sachs Group Inc. (GS), Mondelez International Inc. (MDLZ), Visa Inc. (V), and Raytheon Technologies Corp. (RTX).

on retail investors' trading activity. To identify their trades, we apply the approach suggested by Boehmer, Jones, Zhang, and Zhang (2020). All high-frequency transactions obtained from the TAQ database with exchange code "D" and a price just above (below) a rounded penny are identified as retail initiated buy (sell) transactions.⁹ We denote the buy and sell trading volume of retail investors for stock j on day t as $VB_{j,t}$ and $VS_{j,t}$, respectively. The total volume of trades initiated by retail investors is then $V_{j,t} = VB_{j,t} + VS_{j,t}$.

Following Kumar and Lee (2006), stock-specific sentiment is measured by retail investors' daily buy-sell imbalance. The buy-sell imbalance (BSI) of stock j on day t is defined as:

$$BSI_{j,t} = \frac{VB_{j,t} - VS_{j,t}}{VB_{j,t} + VS_{j,t}}. \quad (3.11)$$

The aggregated sentiment measure is defined as:

$$BSI_t = \frac{1}{N} \sum_{j=1}^N BSI_{j,t}, \quad (3.12)$$

where N is the number of stocks in the sample. A possible alternative definition of the aggregated BSI is obtained by first aggregating the buy (sell) volume of all stocks in the sample and then compute the buy-sell imbalance. However, as Kumar and Lee (2006) pointed out, with this alternative definition, the aggregated sentiment measure could be strongly

⁹As pointed out by Boehmer et al. (2020), the SEC adopted a tick size pilot program from 2016 to 2018 that affected tick sizes inferring with their methodology. However, the tick size pilot program included only stocks with a market capitalization of less than three billion USD and, therefore, did not affect the stocks in this sample.

influenced by a single stock, which could be problematic during major news releases (e.g., earnings calls).

We define retail investors' attention for stock j on day t as the abnormal daily trading volume:

$$A_{j,t} = \log \left(\frac{V_{j,t}}{\text{med}(\{V_{j,t-s}\}_{s=1}^{252})} \right), \quad (3.13)$$

where the denominator corresponds to the median daily retail trading volume over the previous year. This definition is motivated by the findings of Barber and Odean (2008), which show that trading volume is a good proxy for investors' attention. The aggregated retail investors' attention is defined as:

$$A_t = \frac{1}{N} \sum_{j=1}^N A_{j,t}. \quad (3.14)$$

Finally, asymmetries in the correlation dynamics are captured by the daily DJIA return, denoted as r_t^{DJ} . The open-to-close daily stock returns used for the value-at-risk analysis presented in Section 3.4.2 are obtained from the Center for Research in Security Prices (CRSP). Summary statistics of the three exogenous correlation predictors can be found in Panel B of Table 3.1. Aggregated retail investors' attention is on average positive, but varies considerably over time. The average and median values of retail investors' aggregated sentiment are slightly negative.

Table 3.1: Summary statistics of the data

Panel A: Summary statistics of the realized volatilities and correlations					
	Mean	Std.	$Q_{25\%}$	Median	$Q_{75\%}$
Volatility	0.165	0.102	0.110	0.141	0.188
Correlation	0.297	0.222	0.144	0.298	0.455
Panel B: Summary statistics of the correlation predictors					
	Mean	Std.	$Q_{25\%}$	Median	$Q_{75\%}$
Return DJIA (in %)	0.042	1.096	-0.339	0.056	0.498
Attention	0.043	0.235	-0.100	0.016	0.154
Sentiment	-0.015	0.044	-0.039	-0.014	0.012

NOTE: The table depicts the summary statistics for the variables in the dataset from January 2010 to June 2020. Panel A reports the summary statistics for the realized volatility (annualized) and the realized correlations across all 35 stocks. Panel B summarizes the three correlation predictors, being the daily return of the Dow Jones Index and the aggregated attention and sentiment measures.

3.3.2 Forecasting procedure and model alternatives

We assess the predictive power of retail investors' trading activity by including the aggregated measures of sentiment and attention in the tree-structured HAR-DRD model. Aggregated, rather than firm-specific measures, are used for two reasons. First, exogenous predictors included in the model presented in Section 3.2 affect the entire correlation matrix equally. As such, it is more natural to include aggregated measures. Second, the behavioral models and empirical findings presented in other studies suggest that stock price comovements are influenced by aggregated rather than firm-specific sentiment and attention (see, e.g., Barberis et al., 2005; Drake et al., 2017; Kumar & Lee, 2006). The predictive power of the new model introduced in Section 3.2 is analyzed for four specifications of the exogenous predictors: (i) the daily DJIA return (HAR-DRD Leverage); (ii) aggregated retail investors' attention (HAR-DRD Attention); (iii) aggregated retail investors' sentiment (HAR-DRD

Sentiment); and (iv) all three exogenous predictors simultaneously (HAR-DRD All). The limited number of exogenous predictors considered in this study is motivated by the empirical findings documented in the literature. While retail investors' trading activity has many facets, we focus on their sentiment and attention as these measures have been shown to influence the comovement of stock prices. The analysis of other retail investors' trading activity measures, such as investors' disagreement (Cookson & Niessner, 2020) or other general risk variables like the VIX index, remain a topic for future research.

We use the Akaike information criterion to reduce the number of partitions for all four specifications of our new model. The baseline model is the HAR-DRD model introduced by Oh and Patton (2016). In line with the literature, we include the vector HAR model of realized covariances suggested by Chiriac and Voev (2011), which has been found to have good out-of-sample predictive power (Symitsi, Symeonidis, Kourtis, & Markellos, 2018).¹⁰ All models are estimated every 22 observations (every month) over a rolling five-year window. Each partition cell is required to have at least 50 observations for each stock-pair. The results are robust with respect to both of these choices, as shown in Appendix 3.B.¹¹ The

¹⁰Formally, the model is defined as $\text{vech}(Z_{t+1}) = \phi_0 + \phi_1 \cdot \text{vech}(Z_t) + \phi_2 \cdot \frac{1}{5} \sum_{s=0}^4 \text{vech}(Z_{t-s}) + \phi_3 \cdot \frac{1}{22} \sum_{s=0}^{21} \text{vech}(Z_{t-s})$, where Z_t is the Cholesky decomposition of RCov_t . The parameters are estimated with OLS. The 1-day ahead forecast for the realized covariance matrix is obtained by applying the reverse of the Cholesky decomposition to $\hat{Z}_{t+1|t}$ (see Chiriac & Voev, 2011). The 1-day ahead forecast for the realized correlation matrix is then $\widehat{\text{RCorr}}_{t+1|t} = \text{diag}(\widehat{\text{RCov}}_{t+1|t})^{-1/2} \widehat{\text{RCov}}_{t+1|t} \text{diag}(\widehat{\text{RCov}}_{t+1|t})^{-1/2}$.

¹¹In Appendix 3.B, we report the results for a rolling window of four and a half years and five and a half years (see Tables 3.6 and 3.7), and for partition cells with at least 10 and 100 observations for each stock-pair (see Tables 3.8 and 3.9). The results are qualitatively unaffected by these changes.

models are compared at the 1-day, 5-day, 10-day, and 22-day prediction horizons, where the h -period covariance matrix is defined as the sum of h daily covariance matrices. As described in Section 3.2, multi-step ahead predictions are obtained from direct forecasts.¹²

We evaluate the predictive performance of the models in terms of the (squared) Euclidean distance (see Laurent, Rombouts, & Violante, 2013). Since the main focus is on the prediction of stock return correlations, we compute the losses directly for the realized correlation matrices. The loss function is then defined as:

$$L_{t+h|t}^E = \text{vech}(e_{t+h|t})' \text{vech}(e_{t+h|t}), \quad (3.15)$$

where $e_{t+h|t} = \text{RCorr}_{t+h} - \widehat{\text{RCorr}}_{t+h|t}$ is the matrix of forecasting errors, and $\text{vech}(\cdot)$ is the half-vectorization operator. Another popular loss function that satisfies the assumptions for consistent model ranking is the Frobenius distance (see Laurent et al., 2013). Since we evaluate the predictive accuracy for the correlation matrix and the diagonal elements of the forecast error matrix $e_{t+h|t}$ are all zero, the Euclidean and Frobenius distances are equivalent.

The economic value of the correlation predictions is assessed by comparing their accuracy in terms of value-at-risk forecasts, where the value-at-risk is defined as $\mathbb{P}(r_{t+h} \leq \text{VaR}_{t+h}^\gamma) = \gamma$. Following Fengler and

¹²Table 3.10 in Appendix 3.B reports the results for multi-step ahead predictions obtained recursively from 1-day ahead predictions. The results are qualitatively unaffected by the choice of a multi-period prediction approach. However, to obtain recursive forecasts from the tree-structured models, we have to assume that the model “stays” in the same partition cell for all prediction horizons.

Okhrin (2016), we forecast the 1%, 5%, and 10% value-at-risks of an equally weighted portfolio containing the 35 stocks in our sample. The portfolio is held from 9:30 AM to 4:00 PM, meaning that at the end of each trading day all positions are closed. The forecasts of the portfolio's variance are determined as:

$$\hat{\sigma}_{t+h|t}^2 = \frac{1}{N^2} \mathbf{1}' \widehat{\text{RCov}}_{t+h|t} \mathbf{1}, \quad (3.16)$$

where $\mathbf{1}$ is a vector of ones, N denotes the number of stocks, and $\widehat{\text{RCov}}_{t+h|t}$ is the h -period covariance matrix forecast of one of the six considered models. Value-at-risk forecasts are then obtained using filtered historical simulations with 10,000 replications (see Barone-Adesi, Bourgoin, & Giannopoulos, 1998). The accuracy of the value-at-risk forecasts is evaluated with the quantile loss function (see González-Rivera, Lee, & Mishra, 2004):

$$L_{t+h|t}^V(\gamma) = \left(\gamma - f_{t+h|t}^\gamma \right) \left(r_{t+h} - \widehat{\text{VaR}}_{t+h|t}^\gamma \right) \quad (3.17)$$

for $\gamma = 1\%, 5\%, 10\%$, where $f_{t+h|t}^\gamma = \mathbb{1}_{[r_{t+h} < \widehat{\text{VaR}}_{t+h|t}^\gamma]}$. To formally evaluate whether the losses differ significantly across the six analyzed models, we apply the model confidence set (MCS) of Hansen, Lunde, and Nason (2011). For each loss function, the MCS approach determines a set of models consisting of the best forecasting models with a given confidence. More precisely, we determine the 90% and the 70% model confidence

sets.¹³

3.4 Empirical results

3.4.1 Full-sample estimation results

Before analyzing the forecasting accuracy of the models, we summarize the full-sample estimation results for the 1-day horizon. Panel A of Table 3.2 reports the summary statistics across the 35 stocks of the estimated parameters of the variance model defined in Equation (3.3). The first four columns report the mean, standard deviation, median, and 25%- and 75%-quantiles of the estimated parameters. The last two columns report the number of observations and summary statistics of the Akaike information criterion. Panel B of Table 3.2 summarizes the estimates of the correlation models. The first column shows the empirical mean, and the standard deviation in parentheses, of the average correlation across all stock-pairs. Columns two through four in Panel B report the parameter estimates, as well as block-bootstrapped standard deviations in parentheses (10,000 replications with a block-length of 22). The number of observations and the Akaike information criterion are shown in the last two columns of Panel B, respectively. For the four tree-structured HAR-DRD models, Panel B in Table 3.2 reports the estimates and the number of observations for each partition cell separately.

¹³The model confidence sets reported in this paper are based on the T_R test statistic, for which the distribution is estimated with 10,000 replications of a moving-block bootstrap. For more details about the implementation of the model confidence set approach, refer to Hansen et al. (2011).

