Investor Sentiment and the Dispersion of Stock Returns: Evidence Based on the Social Network of Investors

Alya AL-NASSERI[a], Faek MENLA ALI[b*], and Allan TUCKER[c]

[a] *Higher College of Technology, Muscat, Sultanate of Oman*
[b] *University of Sussex Business School, University of Sussex, Brighton BN1 9SL, UK*
[c] *Department of Information Systems, Brunel University London, Uxbridge UB8 3PH, UK*

Abstract

This paper extracts an investor sentiment indicator for the 30 DJIA stocks based on the textual classification of 289,024 online tweets posted on the so-called StockTwits, and examines its contemporaneous and predictability effects on the dispersion of stock returns using the quantile regression technique. We find that both contemporaneous and predictability effects of sentiment are heterogeneous throughout the return distribution. Specifically, sentiment is positively contemporaneously associated with stock returns at higher quantiles. However, it is a strong negative predictor of future returns at lower quantiles. Overall, our findings are broadly consistent with most behavioural theories and show that sentiment mainly affects the valuation of assets in extreme market conditions.

Keywords: Investor sentiment, StockTwits, Stock returns, Quantile regression

JEL Classification: C21, G02, G14

☆ We would like to thank the editor and the anonymous referee for the valuable suggestions and comments on an earlier version of this paper. We are grateful for their useful comments and suggestions to Professor Sir David Hendry, Professor Wolfgang Karl Härdle, and participants of the Portsmouth-Fordham Conference on "Banking and Finance", 24-25 September 2016, the University of Portsmouth, the OxMetrics User Conference, Cass Business School, 10-11 September 2018, and the Workshop on "Recent Developments in Financial Data Science and Econometrics", 26-27 September 2018, the University of Loughborough, as well as the seminar at the University of Durham, 24 October 2018.

* Corresponding author (Faek Menla Ali): 208 Jubilee Building, University of Sussex Business School, University of Sussex, Brighton BN1 9SL, UK. Tel: +44(0)1273 872969. *Email address*: F.Menla-Ali@sussex.ac.uk.

1. Introduction

Investor sentiment has long been thought of as an underlying driver of asset price dynamics. However, it was the study by De Long et al. (1990), which explicitly featured the role of sentiment in financial markets documenting that sentiment can cause asset prices to diverge from their fundamental values. Prompted by this theoretical work, a large body of empirical literature has indeed been developed over the years to provide direct evidence for the significant effects of sentiment; this includes, among many others, Baker and Wurgler (2006, 2007) for exploring such effects on assets whose valuations are highly subjective and difficult to arbitrage confirming that when sentiment is high, subsequent returns are low for stocks of extreme growth and distressed firms,[1] Stambaugh et al. (2012) for analysing its effects on a broad range of anomalies in cross-sectional stock returns finding that each anomaly is stronger (its long-short strategy is more profitable) following high levels of sentiment, and Garcia (2013) for showing its significantly larger effects during recessions compared to expansions, and recently Da et al. (2015) for revealing its role in predicting return reversals and the transitory component of its volatility.

However, the fact remains that investor sentiment is not straightforward to measure, with the extant literature to have used various sentiment indicators (see Baker and Wurglar, 2007; Zhou, 2018; for a comprehensive overview of this literature). These indicators include the country fund discount (Bodurtha et al., 1995), household data (Kelly, 1997), the consumer confidence (Schmeling, 2009), the investor intelligence survey (Lee et al., 2002; Brown and Cliff, 2004; Menkhoff and Rebitzky, 2008; Kurov, 2010), the composite index based on the common variation in six underlying proxies for sentiment (namely, the closed-end fund discount,[2] the number and average first-day returns on IPOs, trading volume measured by the share turnover on an exchange, the equity share in total new issues and the dividend premium) (Baker and Wurgler, 2006, 2007),[3] and more recently retail trading correlations around stock splits (Kumar et al., 2013), the equity mutual fund flow (Chiu and Kini, 2014), and shipping sentiment proxies (Papapostolou et al., 2014), among others.

Moreover, despite the rise of online news media and social networking platforms in recent years, relatively little work so far has been carried out on the role of investor sentiment expressed on these platforms in asset price dynamics. Nonetheless, some exceptions, which analyse and find

[1] Specifically, using various sentiment proxies, their empirical findings based on US data showed that when beginning-of-period proxies for sentiment are low (high), subsequent returns are relatively high (low) for small stocks, young stocks, high volatility stocks, unprofitable stocks, non-dividend-paying stocks, extreme growth stocks, and distressed stocks.

[2] The closed-end fund discount was also used by Lee et al. (1991), Swaminathan (1996) and Neal and Wheatley (1998), among others.

[3] Baker et al. (2012), on the other hand, constructed a composite sentiment index at country level (for Canada, France, Germany, Japan, the UK and the US) based on the following four sentiment proxies: The volatility premium, the number and average first-day returns on IPOs, and market turnover. They, then, decomposed the six sentiment indices at country level into a single "global" index and six "local" indices.

evidence of significant effects of online sentiment indicators, extract such indicators from Internet message boards (Antweiler and Frank, 2004), news media articles (see, e.g., Fang and Peress, 2009; Tetlock, 2007; Garcia, 2013; Yuan, 2015; among others), Twitter posts (see, e.g., Zhang et al., 2011; Bollen, 2011; Mao et al., 2011; Zhang et al., 2016; Sprenger et al., 2014; Cookson and Niessner, 2020), and Internet search data (see, e.g., McLaren and Shanbhogue, 2011; Joseph et al., 2011; Saxa, 2014; Da et al., 2015). Bukovina (2016) provides an overview of this growing literature.

The present paper examines the impact of Twitter investor sentiment on stock returns, and makes a twofold contribution. First, online sentiment indicators have just gained popularity and only a handful of studies recently analyse the impact of such indicators on stock market features, hence our work aims to provide fresh evidence to this growing literature. To this end, we employ a sentiment indicator extracted from tweets posted on the so-called StockTwits platform, which is designed for sharing financial information among its users who are mostly investors, traders, entrepreneurs, and alike. Specifically, we filter out 289,024 online tweets posted on this platform that directly reference the 30 DJIA stocks over April 4, 2012 to April 5, 2013 period and then use various classification algorithms from computational linguistics (namely, the Naive Bayes, the Decision Trees, and Support Vector Machines (SVMs)) to classify such tweets into three distinct classes M^c, where c ∈ {Buy, Hold, Sell}; finally, a bullishness of tweets is calculated, as the relative dominance of buy vs sell tweet signals of a stock on a given day *t*, with its shift being used as a proxy for investor sentiment. Most related previous research which analysed Twitter posts use arbitrary subsamples of all tweets posted on its public timeline as a proxy for investor sentiment (see, e.g., Zhang et al., 2011; Bollen, 2011; Mao et al., 2011; Zhang et al., 2016; among others). Therefore, our StockTwits sentiment indicator is relatively more informative, as it does not contain the random noise present on Twitter's public timeline and captures directly financial information that may impact on the investors' decision-making. Some other distinct features of StockTwits platform are that its posted tweets are real-time reflecting investors' true opinions and their up-to-date beliefs as financial events unfold, are mainly by professional investors of different types (e.g., technical, fundamental, etc.), and include both information sources and interpretation of market information, among others (see, e.g., Section 2 and Cookson and Niessner, 2020 for an overview of the features of StockTwits platform). To the best of our knowledge, the few exceptions that have used StockTwits posts in stock return predictions include Sprenger et al. (2014) and Oh and Sheng (2011). The former employ data on the companies listed on the S&P 100 index from January 1, 2010 to June 30, 2010, whereas the latter analyse data related to companies listed on the Nasdaq and NYSE for four months only, i.e., May 11, 2010 to August 7, 2010. Then, this paper differs from them in terms of a non-overlapping time window and the more informative employed sentiment indicator calculated from the considered data set using different classification algorithms from computational linguistics. Also, the paper focuses on the 30 DJIA highly liquid stocks particularly commented on StockTwits. Despite Baker and Wurgler (2006) argued, and their empirical findings confirmed, that large firms are less affected by sentiment, recent work suggest that the role of sentiment in large firms is indeed at play (see, e.g., Garcia, 2013; Ahmad et al., 2019; Lartey et al., 2020).[4] Therefore, our sample of large firms offers a good opportunity to provide further evidence on how sentiment affects such firms in a period of social media's growing influence on financial markets. Moreover, since our firms generate a great buzz and are particularly referenced in posts on StockTwits at a higher frequency basis, our sentiment measure is constructed and its effect is analysed at daily frequency.[5]

Second, the literature mostly ignored the contemporaneous association between sentiment and returns and focused on sentiment as a predictor of asset price dynamics. This has been the case as most behavioural finance theories indicate that sentiment is a strong negative predictor of future returns on grounds that it drives assets from their fundamental values: When sentiment is high (low), stocks gain relatively low (high) subsequent expected returns (see, e.g., Baker and Wurgler, 2006, 2007; Jiang et al., 2019; among others). In this paper, we shed light on both contemporaneous and predictability effects of sentiment. Moreover, unlike previous related studies which confine the analysis to the conditional mean of the return distribution, i.e., using Ordinary Least Squares (OLS) (see, e.g., Antweiler and Frank, 2004; Sprenger et al., 2014; Jiang et al., 2019; among others), our chosen econometric framework is the quantile regression (QR) technique, introduced by Koenker and Bassett (1978), where such a framework is further flexible to analyse whether contemporaneous and predictability effects of sentiment remain the same throughout the return distribution. Intuitively, online sentiment effects on stock returns may differ throughout the return distribution, since the volume or content of tweets expressing sentiment among investors may not be the same around extreme market conditions compared to normal times; moreover, due to limits to arbitrage, mispricing from sentiment could be strong in extreme market conditions (i.e., low/ high return distribution quantiles), thereby its negative predictability of future returns may be more pronounced under such conditions.[6] All in all,

[4] The empirical evidence by Garcia (2013) is based on the aggregate DJIA index, whereas Ahmad et al. (2019) analyse the time-varying relationship between media-expressed firm-specific tone and firm-level returns for 20 large US firms over the 10-year period from January 2001 to December 2010. The role of sentiment in large firms is also supported by Lartey et al. (2020), who found that the negative link between CEO's market sentiment and corporate innovation is particularly strengthened for such firms.

[5] This is compared to Baker and Wurgler (2006)' sentiment proxies which were annual over the period of 1962 to 2001. Note that Baker and Wurgler found that sentiment has broad effects, but the effect for small firms is relatively strong. The focus of this paper is not to test Baker and Wurgler' prediction in terms of how sentiment affects large versus small firms. Rather, it aims to provide evidence on firms which are particularly commented on StockTwits. Despite our large firms are unlikely to have binding constraints on arbitrage or other market frictions in general, Shleifer and Vishny (1997) argue that arbitrage can be quite ineffective in extreme circumstances, hence sentiment may particularly be at play for our firms under such circumstances, and indeed our results confirm this prediction.

[6] Despite our large firms are unlikely to have binding constraints on arbitrage or other market frictions in general, Shleifer and Vishny (1997) argue that arbitrage can be quite ineffective in extreme circumstances, hence sentiment may particularly be at play for our firms under such circumstances.

understanding how returns during different market conditions react to sentiment may indeed provide important insights for investors with the aim of maximising returns and minimising risk.

We find that both contemporaneous and predictability effects of sentiment exert a heterogeneous pattern throughout the return distribution. Specifically, sentiment is positively contemporaneously associated with stock returns at higher quantiles. However, it is a strong negative predictor of future returns at lower quantiles. Overall, our findings are broadly consistent with most behavioural theories and show that sentiment mainly affects the valuation of assets in extreme market conditions.

The remainder of this paper is organised as follows. Section 2 discusses the features of StockTwits platform and the predictive power of its tweet sentiment in forecasting stock returns. Section 3 presents the data set, the classification method of StockTwits messages, and the bullishness measure employed. Section 4 outlines the empirical methodology and the hypotheses tested. Section 5 presents and discusses the empirical results. Finally, Section 6 provides the concluding remarks.

2. StockTwits as a sentiment indicator

StockTwits is a financial communication platform where more than 300,000 users (as of September 2013), who are mostly investors, traders, entrepreneurs, and alike, share investment ideas and opinions about individual stocks and asset markets; today, StockTwits has grown considerably with more than five million registered members and more than seven million messages tweeted monthly.[7] The distinct features of the StockTwits platform provide great support for the role of sentiment expressed on this platform and its predictive ability in forecasting stock price movements. Such features include the high volume of message posts, the real-time message streams, and the succinctness that leads to the efficient diffusion of information among investors (Java et al., 2007; Bollen et al., 2011).

The platform has Twitter-type style, where its posted 140-character limit messages can easily be read and followed by its users, resulting in posts that are to the point without much of the noise found in traditional news articles whose length may cause investors to ignore parts of such articles. Its posts are also advantageous to bulletin boards per each company used by Antweiler and Frank (2004), since the latter result in outdated information in the absence of new posts.[8] Claburn (2009) and Cookson and Niessner (2020) argue that as messages are generally being posted on StockTwits just before an event occurs, this implies that the platform contains real-time information reflecting up-to-date investors' beliefs that are important for making investment decisions.

[7] For more details, see "About StockTwits" at https://about.stocktwits.com/

[8] Bulletin boards, also referred to as message boards, are organised online forums enabling users to read and post information on specific firms and investment-related topics (Wysocki, 1998).

Moreover, unlike Twitter which provides people with the opportunity to tweet about their daily routine, the StockTwits platform is designed for sharing financial information among its users. It follows that, compared to the large scale tweet posts on Twitter's public timeline which contain random noise, the posts are financial-related ones directly related to stock market features; some commentators described the messages posted on this platform as "the modern version of traders shouting in the pits" (BusinessWeek, 2009). Moreover, the high volume of messages posted on Twitter's public timeline every day on a variety of topics make it difficult to extract tweets related to certain companies (e.g., Microsoft ($MSFT) and Apple ($AAPL), etc.), as the names of some companies are extensively mentioned for purposes other than stock investment-related discussions (Ruiz et al., 2012). For example, Apple is a name that is frequently used for spamming purposes (e.g., "Win a free iPhone" scams).

Further, since the platform also enables its users to classify their financial tweets as bullish, bearish, or leave them unclassified (default option), it may provide valuable signals about market movements.[9] Indeed, various trading articles and news stories argue that financial microblogs have a significant impact on stock markets; for example, TIME (2009) puts it "communities of active investors and day traders who are sharing opinions and in some case sophisticated research about stocks, bonds and other financial instruments will actually have the power to move share prices […] making Twitter-based input as important as any other data to the stock". Moreover, Bloomberg (2010) stated that financial professionals have developed the Twitter based trading systems in order to alter users' investment sentiment; in fact, Bloomberg has incorporated Twitter messages including those of StockTwits into their terminals now.

Another useful feature of StockTwits is that its messages are mainly posted by professional investors and likely contain their true beliefs. Cookson and Niessner (2020) point out that investors with different investment approaches (e.g., fundaments, technical, etc.) post on the platform, where the corresponding posted tweets indicate important informational content or interpretation of market information as per investment approach. The authors also show that sentiment reactions of investors to different types of events are consistent with their self-ascribed investment philosophies. Finally, they argue that since StockTwits users cannot subsequently withdraw their posts on the platform as per its policy, this further implies that the posted messages likely reflect the true or reliable opinions of investors.

[9] Our Supplementary Appendix A provides more details regarding the attributes of StockTwits data (Tables A1 and A3) and reports a few typical examples of the StockTwits messages in their original format (Table A2).

3. Data set

In this Section, we describe our data and outline the classification method of the StockTwits messages, then present the bullishness index, constructed from the messages related to the 30 DJIA companies, which is used as a basis for our empirical analysis of the effect of sentiment on stock returns.

3.1. StockTwits data

We examine the relationship between investor sentiment and stock returns utilising online tweets posted on the so-called StockTwits. Specifically, we aim to analyse StockTwits posts that directly reference the DJIA stocks over the period of April 4, 2012 to April 5, 2013, retrieved via its Application Programming Interface (API).[10] StockTwits posts obtained are pre-processed, where posts with more than one ticker, those without any ticker, and those not related to companies on the DJIA index are removed, leaving us with a total of 289,024 relevant posts. That is, these posts consist of 30 stock tickers of the DJIA index containing the dollar-tagged ticker symbol: 27 of them are listed on the NYSE (contributing about 227,194 of the total posts) and the other three are listed on the Nasdaq (contributing about 61,830 of the total posts).

Following Antweiler and Frank (2004), messages are aligned with US market trading hours; in particular, messages posted after 4.00pm (the market closing time) are combined with those of the following trading day, as the effect of these posts on the market indicators can only appear on that day.

Figure 1 shows the evolution of tweet messages, where panels A, B, and C display such messages over the days of our sample period of one year, the days of the week, and the hours of the day, respectively. A graphical inspection suggests that the considered posts for the DJIA stocks are reasonably stable over the considered sample period. Nonetheless, some increase in the volume of posts is observed during the early summer and the autumn months (i.e., Halloween) as well as around Christmas time and the New Year’s Eve (see panel A), suggesting that people tend to post more actively during these special occasions. Moreover, consistent with previous studies (e.g., Oh and Sheng, 2001), the volume of tweets posted during working days is high (i.e., reaching a peak on Thursdays), as opposed to the low volume of posts observed during the weekends (see panel B).