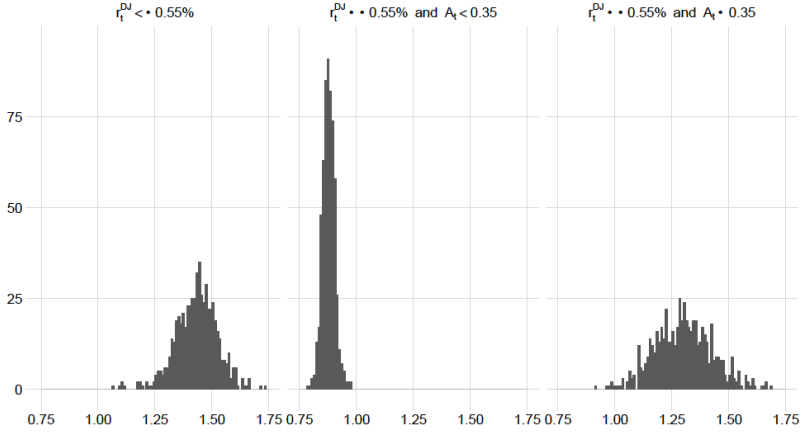
Table 3.2: Full sample estimation results of the variance and correlation models

Panel A: Summary statistics of the parameter estimates of the variance HAR models						
	α_j	$\beta_j^{(d)}$	$\beta_j^{(w)}$	$\beta_j^{(m)}$	Num. obs.	AIC
Mean	-0.821	0.397	0.347	0.168	2,527	6.483
Standard deviation	0.200	0.033	0.043	0.029		0.105
25%-quantile	-0.952	0.376	0.326	0.147		6.413
Median	-0.851	0.396	0.356	0.160		6.475
75%-quantile	-0.653	0.413	0.373	0.189		6.522
Panel B: Estimation results of the HAR-DRD model and its extensions						
	$\overline{\text{RCorr}}_i$	a_i	b_i	c_i	Num. obs.	AIC
<u>HAR-DRD</u>	0.297 (0.064)	0.156 (0.013)	0.398 (0.029)	0.275 (0.037)	2,527	10.776
<u>HAR-DRD Leverage</u>						
$\mathcal{R}_1 : r_t^{DJ} < -0.55\%$	0.421 (0.060)	0.252 (0.028)	0.388 (0.054)	0.091 (0.065)	460	10.760
$\mathcal{R}_2 : r_t^{DJ} \geq -0.55\%$	0.270 (0.065)	0.110 (0.010)	0.391 (0.028)	0.287 (0.034)	2,067	
<u>HAR-DRD Attention</u>						
$\mathcal{R}_1 : A_t < -0.08$	0.279 (0.066)	0.076 (0.015)	0.255 (0.036)	0.482 (0.044)	725	10.766
$\mathcal{R}_2 : A_t < 0.27$ and $A_t \geq -0.08$	0.283 (0.065)	0.153 (0.013)	0.380 (0.029)	0.275 (0.041)	1,495	
$\mathcal{R}_3 : A_t \geq 0.27$	0.413 (0.062)	0.290 (0.040)	0.486 (0.051)	0.025 (0.066)	307	
<u>HAR-DRD Sentiment</u>						
$\mathcal{R}_1 : BSI_t < 0.012$	0.281 (0.064)	0.151 (0.014)	0.344 (0.029)	0.325 (0.034)	1,886	10.775
$\mathcal{R}_2 : BSI_t \geq 0.012$	0.344 (0.064)	0.166 (0.022)	0.509 (0.046)	0.138 (0.059)	641	
<u>HAR-DRD All</u>						
$\mathcal{R}_1 : r_t^{DJ} < -0.55\%$	0.421 (0.060)	0.252 (0.028)	0.388 (0.054)	0.091 (0.065)	460	10.758
$\mathcal{R}_2 : r_t^{DJ} \geq -0.55\%$ and $A_t < 0.35$	0.263 (0.065)	0.103 (0.010)	0.357 (0.026)	0.312 (0.036)	1,939	
$\mathcal{R}_3 : r_t^{DJ} \geq -0.55\%$ and $A_t \geq 0.35$	0.383 (0.068)	0.179 (0.055)	0.561 (0.077)	0.094 (0.083)	128	

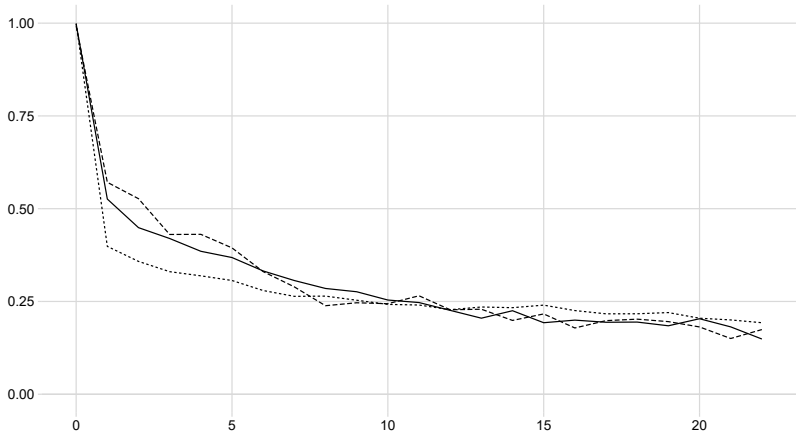
NOTE: The table reports the full-sample estimation results of the variance and correlation models. Panel A shows the summary statistics of the estimated parameters and the Akaike information criterion for the HAR models of log realized variance across the 35 stocks. Panel B reports in the first column the empirical mean, and the standard deviation in parentheses, of the average correlation across all stock-pairs. Columns two through four in Panel B report the parameter estimates, as well as block-bootstrapped standard deviations in parentheses (10,000 replications with a block-length of 22). The number of observations and the Akaike information criterion are shown in the last two columns of Panel B, respectively.

The estimates of the HAR-DRD Leverage model reflect the classical asymmetries documented in the literature. On days where the DJIA return is below -0.55% (corresponding to its 18%-quantile), the correlations are on average higher and more persistent. For the HAR-DRD Attention model, we identify three partition cells. The estimates imply a low persistence in the realized correlations during periods of low aggregated retail investors' attention (A_t below -0.08). When the aggregated attention of retail traders is between its 28%- and 87%-quantile, the estimates are close to those of the baseline HAR-DRD model. During periods of high attention, both the average and the persistence of the realized correlations are considerably higher. These estimation results are in line with the empirical findings documented by, among others, Drake et al. (2017), where higher aggregated attention is associated with larger stock return comovements. For the HAR-DRD Sentiment model, two partition cells identified by the 75%-quantile of the aggregated buy-sell imbalance are estimated. During periods of high aggregated sentiment ($BSI_t \geq 0.012$), the average correlation between the stock returns is higher, as predicted by sentiment-based theories of price comovements (see, e.g., Kumar & Lee, 2006).

Finally, when including all three exogenous variables in the tree-structured HAR-DRD model, our estimation procedure identifies (i) an extreme market downturn period, (ii) a low attention period, and (iii) a period of high retail investors' attention. As for the HAR-DRD Leverage model, the correlations are, on average, higher and more persistent during market downturns. However, the estimation results suggest that



(a) Relative average realized correlations by partition cell



(b) Average empirical autocorrelation of the realized correlations by partition cell

Figure 3.1: Relative averages and autocorrelation functions of the realized correlations

NOTE: The figure shows the relative average (a) and the sample autocorrelation (b) of the realized correlations for three partition cells. More precisely, Figure 3.1a depicts the average correlation of each stock-pair across the three partition cells relative to the stock-pair's average correlation over the entire sample. Figure 3.1b shows the average (across all stock-pairs) autocorrelation function for the three partition cells. The solid line depicts the ACF for days where the DJIA return (r_t^{DJ}) is smaller than -0.55% , the dotted line shows the ACF for days where $r_t^{DJ} \geq -0.55\%$ and the aggregated retail attention (A_t) is below 0.35, and the dashed line depicts the ACF for days where $r_t^{DJ} \geq -0.55\%$ and $A_t \geq 0.35$.

even when the DJIA return is larger than -0.55% , the average realized correlation could be considerably higher than its sample average. This finding is depicted graphically in Figure 3.1. The figure shows in the upper panel (Figure 3.1a) a histogram of the average realized correlation of each stock-pair computed over the three partition cells relative to the respective average correlation computed over the entire sample. A value of one indicates that the average correlation computed over a specific partition cell corresponds to its average computed over the entire sample. Figure 3.1b depicts the average, across all stock-pairs, autocorrelation function (ACF) for each of the three partition cells. The solid line depicts the ACF for days where the DJIA return is smaller than -0.55% , the dotted line shows the ACF for days where $r_t^{DJ} \geq -0.55\%$ and the aggregated retail attention is below 0.35, and the dashed line depicts the ACF for days where $r_t^{DJ} \geq -0.55\%$ and $A_t \geq 0.35$. Figure 3.1a shows that during periods characterized by extreme negative returns or by extremely high retail investors' attention, the stock returns, on average, correlate more. Similarly, at least for the first five lags, the realized correlations' autocorrelations are, on average, higher during market downturns or periods of high aggregated attention. These full-sample estimation results suggest that retail investors' attention could have predictive power for realized correlations, even when controlling for the well-known leverage effect.

3.4.2 Out-of-sample prediction results

Turning our attention to the out-of-sample results, Table 3.3 summarizes the forecasting accuracy of the six models. The table reports the average loss of each model relative to that of the HAR-DRD model. For each forecast horizon and loss function, the lowest loss is highlighted in boldface. Moreover, models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, ..., Friday-to-Friday. In Table 3.3, we report the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

The HAR model suggested by Chiriac and Voev (2011) is not able to outperform the HAR-DRD model of Oh and Patton (2016), except for the 1% value-at-risk forecasts at the 5-day and 10-day horizons. Accounting for retail investors' sentiment does not improve the accuracy of the HAR-DRD model. The HAR-DRD Sentiment model obtains, at all prediction horizons, a slightly higher average loss. The HAR-DRD Leverage model achieves the best 10% value-at-risk forecast at the 5-day horizon. For all other loss functions and prediction horizons, the best prediction accuracy is achieved by either the attention-based model or the HAR-DRD All model. In terms of the Euclidean distance, the HAR-DRD All model has the highest predictive accuracy up to the 10-day horizon.

The model confidence results confirm this finding. In particular, in

Table 3.3: Realized correlations and value-at-risk forecasting results

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Atten- tion	HAR- DRD Senti- ment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.188	1.000	0.985 ^{a, b}	0.986 ^{a, b}	1.001	0.984 ^{a, b}
VaR _{1%}	1.067	1.000	0.989 ^a	0.973 ^{a, b}	1.006	0.982 ^{a, b}
VaR _{5%}	1.020 ^a	1.000 ^a	0.998 ^a	0.992 ^{a, b}	1.001 ^a	0.997 ^a
VaR _{10%}	1.005 ^a	1.000 ^a	1.000 ^a	0.995 ^{a, b}	1.000 ^a	0.998 ^a
<u>5-day ahead prediction</u>						
Euclidean	1.419	1.000 ^{a³ b¹}	0.959 ^{a⁴ b⁴}	0.975 ^{a⁴ b³}	1.003 ^{a²}	0.953 ^{a⁵ b⁵}
VaR _{1%}	0.987 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.982 ^{a⁵ b⁵}	0.997 ^{a⁵ b⁵}	1.006 ^{a⁵ b³}	0.982 ^{a⁵ b⁵}
VaR _{5%}	1.013 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.996 ^{a⁵ b⁵}	0.999 ^{a⁵ b⁵}	1.002 ^{a⁵ b⁴}	0.995 ^{a⁵ b⁵}
VaR _{10%}	1.024 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.998 ^{a⁵ b⁵}	1.001 ^{a⁵ b⁴}	1.002 ^{a⁵ b⁴}	0.998 ^{a⁵ b⁵}
<u>10-day ahead prediction</u>						
Euclidean	1.491	1.000 ^{a⁹ b⁷}	0.970 ^{a¹⁰ b⁷}	0.979 ^{a¹⁰ b⁸}	1.007 ^{a⁹ b⁶}	0.948 ^{a¹⁰ b¹⁰}
VaR _{1%}	0.931 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.986 ^{a⁹ b⁸}	0.988 ^{a¹⁰ b¹⁰}	1.011 ^{a¹⁰ b⁶}	0.981 ^{a¹⁰ b⁹}
VaR _{5%}	1.007 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.997 ^{a⁹ b⁹}	1.001 ^{a⁹ b⁹}	1.004 ^{a¹⁰ b⁸}	0.995 ^{a¹⁰ b¹⁰}
VaR _{10%}	1.028 ^{a¹⁰ b⁹}	1.000 ^{a¹⁰ b¹⁰}	0.999 ^{a¹⁰ b¹⁰}	1.002 ^{a¹⁰ b⁹}	1.002 ^{a¹⁰ b¹⁰}	0.997 ^{a¹⁰ b¹⁰}
<u>22-day ahead prediction</u>						
Euclidean	1.575	1.000 ^{a²¹ b¹⁵}	0.999 ^{a²² b¹⁷}	0.966 ^{a²² b²¹}	1.007 ^{a²⁰ b¹⁷}	0.982 ^{a²² b¹⁹}
VaR _{1%}	1.060 ^{a¹⁷ b¹³}	1.000 ^{a²² b¹⁸}	1.010 ^{a²² b¹⁷}	0.998 ^{a²² b¹⁸}	1.006 ^{a²² b²⁰}	1.004 ^{a²² b¹⁶}
VaR _{5%}	1.028 ^{a²⁰ b¹⁷}	1.000 ^{a²¹ b¹⁸}	1.006 ^{a²² b¹⁵}	0.999 ^{a²² b¹⁷}	1.001 ^{a²² b²¹}	1.001 ^{a²² b²⁰}
VaR _{10%}	1.027 ^{a²¹ b²⁰}	1.000 ^{a²¹ b¹⁸}	1.003 ^{a¹⁸ b¹⁵}	1.000 ^{a²² b¹⁹}	1.000 ^{a²² b²⁰}	1.000 ^{a²² b²²}

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio. The models are estimated every 22 observations over a rolling window of five years. Multi-period forecasts are obtained by direct predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, ..., Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

terms of the Euclidean distance, the HAR model is never, and the HAR-DRD All model is almost always, included in the 90% and 70% MCS. Similarly, the leverage- and the attention-based models are most often included in both model confidence sets when measuring the forecasting accuracy with the Euclidean distance. The model confidence sets become

larger for longer prediction horizons and often included the baseline HAR-DRD model and all its tree-structured extensions. The value-at-risk forecasts indicate significant differences between the models only at the 1-day horizon. For 1-day ahead predictions, the only model included in the 70% model confidence set is the HAR-DRD Attention model, except for the 1% value-at-risk where also the HAR-DRD All model is included.