It is also evident that tweet posts are concentrated between 10.00am and 5.00pm (see panel C), hence more tweets are posted during the market hours. That is, tweet postings differ throughout the day between the opening and the closing times of the market compared to the rest of the day. There

[10] The primary data are obtained from Stocktwits.com (http://www.stocktwits.com) (for more details on the StockTwits features, see our Supplementary Appendix A). The DJIA stocks are highly discussed on SockTwits, where high volumes of messages with reference to such stocks are posted every day. However, handling the massive amount of such tweet messages in terms of filtering out the relevant tweets for each of the 30 stocks requires a considerable amount of time and effort. Hence, considering one-year StockTwits data is quite reasonable compared to other related studies such as Oh and Sheng (2011) and Sprenger et al. (2014), who collected StockTwits data for three and six months, respectively.

are several possible explanations for this pattern. First, tweet volume following the opening time of the market may be influenced by the actual trading activities and the real-time market fluctuations. Second, other market forces such as recommendations by analysts and other financial advisors, who are likely to be active during the market hours, may strongly affect the volume of posted tweets. The tweet volume after the market closes, by contrast, is more likely to be based on investors' logical and intuitive analysis of the financial information available to them.

[Insert Figure1 about here]

3.2. The classification method

To construct our investor sentiment indicator, we initially need to classify messages as sell, hold, or buy. To this end, to manage the huge amount of the collected messages, a random sample of 2,892 tweet messages is first selected and manually classified.[11] Note that such a sample of messages has been randomly selected from all companies included in the analysis and from different periods. To make sure that our random sample is not biased and is a good representative one, a systematic random sampling technique is used. The corpus of manually labelled posts is picked according to two simple rules: (i) The sample of tweet posts selected comprises an equal number of posts from all 30 companies of the DJIA index (approximately 96 tweet posts per company), and (ii) The sample is collected from different periods to avoid any bias in the selection process (i.e., tweets are selected from different months of the year, from different days of the week, and from different times of the day).

The manual classification of the 2,892 messages is performed by the researchers (the primary judge) based on the Harvard IV-4 classification dictionary.[12] [13] Moreover, consistent with most studies based on text classification methods using manual training data sets (e.g., Antweiler and Frank, 2004; Sprenger et al., 2014), a second judge has further worked independently to perform a manual classification of the same training data set using the coding scheme reported in Table 1 to achieve greater reliability and consensus regarding the classification carried out.[14]

[Insert Table 1 about here]

[11] Following most text classification methods using a manual training set (see, e.g., Antweiler and Frank, 2004; and Sprenger et al., 2014), we manually classify 1% of the total StockTwits data (i.e., 2,892 tweets) as sell, buy and hold.

[12] In the General Inquirer's Harvard IV-4 classification dictionary, more than 4,000 emotional words are tagged and classified as either positive or negative. Since a bull message indicates that an investor is optimistic and provides a 'buy' signal to the market participants, it is therefore likely to associate positive emotions with the 'buy' class. On the other hand, when an investor posts a bear message, this indicates that the investor is pessimistic and sends a 'sell' signal to other market participants. The 'hold' class is more likely to contain an equal balance of positive and negative emotions.

[13] Many studies in finance that use textual analysis to quantify information in news stories or social media content have mainly used the Harvard IV-4 dictionary for classification tasks (see, e.g., Tetlock, 2007; Engelberg, 2008; Kothari et al., 2009; Loughran and McDonald, 2011; Boudoukh et al., 2013), although Loughran and McDonald (2011) question its suitability in financial contexts. In this paper, the General Inquirer's Harvard-IV-4 classification dictionary of emotional words is used to add each occurrence of emotional words in a message to the bag of words; in this way, we build on the results of Tetlock et al. (2008), who found that the fraction of (negative) emotional words in firm-specific news stories can predict firms' accounting earnings and stock returns.

[14] Using Cohen's Kappa statistics, we find high inter-rater reliability of 81.4%.

Finally, to implement the classification exercise of the messages conditional on our training data set, we use various classification algorithms from computational linguistics. In the normal setting, the machine learning algorithms are designed for maximising the classification accuracy and minimising the error rate (Kukar and Kononenko, 1998). Previous studies used one or two algorithms for classifying the messages, with the Naïve Bayes being the primarily used algorithm (e.g., Antweiler and Frank, 2004; Sprenger et al., 2014). However, this technique has limitations in sentiment classification as it assumes the conditional independence of words in documents. By contrast, we take a more comprehensive approach by comparing the classification results of three different machine learning algorithms, namely the Naïve Bayes, the Decision Trees, and SVMs. By training the selected sample of StockTwits messages in Weka software and using these algorithms, we find that the (Random Forest) Decision Tree classifier has a higher accuracy rate than the Naïve Bayes and SVMs (for more details, see our Supplementary Appendices B, C, and D).[15]

3.3. The Decision Tree classifier

As established earlier in this paper, we use the Random Forest Decision Tree classifier of the Weka machine learning package (see Hall et al., 2009). Inputs to the model come from the training corpus of 2,892 tweets (the representative sample), which are manually coded based on the Harvard IV-4 dictionary and labelled as buy (1), hold (0), and sell (-1) signals. The results of the percentage allocation of the manual classifications of tweet messages into the three distinct classes are shown in Table 2, which shows that roughly half of these messages are 'buy' signals (47.06%). The 'sell' signals, by contrast, account for roughly one third of the messages (32.54%), whereas the 'hold' signals represent one-fifth of them (20.40%).

[Insert Table 2 about here]

Hence, StockTwits posts seem to be more balanced in terms of their distribution of 'buy' vs 'sell' signals compared to Internet message boards, where the ratio of 'buy' vs 'sell' signals appears to be unbalanced, ranging from 7:1 (e.g., Dewally, 2003) to 5:1 (e.g., Antweiler and Frank, 2004). Moreover, the finding that 'hold' messages constitute a relatively small percentage of 20.40% is not consistent with that of Sprenger et al. (2014), who found that almost half of the messages manually classified are 'hold' signals.

All in all, the small proportion of 'hold' signals indicates that little noise is involved in the StockTwits posts on the DJIA stocks. On the other hand, the higher distribution of 'buy' and 'sell' messages may provide evidence of the existence of more relevant financial information in such posts. Amazingly, the greater proportion of 'buy' messages, consistent with the study by Dewally (2003),

[15] Future work could employ the ensemble technique to construct the sentiment indicator from the data set, which combines the outputs of several base classification algorithms to form an integrated output (see, e.g., Xia et al., 2011; Fersini et al., 2014).

may serve as a proxy for positive investor sentiment expecting stock prices to rise as investors become more bullish and optimistic, hence demanding more of these stocks for their portfolios.

To better understand the nature of the classified messages, it is helpful to look at some examples. Table 3 provides a few typical examples of manually classified tweets from the training data set including manual coding. Table 4, on the other hand, reports the classification accuracy by class. By using the Decision Tree classifier, the overall in-sample classification accuracy is 66.7%. This is considered a good percentage giving a random chance of 33% for the three classes (i.e., buy, sell and hold).

[Insert Tables 3 and 4 about here]

Further, to make sure that our classification accuracy is good enough, we perform out-of-sample testing. In Weka, training (the manually labelled data set) on the first ten months of the year and testing on the remaining two months generate two separate data sets using the supplied test options: the training (in-sample) set and the testing (hold-out) one, which are 1,953 and 939 instances, respectively. As is shown in Table 5, using the supplied test set by training on the entire first ten months' corpus (from April 2012-January 2013) while testing on the remaining two months' corpus (February 2013-March 2013) yielded an accuracy of 60.50%.

Also, the findings of the two methods of training and testing show that both the automatic percentage split using one data set and the supplied test set using two separate training and testing sets consistently yielded accuracy levels somewhere between 60% and 66%. Since the supplied test set methods (training and testing sets) are considered more reliable than randomly split data sets, the accuracy rate achieved by the supplied test set of 60.50%, which is still in the range of the accuracy interval of the percentage split (60-66%), will, therefore, be used to apply the classification results to the entire population of StockTwits data.

Table 6 provides a comparison of the manual classification of hold-out messages and the automated classification of the Random Forest algorithm. The results suggest that the Random Forest algorithm performs reasonably well, as indicated by the relatively small numbers of misclassifications in each sentiment class.

Finally, Table 7 shows the assigned labels for the full set of StockTwits posts. The reported posts are 140,350, 26,157, and 122,517 for 'buy', 'hold', and 'sell' classes, respectively. Following Antweiler and Frank (2004) and many others in this literature, the 'hold' posts are removed from the analysis as they are considered noise and convey neutral opinions, while posts with relative sentiment only (i.e., 140,350 + 122,517 = 262,867) remain useful and are employed for the analysis.

[Insert Tables 5 to 7 about here]

3.4. The bullishness index

Since there are hundreds of daily messages posted every day, these messages need to be aggregated to link them with market movements on a daily basis. Nonetheless, days without any tweets (i.e., silent periods) are replaced with zeros, in the spirit of Antweiler and Frank (2004).[16] Moreover, seeing that the classification algorithm classified the whole tweet messages into three distinct classes M^c, where c ∈ {Buy, Hold, Sell}, the bullishness of messages can be considered, which is a tweet sentiment measure used to aggregate the different message classes in a given time interval. In this paper, we follow Antweiler and Frank (2004) and define bullishness, denoted by B_t,[17] as follows:

$$B_t = M_t^{Buy} - M_t^{Sell}, \tag{1}$$

where M_t^{Buy} and M_t^{Sell} indicate the total number of messages conveying 'buy' and 'sell' signals on day t, respectively. Because a markedly substantial number of messages are tweeted daily, we normalise B_t as it will assist the model's estimation. More specifically, as B_t may contain negative values and to consider such values, the following formula of normalisation is considered:

$$B_{it}^* = \frac{(B_{it} - minB_i\ )}{(minB_i\ - maxB_i\ )}, \tag{2}$$

where B_{it}^* is the normalised bullishness of stock i at time t, and $maxB_i$ and $minB_i$ indicate respectively the maximum and minimum values of the bullishness measure of stock i over the sample period.[18] Note that the normalised bullishness is homogenous of degree between zero and one, in line with the bullishness measure used by Antweiler and Frank (2004). Also, our measure is similar to the investor sentiment index of Wang (2001), who proxied sentiment by different types of traders taking into account their minimum and maximum aggregated positions.

Finally, it is worth noting that since the 'buy' and 'sell' messages indicate that an investor is being bullish and bearish respectively, it is likely that the 'buy' ('sell') message will be associated with a bullish (bearish) investor. The bullishness index, calculated at the end of each day, then represents the dominance of a particular sentiment (buy or sell) on a stock.[19]

---

[16] Empirical studies suggest two possible ways of dealing with the missing observations in the data set either by replacing the missing periods with the medians of the respective measures or by filling those missing values with zeros.

[17] The bullishness measure excludes the number of messages expressing the 'hold' sentiment: M_t^{Hold}. The reason for excluding the 'hold' messages is that this type of message holds neutral opinions, and hence it does not have an effect on the bullishness measure. Moreover, in most cases this set of messages may contain some amount of noise that may bias and distort the bullishness signal (Antweiler and Frank, 2004).

[18] The maximum and minimum values of bullishness in Eq. (2) will be different for each stock of the DJIA index in the panel series.

[19] Future research may further classify the posted messages by some other attributes (e.g., by the attributes of the StockTwits users, etc.), so sentiment due to noise trading, mood swings, overconfidence and other irrational investor behaviours could be analysed.

3.5. Stock returns

We employ daily closing prices of the DJIA stocks over the period April 4, 2012 to April 5, 2013, obtained from Bloomberg. There are no extraordinary market conditions reported during this specific period, therefore it represents a good base test for the analysis. In this paper, the focus is placed on the DJIA index to adequately reflect the US stock market. The DJIA is a price-weighted average of 30 blue-chip stocks traded on the NYSE and the Nasdaq. Regardless of the limitations in the composition and the structure of the index, it is nevertheless the most widely followed and reported stock index (Lee et al., 2002). Further, the DJIA index is particularly suitable because it constitutes the largest market capitalisation of the industrial companies in the US equity market. As of 2012, the 30 stocks that make up the index have a market capitalisation of $3,896.24 billion and comprise about 20.86 percent of the total market value of the US stock market.[20] Therefore, focusing on large and highly liquid firms will probably reduce the problems associated with non-concurrent trading (Rudd, 1979). This, in fact, makes the DJIA a reasonably valuable index for representing short-term market movements.

In addition, since companies listed on the DJIA are actively traded ones, their stocks generate a great 'buzz' on social media networks. That is, such stocks are heavily discussed on StockTwits and have a very high volume of tweet messages. Moreover, since the impact of sentiment on returns regarding large firms has drawn relatively less attention in the literature, our paper is an attempt to provide direct evidence in this regard.

Finally, as for the return series of the DJIA stocks (denoted as R_{it} for stock *i* at time *t*), they are computed by taking the first differences of the logarithm of the daily closing prices, multiplied by 100.

4. The methodology

4.1. The OLS model

The primary aim of this paper is to examine the impact of our online sentiment indicator on stock returns. The benchmark model commonly used in the literature is specified as

$$R_{it} = \alpha_i + \gamma L^s \Delta B_{it}^* + \lambda_1 L^s lnMSG_{it} + \lambda_2 MKT_t + \lambda_3 NWK_t + \varepsilon_{it}, \tag{3}$$

where R_{it} is the daily stock returns for stock *i* at time *t*; ΔB_{it}^* is the shift in bullishness measure, which proxies investor sentiment, for stock *i* at time *t*;[21] $lnMSG_{it}$ is the log of StockTwits messages of stock

[20] These figures are of authors' calculations.

[21] Following Lee et al. (2002), we use changes or shifts in bullishness, computed as $\Delta B_{it}^* = B_{it}^* - B_{it-1}^*$.

i at time t to capture the volume of messages posted for each stock (see, e.g., Antweiler and Frank, 2004; Sprenger et al., 2014); MKT_t is the index returns of the DJIA to capture the overall market-wide effects,[22] NWK_t is a dummy variable for the first day of the trading week to capture the potential Monday return anomaly effect,[23] and ε_{it} is the error term of stock i at time t. Note that L_s is an s-lag operator, more precisely we estimate two separate models by setting $s=0$ (i.e., contemporaneous effects), and then by setting $s=1$ (predictability or lead-lag effects).[24]

4.2. The quantile regression

The OLS analysis in the previous subsection considers sentiment effects on the average (mean) returns. However, sentiment may impact parts of the return distribution rather than the mean. This is especially the case, since the volume or content of online posts conveying sentiment among investors may not be the same around extreme market conditions (i.e., extreme positive/negative returns) compared to normal times; and also, sentiment negative predictability of future returns may be more pronounced, since it may cause stronger mispricing due to limits to arbitrage, under such conditions.

Moreover, given the stylised fact that financial returns are not normally distributed (see also Table 8 for our case), the QR has several advantages, which, in turn, can address some of the potential pitfalls of earlier studies. First, the QR technique provides more robust estimates compared to those obtained by the OLS, since such a model is unresponsive to the effect of the outliers in the data and also to the non-normal distribution feature of the error term. Second, as documented in Chevapatrakul (2015) among others, an assumption concerning the distribution of the error term is not required, given the semiparametric nature of the QR. As shown by Feng et al. (2008), Ma and Pohlman (2008), Huang et al. (2008), Baur et al. (2012), Alagidede and Panagiotidis (2012), and Chevapatrakul (2015) among many others, the QR has been particularly appropriate for modelling stock returns.

Consequently, we revisit the relationship between sentiment and returns by using the QR technique, developed by Koenker and Bassett (1978). The conditional quantile function takes the following form

$$q_\tau(R_{it}|L^s\Delta B^*_{it}, L^s lnMSG_{it}, MKT_t, NWK_t) = \alpha_\tau + \gamma_\tau L^s\Delta B^*_{it} + \lambda_{1\tau}L^s lnMSG_{it} + \lambda_{2\tau}MKT_t + \lambda_{3\tau}NWK_t, \quad (4)$$

where τ denotes the τ-th conditional quantile of stock returns (i.e., $\tau \in (0,1)$). Hence, γ_τ refers to the

[22] In this way, we analyse the role of investor sentiment in the context of the so-called market model (see, e.g., Stapleton and Subrahmanyam, 1983).
[23] Monday effect indicates that stock returns tend to be lower or negative on Mondays relative to other days of the week (see Thaler, 1987, and references therein, for its early documentation).
[24] For a random process Y_t, $L^sY_t = Y_t$ when s=0, and $L^sY_t = Y_{t-1}$ when s=1.

parameter estimate of the shift in bullishness on a specific τ-th conditional quantile of returns. The QR model is also estimated separately by setting $s=0$ and then by setting $s=1$ (i.e., contemporaneous vs predictability effects of sentiment).

The parameters in Eq. (4) are estimated using linear programming techniques (see Koenker and D'Orey, 1987)[25] by solving the following minimisation problem:

$$\min_{\alpha_\tau,\gamma_\tau,\lambda_{1\tau},\lambda_{2\tau},\lambda_{3\tau}} \sum_{t=1}^{T} \rho_\tau(R_{it} - \alpha_\tau - \gamma_\tau \Delta B_{it}^{*} - \lambda_{1\tau} lnMSG_{it} - \lambda_{2\tau} MKT_t - \lambda_{3\tau} NWK_t), \text{ for } s=0, \quad (5a)$$

and

$$\min_{\alpha_\tau,\gamma_\tau,\lambda_{1\tau},\lambda_{2\tau},\lambda_{3\tau}} \sum_{t=2}^{T} \rho_\tau(R_{it} - \alpha_\tau - \gamma_\tau \Delta B_{it-1}^{*} - \lambda_{1\tau} lnMSG_{it-1} - \lambda_{2\tau} MKT_t - \lambda_{3\tau} NWK_t), \text{ for } s=1, \quad (5b)$$

where $\rho_\tau(z)$ refers to the check function given by $\rho_\tau(z) = z(\tau - \mathbf{1}_{\{z\le 0\}})$, with $\mathbf{1}_{\{z\le 0\}}$ being the indicator function taking only two values: $\mathbf{1}_{\{z\le 0\}}=1$ if $z \le 0$, and 0 otherwise. In this way, $\rho_\tau(z)$ gives different weights for positive and negative residuals depending on the value of τ (see Koenker and Hallock, 2001; Chevapatrakul, 2015). Finally, Eq. (4) is estimated with 9 quantiles (i.e., τ = 0.05, 0.1, 0.25…0.95), which are further classified into three different quantile levels: low, medium, and high. The rule of thumb followed in this paper is that a quantile level exerts a statistically significant effect if there are at least two adjacent quantiles that are statistically significant in that corresponding quantile's level. The standard errors are obtained using the bootstrap method.

5. Empirical results and discussion

In this Section, we first report a summary of descriptive statistics, then we provide estimates of the contemporaneous and predictive effects of sentiment on stock returns using the OLS and QR methods. Finally, we check the robustness of our results.

5.1. Descriptive statistics

Table 8 provides the summary statistics on DJIA stock returns and the shift in bullishness. Our sample includes 7,530 observations for the 30 DJIA stocks over 251 trading days. The mean of stock returns is 0.045, while that of the shift in bullishness is 0.001. Further, stock returns are characterised by higher volatility than the shift in bullishness. The highest return observed in the data is 10.49%, whereas the lowest is -10.96%. This range is unusually large for firms included in the DJIA index.[26] The shift in

[25] For more details on the QR techniques, the reader is directed to the surveys by Buchinsky (1998) and Koenker and Hallock (2001).

[26] Winsorisation of returns has also been performed to reduce the effect of outliers in the data. The results (available upon request) showed that the relationship between returns and other studied variables does not change in terms of magnitude and statistical significance.

bullishness, by contrast, ranges from 0.919 to -0.898, which represent the maximum and minimum bullish and bearish sentiments, respectively. Overall, the Jarque-Bera (JB) test statistics imply a rejection of the null hypothesis that returns and the shift in bullishness are normally distributed.

Figure 2 displays the evolution of the normalised bullishness index (panel A) and the daily DJIA index returns (panel B). The index shows an increase in performance of about 2.25%. The bullishness signal also exhibits a gradual increase over the sample period, although such an increase is not significant. Table 9 reports summary statistics for each firm. It is evident that a high volume of messages is generally posted for all firms and that the messages are often bullish. The number of messages ranges from 1,312 for United Technologies Corporation to 35,336 for JP Morgan. Moreover, the reported firm size by market capitalisation (as of 2012) confirms that our sample firms are large. Our largest firm is Exxon Mobil ($389.64 billion) whereas the smallest is Travelers ($27.10 billion). The number of the messages for the former is 7,904 and for the latter is 1,344, indicating that larger firms draw more attention from StockTwits participants.[27]

[Insert Tables 8 and 9 and Figure 2 about here]

5.2. Contemporaneous effects of sentiment

The numerical results of the contemporaneous effects of sentiment using the OLS and the QR, allowing for company fixed effects, are reported in Table 10. The corresponding parameter estimates of γ_τ and their 95% confidence intervals (shaded area) against τ along with the OLS estimate (dashed line) and its 95% confidence interval (dotted lines) are plotted in Figure 3.

[Insert Table 10 and Figure 3 about here]

The OLS results show a positive contemporaneous effect of sentiment, but it is statistically significant at the 10% level (see Table 10). This suggests that sentiment does contain little information in explaining the mean of stock returns. Antweiler and Frank (2004) and Sprenger et al. (2014), by contrast, found highly statistically significant contemporaneous positive effects of sentiment on returns.

Notwithstanding this, the OLS estimate does not tell the whole story and convey little information regarding sentiment effects as compared to that of QR, which shows that such effects vary over the conditional quantiles of the return distribution (see Table 10 or Figure 3). Specifically, the QR results show that estimates of γ_τ are negative (positive) at low (high) quantiles, where they exhibit statistical significance at higher quantiles but not at lower ones. Moreover, regardless of the statistical

[27] The correlation between firm size (as of 2012) and average message volume per firm is 21.6%.

significance, the magnitude of γ_τ generally increases (in either sign) as τ moves out from the medium towards the lower and upper quantiles. This implies that as we move away from the 0.5 percentile towards the tails of the return distribution, the impact of sentiment changes markedly. Thus, sentiment exerts different effects over the return distribution, with such effects becoming stronger at the very extreme quantiles (0.05 and 0.95). The size of these coefficients in absolute value is larger at the higher (i.e., $\gamma_{0.95}$ = +0.880) compared to the lower quantiles (i.e., $\gamma_{0.05}$= -0.343).

Following Buchinsky (1998), we also perform a symmetric quantiles test to examine whether the sentiment-return relationships at the τ-th and $(1-\tau)$-th quantiles are symmetric about the median, i.e., $\gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$. That is, we test whether the following equation

$$\hat{\lambda}_\tau^T = \hat{\gamma}_{(\tau)}^T + \hat{\gamma}_{(1-\tau)}^T - 2\hat{\gamma}_{(0.5)}^T \tag{6}$$

is different from zero. The restriction in Eq. (6) is set for the pair of τ as of (0.05, 0.95), (0.1, 0.9), . . ., (0.45, 0.55). Specifically, we compute a $\chi^2(1)$ test based on the square of the normalised $\hat{\lambda}_\tau^T$ for each pair, with the standard error of $\hat{\lambda}_\tau^T$ being obtained via design matrix bootstrap (see Buchinsky,1998; Chuang et al., 2009). The null hypothesis of symmetric quantile effects cannot be rejected for all τ pairs (see Table 11); thus, sentiment effects are almost all symmetric about the median. The symmetry of these quantiles also provides us with an explanation of the weak effect by the OLS (Table 10), since the corresponding upper and lower quantiles exhibit opposed effects (i.e., positive and negative effects, respectively).

[Insert Table 11 about here]

All in all, our findings broadly indicate a noticeable inverted S-shaped pattern of contemporaneous effects of sentiment across the quantiles of the return distribution: Sentiment exhibits significant positive effects on returns at higher quantiles but negative and insignificant effects, on average, at lower ones. This result implies that the contemporaneous impact of investor sentiment on returns depends on the state of the market (i.e., low vs high quantiles of returns). As far as the control variables are concerned (Table 10), the number of StockTwits messages seems insignificant in the OLS model, consistent with the empirical findings of Antweiler and Frank (2004) and Sprenger et al. (2014); however, it exhibits a similar, but more statistically significant, pattern to sentiment effects in the QR model; specifically, it is negative (positive) and statistically significant in the low (high) quantiles. This further confirms that the QR model provides additional insights compared to the OLS used in previous studies. Finally, unlike the first day of the week effect shown to be insignificant across all quantiles, market returns are significant and positive across all quantiles; these results are in line with those of the OLS which show that market returns are significant (and positive) but the first day of the week effect is not.

5.3. Our sentiment indicator as a predictor of future returns

The regression results for the contemporaneous effects of sentiment presented in the earlier subsection may suffer from endogeneity problem, where such a set up may also imply that sentiment could possibly be driven by returns. Therefore, we also analyse the predictive ability of our sentiment indicator using both the OLS and QR models. Table 12 reports the results of such predictive regressions. Figure 4, on the other hand, plots the corresponding estimates of γ_τ across the quantiles along with the OLS estimate.

Our results show that the estimated slope on the sentiment indicator in the OLS model is negative, albeit it is statistically significant at the 10% level (i.e., γ =-0.165).[28] Most behavioural finance theories indicate that sentiment leads to overvaluation/undervaluation of assets. Hence, compared to Antweiler and Frank (2004) and Sprenger et al. (2014) who found insignificant predictive sentiment effects on returns, this finding of sentiment as a negative predictor of future returns is consistent with these theories: When sentiment is low (high), this leads to high (low) subsequent returns. Our finding is also in line with the empirical finding of Jiang et al. (2019), who found that a manager sentiment index, constructed based on the aggregate textual tone of corporate financial disclosures, is a negative predictor of future returns.

The estimates of the QR further show that the negative slope on sentiment varies across the quantiles. In specific, as we move from the lower quantiles towards the higher ones, the magnitude and statistical significance of this slope diminish. Therefore, sentiment as a negative predictor of returns is statistically and economically significant only in the lower quantiles, implying that investor sentiment mainly affects the valuation of assets in turbulent times. A possible explanation for this finding is that mispricing is stronger in the lower quantiles due to limits to arbitrage, where Shleifer and Vishny (1997) indeed argue that arbitrage can be quite ineffective in extreme circumstances, hence sentiment is particularly at play in such quantiles. As for the symmetry of effects across the quantiles (Table 13), predictive sentiment effects are symmetric about the median at the 5% for all τ pairs, except the (0.45, 0.55) one.

Finally, the predictive impact of the number of messages seems broadly the same as its contemporaneous one. It is insignificant in the OLS model, as also found by Sprenger et al. (2014) but unlike Antweiler and Frank (2004) who found a significantly predictive negative effect instead. However, it is negative (positive) and statistically significant in the low (high) quantiles. As for Monday effect, it seems to be positive and significant in the lower quantiles only now.

[28] Since finance theory suggests a negative sign on γ, one may conduct a one-tailed test (see, e.g., Jiang et al., 2019), of which our effect would be statistically significant at the 5% level.

5.4. Further analysis

In this subsection, we conduct some robustness checks of our earlier findings. Specifically, we first re-estimate our earlier models but using the Fama and French (1993) 3-factor model instead. Next, we check the robustness of our findings by estimating the models across two equally divided sub-samples. Finally, we check for the reverse effect.

5.4.1. Using Fama-French (1993) model

Our earlier results were based on the market model, where sentiment was included as an additional factor along with the number of messages and the Monday dummy. In this subsection, we check the robustness of our findings by employing the Fama and French (1993) 3-factor model instead, so we control for additional risk factors when analysing sentiment effects on returns. That is, we estimate the following OLS model

$$R_{it} - R_{ft} = \alpha_i + \gamma L^s \Delta B^*_{it} + \beta_1 (MKT_t - R_{ft}) + \beta_2\, SMB_{it} + \beta_3 HML_{it} + \lambda_1 L^s lnMSG_{it} + \lambda_2 NWK_t + \varepsilon_{it}, \tag{7}$$

and the following conditional quantile function

$$q_\tau(R_{it} - R_{ft} \mid L^s \Delta B^*_{it}, MKT_t - R_{ft}, SMB_{it}, HML_{it}, L^s lnMSG_{it}, NWK_t) = \alpha_\tau + \gamma_\tau L^s \Delta B^*_{it} + \beta_{1\tau}(MKT_t - R_{ft}) + \beta_{2\tau} SMB_{it} + \beta_{3\tau} HML_{it} + \lambda_{1\tau} L^s lnMSG_{it} + \lambda_{2\tau} NWK_t, \tag{8}$$

where $R_{it} - R_{ft}$ is the excess return for stock *i* at time *t*, $MKT_t - R_{ft}$ is the excess return on the market index at time *t*, SMB_{it} denotes the daily return difference between the return on small size stocks and the return on big size stocks (i.e., size factor) and HML_{it} denotes the daily return difference between the return on high value stocks and the return on low value stocks (i.e., value factor). The notations of the rest of the variables remain the same.

Note that both models, Eqs. (7) and (8), are also estimated by considering contemporaneous and predictability effects of sentiment (i.e., by setting *s*=0, 1). The parameters in Eq. (8), on the other hand, are now estimated by solving the following minimisation problem:

$$\min_{\alpha_\tau,\gamma_\tau,\beta_{1\tau},\beta_{2\tau},\beta_{3\tau},\lambda_{1\tau},\lambda_{2\tau}} \sum_{t=1}^{T} \rho_\tau (R_{it} - \alpha_\tau - \gamma_\tau \Delta B^*_{it} - \beta_{1\tau}(MKT_t - R_{ft}) - \beta_{2\tau} SMB_{it} - \beta_{3\tau} HML_{it} - \lambda_{1\tau} lnMSG_{it} - \lambda_{2\tau} NWK_t), \text{ for } s=0, \tag{9a}$$

and

$$\min_{\alpha_\tau,\gamma_\tau,\beta_{1\tau},\beta_{2\tau},\beta_{3\tau},\lambda_{1\tau},\lambda_{2\tau}} \sum_{t=2}^{T} \rho_\tau\big(R_{it} - \alpha_\tau - \gamma_\tau \Delta B^*_{it-1} - \beta_{1\tau}(MKT_t - R_{ft}) - \beta_{2\tau} SMB_{it} - \beta_{3\tau} HML_{it} - \lambda_{1\tau} lnMSG_{it-1} - \lambda_{2\tau} NWK_t\big), \text{ for } s\text{=1}. \quad (9b)$$

Overall, employing the Fama and French (1993) model based on OLS and QR methods further confirms our previous findings of both the contemporaneous and predictive effects of sentiment (see Tables 14 and 16 and Figures 5 and 6). Specifically, the OLS estimates suggest that the contemporaneous impact of sentiment on excess returns is positive but statistically insignificant, whereas the QR estimates show that such an impact of sentiment is positive and significant at higher quantiles only (see Table 14 or Figure 5). As for the predictive effects, our OLS results show that the slope on sentiment is negative and statistically significant at the 5% level now, where this negative effect of sentiment is only significant in the lower quantiles (see Table 16 or Figure 6). Tables 15 and 17 provide symmetry tests of contemporaneous and predictive effects of sentiment across the quantiles respectively, where the null of sentiment symmetric effects across the quantiles (for both contemporaneous and predictive effects) cannot be rejected for all pairs of τ, at the 5% level.

[Insert Tables 14 to 17, and Figures 5 and 6 about here]

In regard to the control variables (Tables 14 and 16), the number of messages is negative (positive) and statistically significant in the low (high) quantiles in both contemporaneous and predictive models, albeit its OLS estimate is positive and now significant in the former but not in the latter model. Monday effect remains weak as earlier, although, in the predictive regressions only, it shows some statistical significance in the upper (negative) compared to the lower (positive) quantiles. On the other hand, the estimates of Fama and French factors across both contemporaneous and predictive models are as expected, with the first two factors (excess return on the market and size) being stable across the quantiles whilst the third (value) factor being mostly significant on the two sides of the return distribution.

5.4.2. Sub-sample estimation

We further check the robustness of our results by estimating sentiment effects across roughly two equally divided sub-samples: April 4, 2012 - September 28, 2012 and October 1, 2012 - April 5, 2013. The sub-sample parameter estimates of γ_τ based on the market model, Eqs. (3) and (4), are plotted in Figures 7 and 8 for the contemporaneous and predictive effects of sentiment, respectively, whereas the corresponding estimates of γ_τ based on the Fama-French model, Eqs. (7) and (8), are respectively displayed in Figures 9 and 10.[29]

In a broad sense, these results further confirm our previous findings on both the contemporaneous and predictive effects of sentiment. For example, contemporaneous effects across both sub-samples are negative (positive) at low (high) quantiles, and they are statistically significant at higher but not lower quantiles, as our earlier findings. These effects based on the OLS have some discrepancies though, where they are positive (negative) and (not) significant in the second (first) sub-sample (see Figure 7). As for the predictive effects, they are negative and significant in the lower quantiles only as confirmed earlier; such effects across the two sub-samples are also negative when the OLS is used, albeit, at the 10% level, they are significant for the first but not for the second sub-sample (see Figure 8). Finally, when the Fama-French model is instead used in the estimation, contemporaneous and predictive effects across both sub-samples broadly exhibit a similar pattern as those based on the market model (see Figures 9 and 10), consistent with our earlier findings.

[Insert Figures 7 to 10 about here]

5.4.3. The reverse effect

We have further checked the robustness of our findings by estimating the reverse effect or causality. The results (unreported) of both the OLS and QR models showed no effect of stock returns on the shift in bullishness (sentiment).

6. Conclusions

In this paper, we extract a Twitter sentiment indicator for the 30 DJIA stocks over the period of April 4, 2012 to April 5, 2013 based on the textual classification of 289,024 online tweets posted on the so-called "StockTwits". The textual classification is conducted using various algorithms from computational linguistics (i.e., the Naive Bayes, the Decision Trees, and SVMs). Then, we analyse both the contemporaneous and predictive effects of our sentiment indicator on returns using the OLS and QR techniques, where the latter technique is flexible to investigate whether sentiment has different effects on returns across the quantiles of the return distribution.

We find that contemporaneous effects of sentiment on returns exhibit a distinctive inverse S-shaped pattern across the quantiles of the conditional return distribution: Sentiment effects on returns are negative (positive) at low (high) quantiles, albeit such effects are significant at high quantiles but not at low ones. Predictive effects of sentiment, on the other hand, show that sentiment is a strong

[29] The numerical results for this sub-section are available upon request.

negative predictor of returns in the lower quantiles only, implying that sentiment mainly affects the valuation of assets in turbulent times.

Our empirical results are robust when using the Fama and French (1993) 3-factor model, and have various implications for related research in finance. For example, they confirm the results of previous related studies in that sentiment indicators that are extracted from news stories and social media big data convey valuable information that can be used in the prediction of asset prices. Further, since our findings are consistent with most behavioural finance theories, which emphasise on the role of investors' psychology, emotions, preferences, and mistaken beliefs in asset prices, the proposed asset pricing models in financial markets could incorporate such behavioural aspects to better explain asset returns.

References

Ahmad, K., Han, J., Hutson, E., Kearney, C. and Liu, S. (2019) Media-expressed negative tone and firm-level stock returns, *Journal of Corporate Finance* **37**, 152–172.

Alagidede, P. and Panagiotidis, T. (2012) Stock returns and inflation: Evidence from quantile regressions, *Economics Letters* **117**(1), 283-286.

Antweiler, W. and 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.