The results clearly show that the tree-structured HAR-DRD model has predictive power and can outperform the baseline model of Oh and Patton (2016). In particular, we find that retail investors' attention has predictive power for the comovement of stock prices, and in some situations, even more than the leverage effect. The predictive differences are more pronounced in terms of the Euclidean distance rather than value-at-risks. In this study, we focus on improving the predictive accuracy for realized correlations and allow exogenous variables only to influence the correlation dynamics. The predictive accuracy in terms of value-at-risks, however, also depends on the volatility forecasts. Since the same volatility predictions are used for the HAR-DRD model and all its tree-structured extensions, differences in the forecasts of the realized covariance matrix are small.

3.4.3 Prediction results during the COVID-19 pandemic

In the first half of 2020, the number of retail investors in the U.S. market increased considerably. The largest U.S. online brokers, such as Robinhood, TD Ameritrade, E-Trade, and Schwab, saw a considerable increase

in users during the COVID-19 pandemic. For instance, Figure 3.2 depicts the average number of trading accounts on the platform Robinhood that hold at least one share of the companies in our sample.¹⁴ Starting from late February, the number of accounts holding at least one DJIA stock increased exponentially. During this period, it is likely that price comovements are driven even more by investors' behavior, and that retail traders' sentiment and attention measures have more predictive power for stock return correlations.

This hypothesis is investigated by focusing on the prediction results during the COVID-19 pandemic. Table 3.4 reports the forecasting results evaluated over the first half of 2020, a period characterized by a consid-

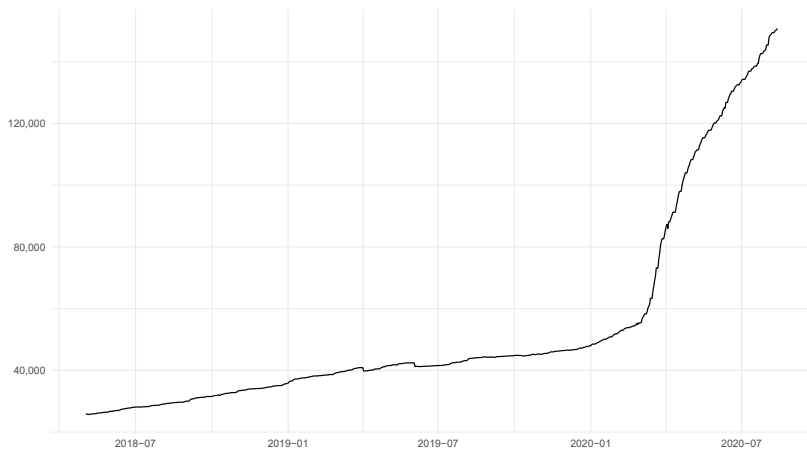


Figure 3.2: Average number of Robinhood accounts holding at least one stock from the sample

NOTE: The figure depicts the average (across the 35 stocks in our sample) number of accounts on the Robinhood trading platform that hold at least one share of the companies in the sample. The data is provided by <https://robintrack.net>.

¹⁴Robinhood Markets Inc. is an online trading platform that offers investors the possibility to trade in stocks, ETFs, and options on their website or a mobile app (see <https://robinhood.com>).

erable market downturn from February to the end of March, and a fast upturn from April to June. Overall, the best performing models are again the tree-structured HAR-DRD models enhanced with either the DJIA return, the aggregated measure of retail investors' attention, or all exoge-

Table 3.4: Realized correlations and value-at-risk forecasting results from January to June 2020

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Atten- tion	HAR- DRD Sentiment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.135	$1.000^{a, b}$	$0.980^{a, b}$	$0.942^{a, b}$	$0.997^{a, b}$	$0.975^{a, b}$
VaR _{1%}	1.149	1.000^a	$0.980^{a, b}$	$0.948^{a, b}$	1.006^a	$0.946^{a, b}$
VaR _{5%}	$1.040^{a, b}$	$1.000^{a, b}$	$1.005^{a, b}$	$0.982^{a, b}$	$0.999^{a, b}$	$0.997^{a, b}$
VaR _{10%}	$1.009^{a, b}$	$1.000^{a, b}$	$1.003^{a, b}$	$0.988^{a, b}$	$1.001^{a, b}$	$0.995^{a, b}$
<u>5-day ahead prediction</u>						
Euclidean	$1.143^{a^5, b^1}$	$1.000^{a^5, b^4}$	$0.900^{a^5, b^5}$	$0.892^{a^5, b^4}$	$1.009^{a^5, b^3}$	$0.902^{a^4, b^4}$
VaR _{1%}	$1.099^{a^2, b^1}$	$1.000^{a^2, b^4}$	$0.995^{a^2, b^3}$	$1.004^{a^2, b^1}$	$1.000^{a^5, b^2}$	$0.997^{a^2, b^1}$
VaR _{5%}	1.129^{a^1}	$1.000^{a^5, b^5}$	$0.992^{a^2, b^5}$	$0.996^{a^5, b^4}$	$1.002^{a^2, b^5}$	$0.993^{a^4, b^3}$
VaR _{10%}	$1.121^{a^2, b^1}$	$1.000^{a^5, b^5}$	$0.996^{a^5, b^5}$	$1.007^{a^4, b^3}$	$1.001^{a^5, b^5}$	$0.995^{a^5, b^4}$
<u>10-day ahead prediction</u>						
Euclidean	$1.097^{a^9, b^7}$	$1.000^{a^{10}, b^8}$	$0.967^{a^{10}, b^8}$	$0.967^{a^8, b^8}$	$0.999^{a^9, b^7}$	$0.907^{a^{10}, b^{10}}$
VaR _{1%}	$1.111^{a^7, b^6}$	$1.000^{a^{10}, b^9}$	$1.019^{a^7, b^4}$	$0.997^{a^5, b^3}$	$1.005^{a^{10}, b^7}$	$0.996^{a^6, b^5}$
VaR _{5%}	$1.195^{a^6, b^3}$	$1.000^{a^{10}, b^9}$	$1.020^{a^8, b^6}$	$1.026^{a^3, b^3}$	$1.002^{a^{10}, b^8}$	$1.017^{a^2, b^5}$
VaR _{10%}	$1.196^{a^7, b^2}$	$1.000^{a^{10}, b^{10}}$	$1.018^{a^8, b^5}$	$1.018^{a^5, b^4}$	$1.002^{a^{10}, b^9}$	$1.016^{a^6, b^6}$
<u>22-day ahead prediction</u>						
Euclidean	$1.106^{a^{17}, b^{11}}$	$1.000^{a^{20}, b^9}$	$1.012^{a^{17}, b^8}$	$0.865^{a^{22}, b^{19}}$	$0.989^{a^{19}, b^8}$	$0.897^{a^{21}, b^{18}}$
VaR _{1%}	$1.460^{a^{10}}$	$1.000^{a^{20}, b^{19}}$	$1.035^{a^{20}, b^{13}}$	$0.983^{a^{11}, b^9}$	$1.017^{a^{21}, b^{18}}$	$0.997^{a^{14}, b^8}$
VaR _{5%}	$1.435^{a^{13}, b^2}$	$1.000^{a^{20}, b^{19}}$	$1.024^{a^{21}, b^{14}}$	$0.997^{a^{13}, b^8}$	$1.002^{a^{21}, b^{18}}$	$1.021^{a^{14}, b^9}$
VaR _{10%}	$1.406^{a^{12}, b^3}$	$1.000^{a^{20}, b^{17}}$	$1.020^{a^{20}, b^{15}}$	$0.999^{a^{15}, b^{13}}$	$1.002^{a^{21}, b^{18}}$	$1.016^{a^{16}, b^{12}}$

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio from January to June 2020. The models are estimated every 22 observations over a rolling window of five years. Multi-period forecasts are obtained by direct predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, $\dots$, Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

nous predictors simultaneously. Interestingly, both the attention- and the sentiment-based models achieve a sizable better predictive performance relative to the baseline model during the COVID-19 pandemic, compared to the results for the entire out-of-sample period. For most prediction horizons and loss functions, the best predictive performance is achieved by the HAR-DRD Attention model. Moreover, the HAR-DRD Sentiment model outperforms the baseline model in terms of the Euclidean distance at the 1-day, 10-day, and 22-day horizons. The model confidence set results are more difficult to interpret since the the number of observations for this time period is low, especially for the multi-period predictions.

In summary, the increase of retail investors in the market underpins the importance of accounting for their behavior when predicting stock price comovements. Moreover, the results reported in Table 3.4 show that the tree-structured HAR-DRD models have predictive power during turbulent market conditions.

3.5 Conclusion

This study analyzes whether the sentiment-based theories of price comovements have out-of-sample predictive power for realized correlations. First, we propose an extension of the realized covariance model introduced by Oh and Patton (2016) to allow exogenous predictors to have a non-linear influence on the model's predictions while still ensuring the forecasted matrices' positive definiteness. Second, we construct sentiment and attention measures using retail investors' transaction data. Together

with the daily DJIA return, we analyze the predictive out-of-sample power of these measures for 35 stocks that are or have been part of the Dow Jones Industrial Average Index.

The findings show that (i) our model significantly outperforms other realized covariance and correlation models, (ii) aggregated retail investors' attention has predictive power for stock return correlations, (iii) in particular at longer forecasting horizons and during the COVID-19 pandemic, (iv) whereas retail investors' sentiment only improves the forecasting accuracy of realized correlations during the first half of 2020.

Our results have several implications for both the academic literature and practitioners. It is crucial for practitioners to accurately forecast stock price comovements for managing a portfolio's risk, even more so during volatile times. The increasing number of retail investors in the market underlines the importance of accounting for their behavior when predicting stock return correlations. The model introduced in this study may serve as a building block for future work, allowing researchers to further investigate the role of retail investors, or to analyze the predictive power of other exogenous variables for realized correlations and variances. Moreover, it would be of interest to further extend the presented model and allow for smooth transitions between the partition cells, or consider the use of bagging approaches as an alternative to the pruning strategy used in this paper.

Acknowledgements

This research project is funded by the Swiss National Science Foundation (Grant 100018-172685). I would like to thank Francesco Audrino, Simon Behrendt, Cédric Müller, Gabriel Okasa, Andrin Pelican, and Fabio Sigrist for their helpful discussions and comments.

References

- Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), 43–76.
- Ang, A., & Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3), 443–494.
- Audrino, F., & Bühlmann, P. (2001). Tree-structured generalized autoregressive conditional heteroscedastic models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4), 727–744.
- Audrino, F., & Corsi, F. (2010). Modeling tick-by-tick realized correlations. *Computational Statistics & Data Analysis*, 54(11), 2372–2382.
- Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36(2), 334–357.
- Audrino, F., & Trojani, F. (2006). Estimating and predicting multivariate volatility thresholds in global stock markets. *Journal of Applied Econometrics*, 21(3), 345–369.
- Audrino, F., & Trojani, F. (2011). A General Multivariate Threshold GARCH Model With Dynamic Conditional Correlations. *Journal of Business & Economic Statistics*, 29(1), 138–149.
- Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and

- Institutional Investors. *The Review of Financial Studies*, 21(2), 785–818.
- Barberis, N., & Shleifer, A. (2003). Style investing. *Journal of Financial Economics*, 68(2), 161–199.
- Barberis, N., Shleifer, A., & Wurgler, J. (2005). Comovement. *Journal of Financial Economics*, 75(2), 283–317.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: trades and quotes. *The Econometrics Journal*, 12(3), C1–C32.
- Barndorff-Nielsen, O. E., & Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3), 885–925.
- Barone-Adesi, G., Bourgoin, F., & Giannopoulos, K. (1998). Don't look back. *Risk*, 11, 100–103.
- Bauer, G. H., & Vorkink, K. (2011). Forecasting multivariate realized stock market volatility. *Journal of Econometrics*, 160(1), 93–101.
- Boehmer, E., Jones, C. M., Zhang, X., & Zhang, X. (2020). *Tracking retail investor activity*. (Working Paper)
- Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2018). Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions. *Journal of Econometrics*, 207(1), 71–91.
- Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2020). Multivariate leverage effects and realized semicovariance GARCH models. *Journal of Econometrics*, 217(2), 411–430.

- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32.
- Breiman, L., Friedman, J., Stone, C., & Olshen, R. (1984). *Classification and regression trees*. Taylor & Francis.
- Cappiello, L., Engle, R. F., & Sheppard, K. (2006, 09). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics*, 4(4), 537–572.
- Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6), 922–947.
- Cookson, J. A., & Niessner, M. (2020). Why Don’t We Agree? Evidence from a Social Network of Investors. *Journal of Finance*, 75(1), 173–228.
- Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics*, 7(2), 174–196.
- de Goeij, P., & Marquering, W. (2004). Modeling the Conditional Covariance Between Stock and Bond Returns: A Multivariate GARCH Approach. *Journal of Financial Econometrics*, 2(4), 531–564.
- Drake, M. S., Jennings, J., Roulstone, D. T., & Thornock, J. R. (2017). The Comovement of Investor Attention. *Management Science*, 63(9), 2847–2867.
- Engle, R. (2002). Dynamic conditional correlation. *Journal of Business & Economic Statistics*, 20(3), 339–350.
- Fengler, M. R., & Okhrin, O. (2016). Managing risk with a realized copula parameter. *Computational Statistics & Data Analysis*, 100,

131–152.

- Francq, C., & Zakoïan, J.-M. (2012). QML estimation of a class of multivariate asymmetric GARCH models. *Econometric Theory*, 28(1), 179–206.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2006). Predicting volatility: getting the most out of return data sampled at different frequencies. *Journal of Econometrics*, 131(1), 59–95.
- González-Rivera, G., Lee, T.-H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting*, 20(4), 629–645.
- Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The Model Confidence Set. *Econometrica*, 79(2), 453–497.
- Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Huang, S., Huang, Y., & Lin, T.-C. (2019). Attention allocation and return co-movement: Evidence from repeated natural experiments. *Journal of Financial Economics*, 132(2), 369–383.
- Kumar, A., & Lee, C. M. (2006). Retail Investor Sentiment and Return Comovements. *The Journal of Finance*, 61(5), 2451–2486.
- Laurent, S., Rombouts, J. V., & Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models. *Journal of Econometrics*, 173(1), 1–10.
- McAleer, M., Hoti, S., & Chan, F. (2009). Structure and Asymptotic Theory for Multivariate Asymmetric Conditional Volatility. *Econo-*

metric Reviews, 28(5), 422–440.

- Oh, D. H., & Patton, A. J. (2016). High-dimensional copula-based distributions with mixed frequency data. *Journal of Econometrics*, 193(2), 349–366. (The Econometric Analysis of Mixed Frequency Data Sampling)
- Symitsi, E., Symeonidis, L., Kourtis, A., & Markellos, R. (2018). Covariance forecasting in equity markets. *Journal of Banking & Finance*, 96, 153–168.
- Tse, Y. K., & Tsui, A. K. C. (2002). A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model With Time-Varying Correlations. *Journal of Business & Economic Statistics*, 20(3), 351–362.

Appendix

3.A Bias correction

Oh and Patton's (2016) model separately defines the dynamics of realized variances and correlations. Forecasts for the entire realized covariance matrix are obtained through the transformation outlined in Equation (3.5). Due to the non-linear nature of the aggregation of volatility and correlation predictions, the realized covariance matrix forecasts are biased. We propose an empirical bias correction to assess the severity of the bias for the out-of-sample realized covariance predictions. Chiriac and Voev (2011), as well as Bauer and Vorkink (2011), proposed empirical corrections for the bias induced by the Cholesky factorization and the matrix logarithm. They estimate an empirical bias correction factor from the ratio between the true and fitted realized volatility. We extend this idea to the entire covariance matrix and define the empirical bias correction factor as:

$$bc_{ji} = \text{med} \left(\{ \nu_{ji,t+1|t} \}_{t=\tilde{T}-W}^{\tilde{T}-1} \right) \quad \text{with} \quad \nu_{ji,t+1|t} = \frac{\text{RCov}_{ji,t+1|t}}{\widehat{\text{RCov}}_{ji,t+1|t}} \quad (3.18)$$

for $j = 1, \dots, N$ and $i = 1, \dots, N$, where $\tilde{T}$ is the last observation of the rolling estimation window of length W . The bias correction factors are estimated for the HAR-DRD model and its tree-structured extensions. Bias-corrected forecasts are obtained as $\widehat{\text{RCov}}_{ji,t+1|t}^{\text{bc}} = bc_{ji} \widehat{\text{RCov}}_{ji,t+1|t}$. Panel A in Table 3.5 reports the summary statistics of the estimated bias

Table 3.5: Summary of the results with empirical bias correction

Panel A: Summary statistics of the empirical bias correction factor					
	HAR- DRD	HAR- DRD Leverage	HAR- DRD Attention	HAR- DRD Senti- ment	HAR- DRD All
Mean	0.947	0.942	0.942	0.948	0.947
Standard deviation	0.024	0.023	0.024	0.024	0.024
25%-quantile	0.933	0.927	0.927	0.933	0.932
Median	0.950	0.944	0.943	0.949	0.949
75%-quantile	0.964	0.959	0.959	0.964	0.964
Panel B: Realized covariance forecast results with bias correction					
	HAR- DRD	HAR- DRD Leverage	HAR- DRD Attention	HAR- DRD Senti- ment	HAR- DRD All
Euclidean (RCov)	1.052*	1.055*	1.055*	1.052*	1.057*

NOTE: The table summarizes the results of the empirical bias correction of the HAR-DRD model of Oh and Patton (2016) as well as its tree-structured extensions. More precisely, Panel A reports the summary statistics of the empirical bias correction factors estimated over a rolling window of five years. The empirical bias correction factor is defined as the median of the ratio between the true and the fitted 1-day ahead realized covariance. Panel B reports the loss, in terms of the Euclidean distance, of 1-day ahead bias-corrected realized covariance predictions. The reported losses are relative to the corresponding loss obtained without applying the empirical bias correction, where a value larger than one indicates that the bias correction reduces the model's predictive accuracy. Results for which Diebold-Mariano two-sided tests reject the hypothesis of equal forecasting accuracy at the 1%, 5%, and 10% levels, after applying the Bonferroni-Holm correction (Holm, 1979), are marked with ***, **, and *, respectively.

correction factors. The results indicate that the non-linear transformation outlined in Equation (3.5) induces a bias. The bias correction factors are of a similar magnitude as those reported by Chiriac and Voev (2011) for the Cholesky decomposition. In Panel B of Table 3.5, the accuracy of the bias-corrected forecasts for 1-day ahead realized covariance predictions are presented. We evaluate the predictive accuracy in terms of the Euclidean distance.¹⁵ The reported losses are relative to the corresponding loss obtained without applying the empirical bias correction, where a value larger than one indicates that the bias correction reduces the model's pre-

¹⁵Note that, Table 3.5 reports the Euclidean loss for the realized covariance matrix, while Section 3.4.2 evaluates the predictive accuracy for the realized correlation matrix.

dictive accuracy. The forecasting results clearly show that the empirical bias correction does not improve the accuracy thereof. On the contrary, for all HAR-DRD specifications, the uncorrected versions outperform the bias-corrected forecasts and we can reject the null hypothesis of equal predictive accuracy at the 10% level with Bonferroni-Holm corrected p-values (Holm, 1979).

The bias appears to be small in magnitude and a simple empirical correction does not improve the forecasting results of the HAR-DRD model or its tree-structured extensions. Therefore, this study does not correct the bias induced by the non-linearity of Equation (3.5) and leaves the derivation of an analytical bias correction for future research.

3.B Robustness

Table 3.6: Realized correlations and value-at-risk forecasting results for models estimated over a rolling window of four and a half years

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Attention	HAR- DRD Sentiment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.190	1.000	0.983 ^{a, b}	0.987 ^{a, b}	1.000	0.982^{a, b}
VaR _{1%}	1.063	1.000	0.984	0.969^{a, b}	1.002	0.981
VaR _{5%}	1.019	1.000	0.997	0.991^{a, b}	1.001	0.996
VaR _{10%}	1.005 ^a	1.000	0.999 ^a	0.995^{a, b}	1.000	0.999 ^a
<u>5-day ahead prediction</u>						
Euclidean	1.423	1.000 ^{a⁵ b²}	0.964 ^{a⁵ b⁴}	0.980 ^{a⁵ b²}	1.002 ^{a⁴ b²}	0.956^{a⁵ b⁵}
VaR _{1%}	0.978^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.987 ^{a⁵ b⁴}	0.995 ^{a⁵ b⁴}	1.005 ^{a⁵ b⁴}	0.992 ^{a⁵ b⁴}
VaR _{5%}	1.016 ^{a⁵ b⁴}	1.000 ^{a⁵ b⁴}	0.997^{a⁵ b⁴}	1.000 ^{a⁵ b⁴}	1.001 ^{a⁵ b⁴}	0.999 ^{a⁴ b³}
VaR _{10%}	1.020 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	1.000^{a⁵ b⁴}	1.000 ^{a⁵ b⁵}	1.001 ^{a⁵ b⁴}	1.000 ^{a⁵ b⁴}
<u>10-day ahead prediction</u>						
Euclidean	1.497	1.000 ^{a⁹ b⁷}	0.968 ^{a¹⁰ b⁸}	0.984 ^{a⁹ b⁷}	1.009 ^{a⁷ b⁶}	0.955^{a¹⁰ b¹⁰}
VaR _{1%}	0.932^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b⁹}	0.985 ^{a¹⁰ b⁷}	0.991 ^{a¹⁰ b⁹}	1.007 ^{a¹⁰ b⁷}	0.976 ^{a¹⁰ b¹⁰}
VaR _{5%}	1.004 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.998 ^{a⁹ b⁸}	0.998 ^{a¹⁰ b¹⁰}	1.003 ^{a⁹ b⁹}	0.992^{a¹⁰ b¹⁰}
VaR _{10%}	1.023 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	1.001 ^{a⁹ b⁹}	1.001 ^{a¹⁰ b¹⁰}	1.002 ^{a¹⁰ b⁸}	0.997^{a¹⁰ b¹⁰}
<u>22-day ahead prediction</u>						
Euclidean	1.586	1.000 ^{a¹⁹ b¹³}	0.999 ^{a²¹ b¹⁵}	0.959^{a²² b²¹}	1.007 ^{a¹⁹ b¹³}	0.971 ^{a²² b²⁰}
VaR _{1%}	1.039 ^{a¹⁴ b¹¹}	1.000 ^{a²² b²⁰}	0.994^{a²² b¹⁵}	0.999 ^{a²² b¹⁸}	1.006 ^{a²² b²²}	0.998 ^{a²² b¹⁷}
VaR _{5%}	1.025 ^{a²⁰ b¹⁷}	1.000^{a²² b²¹}	1.002 ^{a²⁰ b¹⁶}	1.000 ^{a²² b¹⁷}	1.001 ^{a²² b²²}	1.000 ^{a²⁰ b¹⁹}
VaR _{10%}	1.022 ^{a²² b²⁰}	1.000^{a²² b²¹}	1.001 ^{a²⁰ b¹⁶}	1.000 ^{a²² b²¹}	1.000 ^{a²² b²²}	1.000 ^{a²² b²¹}

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio. The models are estimated every 22 observations over a rolling window of four and a half years. Multi-period forecasts are obtained by direct predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts *a* and *b*, respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, . . . , Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts *a* and *b* the number of times a model is included in the respective model confidence set.