Baker, M. and Wurgler, J. (2006) Investor sentiment and the cross-section of stock returns, *The Journal of Finance* **61**(4), 1645-1680.

Baker, M. and Wurgler, J. (2007) Investor sentiment in the stock market, *The Journal of Economic Perspectives* **21**(2), 129-151.

Baker, M., Wurgler, J. and Yuan, Y. (2012) Global, local, and contagious investor sentiment, *Journal of Financial Economics* **104**(2), 272-287.

Barber, B.M. and 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.

Baur, D.G., Dimpfl, T. and Jung, R.C. (2012) Stock return autocorrelations revisited: A quantile regression approach, *Journal of Empirical Finance* **19**(2), 254-265.

Bloomberg, 2010. Hedge fund will track Twitter to predict stock moves, *Bloomberg, Online Edition (author Jack Jordan)*, December 22.

Bodurtha, J.N., Kim, D-S. and Lee, C.M.C. (1995) Closed-end country funds and US market sentiment, *The Review of Financial Studies* **8**(3), 879-918.

Brown, G.W. and Cliff, M.T. (2004) Investor sentiment and the near-term stock market, *Journal of Empirical Finance* **11**(1), 1-27.

Brown, G.W. and Cliff, M.T. (2005) Investor sentiment and asset valuation, *The Journal of Business* **78**(2), 405-440.

Bollen, J., Mao, H. and Zeng, X. (2011) Twitter mood predicts the stock market, *Journal of Computational Science* **2**(1), 1-8.

Buchinsky, M. (1998) Recent advances in quantile regression models: A practical guideline for empirical research, *Journal of Human Resources* **33**(1), 88-126.

Bukovina, J. (2016) Social media big data and capital markets-An overview, *Journal of Behavioral and Experimental Finance* **11**, 18–26.

BusinessWeek, 2009, StockTwits may change how you trade, *BusinessWeek*, *Online Edition (author Max Zeledon)*, February 11.

Chevapatrakul, T. (2015) Monetary environments and stock returns: International evidence based on the quantile regression technique, *International Review of Financial Analysis* **38**, 83-108.

Chiu, H.H. and Kini, O. (2014) Equity issuances, equity mutual fund flows, and noise trader sentiment, *Review of Finance* **18**(2), 749-802.

Chuang, C-C., Kuan, C-M. and Lin, H-Y. (2009) Causality in quantiles and dynamic stock return–volume relations, *Journal of Banking and Finance* **33**, 1351–1360.

Chung, S-L., Hung, C-H. and Yeh, C-Y. (2012) When does investor sentiment predict stock returns?, *Journal of Empirical Finance* **19**(2), 217–240.

Claburn, T. (2009) Twitter growth surges 131% in March, *Information Week*. Retrieved 25 Oct, 2013 from:
http://www.informationweek.com/news/internet/social_network/showArticle.jhtml?articleID=216500968

Cookson, J.A. and Niessner, M. (2020) Why don’t we agree? evidence from a social network of investors, *The Journal of Finance* **75**(1), 173-228.

Da, Z., Engelberg, J. and Gao, P. (2015) The sum of all FEARS investor sentiment and asset prices, *The Review of Financial Studies* **28**(1), 1-32.

De Bondt, W.P.M. (1993) Betting on trends: Intuitive forecasts of financial risk and return, *International Journal of Forecasting* **9**(3), 355-371.

De Long, J.B., Shleifer, A., Summers, L.H. and Waldmann, R.J. (1990) Noise trader risk in financial markets, *The Journal of Political Economy* **98**(4), 703-738.

Dewally, M. (2003) Internet investment advice: Investing with a rock of salt, *Financial Analysts Journal* **59**(4), 65-77.

Engelberg, J. (2008) Costly information processing: Evidence from earnings announcements, Working Paper, University of North Carolina.

Fama, E.F. and French, K.R. (1993) Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* **33**, 3-56.

Fang, L. H. and Peress, J. (2009) Media coverage and the cross-section of stock returns. *The Journal of Finance* **64**(5), 2023.2052.

Feng, Y., Chen, R. and Basset, G.W. (2008) Quantile momentum, *Statistics and Its Interface* **1**, 243-254.

Fersini, E., Messina, E. and Pozzi, F.A. (2014) Sentiment analysis: Bayesian ensemble Learning, *Decision support systems* **68**, 26-38.

Garcia, D. (2013) Sentiment during recessions, *The Journal of Finance* **68**(3), 1267-1300.

Gervais, S. and Odean, T. (2001) Learning to be overconfident, *The Review of Financial Studies* **14**(1), 1-27.

Hall, M., Frank, E., Holmes, G. Pfahringer, B., Reutemann, P. and Witten, I.H. (2009) The WEKA data mining software: an update, *ACM SIGKDD Explorations Newsletter* **11**(1), 10-18.

Huang, C-H., Yang, D-X. and Chuang, Y-T. (2008) Application of wrapper approach and composite classifier to the stock trend prediction, *Expert Systems with Applications* **34**(4), 2870-2878.

Jansen, D.W. and Tsai, C-L. (2010) Monetary policy and stock returns: Financing constraints and asymmetries in bull and bear markets, *Journal of Empirical Finance* **17**(5), 981-990.

Java, A., Song, X., Finin, T. and Tseng, B. (2007) Why we twitter: Understanding microblogging usage and communities, *Proceedings of the joint 9th WEBKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis,* Springer, pp. 56-65.

Jiang, F., Lee, J., Martin, X. and Zhou, G. (2019) Manager sentiment and stock returns, *Journal of Financial Economics* **132**, 126–149.

Joseph, K., Wintoki, M.B. and Zhang, Z. (2011) Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search, *International Journal of Forecasting* **27**, 1116‑1127.

Kelly, M. (1997) Do noise traders influence stock prices?, *Journal of Money, Credit, and Banking* **29**(3), 351-363.

Koenker, R. and Bassett, G. (1978) Regression quantiles, *Econometrica* **46**(1), 33-50.

Koenker, R. and D'Orey, V. (1987) Computing regression quantiles, *Applied Statistics* **36**, 383–393.

Koenker, R. and Hallock, K.F. (2001) Quantile regression, *Journal of Economic Perspectives* **15**(4), 143–156.

Kukar, M. and Kononenko, I. (1998) Cost-sensitive learning with neural networks, *Proceedings of the Thirteenth European Conference on Artificial Intelligence*, 445-449.

Kumar, A., Page, J.K. and Spalt, O.G. (2013) Investor sentiment and return comovements: Evidence from stock splits and headquarters changes, *Review of Finance* **17**(3), 921-953.

Kothari, S. P., Li, X. and Short, J.E. (2009) The effect of disclosures by management, analysts, and financial press on the equity cost of capital: a study using content analysis, *Accounting Review* **84**, 1639–1670.

Kurov, A. (2010) Investor sentiment and the stock market's reaction to monetary policy, *Journal of Banking & Finance* **34**(1), 139-149.

Lartey, T., Danso, A. and Owusu-Agyei, S. (2020) CEOs' market sentiment and corporate innovation: The role of financial uncertainty, competition and capital intensity, *International Review of Financial Analysis* **72**, 101581.

Lee, C.M.C., Shleifer, A. and Thaler, R.H. (1991) Investor sentiment and the closed-end fund puzzle, *The Journal of Finance* **46**(1), 75-109.

Lee, W.Y., Jiang, C.X. and Indro, D.C. (2002) Stock market volatility, excess returns, and the role of investor sentiment, *Journal of Banking & Finance* **26**(12), 2277-2299.

Li, Q. and Shah, S. (2017) Learning stock market sentiment lexicon and sentiment-oriented word vector from StockTwits, *Proceedings of the 21st Conference on Computational Natural Language Learning,* pp. 301-310.

Loughran, T. and McDonald, B. (2011) When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks, *The Journal of Finance* **66**, 35–65.

Ma, L. and Pohlman, L. (2008) Return forecasts and optimal portfolio construction: A quantile regression approach, *The European Journal of Finance* **14**(5), 409-425.

McLaren, N. and Shanbhogue, R. (2011) Using Internet search data as economic indicators. *Bank of England Quarterly Bulletin* **51**(2), 134.140.

Mao, H., Counts, S. and Bollen, J. (2011) Predicting financial markets: Comparing survey, news, twitter and search engine data. Cornell University Library. ArXiv Preprint: https://arxiv.org/abs/1112.1051

Menkhoff, L. and Rebitzky, R.R. (2008) Investor sentiment in the US-dollar: Longer-term, non-linear orientation on PPP, *Journal of Empirical Finance* **15**(3), 455-467.

Miller, E.M. (1977) Risk, uncertainty, and divergence of opinion, *The Journal of Finance* **32**(4), 1151-1168.

Neal, R. and Wheatley, S.M. (1998) Do measures of investor sentiment predict returns?, *Journal of Financial and Quantitative Analysis* **33**(4), 523-547.

Odean, T. (1998) Are investors reluctant to realize their losses?, *The Journal of Finance* **53**(5),1775-1798.

Odean, T. (1999) Do investors trade too much?, *The American Economic Review,* **89**(5), 1279-1298.

Oh, C. and Sheng, O.R.L. (2011) Investigating predictive power of stock micro blog sentiment in forecasting future stock price directional movement, In: *The 32nd International Conference on Information Systems, Shanghai, China.*

Papapostolou, N.C., Nomikos, N.K., Pouliasis, P.K. and Kyriakou, I. (2014) Investor sentiment for real assets: The case of dry bulk shipping market, *Review of Finance* **18**(4), 1507-1539.

Rudd, A.T. (1979) The revised Dow Jones Industrial Average: New wine in old bottles?, *Financial Analysts Journal* **35**(6), 57-63.

Ruiz, E.J., Hristidis, V., Castillo, C., Gionis, A. and Jaimes, A. (2012) Correlating financial time series with micro-blogging activity, *Proceedings of the fifth ACM international conference on Web search and data mining* 2012, ACM, 513-522.

Saxa, B. (2014) Forecasting mortgages: Internet search data as a proxy for mortgage credit demand. CNB working paper series 14/2014.

Schmeling, M. (2009) Investor sentiment and stock returns: Some international evidence, *Journal of Empirical Finance* **16**(3)**,** 394-408.

Shleifer, A. and Vishny, R.W. (1997) The limits of arbitrage, *The Journal of Finance* **52**(1), 35-55.

Simon, D.P. and Wiggins, R.A. (2001) S&P futures returns and contrary sentiment indicators, *Journal of Futures Markets* **21**(5), 447-462.

Sprenger, T.O., Tumasjan, A., Sandner, P.G. and Welpe, I.M. (2014) Tweets and trades: The information content of stock microblogs, *European Financial Management* **20**(5), 926-957.

Stambaugh, R.F., Yu, J. and Yuan, Y. (2012) The short of it: Investor sentiment and anomalies, *Journal of Financial Economics* **104**(2), 288-302.

Stapleton, R.C. and Subrahmanyam, M.G. (1983) The market model and capital asset pricing theory: A note, The Journal of Finance 38(5), 1637-1642.

Swaminathan, B. (1996) Time-varying expected small firm returns and closed-end fund discounts, *The Review of Financial Studies* **9**(3), 845-887.

Tetlock, P.C. (2007) The role of media in the stock market. *The Journal of Finance* **62**(3), 1139-1168.

Tetlock, P.C., Saar-Tsechansky, M. and Macskassy, S. (2008) More than words: Quantifying language to measure firms' fundamentals, *The Journal of Finance* **63**(3), 1437-1467.

Thaler, R. (1987). Anomalies: Seasonal movements in security prices II: Weekend, holiday, turn of the month, and intraday effects. *The Journal of Economic Perspectives* **1**(2), 169-177.

TIME, 2009, Turning Wall Street on its head, TIME magazine, Online Edition (author Douglas McIntyre), May 29.

Verma, R. and Verma, P. (2007) Noise trading and stock market volatility, *Journal of Multinational Financial Management* **17**(3), 231-243.

Wang, C. (2001) Investor sentiment and return predictability in agricultural futures markets, *Journal of Futures Markets* **21**(10), 929-952.

Wysocki, P.D. (1998) Cheap talk on the web: The determinants of postings on stock message boards, University of Michigan Business School Working Paper No. 98025.

Xia, R., Zong, C. and Li, S. (2011) Ensemble of feature sets and classification algorithms for sentiment classification, *Information Sciences* **181**(6), 1138-1152.

Yuan, Y. (2015) Market-wide attention, trading, and stock returns, *Journal of Financial Economics* **116**(3), 548–564.

Zhang, X., Fuehres, H. and 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.

Zhang, J.L., Härdle, W.K., Chen, C.Y. and Bommes, E. (2016) Distillation of news flow into analysis of stock reactions, *Journal of Business & Economic Statistics* **34**(4), 547-563.

Table 1: The coding scheme for manually labelled tweets

This table provides the general rules that the primary coder (the authors) and the independent coder applied for manually coding the StockTwits messages. Such general rules are used as input data in the text-processing model.

The General Rules Applied in Labelling the StockTwits Data
(i) If the tweet post contains external links to long articles or numerical charts about the stocks, it is generally marked as neutral. The content of the article and the information revealed by the chart are not taken into account.
(ii) Buy, hold, or sell labels are only given when the sentiment can be explicitly speculated from the tweet.
(iii) Tweets with question marks are generally marked as neutral.
(iv) Simple summarisations of the stock performance by the end of the day are not taken into consideration.
(v) If the user reports a loss in a subjective way instead of reporting numbers, it is fair to assume that the user has a negative feeling towards the stock and vice versa.
(vi) If a tweet post contains company names (Apple, Google, Microsoft) or any other neutral words (such as day, report, look, watch, …, etc.), it is generally marked as a hold message.
(vii) All positive words/emotions in a tweet message indicate linguistic bullishness (e.g., strong, high, happy, earn, …, etc.) and will, therefore, be marked as a buy signal.
(viii) Sell messages contain corresponding bearish words (e.g., loss, weak, low, fall, decline, down, etc.); therefore, all negative words/emotions in a tweet message indicate linguistic bearishness and are commonly marked as a sell signal.
(ix) Normally tweet posts containing a balance of positive and negative words will be classified as hold messages.
(x) A tweet post containing a mixture of positive and negative emotional words will be assigned to the correct class based on the probability value assigned to each class, where the message will be assigned to the class of high probability. For example, if a tweet message contains 65% positive, 20% negative, and 15% neutral words, the message will be classified as a buy message since positive words, which are more likely to be associated with the buy signal, dominate the message.

Table 2: The manual classification of StockTwits messages

This table presents the percentage of the random sample of 2,892 hand-coded messages of StockTwits allocated to three sentiment classes, namely buy, hold, or sell. The percentage distributions, as well as the number of messages, are reported under each of these sentiment classes.

The Number and Percentage Distribution of the Manual Classification of Posts				
Class	**Buy**	**Hold**	**Sell**	**Total**
Number	1,361	590	941	2,892
Percentage	47.06%	20.40%	32.54%	100%

Table 3: Sample tweets from training data set with manual classification

Tweets are randomly selected and are shown in their original format. By looking at the most common words associated with each class, we find that some general features occur almost frequently in all of the three classes (e.g., figures, ticker names, and external links). However, beyond these universal features, there is a unique pattern that reasonably distinguishes the sentiment of messages as per the three classes. For example, positive words such as “good” and “high” are the most common words likely to be found in buy messages. Financial words such as “buy”, “long” and “call” also give a clear indication in the financial context that the investors expected a particular stock to rise. In contrast, the most common words that are more likely to appear in sell messages are the negative words such as “down”, “ugly”, “break” and “low”, as well as words such as “sell”, “put”, “loss” and “short”, which give a clear indication that users expected the discussed stock to fall. These results match those observed in earlier studies by Tetlock et al. (2008) and Sprenger et al. (2014). However, if the tweet message contains external links of long articles or charts about the stocks in which more neutral words are presented such as product names (e.g., “Aircraft”, “BigMac”, “Window7”), it is generally labelled as hold. Therefore, in hold messages, the positive and negative words are much more equitable; that is, neutral words dominate such messages.

Sample Tweets (Training Data Set)	**Manual Classification**
"Our highest long as of today low $JPM and $BAC.//LOL"	Buy
"$xom $intc $dvn $ko $cm $ftse some analysis on these charts"	Hold
"Short $NKE http://chart.ly/jmbomde"	Sell
"$KO http://stks.co/3OvK Breaks yesterdays high will add! Bullish"	Buy
"$CAT In again for giggles at 81.16... 2 Dec 80 Put for DCA of $2.32 Average entry at 81.25"	Sell
"$SBUX The Starbucks Trade http://stks.co/nDQ6 $DNKN $MCD $ARCO $GMCR"	Hold
"$T good stock for buying... http://stks.co/t04i"	Buy
"$GS we are buyers on dips. (Shares and long term calls)"	Buy
“$NKE down over 2% now. Making new lows trying to break $97"	Sell
"$CAT Looks ugly down there http://chart.ly/gk4hhm8"	Sell

Table 4: Classification accuracy by class using Random Forest classifier (10- fold cross-validation)

This table shows the classification accuracy by class using the 10-fold cross-validation method. In this method the training set is split into 10 parts of equal size, each of which is classified based on a model trained on the remaining 9 of the data set. True positives (or precision) represent, for example, the share of messages classified as sell, which are labelled as such in the training data set. False positives are messages incorrectly classified as sell. The recall represents the share of all messages of a particular class, which are classified correctly. The F-measure combines precision and recall. The ROC area measures the quality of the trade-off between true and false positives.