Table 3.7: Realized correlations and value-at-risk forecasting results for models estimated over a rolling window of five and a half years

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Atten- tion	HAR- DRD Senti- ment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.178	1.000	0.986 ^{a, b}	0.986 ^{a, b}	1.002	0.986 ^{a, b}
VaR _{1%}	1.073	1.000	0.986 ^a	0.970 ^{a, b}	1.006	0.986 ^a
VaR _{5%}	1.022	1.000	0.999 ^a	0.991 ^{a, b}	1.001	1.000 ^a
VaR _{10%}	1.005 ^a	1.000 ^a	1.001 ^a	0.995 ^{a, b}	1.001 ^a	1.001 ^a
<u>5-day ahead prediction</u>						
Euclidean	1.405	1.000 ^{a⁴ b³}	0.965 ^{a⁵ b⁴}	0.981 ^{a⁵ b⁴}	1.003 ^{a⁵ b³}	0.967 ^{a⁵ b⁵}
VaR _{1%}	0.990 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.987 ^{a⁵ b³}	0.994 ^{a⁴ b⁴}	1.009 ^{a⁵ b³}	0.991 ^{a⁵ b³}
VaR _{5%}	1.016 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁴}	0.997 ^{a⁵ b⁴}	1.001 ^{a⁵ b⁴}	1.002 ^{a⁵ b⁴}	1.000 ^{a⁴ b³}
VaR _{10%}	1.023 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.998 ^{a⁵ b⁵}	1.002 ^{a⁵ b³}	1.001 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁴}
<u>10-day ahead prediction</u>						
Euclidean	1.481	1.000 ^{a⁹ b⁷}	0.962 ^{a⁹ b⁹}	0.989 ^{a⁹ b⁸}	1.007 ^{a⁹ b⁷}	0.954 ^{a¹⁰ b¹⁰}
VaR _{1%}	0.918 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.985 ^{a⁹ b⁹}	0.984 ^{a¹⁰ b¹⁰}	1.005 ^{a¹⁰ b⁹}	0.980 ^{a¹⁰ b⁹}
VaR _{5%}	1.005 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.998 ^{a⁹ b⁹}	0.997 ^{a¹⁰ b⁹}	1.003 ^{a¹⁰ b⁹}	0.995 ^{a¹⁰ b¹⁰}
VaR _{10%}	1.027 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b⁷}	0.999 ^{a⁹ b⁹}	1.000 ^{a¹⁰ b⁹}	0.998 ^{a¹⁰ b¹⁰}
<u>22-day ahead prediction</u>						
Euclidean	1.567	1.000 ^{a²¹ b¹⁷}	1.001 ^{a²¹ b¹⁷}	0.971 ^{a²⁰ b¹⁹}	1.009 ^{a²⁰ b¹⁷}	0.960 ^{a²² b²²}
VaR _{1%}	1.077 ^{a¹⁶ b¹⁰}	1.000 ^{a²² b²⁰}	1.007 ^{a²² b¹⁷}	0.998 ^{a²² b¹⁷}	1.009 ^{a²² b¹⁹}	1.003 ^{a²² b²⁰}
VaR _{5%}	1.024 ^{a²⁰ b¹⁸}	1.000 ^{a²² b¹⁹}	1.002 ^{a²⁰ b¹⁸}	1.000 ^{a²¹ b¹⁹}	1.001 ^{a²² b¹⁹}	0.995 ^{a²² b²²}
VaR _{10%}	1.027 ^{a²² b²¹}	1.000 ^{a²² b²¹}	1.002 ^{a²⁰ b¹⁹}	1.000 ^{a²² b¹⁹}	1.001 ^{a²² b²⁰}	0.996 ^{a²² b²²}

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio. The models are estimated every 22 observations over a rolling window of five and a half years. Multi-period forecasts are obtained by direct predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, ..., Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

Table 3.8: Realized correlations and value-at-risk forecasting results with at least 10 observations per stock-pair in each partition cell

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Attention	HAR- DRD Sentiment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.188	1.000	0.985 ^{a, b}	0.986 ^{a, b}	1.001	0.985 ^{a, b}
VaR _{1%}	1.067	1.000	0.990 ^a	0.973 ^{a, b}	1.006	0.983 ^{a, b}
VaR _{5%}	1.020 ^a	1.000 ^a	0.999 ^a	0.992 ^{a, b}	1.001 ^a	0.998 ^a
VaR _{10%}	1.005 ^a	1.000 ^a	1.000 ^a	0.995 ^{a, b}	1.000 ^a	0.999 ^a
<u>5-day ahead prediction</u>						
Euclidean	1.419	1.000 ^{a², b²}	0.958 ^{a⁴, b⁴}	0.978 ^{a⁴, b⁴}	1.002 ^{a³, b¹}	0.953 ^{a⁵, b⁵}
VaR _{1%}	0.987 ^{a⁵, b⁵}	1.000 ^{a⁵, b⁵}	0.977 ^{a⁵, b⁵}	1.000 ^{a², b⁴}	1.006 ^{a⁵, b³}	0.978 ^{a², b⁴}
VaR _{5%}	1.013 ^{a⁵, b⁵}	1.000 ^{a⁵, b⁵}	0.996 ^{a⁵, b⁵}	1.001 ^{a⁵, b⁵}	1.002 ^{a⁵, b⁵}	0.995 ^{a⁵, b⁵}
VaR _{10%}	1.024 ^{a⁵, b⁵}	1.000 ^{a⁵, b⁵}	0.999 ^{a⁵, b⁵}	1.003 ^{a⁵, b⁵}	1.001 ^{a⁵, b⁵}	0.999 ^{a⁵, b⁵}
<u>10-day ahead prediction</u>						
Euclidean	1.491	1.000 ^{a⁸, b⁸}	0.970 ^{a⁹, b⁹}	0.972 ^{a², b⁹}	1.007 ^{a⁸, b⁷}	0.949 ^{a¹⁰, b¹⁰}
VaR _{1%}	0.931 ^{a¹⁰, b¹⁰}	1.000 ^{a¹⁰, b¹⁰}	0.981 ^{a⁹, b⁸}	0.987 ^{a¹⁰, b⁹}	1.010 ^{a¹⁰, b⁷}	0.981 ^{a¹⁰, b⁹}
VaR _{5%}	1.007 ^{a¹⁰, b¹⁰}	1.000 ^{a¹⁰, b¹⁰}	0.997 ^{a⁹, b⁹}	1.000 ^{a¹⁰, b⁸}	1.003 ^{a¹⁰, b⁹}	0.996 ^{a¹⁰, b¹⁰}
VaR _{10%}	1.028 ^{a¹⁰, b⁹}	1.000 ^{a¹⁰, b¹⁰}	0.999 ^{a¹⁰, b¹⁰}	1.002 ^{a¹⁰, b⁸}	1.001 ^{a¹⁰, b¹⁰}	0.998 ^{a¹⁰, b¹⁰}
<u>22-day ahead prediction</u>						
Euclidean	1.575	1.000 ^{a²¹, b²⁰}	0.997 ^{a²², b²⁰}	0.974 ^{a²², b²²}	1.007 ^{a²¹, b¹⁹}	0.987 ^{a²², b²¹}
VaR _{1%}	1.060 ^{a¹⁷, b¹²}	1.000 ^{a²², b¹⁷}	1.003 ^{a²², b¹⁸}	1.003 ^{a²², b¹⁴}	1.006 ^{a²², b¹⁹}	1.010 ^{a²², b¹⁶}
VaR _{5%}	1.028 ^{a²⁰, b¹⁸}	1.000 ^{a²¹, b²⁰}	1.004 ^{a²², b¹⁵}	1.004 ^{a²¹, b¹⁶}	1.001 ^{a²², b²¹}	1.005 ^{a²¹, b¹⁷}
VaR _{10%}	1.027 ^{a²¹, b²⁰}	1.000 ^{a²¹, b²¹}	1.002 ^{a²¹, b¹⁸}	1.004 ^{a²², b¹⁹}	1.000 ^{a²², b²²}	1.003 ^{a²², b²⁰}

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio. The models are estimated every 22 observations over a rolling window of five years and a minimum of 10 observations per stock-pair is required in each partition cell. Multi-period forecasts are obtained by direct predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, ..., Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

Table 3.9: Realized correlations and value-at-risk forecasting results with at least 100 observations per stock-pair in each partition cell

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Attention	HAR- DRD Sentiment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.188	1.000	0.984 ^{a, b}	0.986 ^{a, b}	1.001	0.984 ^{a, b}
VaR ₁ %	1.067	1.000	0.985 ^{a, b}	0.973 ^{a, b}	1.006	0.983 ^{a, b}
VaR ₅ %	1.020 ^a	1.000 ^a	0.998 ^a	0.992 ^{a, b}	1.001 ^a	0.998 ^a
VaR ₁₀ %	1.005 ^a	1.000 ^a	1.000 ^a	0.995 ^{a, b}	1.000 ^a	1.000 ^a
<u>5-day ahead prediction</u>						
Euclidean	1.419	1.000 ^{a³ b²}	0.963 ^{a⁴ b⁴}	0.972 ^{a⁴ b⁴}	1.000 ^{a³ b¹}	0.958 ^{a⁴ b⁴}
VaR ₁ %	0.987 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁴}	0.982 ^{a⁵ b⁵}	0.997 ^{a⁵ b⁵}	1.004 ^{a⁵ b⁴}	0.986 ^{a⁵ b⁵}
VaR ₅ %	1.013 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.996 ^{a⁵ b⁵}	0.999 ^{a⁵ b⁵}	1.001 ^{a⁵ b⁵}	0.996 ^{a⁵ b⁵}
VaR ₁₀ %	1.024 ^{a⁵ b⁵}	1.000 ^{a⁵ b⁵}	0.998 ^{a⁵ b⁵}	1.001 ^{a⁵ b⁵}	1.001 ^{a⁵ b⁵}	0.999 ^{a⁵ b⁵}
<u>10-day ahead prediction</u>						
Euclidean	1.491	1.000 ^{a⁷ b⁷}	0.972 ^{a¹⁰ b⁸}	0.975 ^{a¹⁰ b⁸}	1.004 ^{a⁷ b⁷}	0.958 ^{a¹⁰ b¹⁰}
VaR ₁ %	0.931 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.987 ^{a¹⁰ b⁸}	0.986 ^{a¹⁰ b¹⁰}	1.011 ^{a¹⁰ b⁶}	0.989 ^{a¹⁰ b⁸}
VaR ₅ %	1.007 ^{a¹⁰ b¹⁰}	1.000 ^{a¹⁰ b¹⁰}	0.997 ^{a¹⁰ b⁹}	0.999 ^{a¹⁰ b⁹}	1.003 ^{a¹⁰ b¹⁰}	0.998 ^{a¹⁰ b¹⁰}
VaR ₁₀ %	1.028 ^{a¹⁰ b⁹}	1.000 ^{a¹⁰ b⁹}	0.999 ^{a¹⁰ b¹⁰}	1.001 ^{a⁹ b⁹}	1.001 ^{a¹⁰ b⁹}	0.998 ^{a¹⁰ b⁹}
<u>22-day ahead prediction</u>						
Euclidean	1.575	1.000 ^{a²¹ b¹⁶}	0.994 ^{a²² b¹⁸}	0.966 ^{a²² b²²}	1.007 ^{a²⁰ b¹⁷}	0.972 ^{a²² b²⁰}
VaR ₁ %	1.060 ^{a¹⁷ b¹³}	1.000 ^{a²² b¹⁸}	1.010 ^{a²² b¹⁶}	0.998 ^{a²² b¹⁹}	1.006 ^{a²² b²¹}	1.006 ^{a²² b¹⁹}
VaR ₅ %	1.028 ^{a¹⁹ b¹⁷}	1.000 ^{a²¹ b¹⁸}	1.004 ^{a²¹ b¹⁵}	0.999 ^{a²¹ b¹⁷}	1.001 ^{a²² b²¹}	1.006 ^{a²¹ b²⁰}
VaR ₁₀ %	1.027 ^{a²¹ b²⁰}	1.000 ^{a²¹ b¹⁷}	1.001 ^{a²⁰ b¹⁸}	1.000 ^{a²¹ b¹⁹}	1.000 ^{a²² b¹⁹}	0.999 ^{a²² b²¹}

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio. The models are estimated every 22 observations over a rolling window of five years and a minimum of 100 observations per stock-pair is required in each partition cell. Multi-period forecasts are obtained by direct predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, ..., Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

Table 3.10: Realized correlations and value-at-risk forecasting results obtained from iterative predictions