Classification Accuracy by Class Using Decision Tree (Random Forest) Classifier						
10 -Fold Cross-Validation						
Class	**True Positives**	**False Positives**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Buy	78.40%	30.50%	69.60%	78.40%	73.70%	80.00%
Hold	63.90%	10.20%	61.60%	63.90%	62.70%	85.90%
Sell	51.50%	13.40%	65.00%	51.50%	57.50%	75.70%
Weighted Average	66.70%	20.80%	66.50%	66.70%	66.20%	79.80%

Table 5: Classification accuracy by class using Random Forest classifier (supplied test set)

This table shows the classification accuracy by class using the supplied test method. In this method two separate data sets are used: the training (in-sample) set and the testing (hold-out) set. True positives (or precision) represent, for example, the share of messages classified as sell, which are labelled as such in the training data set. False positives are messages incorrectly classified as sell. The recall represents the share of all messages of a particular class, which are classified correctly. The F-measure combines precision and recall. The ROC area measures the quality of the trade-off between true and false positives.

Classification Accuracy by Class Using Decision Tree (Random Forest) Classifier						
Supplied Test Set						
Class	**True Positive**	**False Positives**	**Precision**	**Recall**	**F-Measure**	**ROC Area**
Buy	63.31%	32.0%	69.10%	63.10%	66.00%	68.30%
Hold	36.90%	4.60%	56.50%	36.90%	44.70%	78.00%
Sell	66.10%	30.70%	51.50%	66.10%	57.90%	71.00%
Weighted Average	60.50%	27.80%	61.50%	60.50%	60.04%	70.50%

Table 6: Random Forest classification accuracy of the supplied test set and the overall classification distribution

This table provides the buy-hold-sell matrix entries of the hold-out sample (939 messages) and the prediction accuracy of the classification algorithm concerning the training (in-sample) set of 1,953 messages. The total rows of the table show the actual share of 939 hand-coded messages that were classified as buy, hold or sell, whereas the total columns represent the share of the messages that were automatically classified as per class by the algorithm. The last row provides summary statistics of the percentage distribution of the out-of-sample classification of each class that will then be deployed and aggregated for the daily ticker level analysis.

Classified by Algorithm				
Class	**Buy**	**Hold**	**Sell**	**Manual Classification**
Buy	315	26	158	**499**
Hold	47	48	35	**130**
Sell	94	11	205	**310**
Total Classified by Algorithm	**456**	**85**	**398**	**939**
% Classification by Algorithm as per Class	48.56%	9.05%	42.39%	100%

Table 7: The Overall distribution of the total StockTwits posts of the 30 DJIA stocks

This table shows the overall distribution of the entire StockTwits data employed. The first percentage column indicates the actual proportion of the manually coded messages that were classified as buy, hold, or sell. The second percentage column, in contrast, shows the automatic classification. The last column shows the overall distributions of the StockTwits posts based on the automatic percentage classifications per sentiment class.

Total Distribution of the Manual and Automatic Classification			
Class	**Manual Classification (in %)**	**Automatic Classification (in %)**	**Total Tweets per Class**
Buy	47.60	48.56	140,350
Hold	20.40	9.05	26,157
Sell	32.54	42.39	122,517
Total	100%	100%	**289,024**

Table 8: Descriptive Statistics

This table shows the summary statistics of the pooled sample. The total number of observations is 7,530, with a sample period from April 4, 2012 to April 5, 2013.

Variables	Minimum	Maximum	Mean	Std. Deviation	JB
R_{it}	-10.963	10.493	0.045	1.193	7752.90 [0.000]
ΔB_{it}^{*}	-0.898	0.919	0.001	0.122	23609.03 [0.000]

Notes: R_{it} denotes stock returns calculated as the log difference in the closing prices of each stock over two consecutive trading days. ΔB_{it}^{*} denotes the shift in bullishness which proxies investor sentiment. JB is the Jarque-Bera test for normality; *P*-values are presented in square brackets [.].

Table 9: Summary statistics by company

This table shows some summary statistics for the data used as per company. The sample period is from April 4, 2012 to April 5, 2013. The figures reported for the returns and the bullishness index are the averages, whereas those of the message volume refer to the total tweet posts per company, over the sample period. Company size (in $billion) is as of 2012.

Ticker	**Company**	**Bullishness Index** B_t^*	**Message Volume**	**Returns** $(\boldsymbol{R_t})$	**Size ($billion)**
Axp	American Express Company	0.369	3,165	0.058	63.51
BA	The Boeing Company	0.185	4,867	0.063	56.94
CAT	Caterpillar, Inc.	0.498	18,020	-0.143	58.69
CSCO	Cisco Systems, Inc.	0.090	11,512	-0.002	84.50
CVX	Chevron Corporation	0.449	4,564	0.033	210.51
DD	DuPont de Nemours, Inc.	0.329	1,564	-0.038	38.90
DIS	The Walt Disney Company	0.418	5,948	0.073	93.05
GE	General Electric Company	0.338	6,728	0.063.5154	218.41
GS	The Goldman Sachs Group, Inc.	0.328	28,972	0.104	61.29
HD	The Home Depot	0.214	6,924	0.017	68.22
IBM	INTL Business Machines Corporation (IBM)	0.493	17,372	-0.070	214.03
INTC	Intel Corporation	0.270	16,608	0.102	101.94
JNJ	Johnson & Johnson	0.336	4,988	0.025	194.77
JPM	JPMorgan Chase & Co.	0.246	35,336	0.001	167.25
KO	Coca-Cola Company	0.213	5,716	-0.026	162.00
MCD	McDonald's Corporation	0.373	7,952	0.077	88.44
MMM	3M Company	0.166	1,736	0.035	63.79
MRK	Merck & Co., Inc.	0.319	3,160	-0.070	123.91
MSFT	Microsoft Corporation	0.238	33,700	0.039	256.37
NKE	Nike, Inc.	0.103	10,620	0.134	49.54
PFE	Pfizer, Inc.	0.386	4,460	0.115	182.47
PG	Procter & Gamble Company	0.458	3,228	0.089	168.31
T	American Telephone & Telegraph (AT&T)	0.096	9,832	0.076	188.14
TRV	The Travelers Companies, Inc.	0.467	1,344	0.032	27.10
UNH	UnitedHealth Group, Inc.	0.180	2,636	0.023	55.27
UTX	United Technologies Corporation (UTC)	0.258	1,312	0.090	75.35
V	Visa, Inc.	0.295	11,224	0.085	100.35
VZ	Verizon Communications, Inc.	0.463	6,892	0.108	123.69
WMT	Walmart, Inc.	0.206	10,740	0.047	209.72
XOM	Exxon Mobil Corporation	0.468	7,904	0.058	389.64

Notes: R_t denotes stock returns which are calculated as the log difference in the closing prices of each stock over two consecutive trading days. B_t^* denotes the bullishness index of a particular company.

Table 10: Contemporaneous effects of sentiment

The estimated OLS and QR models are respectively specified as $R_{it} = \alpha_i + \gamma\Delta B_{it}^* + \lambda_1 lnMSG_{it} + \lambda_2 MKT_t + \lambda_3 NWK_t + \varepsilon_{it}$ and $q_\tau(R_{it}|\Delta B_{it}^*, lnMSG_{it}, MKT_t, NWK_t) = \alpha_\tau + \gamma_\tau\Delta B_{it}^* + \lambda_{1\tau} lnMSG_{it} + \lambda_{2\tau} MKT_t + \lambda_{3\tau} NWK_t$, where R_{it} is the returns and ΔB_{it}^* is the shift in the bullishness index (the proxy of investor sentiment). We also control for the volume of StockTwits messages ($lnMSG_{it}$), market returns (NWK_t), and a dummy for the first day of the trading week (NWK_t). The total number of observations is 7,530, with a sample period from April 4, 2012 to April 5, 2013.

Level	OLS	Low			Medium			High		
τ		0.05	0.10	0.25	0.40	0.50	0.60	0.75	0.90	0.95
α	0.0051	-0.955***	-0.702***	-0.328***	-0.104***	0.029	0.162***	0.410	0.555***	0.703***
	(0.029)	(0.042)	(0.033)	(0.023)	(0.022)	(0.021)	(0.021)	(0.022)	(0.135)	(0.273)
γ	0.149*	-0.343**	-0.002	0.189	0.154	0.139	-0.009	0.089	0.542***	0.880***
	(0.090)	(0.153)	(0.173)	(0.143)	(0.101)	(0.096)	(0.090)	(0.090)	(0.159)	(0.233)
λ_1	-0.0011	-0.158***	-0.113***	-0.055***	-0.020***	-0.007	-0.010	0.035***	0.091***	0.141***
	(0.010)	(0.304)	(0.011)	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)	(0.012)	(0.017)
λ_2	0.998***	1.005***	0.997***	0.984***	0.978***	0.969***	0.969***	0.956***	0.980***	1.031***
	(0.014)	(0.024)	(0.021)	(0.017)	(0.016)	(0.015)	(0.015)	(0.017)	(0.022)	(0.024)
λ_3	0.012	-0.047	0.010	0.002	0.005	-0.001	-0.022	-0.019	-0.005	0.019
	(0.025)	(0.064)	(0.043)	(0.027)	(0.023)	(0.021)	(0.022)	(0.028)	(0.041)	(0.056)
R^2	0.393									
Pseudo-R^2		0.212	0.228	0.242	0.252	0.257	0.258	0.253	0.241	0.225

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Standard errors are represented in parentheses.

Table 11: Testing symmetry of contemporaneous effects of sentiment across quantiles

τ Pair	(0.05, 0.95)	(0.10, 0.90)	(0.15, 0.85)	(0.20, 0.80)	(0.25, 0.75)	(0.30, 0.70)	(0.35, 0.65)	(0.40, 0.60)	(0.45, 0.55)
$H_0: \gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$	0.083 (0.319)	0.275 (0.227)	0.054 (0.217)	0.098 (0.167)	-0.001 (0.150)	0.001 (0.134)	-0.062 (0.109)	-0.133 (0.087)	-0.116 (0.061)

Notes: Each value is a test statistic for the hypothesis that the quantile effects are symmetric about the median, i.e., $\gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$. Standard errors are in parentheses.

Table 12: Predictive effects of sentiment

The estimated OLS and QR models are respectively specified as $R_{it} = \alpha_i + \gamma\Delta B^*_{it-1} + \lambda_1 lnMSG_{it-1} + \lambda_2 MKT_t + \lambda_3 NWK_t + \varepsilon_{it}$ and $q_\tau(R_{it}|\Delta B^*_{it-1}, L^s lnMSG_{it-1}, MKT_t, NWK_t) = \alpha_\tau + \gamma_\tau\Delta B^*_{it-1} + \lambda_{1\tau} lnMSG_{it-1} + \lambda_{2\tau} MKT_t + \lambda_{3\tau} NWK_t$, where R_{it} is the returns and ΔB^*_{it-1} is the lagged shift in the bullishness index (the proxy of investor sentiment). We also control for the lagged volume of StockTwits messages ($lnMSG_{it-1}$), market returns (NWK_t), and a dummy for the first day of the trading week (NWK_t). The total number of observations is 7,530, with a sample period from April 4, 2012 to April 5, 2013.

Level	**OLS**	**Low**			**Medium**			**High**		
τ		**0.05**	**0.10**	**0.25**	**0.40**	**0.50**	**0.60**	**0.75**	**0.90**	**0.95**
α	-0.005	-1.168***	-0.783***	-0.356***	-0.090***	0.048**	0.184***	0.428***	0.886***	1.195***
	(0.028)	(0.065)	(0.036)	(0.023)	(0.021)	(0.021)	(0.019)	(0.023)	(0.031)	(0.042)
γ	-0.165*	-0.353*	-0.357***	-0.263***	-0.124	-0.092	-0.032	-0.150*	-0.086	0.055
	(0.090)	(0.244)	(0.088)	(0.118)	(0.080)	(0.087)	(0.091)	(0.090)	(0.277)	(0.175)
λ_1	0.001	-0.084***	-0.087***	-0.045***	-0.025***	-0.016**	0.001	0.023***	0.053***	0.089***
	(0.010)	(0.022)	(0.011)	(0.008)	(0.008)	(0.007)	(0.007)	(0.008)	(0.012)	(0.018)
λ_2	0.998***	1.031***	1.007***	0.993***	0.976***	0.969***	0.970***	0.957***	0.992***	1.028***
	(0.014)	(0.027)	(0.022)	(0.017)	(0.016)	(0.015)	(0.015)	(0.017)	(0.023)	(0.029)
λ_3	0.010	0.121**	0.078*	0.020	0.005	-0.002	-0.026	-0.019	-0.041	0.068
	(0.025)	(0.062)	(0.044)	(0.027)	(0.022)	(0.021)	(0.022)	(0.028)	(0.045)	(0.067)
R^2	0.391									
Pseudo-R^2		0.217	0.236	0.244	0.252	0.257	0.258	0.253	0.241	0.226

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Standard errors are represented in parentheses.

Table 13: Testing symmetry of predictive effects of sentiment across quantiles

τ Pair	**(0.05, 0.95)**	**(0.10, 0.90)**	**(0.15, 0.85)**	**(0.20, 0.80)**	**(0.25, 0.75)**	**(0.30, 0.70)**	**(0.35, 0.65)**	**(0.40, 0.60)**	**(0.45, 0.55)**
$H_0: \gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$	-0.114 (0.303)	-0.258 (0.279)	-0.111 (0.203)	-0.133 (0.154)	-0.229* (0.133)	-0.191* (0.114)	-0.124 (0.097)	0.028 (0.078)	0.112** (0.055)

Notes: Each value is a test statistic for the hypothesis that the quantile effects are symmetric about the median, i.e., $\gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$. Standard errors are in parentheses. . * and ** denote statistical significance at the 10% and 5% levels, respectively.

Table 14: Contemporaneous effects of sentiment based on the Fama-French model

The estimated OLS and QR models are respectively specified as $R_{it} - R_{ft} = \alpha_i + \gamma\Delta B^*_{it} + \beta_1(MKT_t - R_{ft}) + \beta_2\, SMB_{it} + \beta_3 HML_{it} + \lambda_1 lnMSG_{it} + \lambda_2 NWK_t + \varepsilon_{it}$ and $q_\tau(R_{it} - R_{ft}\,|\Delta B^*_{it}, MKT_t - R_{ft}, SMB_{it}, HML_{it}, lnMSG_{it}, NWK_t) = \alpha_\tau + \gamma_\tau\Delta B^*_{it} + \beta_{1\tau}(MKT_t - R_{ft}) + \beta_{2\tau}SMB_{it} + \beta_{3\tau}HML_{it} + \lambda_{1\tau}lnMSG_{it} + \lambda_{2\tau}NWK_t$, where $R_{it} - R_{ft}$ is the excess return for stock *i* at time *t*, $MKT_t - R_{ft}$ is the excess return on the market index at time *t*, SMB_{it} denotes the daily return difference between the return on small size stocks and the return on big size stocks (i.e., size factor), HML_{it} denotes the daily return difference between the return on high value stocks and the return on low value stocks (i.e., value factor), and ΔB^*_{it} is the shift in the bullishness index (the proxy of investor sentiment). We also control for the volume of StockTwits messages ($lnMSG_{it}$), and a dummy for the first day of the trading week (NWK_t). The total number of observations is 7,530, with a sample period from April 4, 2012 to April 5, 2013.

Level	**OLS**	**Low**			**Medium**			**High**		
τ		**0.05**	**0.10**	**0.25**	**0.40**	**0.50**	**0.60**	**0.75**	**0.90**	**0.95**
α	-0.018 (0.029)	-0.997*** (0.043)	-0.744*** (0.034)	-0.363*** (0.024)	-0.100*** (0.022)	0.020 (0.022)	0.168*** (0.021)	0.398 (0.023)	0.764*** (0.030)	1.033*** (0.054)
γ	0.112 (0.091)	-0.332* (0.208)	-0.026 (0.218)	-0.015 (0.126)	0.095 (0.106)	0.030 (0.107)	0.024 (0.097)	0.156* (0.095)	0.429*** (0.157)	0.786*** (0.260)
β_1	0.907*** (0.014)	0.934*** (0.026)	0.897*** (0.022)	0.891*** (0.016)	0.896*** (0.015)	0.883*** (0.015)	0.884*** (0.015)	0.854*** (0.017)	0.906*** (0.022)	0.925*** (0.029)
β_2	-0.228*** (0.029)	-0.313*** (0.075)	-0.218*** (0.047)	-0.167*** (0.032)	-0.213*** (0.027)	-0.204*** (0.028)	-0.201*** (0.027)	-0.194*** (0.037)	-0.243*** (0.045)	-0.213*** (0.079)
β_3	0.076*** (0.030)	0.144*** (0.055)	0.141** (0.058)	0.094*** (0.034)	0.029 (0.031)	0.042 (0.031)	0.035 (0.032)	0.074** (0.035)	0.073* (0.044)	0.030 (0.062)
λ_1	0.006*** (0.010)	-0.150*** (0.015)	-0.104*** (0.012)	-0.050*** (0.008)	-0.025*** (0.008)	-0.004 (0.008)	0.008 (0.007)	0.038*** (0.009)	0.101*** (0.011)	0.154*** (0.020)
λ_2	-0.006 (0.026)	-0.055 (0.066)	-0.013 (0.043)	-0.002 (0.027)	-0.030 (0.022)	-0.026 (0.023)	-0.044 (0.023)	-0.046 (0.029)	-0.002 (0.042)	-0.027 (0.053)
R^2	0.383									
Pseudo-R^2		0.227	0.236	0.240	0.245	0.248	0.249	0.247	0.244	0.236

Notes: *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. Standard errors are represented in parentheses.

Table 15: Testing symmetry of contemporaneous effects of sentiment across quantiles based on the Fama-French model

τ Pair	(0.05, 0.95)	(0.10, 0.90)	(0.15, 0.85)	(0.20, 0.80)	(0.25, 0.75)	(0.30, 0.70)	(0.35, 0.65)	(0.40, 0.60)	(0.45, 0.55)
$H_0: \gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$	0.394 (0.342)	0.342 (0.227)	0.354* (0.210)	0.271 (0.127)	0.081 (0.157)	0.129 (0.140)	0.090 (0.119)	0.058 (0.097)	-0.018 (0.069)

Notes: Each value is a test statistic for the hypothesis that the quantile effects are symmetric about the median, i.e., $\gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$. Standard errors are in parentheses.

Table 16: Predictive effects of sentiment based on the Fama-French model

The estimated OLS and QR models are respectively specified as $R_{it} - R_{ft} = \alpha_i + \gamma \Delta B^*_{it-1} + \beta_1(MKT_t - R_{ft}) + \beta_2 SMB_{it} + \beta_3 HML_{it} + \lambda_1 lnMSG_{it-1} + \lambda_2 NWK_t + \varepsilon_{it}$ and $q_\tau(R_{it} - R_{ft} \,|\Delta B^*_{it-1}, MKT_t - R_{ft}, SMB_{it}, HML_{it}, lnMSG_{it-1}, NWK_t) = \alpha_\tau + \gamma_\tau \Delta B^*_{it-1} + \beta_{1\tau}(MKT_t - R_{ft}) + \beta_{2\tau} SMB_{it} + \beta_{3\tau} HML_{it} + \lambda_{1\tau} lnMSG_{it-1} + \lambda_{2\tau} NWK_t$, where $R_{it} - R_{ft}$ is the excess return for stock *i* at time *t*, $MKT_t - R_{ft}$ is the excess return on the market index at time *t*, SMB_{it} denotes the daily return difference between the return on small size stocks and the return on big size stocks (i.e., size factor), HML_{it} denotes the daily return difference between the return on high value stocks and the return on low value stocks (i.e., value factor), and ΔB^*_{it-1} is the lagged shift in the bullishness index (the proxy of investor sentiment). We also control for the lagged volume of StockTwits messages ($lnMSG_{it-1}$), and a dummy for the first day of the trading week (NWK_t). The total number of observations is 7,530, with a sample period from April 4, 2012 to April 5, 2013.

Level	**OLS**	**Low**			**Medium**			**High**		
τ		**0.05**	**0.10**	**0.25**	**0.40**	**0.50**	**0.60**	**0.75**	**0.90**	**0.95**
α	-0.034	-1.215***	-0.833***	-0.397***	-0.103***	0.03*	0.171***	0.435***	0.852***	1.186***
	(0.028)	(0.071)	(0.037)	(0.024)	(0.021)	(0.020)	(0.020)	(0.023)	(0.033)	(0.053)
γ	-0.180**	-0.340*	-0.358**	-0.199*	-0.143	-0.091	-0.083	-0.155	-0.207	-0.076
	(0.091)	(0.215)	(0.096)	(0.125)	(0.094)	(0.088)	(0.086)	(0.107)	(0.208)	(0.149)
β_1	0.906***	0.938***	0.898***	0.903***	0.893***	0.882***	0.884***	0.867***	0.991***	0.928***
	(0.014)	(0.027)	(0.023)	(0.016)	(0.015)	(0.015)	(0.015)	(0.017)	(0.021)	(0.029)
β_2	-0.227***	-0.311***	-0.224***	-0.177***	-0.211***	-0.199***	-0.204***	-0.213***	-0.235***	-0.181***
	(0.029)	(0.066)	(0.045)	(0.035)	(0.027)	(0.027)	(0.027)	(0.037)	(0.050)	(0.066)
β_3	0.078***	0.225***	0.117**	0.083**	0.026	0.046	0.037	0.068*	0.094*	0.075
	(0.030)	(0.071)	(0.054)	(0.035)	(0.030)	(0.031)	(0.032)	(0.035)	(0.049)	(0.062)
λ_1	0.013	-0.074***	-0.071***	-0.036***	-0.036***	-0.008	0.006	0.023***	0.068***	0.088***
	(0.203)	(0.022)	(0.011)	(0.009)	(0.008)	(0.007)	(0.007)	(0.009)	(0.013)	(0.019)
λ_2	-0.012	0.080	0.074*	0.020	-0.026	-0.026	-0.050**	-0.065***	-0.047	-0.092*
	(0.025)	(0.066)	(0.038)	(0.027)	0.022	(0.023)	(0.023)	(0.028)	(0.039)	(0.055
R^2	0.382									
Pseudo-R^2		0.214	0.232	0.239	0.245	0.247	0.245	0.247	0.235	0.220

Notes: *, ** and *** denote statistical significance at the 10%, 5% and 1% levels, respectively. Standard errors are represented in parentheses.

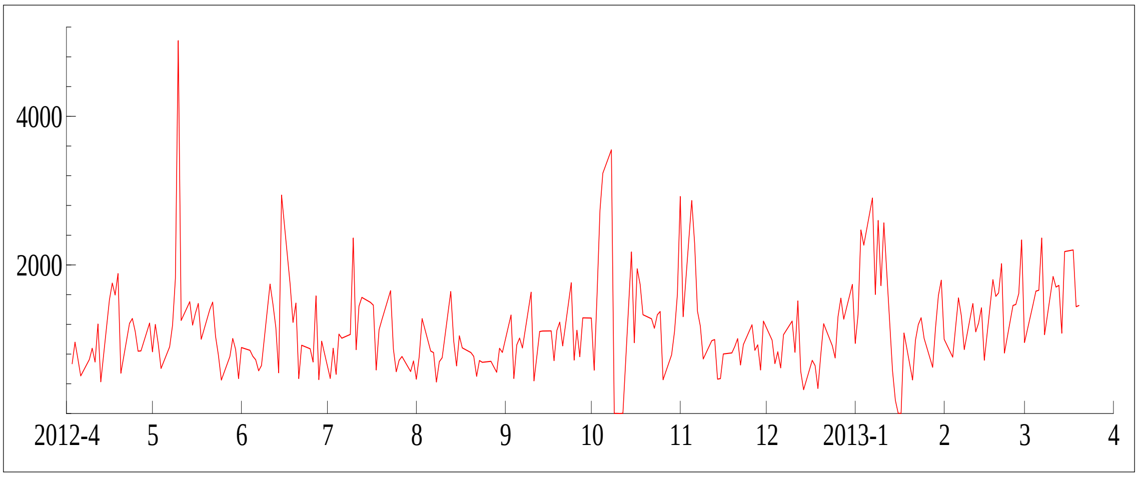
Table 17: Testing symmetry of predictive effects of sentiment across quantiles based on the Fama-French model

τ Pair	(0.05, 0.95)	(0.10, 0.90)	(0.15, 0.85)	(0.20, 0.80)	(0.25, 0.75)	(0.30, 0.70)	(0.35, 0.65)	(0.40, 0.60)	(0.45, 0.55)
$H_0: \gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$	-0.236 (0.272)	-0.384* (0.229)	-0.276 (0.181)	-0.133 (0.160)	-0.173 (0.142)	-0.158 (0.114)	-0.087 (0.096)	-0.044 (0.079)	0.063 (0.056)

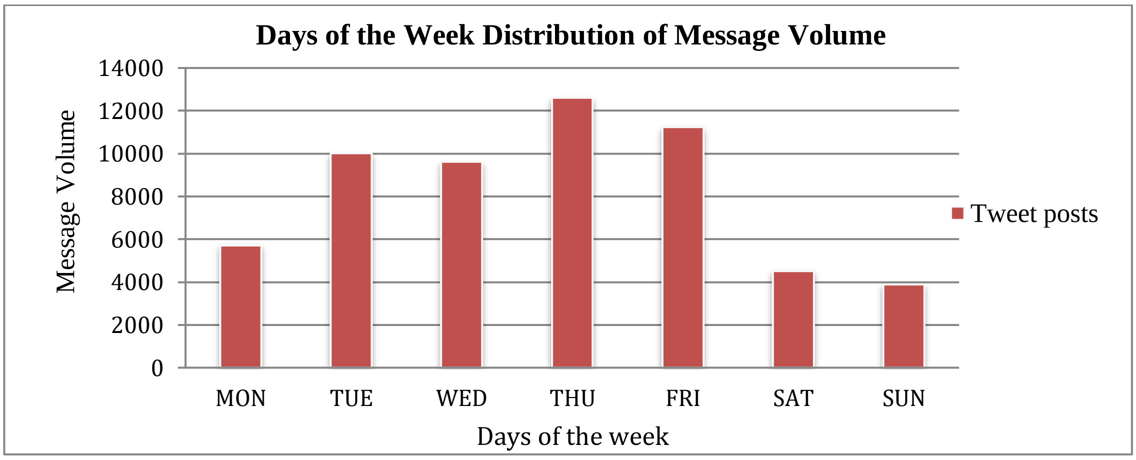
Notes: Each value is a test statistic for the hypothesis that the quantile effects are symmetric about the median, i.e., $\gamma_{(\tau)} + \gamma_{(1-\tau)} = 2\gamma_{(0.50)}$. Standard errors are in parentheses. * denotes statistical significance at the 10% level.

Figure 1: The distribution of StockTwits posts

Panel A: The evolution of daily StockTwits messages (posting activity of the 30 companies of the DJIA index combined)



Panel B: The distribution of StockTwits posts throughout the week (average posts of all companies in the sample are considered across days of the week)



Panel C: The percentage of posts of the 30 companies of the DJIA index during the daytime

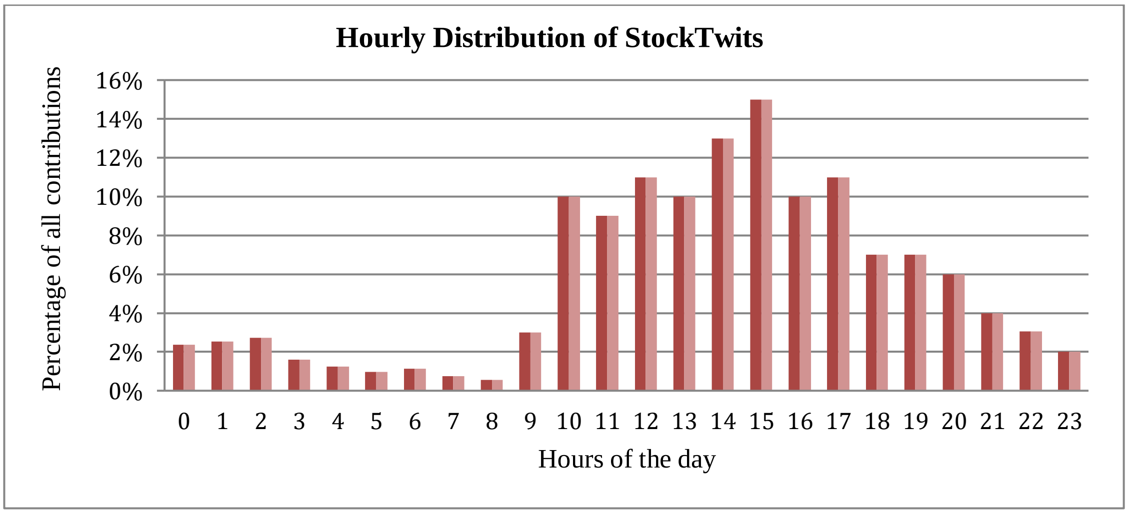
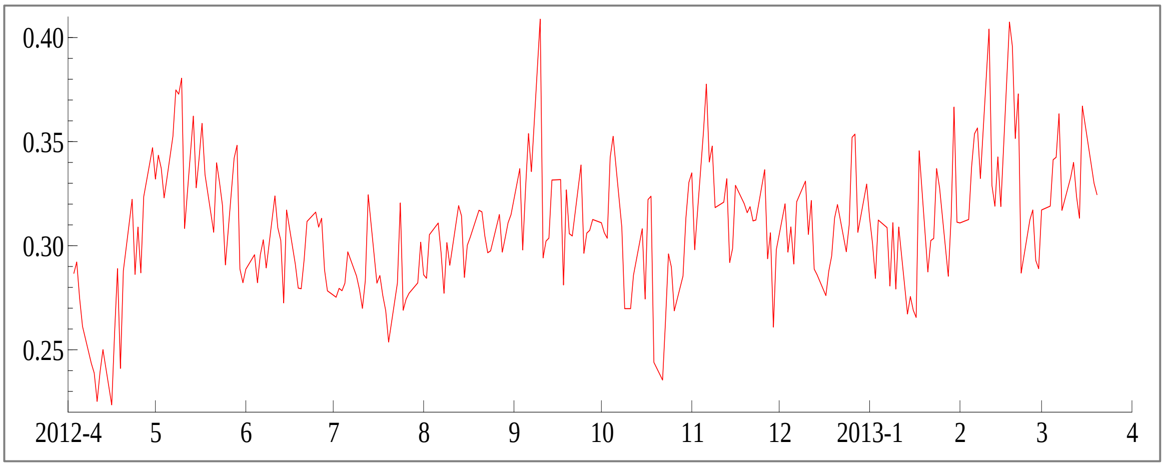


Figure 2: The evolution of the DJIA index and the StockTwits bullishness indicator

This figure displays the evolution of the daily DJIA index returns (lower panel) and the corresponding normalised bullishness indicator (upper panel).

Panel A: The evolution of the normalised bullishness of the DJIA index.



Panel B: The evolution of the daily DJIA index returns.

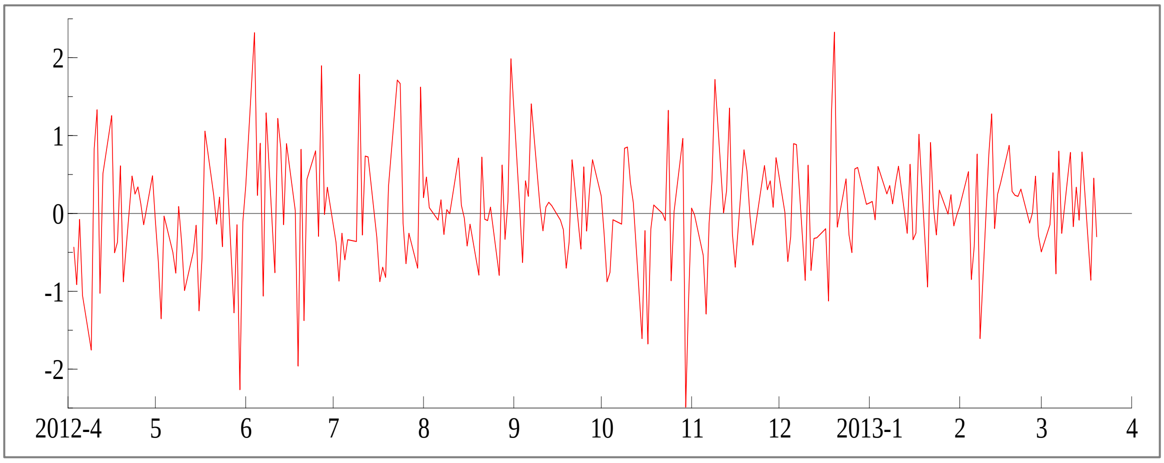


Figure 3: Estimates of contemporaneous effects of sentiment

This figure shows estimates of the contemporaneous effect of the shift in bullishness on stock returns from the OLS and the QR (i.e., Eqs. (3) and (4) for *s*=0). The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model are depicted by the green line along with their 95% confidence intervals (shaded area).

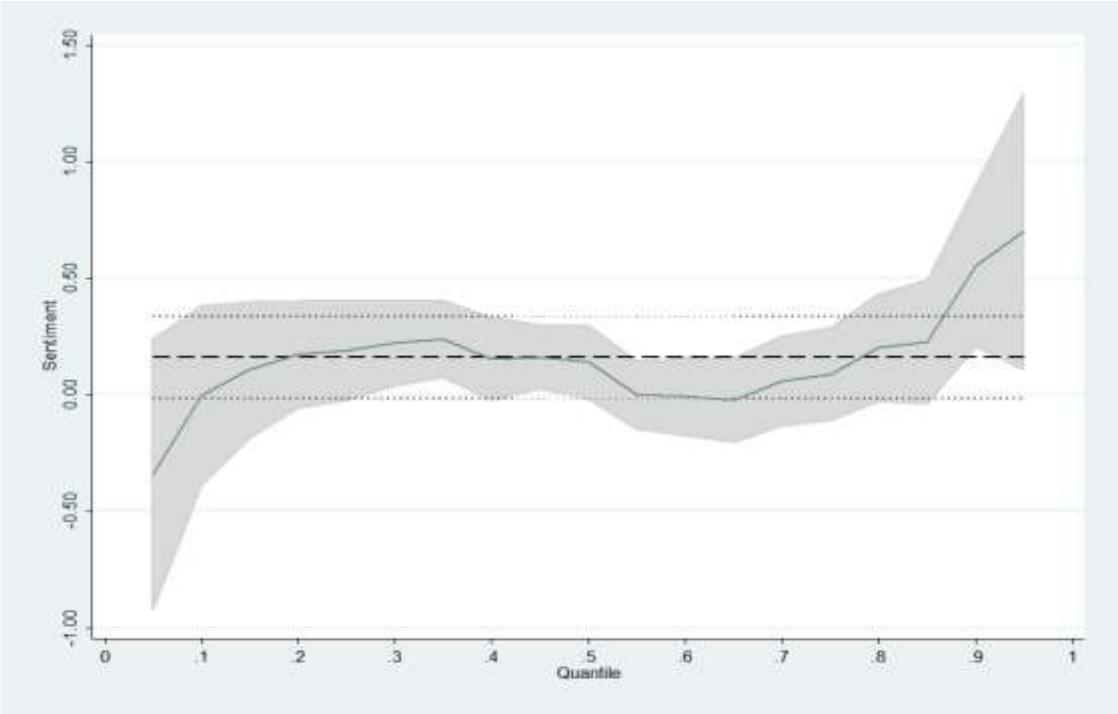


Figure 4: Estimates of predictive effects of sentiment

This figure shows estimates of the predictive effect of the shift in bullishness on stock returns from the OLS and the QR (i.e., Eqs. (3) and (4) for *s*=1). The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model are depicted by the green line along with their 95% confidence intervals (shaded area).