	HAR	HAR- DRD	HAR- DRD Leverage	HAR- DRD Atten- tion	HAR- DRD Sentiment	HAR- DRD All
<u>1-day ahead prediction</u>						
Euclidean	1.188	1.000	0.985 ^{a, b}	0.986 ^{a, b}	1.001	0.984 ^{a, b}
VaR _{1%}	1.069	1.000	0.989 ^a	0.973 ^{a, b}	1.005	0.982 ^{a, b}
VaR _{5%}	1.020 ^a	1.000 ^a	0.998 ^a	0.992 ^{a, b}	1.001 ^a	0.997 ^a
VaR _{10%}	1.005 ^a	1.000 ^a	1.000 ^a	0.995 ^{a, b}	1.001 ^a	0.999 ^a
<u>5-day ahead prediction</u>						
Euclidean	1.403	1.000 ^{a, b}	0.968 ^{a, b}	0.972 ^{a, b}	1.005 ^{a, b}	0.964 ^{a, b}
VaR _{1%}	1.040 ^{a, b}	1.000 ^{a, b}	0.979 ^{a, b}	0.998 ^{a, b}	1.004 ^{a, b}	0.978 ^{a, b}
VaR _{5%}	1.018 ^{a, b}	1.000 ^{a, b}	0.996 ^{a, b}	1.000 ^{a, b}	1.000 ^{a, b}	0.996 ^{a, b}
VaR _{10%}	1.017 ^{a, b}	1.000 ^{a, b}	0.997 ^{a, b}	1.000 ^{a, b}	1.000 ^{a, b}	0.998 ^{a, b}
<u>10-day ahead prediction</u>						
Euclidean	1.477	1.000 ^{a, b}	0.976 ^{a, b}	0.967 ^{a, b}	1.012 ^{a, b}	0.965 ^{a, b}
VaR _{1%}	1.033 ^{a, b}	1.000 ^{a, b}	0.971 ^{a, b}	0.981 ^{a, b}	1.011 ^{a, b}	0.968 ^{a, b}
VaR _{5%}	1.027 ^{a, b}	1.000 ^{a, b}	0.994 ^{a, b}	0.997 ^{a, b}	1.002 ^{a, b}	0.994 ^{a, b}
VaR _{10%}	1.023 ^{a, b}	1.000 ^{a, b}	0.997 ^{a, b}	0.999 ^{a, b}	1.000 ^{a, b}	0.998 ^{a, b}
<u>22-day ahead prediction</u>						
Euclidean	1.580	1.000 ^{a, b}	0.994 ^{a, b}	0.949 ^{a, b}	1.011 ^{a, b}	0.986 ^{a, b}
VaR _{1%}	1.180 ^{a, b}	1.000 ^{a, b}	0.978 ^{a, b}	0.982 ^{a, b}	1.013 ^{a, b}	0.978 ^{a, b}
VaR _{5%}	1.033 ^{a, b}	1.000 ^{a, b}	0.993 ^{a, b}	0.994 ^{a, b}	1.005 ^{a, b}	0.995 ^{a, b}
VaR _{10%}	1.017 ^{a, b}	1.000 ^{a, b}	0.996 ^{a, b}	0.997 ^{a, b}	1.003 ^{a, b}	0.997 ^{a, b}

NOTE: The table reports the 1-day, 5-day, 10-day, and 22-day ahead forecasting results of the realized correlation matrix and the value-at-risk of an equally weighted portfolio. The models are estimated every 22 observations over a rolling window of five years. Multi-period forecasts are obtained by iterating forward 1-day ahead predictions. Forecasts of the correlation matrix are evaluated in terms of the Euclidean distance loss function. Value-at-risks are computed using filtered historical simulations (10,000 replications) and assessed with the quantile loss function. All results are reported relative to the average loss of the HAR-DRD model, and the lowest loss is highlighted in boldface. Models belonging to the 90% and 70% model confidence sets are marked with superscripts a and b , respectively. For multi-period forecasts, we compute the loss functions and MCS for all possible starting points, e.g., for the 5-day horizon Monday-to-Monday, Tuesday-to-Tuesday, ..., Friday-to-Friday. The table reports the average losses across all starting points and beside the superscripts a and b the number of times a model is included in the respective model confidence set.

Chapter 4

Structural breaks in online investor sentiment: A note on the nonstationarity of financial chatter

Finance Research Letters (2020), 35, 101479

Co-author: Simon Behrendt

Abstract

Given the increasing interest in investor sentiment derived from social media platforms, we address one overlooked question – are there structural breaks in online investor sentiment? We cast the problem of break-point estimation in the dynamics of the sentiment series as a model selection problem. Considering 360 stocks, we detect structural breaks in most of the respective online investor sentiment series. A return prediction exercise illustrates the economic significance of the detected structural breaks. Our results call into question the widespread practice of using online investor sentiment series without taking into account the nonstationarity induced by structural breaks.

Keywords: investor sentiment; structural breaks; nonstationarity; group Lasso

4.1 Introduction

It has become increasingly popular among investors to share their opinion about or to comment on companies' stock market performances and perceived prospects on social media platforms such as Twitter and Stock-Twits. Following the initial findings of Antweiler and Frank (2004) and Das and Chen (2007), a considerable amount of research has evolved around the questions of (i) how to estimate investor sentiment, i.e., subjective beliefs about future cash flows and investment risks, from such online sources and (ii) how to augment forecasts of financial variables using these sentiment measures (for a recent survey, see Nardo, Petracco-Giudici, & Naltsidis, 2016). For instance, Behrendt and Schmidt (2018) use Twitter sentiment to augment forecasts of intraday volatility and Audrino, Sigrist, and Ballinari (2020) determine the effectiveness of several sentiment measures for predicting daily stock market volatility. While such time series applications make use of online investor sentiment time series at different frequencies and generally assume stationarity of these series, the previous literature has overlooked one important question – are there structural breaks in online investor sentiment? It is likely that company-related news or general shifts in investors' opinion may change the dynamics of the sentiment series attributed to a given stock and thus introduce subtle kinds of nonstationarity, especially when long time horizons are considered in an empirical analysis (Audrino & Teterewa, 2019). As a result, the usage of such nonstationary time series in models requiring stationarity may be questionable and, consequently, modeling

approaches taking into account changing dynamics in the online investor sentiment series may prove valuable endeavors in empirical applications. In fact, empirical findings based on other measures of latent investor sentiment than those obtained from social media platforms point towards distinguishable effects of investor sentiment during periods of different market conditions. For example, Kurov (2010) finds that the impact of monetary policy news on the stock market and investor sentiment, the latter measured either by the index of Baker and Wurgler (2006, 2007) or by an index composed from the Investor Intelligence survey, is different for bear and bull markets. In addition, Garcia (2013) shows that the predictability of stock returns using investor sentiment estimated from financial news is predominantly restricted to recession periods. However, since there exists no preferred proxy for latent investor sentiment (Baker & Wurgler, 2006, 2007), the results of our empirical analysis, as elaborated upon in detail below, are mostly restricted to the discussion related to online investor sentiment.

We add to the existing literature by investigating changes in the dynamics of daily investor sentiment time series estimated from short messages published on Twitter and StockTwits for 360 stocks over a seven years time period. The problem of structural break detection is cast as a model selection problem and solved by applying the state-of-the-art procedure of Chan, Yau, and Zhang (2014). This two-step procedure allows us to consistently estimate the unknown number and timing of structural breaks in a given sentiment series by assuming that each stationary segment of the series is approximated well by an autoregressive

(AR) process. Indeed, we find evidence that a large number of investor sentiment series contain structural breaks and that these breaks are also dependent on the sentiment source, i.e., Twitter or StockTwits. We illustrate the potential effect of ignoring the detected structural breaks by drawing upon a simple return prediction exercise.

The remainder of the paper is structured as follows: Section 4.2 outlines the two-step procedure for break-point detection of Chan et al. (2014), while Section 4.3 briefly describes our data set and online investor sentiment measures. Results of the structural break estimation are presented in Section 4.4 and Section 4.5 illustrates the economic implications of the detected structural breaks within a simple return prediction exercise. Lastly, Section 4.6 concludes.

4.2 Detecting structural breaks in the dynamics of a time series

Given that AR processes often describe simple stationary time series data reasonably well, we consider a structural break AR process with $(m + 1)$ regimes in our empirical analysis of investor sentiment time series:

$$Y_t = \sum_{j=1}^{m+1} [\beta_j' Y_{t-1} + \sigma(Y_{t-1}, \dots, Y_{t-q}) \varepsilon_t] \mathbf{1}\{t_{j-1} \leq t < t_j\}, \quad (4.1)$$

$$t = 1, 2, \dots, n,$$

where $\mathbf{Y}_{t-1} = (1, Y_{t-1}, \dots, Y_{t-p})'$, $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \dots, \beta_{jp})' \in \mathbb{R}^{p+1}$, $1 = t_0 < t_1 < \dots < t_{m+1} = n+1$, and $\sigma(\cdot)$ denoting a measurable function on $\mathbb{R}^q$, which allows for the possibility of conditionally heteroscedastic errors. Moreover, the $\{\varepsilon_t\}$ are white noise with zero mean and unit variance. Let break-points be indexed by $\{t_1, \dots, t_m\}$, i.e., at time t_j the parameter $\boldsymbol{\beta}_j$ changes to $\boldsymbol{\beta}_{j+1}$. We are now able to rewrite the model in Equation (4.2) as a high dimensional regression model using the following notation: $\mathbf{Y}_n^0 = (Y_1, Y_2, \dots, Y_n)'$, $\boldsymbol{\eta}(n) = (\sigma_1 \varepsilon_1, \sigma_2 \varepsilon_2, \dots, \sigma_n \varepsilon_n)'$, with $\sigma_t = \sigma(Y_{t-1}, \dots, Y_{t-q})$, $\boldsymbol{\theta}(n) = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_n)'$, and $\mathbf{X}_n$ is an $(n \times np)$ -matrix:

$$\mathbf{X}_n = \begin{pmatrix} \mathbf{Y}'_0 & 0 & 0 & \dots & 0 \\ \mathbf{Y}'_1 & \mathbf{Y}'_1 & 0 & \dots & 0 \\ \mathbf{Y}'_2 & \mathbf{Y}'_2 & \mathbf{Y}'_2 & \dots & 0 \\ \vdots & & & & \\ \mathbf{Y}'_{n-1} & \mathbf{Y}'_{n-1} & \mathbf{Y}'_{n-1} & \dots & \mathbf{Y}'_{n-1} \end{pmatrix}, \quad (4.2)$$

with $\mathbf{Y}'_k = (Y_k, Y_{k-1}, \dots, Y_{k-p+1})$. Letting $\boldsymbol{\theta}(n)$ be the vector of all potential parameter changes, we set $\boldsymbol{\theta}_1 = \boldsymbol{\beta}_1$ for the first parameter and

$$\boldsymbol{\theta}_i = \begin{cases} \boldsymbol{\beta}_{j+1} - \boldsymbol{\beta}_j & \text{when } i = t_j \text{ and } t_j \text{ is a break-point,} \\ \mathbf{0} & \text{otherwise,} \end{cases} \quad (4.3)$$

for $i = 2, \dots, n$. The model in Equation (4.2) can now be written as:

$$\mathbf{Y}_n^0 = \mathbf{X}_n \boldsymbol{\theta}(n) + \boldsymbol{\eta}(n). \quad (4.4)$$

The group Lasso estimator of Yuan and Lin (2006) provides a sparse solution for $\boldsymbol{\theta}(n)$:

$$\hat{\boldsymbol{\theta}}(n) = \arg \min_{\boldsymbol{\theta}(n)} \frac{1}{n} \|\mathbf{Y}_n^0 - \mathbf{X}_n \boldsymbol{\theta}(n)\|^2 + \lambda_n \sum_{i=1}^n \|\boldsymbol{\theta}_i\|, \quad (4.5)$$

where λ_n denotes a tuning parameter and $\|\cdot\|$ refers to the Euclidean norm. The m break-points can be estimated by identifying the nonzero $\hat{\boldsymbol{\theta}}_i$ since $\hat{\boldsymbol{\theta}}_i \neq 0$ for an $i \geq 2$ is equivalent to a change in the AR parameter $\hat{\beta}_i$. We estimate Equation (4.5) with the group-LARS algorithm of Yuan and Lin (2006), where we specify the upper bound for the number of breaks k .

Chan et al. (2014) prove consistency of the estimated break-points when m is unknown. Since group Lasso overestimates the true number of break-points in such a setting, active breaks have to be selected in a second step. Following Chan et al. (2014) by defining an information criterion (IC) based on the minimum description length principle of Davis, Lee, and Rodriguez-Yam (2006), we select active breaks from the set of break-point candidates via minimization of the IC. This minimization is repeated in a backward elimination algorithm, shrinking the set of potential break-points down to the set of active breaks.