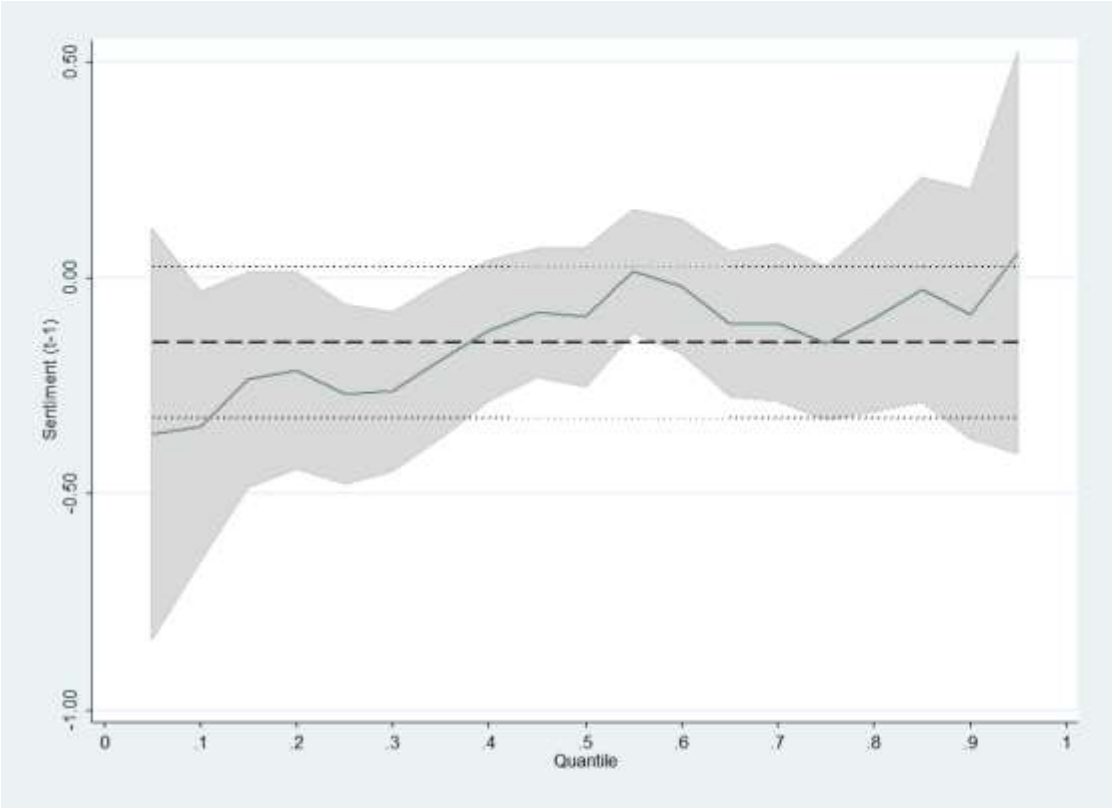


Figure 5: Estimates of contemporaneous effects of sentiment based on the Fama-French model

This figure shows estimates of the contemporaneous effect of the shift in bullishness on stock returns from the OLS and the QR using the Fama and French (1993) 3-factor model (i.e., Eqs. (7) and (8) for *s*=0). The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model are depicted by the green line along with their 95% confidence intervals (shaded area).

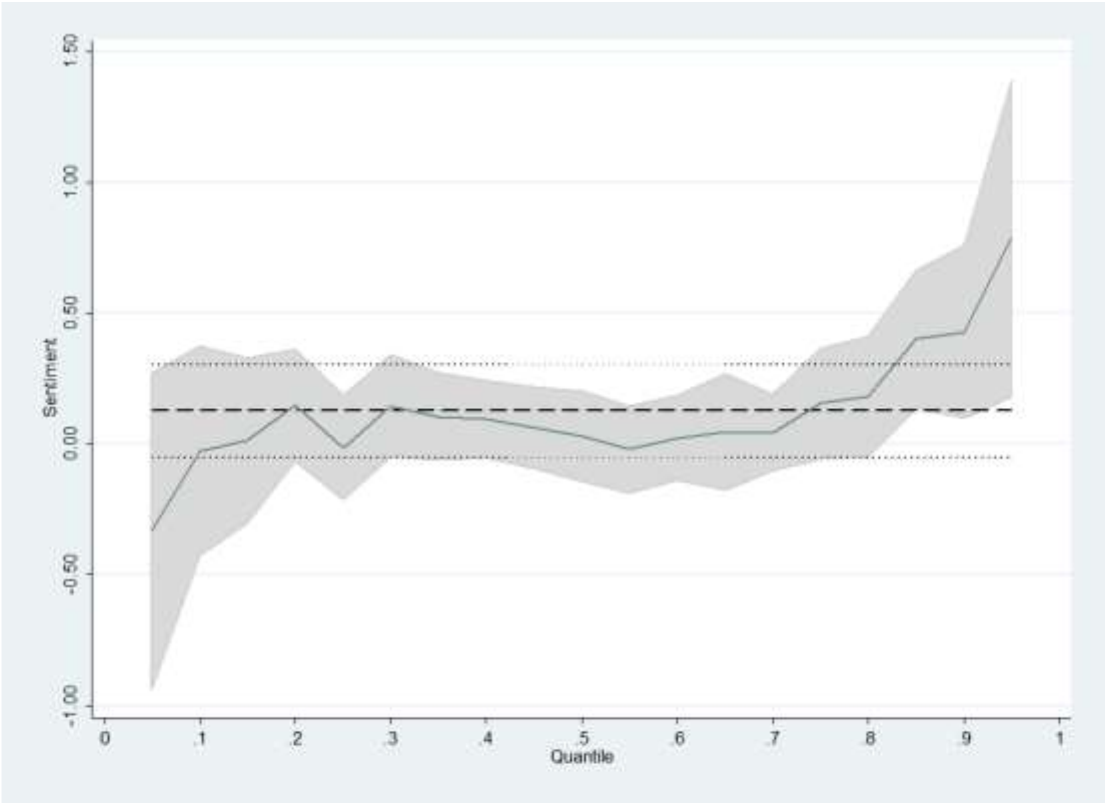


Figure 6: Estimates of predictive effects of sentiment based on the Fama-French model

This figure shows estimates of the predictive effect of the shift in bullishness on stock returns from the OLS and the QR using the Fama and French (1993) 3-factor model (i.e., Eqs. (7) and (8) for s=1). The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model are depicted by the green line along with their 95% confidence intervals (shaded area).

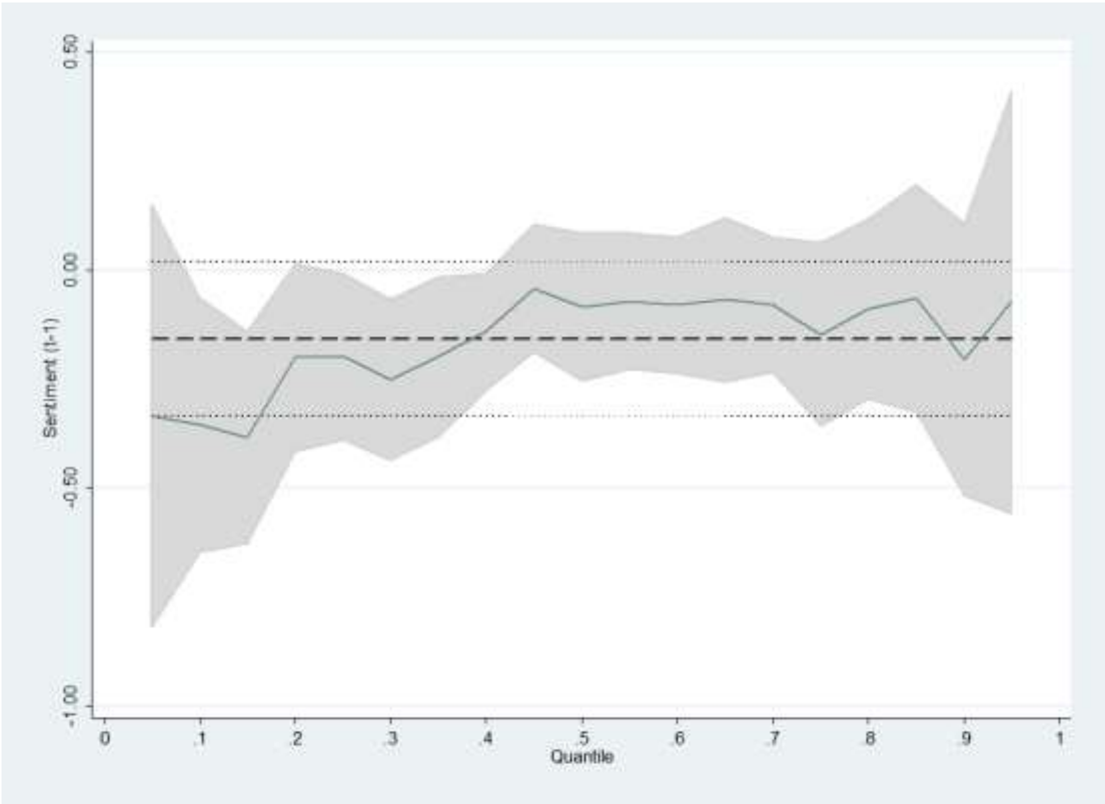


Figure 7: Sub-sample estimates of contemporaneous effects of sentiment

These figures present the sub-sample estimates of the contemporaneous effect of the shift in bullishness on stock returns from the OLS and the QR (i.e., Eqs. (3) and (4) for *s*=0), where the first (second) sub-sample is displayed in the left (right) panel. The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model for each sub-sample are depicted by the green line along with their 95% confidence intervals (shaded area).

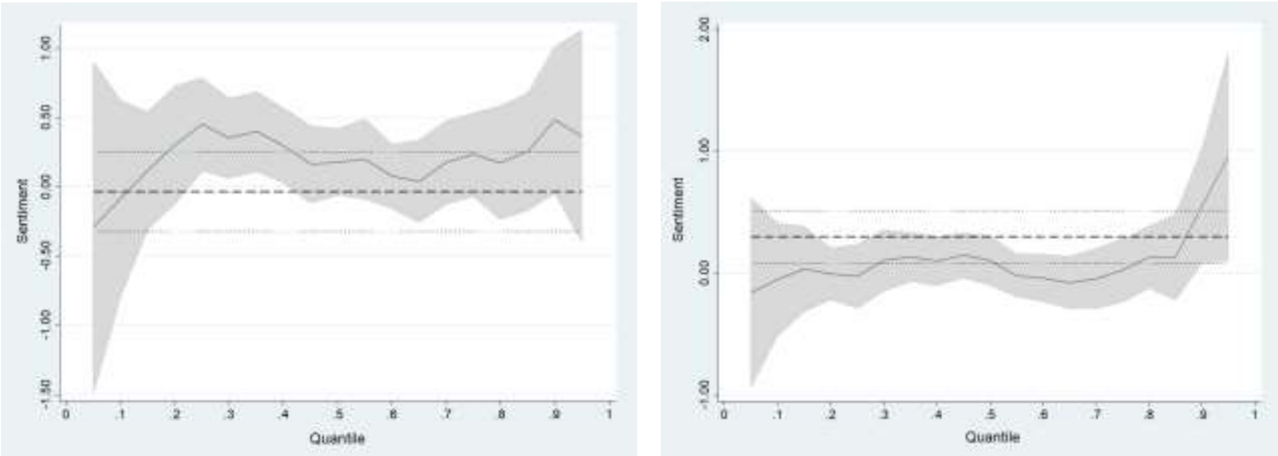


Figure 8: Sub-sample estimates of predictive effects of sentiment

These figures present the sub-sample estimates of the predictive effect of the shift in bullishness on stock returns from the OLS and the QR (i.e., Eqs. (3) and (4) for *s*=1), where the first (second) sub-sample is displayed in the left (right) panel. The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model for each sub-sample are depicted by the green line along with their 95% confidence intervals (shaded area).

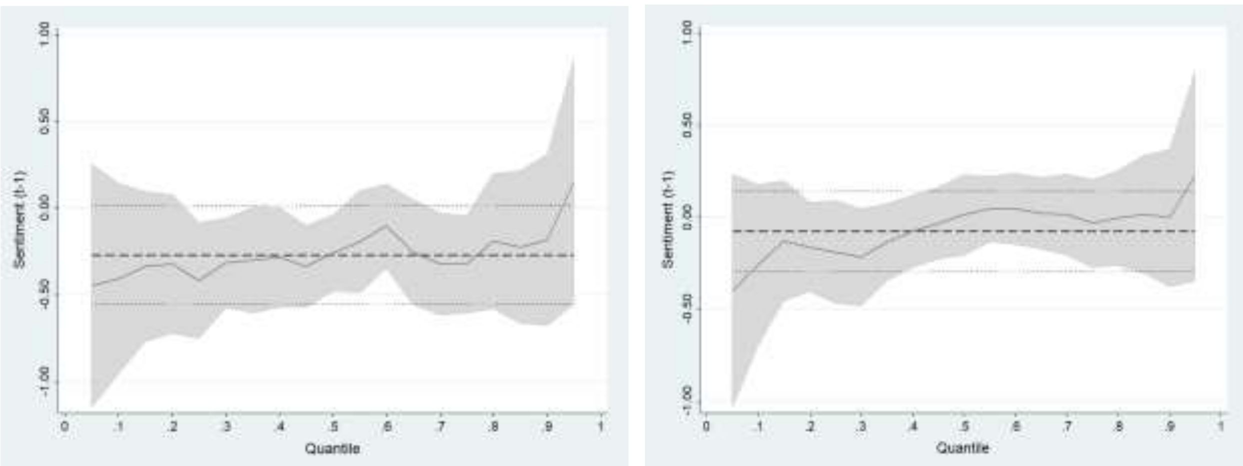


Figure 9: Sub-sample estimates of contemporaneous effects of sentiment based on the Fama-French model

These figures present the sub-sample estimates of the OLS and the QR of the contemporaneous effect of the shift in bullishness on stock returns from the OLS and the QR using the Fama-French model (i.e., Eqs. (7) and (8) for *s*=0), where the first (second) sub-sample is displayed in the left (right) panel. The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model for each sub-sample are depicted by the green line along with their 95% confidence intervals (shaded area).

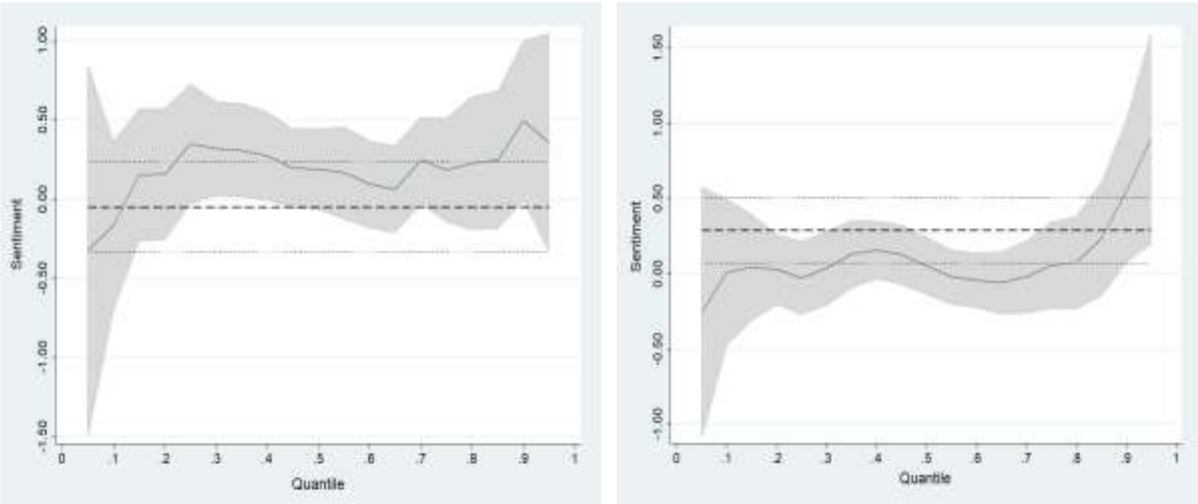
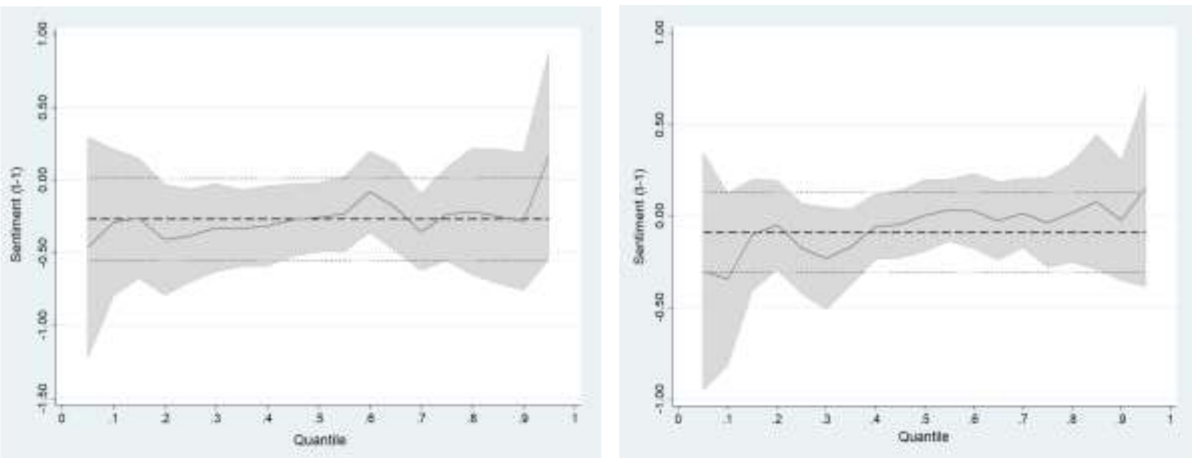


Figure 10: Sub-sample estimates of predictive effects of sentiment based on the Fama-French model

These figures present the sub-sample estimates of the predictive effect of the shift in bullishness on stock returns from the OLS and the QR using the Fama-French model (i.e., Eqs. (7) and (8) for *s*=1), where the first (second) sub-sample is displayed in the left (right) panel. The dashed line represents the OLS estimate along with its 95% confidence interval (dotted lines). The x-axis represents the quantiles of the return distribution (τ= 0.05, 0.1, 0.25,…, 0.75, 0.9, 0.95), while the y-axis indicates the coefficient estimates for the shift in bullishness. The estimated γ_τ parameters of the QR model for each sub-sample are depicted by the green line along with their 95% confidence intervals (shaded area).