4.3 Online investor sentiment data

We estimate daily online investor sentiment from short messages posted on Twitter and StockTwits. While Twitter is a general social media

platform, StockTwits is especially tailored towards traders and investors. Thus, while the topics discussed on Twitter may, in some instances, be of a more general nature, the short messages posted on StockTwits are directly related to financial markets and the performance of stocks. More precisely, for 360 companies that are constantly part of the S&P 500 between January 2011 and December 2017, we collect all messages shared on these social media platforms over that time period which mention a company's name and/or cashtag (the company's ticker symbol preceded by the dollar sign, e.g., "\$INTC" for Intel Corp.). We account for changes in a company's name and ticker over this time period and collect a total of 30,512,412 and 9,887,765 short messages from Twitter and StockTwits, respectively.¹ Following the findings in Ballinari and Behrendt (2019), we make use of the field-specific dictionary L2 for finance-related microblogs outlined in Renault (2017) to estimate the latent investor sentiment from these short messages. We have also considered investor sentiment series estimated using the dictionary of Loughran and McDonald (2011) and the convolutional neural network Deep-MLSA of Deriu et al. (2017). Although the latter dictionary is only optimized for financial content of textual sources and not for the informal speech often encountered in social media short messages, Ballinari and Behrendt (2019) find it to perform reasonably well in financial applications. Deep-MLSA performs very well in the financial applications considered in Ballinari and Behrendt (2019) but, in many cases, it classifies short messages as having a neutral sentiment. Thus, the time series are highly persistent and a larger value

¹A more detailed description of the data, its cleaning, and management, is available from the authors upon request.

for p is often necessary to achieve convergence of the two-step procedure of Chan et al. (2014). Results remain robust qualitatively and are available upon request. Aggregation to a daily frequency is achieved by calculating daily bullishness scores as proposed by Antweiler and Frank (2004). Daily online investor sentiment is calculated from 4pm on the current day to 4pm on the following day, matching the closing of stock markets in the US. Note that we also estimate investor sentiment for non-trading days since investors may also place orders based on information obtained online on weekends and holidays. However, dynamics do not seem to be affected significantly when leaving out non-trading days.

4.4 Structural breaks in online investor sentiment

The results of the procedure of Chan et al. (2014), applied to the online investor sentiment time series for all 360 stocks, are depicted in Figure 4.1 for the Twitter sentiment series and in Figure 4.2 for the StockTwits sentiment series. The figures illustrate for how many stocks a given number of structural breaks, $\hat{m}$, are estimated. Following Chan et al. (2014), we set the upper bound for the number of structural breaks to $k = 10$. We consider different lag lengths p , each $\text{AR}(p)$ depicted by a different panel since for some stocks the sentiment series are highly persistent (due to stale online investor sentiment, mostly classified as being neutral, i.e., coded as zero) and the group-LARS algorithm requires a larger p for the stationary AR model that is active in each regime in order

to converge.² Thus, by considering reasonable lag lengths $1 \leq p \leq 12$ (for example, considering highly persistent return volatility, Chan et al., 2014, use up to 12 lags), we make sure that the algorithm converges for each of the stocks at least for one AR specification and for both sentiment series estimated from short messages published on Twitter and StockTwits. However, note that the non-convergence for some stocks and lag lengths means that the number of stocks does not necessarily add up to 360 for each panel depicted in Figures 4.1 and 4.2. While the maximum number of breaks is higher for the StockTwits sentiment series (5 breaks) than for the Twitter sentiment series (4 breaks), the findings are robust across social media platforms – there exist structural breaks in the autoregressive sentiment dynamics for a large part of the considered stocks. Moreover, for larger p , most sentiment series contain fewer structural breaks. In these cases, we find breaks mostly for these stocks for which the algorithm fails to converge when p is set to a smaller value. Note that our goal here is not to determine the correct p for each online investor sentiment series and segment but instead estimate the number and timing of structural breaks across a range of reasonable lag lengths. If the goal is to also model the time series as a segmented locally stationary AR process in further empirical applications, one should first test for breaks using a reasonably large value of p and then determine the lag length for each segment separately using, for example, the Bayesian Information Criterion (BIC) to select a $p_j \leq p$ for each segment j . The consistent estimation of breaks-points is not affected (see Remark 2.6 of

²To be more precise, in these cases, we encounter a nearly singular matrix in the step concerning the computation of the decent direction in the group-LARS algorithm.

Chan et al., 2014).

Interestingly, on average, we find more structural breaks for stocks with a lower (traditional) news coverage, i.e., there are fewer news releases for these stocks during the considered time period as measured by the proprietary news database RavenPack News Analytics. Moreover, these stocks are among the companies with the smallest market capitalization in our data set. Since we are aiming to detect structural breaks in the dynamics of the online investor sentiment series, the relatively rare news releases for these stocks seem to have a particular impact on the respective online investor sentiment in the sense that the corresponding change in the persistence of the series is large enough to be detected as a structural break.

Figure 4.3 illustrates exemplary investor sentiment series estimated from Twitter and StockTwits together with the corresponding structural breaks that are detected when assuming an $AR(4)$ to be active in each regime for two companies, Southwest Airlines Co. (left panels) and Intel Corp. (right panels). Given that online investor sentiment series do not show a long-memory behavior in many cases, an $AR(4)$ is a reasonable choice for modeling the dynamics of the considered series.³ Considering a single stock, it is obvious that the short messages published on these two social media platforms lead to very different estimates of investor

³In some cases and assuming the absence of structural breaks, the BIC does select a large number of autoregressive terms for some of the considered stocks. For example, the BIC chooses an $AR(8)$ and $AR(10)$ for the online investor sentiment series of Nasdaq Inc., which is based on the L2 dictionary of Renault (2017) and is obtained from short messages published on Twitter and StockTwits, respectively. However, this is due to the fact that for these companies the online investor sentiment is neutral (i.e., it is zero) for a considerable number of days.

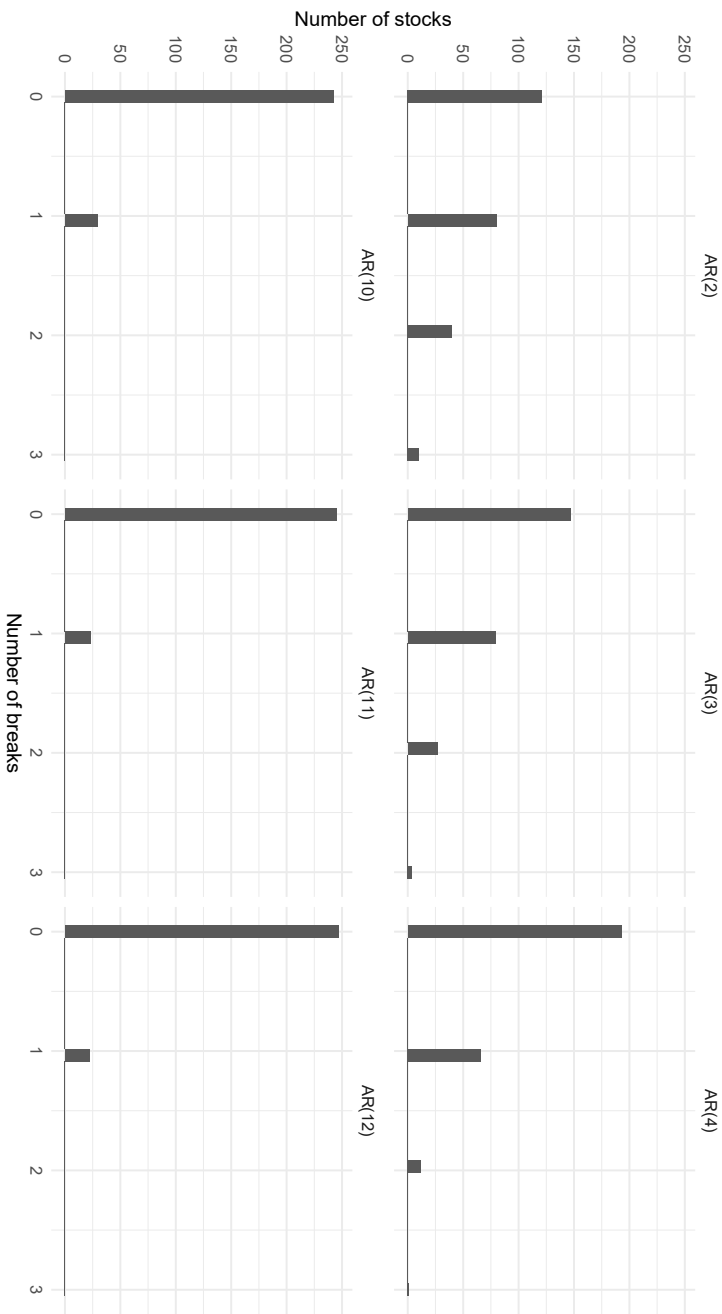


Figure 4.1: Number of structural breaks detected for Twitter online investor sentiment across stocks
NOTE: This figure shows the number of structural breaks detected for the daily investor sentiment derived from Twitter short messages for the 360 analyzed stocks across different lag lengths.

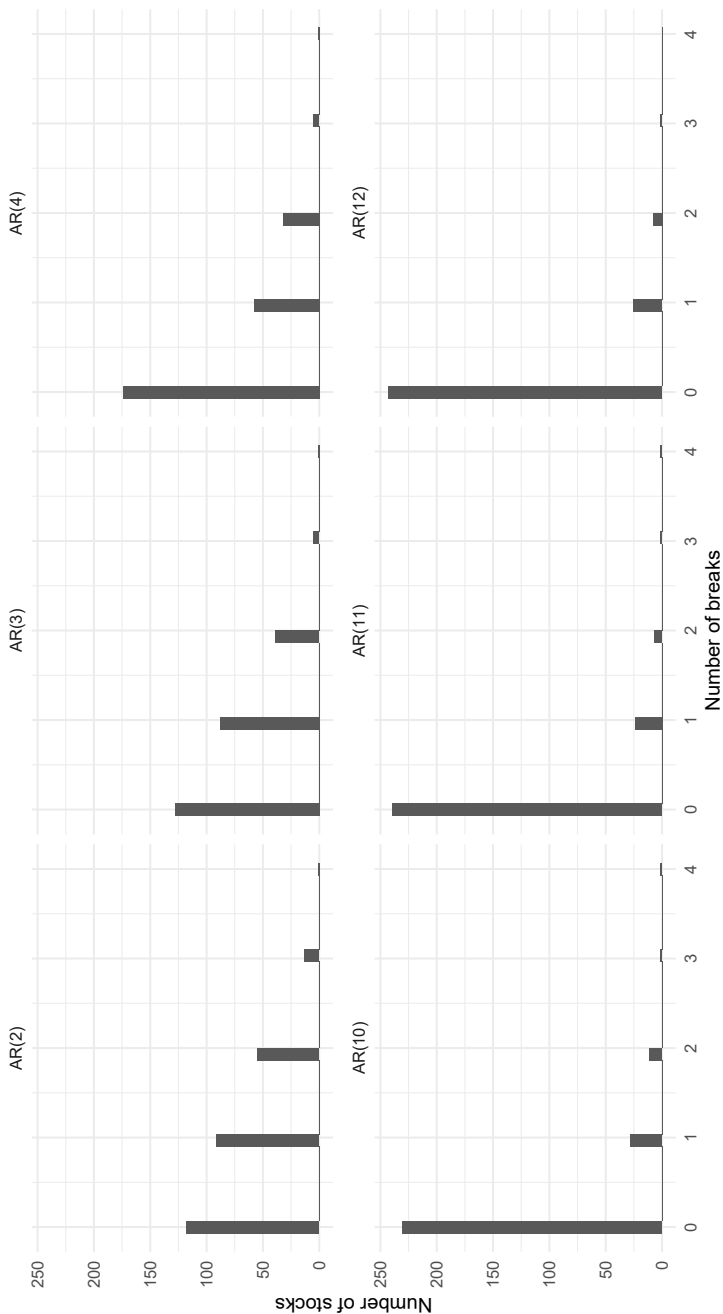


Figure 4.2: Number of structural breaks detected for StockTwits online investor sentiment across stocks
NOTE: This figure shows the number of structural breaks detected for the daily investor sentiment derived from StockTwits short messages for the 360 analyzed stocks across different lag lengths.

sentiment with distinct dynamics in the sentiment series since the number and timing of the estimated structural breaks for the given time series vary across social media platforms. Interestingly, while for Intel Corp. two structural breaks are detected when considering the more general Twitter sentiment, the dynamics seem more stable in the case of the StockTwits sentiment series, where only one structural break is detected. Considering that especially short messages published on Twitter related to such a company as Intel Corp. are not necessarily referring to the respective company's (stock market) performance but may also refer to, for instance, its products, these results seem reasonable.

4.5 The effect of structural breaks: Predicting excess returns

Below, we illustrate exemplary the effect of structural breaks in the autoregressive online investor sentiment dynamics for the stocks of Southwest Airlines Co. and Intel Corp. within the framework of a simple return prediction exercise. To be more precise, we evaluate the predictive power of online investor sentiment estimated from Twitter and StockTwits when forecasting excess returns in a regression framework (i) ignoring structural breaks (i.e., $\hat{m} = 0$) and (ii) separately for each detected regime (i.e., $\hat{m} > 0$) as depicted in Figure 4.3 in the previous section. We control for the market-wide risk-factors introduced by Fama and French (1993) and Carhart (1997), namely lagged (log) market capitalization, market-to-book ratio, excess market returns, and a momentum factor.

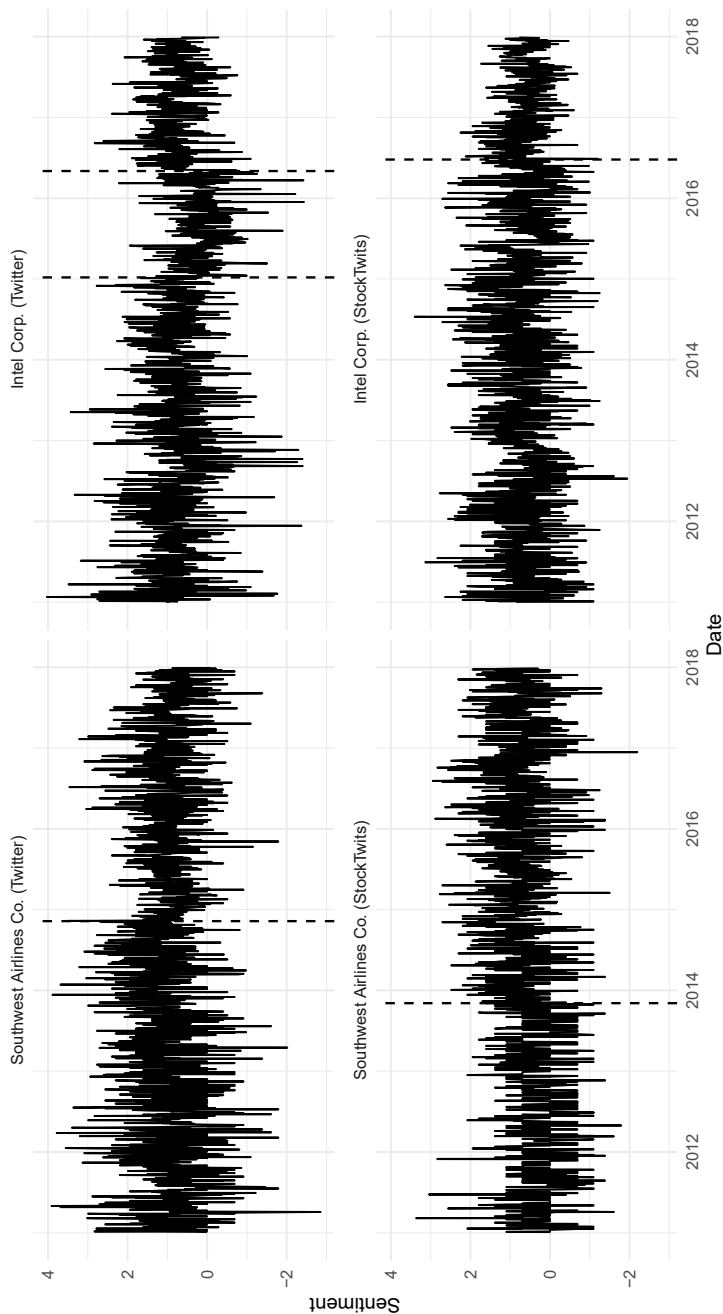


Figure 4.3: Daily Twitter and StockTwits online investor sentiment time series for Southwest Airlines Co. and Intel Corp. **NOTE:** This figure shows the daily online investor sentiment estimated from Twitter and StockTwits messages (solid lines) and the identified structural breaks (dashed vertical lines) for Southwest Airlines Co. and Intel Corp. In this case, p is set to 4, i.e., an AR(4) is active in each regime.

Data are obtained from Compustat and French's website.⁴ For each trading day, we run the following regression:

$$R_{i,t+1} - R_{f,t+1} = \alpha + \sum_{j=1}^{\hat{m}+1} [\beta_j B_{i,t}] \mathbb{1}\{t_{j-1} \leq t < t_j\} + \theta' X_{i,t} + \varepsilon_{i,t+1}, \quad (4.6)$$

where $R_{i,t}$ denotes the raw return in basis points of company i , $i \in \{\text{Southwest Airlines Co., Intel Corp.}\}$, on trading day t , $R_{f,t}$ is the risk-free rate, $B_{i,t}$ the daily online investor sentiment (either estimated from Twitter or StockTwits), and $X_{i,t}$ is the vector of the above-mentioned control variables. Online investor sentiment is scaled such that we can interpret the estimated coefficients as the effect of a one standard deviation increase in sentiment. Statistical inference is based on Newey-West (1987) standard errors.

Table 4.1 depicts the sentiment coefficients β_j for both stocks and social media platforms, Twitter sentiment in Panel A and StockTwits sentiment in Panel B. For the sake of simplicity, we drop the subscript for the sentiment coefficient in the case of $\hat{m} = 0$, i.e., we refer to the respective coefficient as β in the table. As can be seen, by ignoring the presence of structural breaks in Twitter online investor sentiment, we would conclude that sentiment has a significantly positive effect on next day's excess return for Intel Corp. However, between 2015 and 2016 (second regime), the sentiment coefficient estimate for Twitter is negative and not significant. In fact, the predictive effect of sentiment estimated for the first and third regimes is significantly different from that estimated in

⁴See <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french>.

the second regime. A similar pattern can be observed for the StockTwits sentiment of Southwest Airlines Co. While the overall predictive effect appears not to be statistically significant, we observe a negative and not significant estimate for β_1 and a positive and significant estimate for β_2 when estimating the predictive effect separately for the two regimes. These two estimated sentiment coefficients are significantly different from each other. The exemplary results summarised in Table 4.1 also show that return prediction results obtained from sentiment scores vary across social media platforms. We do not find any significant predictive effect of sentiment for the stock returns of Southwest Airlines Co. (Intel Corp.) when constructing the sentiment measure from messages shared on Twitter (StockTwits).

Similar results can also be found for other stocks in our sample. For example, among the companies for which we find at least one structural break in Twitter (StockTwits) online investor sentiment, we detect a significant change in the predictive effect of sentiment for 24% (32%) of the stocks at the 10% significance level. When analysing the contemporaneous relation between stock returns and sentiment, we observe a significant change in the Twitter and StockTwits sentiment coefficient for 40% and 55% of the companies, respectively. Lastly, note that not every structural break necessarily has a significant impact in the context of an economic (forecasting) application (see the discussion about local-to-zero breaks in, for example, Boot & Pick, 2019; Pesaran & Timmermann, 2005, and references cited by these authors).

Table 4.1: Sentiment coefficients for return predictions and their differences

Panel A: Twitter							
	Southwest Airlines Co.			Intel Corp.			
	β	β_1	β_2	β	β_1	β_2	β_3
Estimates	2.17	2.26	1.86	11.80	13.99	-8.96	16.73
		β_1	β_2		β_1	β_2	β_3
t-stat diff.	β_1	-	0.05	β_1	-	1.73	-0.4
in coeffi-	β_2		-	β_2		-	-1.81
cients				β_3			-
Panel B: StockTwits							
	Southwest Airlines Co.			Intel Corp.			
	β	β_1	β_2	β	β_1	β_2	
Estimates	4.00	-		2.67	1.90	8.87	
		13.14	11.47				
		β_1	β_2		β_1	β_2	
t-stat diff.	β_1	-	-2.38	β_1	-	-0.85	
in coeffi-	β_2		-	β_2		-	
cients							

NOTE: This table reports the sentiment coefficients (Estimates) for the stocks of Southwest Airlines Co. and Intel Corp., as obtained from Equation (4.6), for online investor sentiment estimated from two social media platforms, namely Twitter (Panel A) and StockTwits (Panel B). Two scenarios are considered: (i) the case when structural breaks in the sentiment series are ignored (β) and (ii) sentiment coefficients estimated separately for each detected regime ($\beta_j, j = 1, 2, \dots$). The estimated number and timing of structural breaks for each stock and sentiment series are given in Figure 4.3. Statistical inference is based on heteroscedasticity robust standard errors and one-sided significance at the 5% is highlighted by boldfaced numbers. Moreover, we test for statistically significant differences between the β_j (t-stat diff. in coefficients).

4.6 Conclusion

Based on a sample of 360 US-listed stocks and the state-of-the-art breakpoint detection procedure of Chan et al. (2014), we provide empirical evidence that there are structural breaks present in the autoregressive dynamics of most online investor sentiment time series related to these considered stocks. Moreover, the number and timing of these structural breaks are dependent on the sentiment source, i.e., in the case presented here, whether the latent investor sentiment is estimated from short mes-

sages published on Twitter or StockTwits. The latter finding underlines the fact that short messages published on different social media platforms produce very different estimates of investor sentiment. A simple return prediction exercise illustrates the economic significance of the detected structural breaks. Since a large literature is concerned with the estimation of investor sentiment from textual data obtained from online social media platforms and, subsequently, the augmentation of financial variables' forecasts by including these measures of investor sentiment into time series models, it seems important to account for the nonstationarity that is introduced by structural breaks.

Acknowledgements

Funding by the German Research Foundation (Grant PE 2370/2-1) and by the Swiss National Science Foundation (Grant 100018-172685) is greatly appreciated. We thank Alexander Schmidt, Jeannine Polivka, and two anonymous referees for helpful comments. Moreover, we thank Francesco Audrino and Fabio Sigrist for their support in the data acquisition process.

References

- Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance*, 59(3), 1259–1294.
- Audrino, F., Sigrist, F., & Ballinari, D. (2020). The impact of sentiment and attention measures on stock market volatility. *International Journal of Forecasting*, 36(2).
- Audrino, F., & Tetereva, A. (2019). Sentiment spillover effects for US and European companies. *Journal of Banking & Finance*, 106, 542–567.
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4), 1645–1680.
- Baker, M., & Wurgler, J. (2007, June). Investor sentiment in the stock market. *Journal of Economic Perspectives*, 21(2), 129–152.
- Ballinari, D., & Behrendt, S. (2019). *How to gauge investor behavior? A comparison of online investor sentiment measures*. (Working Paper)
- Behrendt, S., & Schmidt, A. (2018). The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance*, 26, 355–367.
- Boot, T., & Pick, A. (2019). Does modeling a structural break improve forecast accuracy? *Journal of Econometrics*, 215(1), 35–59.
- Carhart, M. M. (1997). On persistence in mutual fund performance. *Journal of Finance*, 52(1), 57–82.

- Chan, N. H., Yau, C. Y., & Zhang, R.-M. (2014). Group Lasso for structural break time series. *Journal of the American Statistical Association*, 109(506), 590–599.
- Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, 53(9), 1375–1388.
- Davis, R. A., Lee, T. C. M., & Rodriguez-Yam, G. A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101(473), 223–239.
- Deriu, J., Lucchi, A., De Luca, V., Severyn, A., Müller, S., Cieliebak, M., ... Jaggi, M. (2017). Leveraging large amounts of weakly supervised data for multi-language sentiment classification. In *Proceedings of the 26th international conference on world wide web* (pp. 1045–1052).
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3–56.
- Garcia, D. (2013). Sentiment during recessions. *Journal of Finance*, 68(3), 1267–1300.
- Kurov, A. (2010). Investor sentiment and the stock market’s reaction to monetary policy. *Journal of Banking & Finance*, 34(1), 139–149.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, 66(1), 35–65.
- Nardo, M., Petracco-Giudici, M., & Naltsidis, M. (2016). Walking down Wall Street with a tablet: A survey of stock market predictions

- using the web. *Journal of Economic Surveys*, 30(2), 356–369.
- Newey, W. K., & West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, 28, 777–787.
- Pesaran, M. H., & Timmermann, A. (2005). Small sample properties of forecasts from autoregressive models under structural breaks. *Journal of Econometrics*, 129(1-2), 183–217.
- Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, 84, 25–40.
- Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 68(1), 49–67.

Curriculum Vitae

EDUCATION

Bachelor in Economics, University of St. Gallen, 2016.

Master in Quantitative Economics and Finance, University
of St. Gallen, 2018.

Ph.D. in Economics and Finance, University of St. Gallen, 2020.

WORK EXPERIENCE

Analyst, c-alm AG, 2014–2017.

Research assistant, Faculty of Mathematics and Statistics,
University of St. Gallen, 2017–2020.

TEACHING

Statistics – Exercises, Bachelor in Economics, University of St. Gallen,
2018.

Data Analytics I: Statistics – Exercises, Bachelor in Economics, University
of St. Gallen, 2019–2020.

Data Analytics I: Statistics – Tutorials, Bachelor in Economics, University
of St. Gallen, 2019.

PUBLICATIONS

The Impact of Sentiment and Attention Measures on Stock Market

Volatility, *International Journal of Forecasting*, F. Audrino, F. Sigrist, and D. Ballinari (2020).

Structural breaks in online investor sentiment: A note on the nonstationarity of financial chatter, *Finance Research Letters*, D. Ballinari and S. Behrendt (2020).

PRESENTATIONS

Society of Financial Econometrics, Fudan University, Shanghai, 2019.

Conference on Computational and Financial Econometrics, University of London, 2019.

FUNDING

Swiss National Science Foundation, Grant 100018-172685, 2017–2020.