Supplementary Appendix

Appendix A. StockTwits Data

The primary data for this study were obtained from Stocktwits.com (http://www.stocktwits.com). One year of StockTwits data (252 days) were retrieved via the website's API for the period of April 3, 2012 to April 5, 2013.[30] Specifically, more than 3,541,959 stock micro-blog posts were initially obtained from the API, where messages related to the companies making up the DJIA index were filtered out along with the required information related to each message, such as user ID, the content of the message and the published date and time. A complete list of the required attributes of StockTwits data needed for this study can be found in Table A1. Table A2, on the other hand, shows a few typical examples of the StockTwits messages, which are presented in their original format before pre-processing. The StockTwits API Schema, which describes the full StockTwits data, is provided in Table A3.

Table A1: The list of the required attributes for StockTwits collection

StockTwits Data	Attributes for Collection
ID	StockTwits unique identifier for the message
Body	Message content
Created _at	Date and time when the message was created

[30] In order to download the data, a formal agreement was made between the Research Support and Development Office of Brunel University London and StockTwits firm.

Table A2: Examples of StockTwits messages

ID	Tweet	Date	Time
12488749	"$IBM out half +.50"	11/03/2013	17:30:13
9901572	"$INTC short from Thursday working well. Up 2% with it so far. http://stks.co/mC9s"	09/10/2012	17:12:31
9611602	"$MA $V $AXP just wait until mobile payments overtake cash"	20/09/2012	15:46:30
12158099	"$VZ breaking out through 45 level with volume"	20/02/2013	18:20:52
7503061	"The Cramer on $INTC and $MSFT: http://stks.co/3EI2 (holding both)"	05/04/2012	01:04:16
11147935	"$JPM - Buy 43.50 puts for next week."	22/12/2012	13:11:07
12514291	"Current holdings: $ADP $T $V $ERX $XLU $QCOM $MSFT $ALTR $MUR"	13/03/2013	20:39:09
9562805	"$GS looks good here.....>122 YOU PRESS LONG ......got bull flag? http://stks.co/fBHM"	17/09/2012	22:43:01
10837630	"$MSFT for long term short!!!!!!"	05/12/2012	07:37:10
10171127	"$BA Buying before call with good numbers."	24/10/2012	14:06:15
11677420	"$DIS bearish to downside to 51.50"	25/01/2013	02:57:41
9541300	"$UNH vs. KFT News ~ Dow Swaps Out Kraft for United health ~ http://stks.co/iB6k"	15/09/2012	23:35:36

Table A3: StockTwits's Application Programming Interface (API) Schema

This schema describes the full StockTwits partner level firehose endpoint: http://stocktwits.com/developers/docs/api#streams-all-docs

user:classification	The users classification, if identity is "Official" the classification is either "ir" for the companies Investor Relation department or "pro"for Professionals and Analysts that are designated on StockTwits. An identified "user" canbe classified as "suggested" as a consistent respected contributor to StockTwits
user: followers	The number of people that are following this user
user: following	The number of people this user is following
user: ideas	The number of shared ideas

user:following_stocks	The number of stocks the user is following
user: bio	The users self described biography
user: website_url	A link provided by the user
user:trading_strategy: assets_frequently_traded	The users self described trading strategy describing the users assets frequently traded.Can be any: “Equities”, “Options”,“Forex”,“Futures”, “Bonds”, “PrivateCompanies”

id	StockTwits unique identifier for the message
body	Message content
created_at	Date when the message was created
user: id	StockTwits unique identifier for the user. Messages only have one user
user: username	Username of the user
user: name	Full name of the user
user: avatar_url	Path to the users avatar
user:avatar_url_ssl	SSL path to the users avatar
user: identity	The type of user, either “Official”,“User”

approach	Describing the users approach. Can be one of the following: “None”,“Technical”, “Fundamental”, “Global Macro”, “Momentum”, “Growth”, “Value”
user:trading_strategy: holding_period	The users self described trading strategy describing the users holding period. Can be one of the following: “None”, “Day Trader”,“SwingTrader”, “Position Trader”, “Long TermInvestor”

experience	The users self-described trading strategy describing the users experience. Can be one of the following: “None”, “Novice”,“Intermediate”,“Professional”
source: id	Message source unique identifier. Source is which application the message has originated from. Messages only have one source
source: title	The title of the source application
source: url	Link to the source application
symbols: id	StockTwits symbol internal unique identifier. Messages can have more than one symbol. Daily list of symbols can be downloaded here by date: http://stocktwits.com/symbol-sync/2013-01-30.c sv

symbols: symbol	Public ticker symbol
	Full public title of the ticker symbol
symbols: title	Stock exchange the ticker symbol resides on
symbols: sector	Sector for the ticker symbol. Sector list can be downloaded here: http://stocktwits.com/sectors/StockTwits-sectors -industries.csv
symbols:industry	Industry for the ticker symbol. Industry list can be downloaded here:http://stocktwits.com/sectors/StockTwits-sectors -industries.csv
symbols:trending	True or false flag if the ticker symbol was trending at the time of the message creation
entities:sentiment	Entities are optional. User specified sentiment at time of message creation. If sentiment is set within the message this will be either 0 - “bullish” or 1- “bearish”
entities: chart:thumb	Entities are optional. Path to the charts thumbnail image
entities: chart:original	Entities are optional. Path to the charts original image

entities: chart: url	Entities are optional. URL to the chart page on StockTwits
conversation:parent_ message_id	Conversation are optional. If there is a conversation The parent message for the StockTwits unique identifier
conversation:in_reply_to_m essage_id	Conversations are optional. If the message is a reply to another message the StockTwits unique identifier is represented
conversation:parent	Conversations are optional. True or false value if the message is the parent message that started the conversation
conversation:replies	Conversations are optional. Number of replies at the time of the message creation

Appendix B. Model building in Weka

Several types of models have been made available in Weka, each with different algorithms. The various forms of machine learning algorithms most commonly used include the Bayesian Networks, Decision Trees, Neural Networks, Fuzzy Networks, Support Vector Machine, Genetic Algorithms and many more. However, to keep the scope of this paper more focused, the Bayesian Networks (Naïve Bayes (NB)), Decision Trees (Random Forest (RandF)), and Support Vector Machines (Sequential Minimal Optimisation (SOM)) are used to perform the text analysis tasks. The performances of these three models were then evaluated on the training data and compared to select the best model.

More specifically, we test the three models using two different methodologies, namely testing on the training data set and testing by 10-fold cross-validation. The input for the models comes from a training corpus of 2,892 tweet messages. Ideally, the model should have been trained on more data instances as it is expected that the accuracy of the models will increase when bigger training data sets are handled.

Moreover, the stratified k-fold cross-validation provides a more realistic picture than testing on the full training data sets, i.e., the stratified k-fold cross-validation is considered a more conservative measure for classification accuracy.[31] In addition, it provides the best generalisability and helps overcome the risk of model over-fitting as each of these folds checks whether the learned model over-fits on the validation set (Cawley and Talbot, 2003).

It follows that our focus in the analyses will be on the results concerning the stratified 10-fold cross-validation to strengthen the validity of the results.[32] In line with standard metrics of information retrieval (Whitten and Frank, 2005), recall, precision, and F-measure are the reported measures used to evaluate the performance of the predictive model.

Table B1 presents a consolidated summary of all performance metrics of the three classifiers using 10-fold cross-validation. It is evident that there is no clear winning classifier in terms of the performance evaluation method used, yet the RandF classifier is possibly the best in terms of almost all the metrics.

As shown from Table B1, the 10-fold cross-validation experiments achieved accuracy rates of 66.70%, 62.80%, and 65.20%, where 1,929, 1,815, and 1,887 instances were correctly classified out of 2,892, for RandF, NB and SMO, respectively. It follows that the RandF Decision Tree classifier outperforms the NB and SMO counterparts in predicting investor sentiment class of StockTwits posts (i.e., buy, hold and sell). The weighted averages of the three classes of RandF classifier are also reported in Table B1, achieving 65.50%, 66.70%, and 66.20% for precision, recall, and F-measures, respectively.

Figure B1 shows the graphical representation of the comparative performance of the three discussed classifiers using some of the important measures given in Table B1. A graphical inspection indicates that all classifiers perform markedly the same, while the RandF shows a slightly better performance than the NB and SMO.

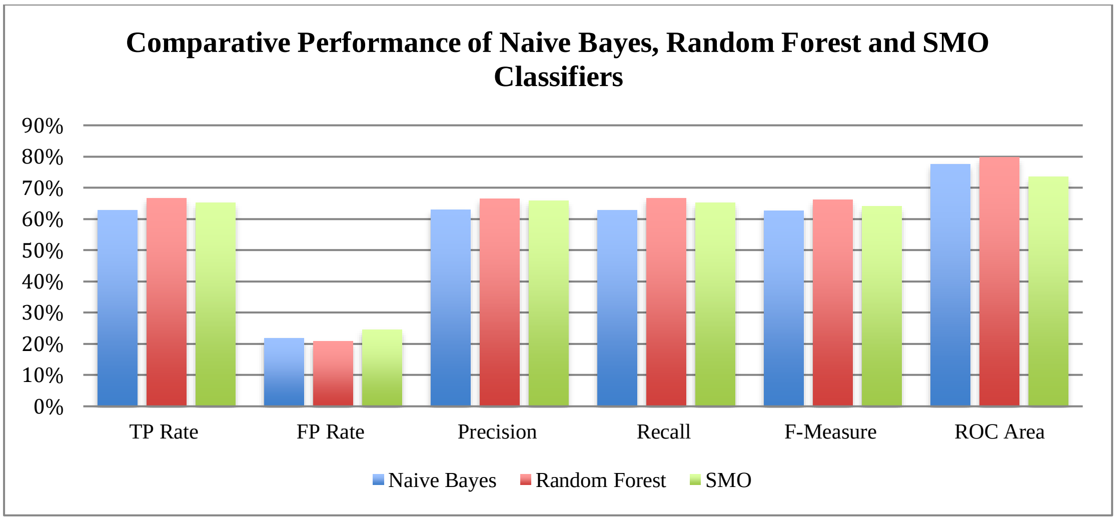
[31] In k-fold cross-validation, the data set is partitioned into k subsets. Then, the cross-validation procedure is repeated k times. Each time, one of the k subsets is used once as the test set and the other k-1 sets are combined to form a training data set.

[32] In this study, we use 10-fold cross-validation as it has proved to be statistically sufficient for the model evaluation method (Witten and Frank, 2005). In 10-fold cross-validation, the data sets are equally partitioned into 10 different subsets. The cross-validation process is repeated 10 times, where each time one of the 10 subsets will be used as a test set and the other 9 subsets will be combined to form the training data sets of the model. All 10 subsets will have an equal opportunity to be used as a test set only once and used as a training data set 9 times. Then, the average error estimates will be calculated across all 10 trails.

Table B1: Summary results of the classification performance evaluation of NB, RandF and SMO

Weighted Average Metrics for (Buy, Hold and Sell) Class	Classifiers		
	Naive Bayes	Random Forest	SMO
Accuracy Rate	62.80%	**66.70%**	65.25%
Correctly Classified Instances	1,815	**1,929**	1,887
Incorrectly Classified Instances	1,077	**963**	1,055
TP Rate	62.80%	**66.70%**	65.20%
FP Rate	21.80%	**20.80%**	24.60%
Precision	62.90%	**66.50%**	65.90%
Recall	62.80%	**66.70%**	65.20%
F-Measure	62.60%	**66.20%**	64.00%
ROC Area	77.60%	**79.80%**	73.60%

Figure B1: Comparative performance of NB, RandF and SMO classifiers



Appendix C. Decision Tree classifiers

In this Section, we describe in detail the underlying method of our Decision Tree text classification. As explained by Quinlan (1993), a Decision Tree algorithm generates Decision Trees or nodes by choosing the most effective attribute that splits each node into sub-nodes, where each node or sub-node holds a class label. Moreover, to evaluate the effectiveness of an attribute in splitting the data, the algorithm uses the Information Gain (IG) criterion in which the attribute with the greatest normalised information gain is chosen to make the decision. The process of splitting the decision nodes continues until no further split is possible. This process safeguards the maximum accuracy of the training data, implying that the data have been classified as close to perfection as possible.

Creating Decision Trees

To form a Decision Tree, the following steps are required:

Step 1: Define the entropy of *x*

$$\mathrm{H}(x) = -\sum_{i}^{k} P_i \log_2(P_i), \tag{B1}$$

where *x* is a random variable with *k* discrete values, distributed according to probability value *P* = (P_1, P_2, P_3,…P_n) of class subset *i*.

Step 2: Calculate the weighted sum of the entropies for each class subset

$$H_s(\mathrm{T}) = \sum_{i=1}^{k} P_i H_s(T_i), \tag{B2}$$

where Pi is the proportion of records in subset i.

Step 3: Calculate the IG

$$IG_s = H(T) - H_s(T), \tag{B3}$$

where the IG is the criterion necessary to choose the most effective attribute to make the decision, as indicated earlier. Then, the selection of attribute at each decision node will be the one with the highest IG.

Text Preparation

Text preparation is considered the initial stage of the textual data mining process. At this stage, pre-processing of textual data is carried out and the selection of input variables or attributes is made. The task of selecting input variables (the so-called "bag of words" approach through a feature selection method) needs to be interactively and collaboratively determined by data mining and human experts (i.e., financial managers) in the domain field of data (i.e., financial data). The guidance of domain experts can help in determining which terms or phrases are more appropriate in textual analysis. These input variables are then coded and put in a format suitable for text data mining (TDM) tasks.

The next step is to apply some pre-processing techniques. Specifically, six pre-processing steps are performed to improve the quality of data input and reduce feature space. First, unnecessary words or noise words with low effectiveness in textual analysis of the data are removed. These so-called stop words include some verbs (e.g., is, are, were, etc.), pronouns, conjunctions, and disjunctions (e.g., a, an, the, of, and, I, etc.). The advantage of removing such words is that the text is cleansed of ineffective words and can be interpreted more effectively and efficiently. That is, the omission of these less informative words improves the accuracy of the results of the text mining process and is considered a common task in most text mining applications (Blair, 1990). While unnecessary words are removed, the addition of other words that are relevant to a particular context (e.g., in this study, company names are proved to be relevant) is also effective in textual data analysis.

Second, the text needs to be re-formatted (e.g., whitespace removal). Third, all tweet data must be converted to lower case. The assumption behind this is that an automated algorithm will, therefore, treat any of these characters separately (e.g., “sell” and “Sell” would be two distinct features). Fourth, the widely used Porter stemmer approach is applied to remove suffixes (or morphological endings) from words. Word stemming is one of the important preprocessing steps to be considered. It refers to the process of bringing words back to their actual form. In other words, it is the process of shortening derived words to their initial roots. For example, words like ‘buys’ and ‘buying’ are stemmed from their base word ‘buy’ (Porter, 1980).

Fifth, Tokenisation[33] must be performed on the database. This is defined as a process of replacing all values, symbols, percentages, hyperlinks, and figures with a token (text). For example, all stock tickers “$ticker” of the companies are replaced with the token “Stock dollar sign” (e.g., “$NKE” denotes the ticker symbol of the Nike company which will be tokenised as “Nike dollar sign”). The characters “$$” or “$$$”, which are most commonly used as abbreviations for the term money, are replaced by a common format “money”, and the “@” sign in tweets is replaced by the text “at”. Sixth, all tweets duplicated by the same user and tweets posted at weekends and on public holidays are removed.

Appendix D. Weka output results for Random Forest (Decision Tree classifier)

This Section presents the automated classification results of Random Forest Decision Tree algorithm.

== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	1929	66.7012 %
Incorrectly Classified Instances	963	33.2988 %
Kappa statistic	0.4657	
Mean absolute error	0.2748	
Root mean squared error	0.3989	
Relative absolute error	65.307 %	
Root relative squared error	86.9758 %	
Coverage of cases (0.95 level)	94.0526 %	
Mean rel. region size (0.95 level)	72.1531 %	

[33] Tokenisation is the process of substituting a sensitive data element with a non-sensitive equivalent, referred to as a token that has no extrinsic or exploitable meaning or value. It is the process of mapping the original data to a token system.

Total Number of Instances 2892

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.784	0.305	0.696	0.784	0.737	0.479	0.800	0.736	buy
0.639	0.102	0.616	0.639	0.627	0.530	0.859	0.616	hold
0.515	0.134	0.650	0.515	0.575	0.409	0.757	0.623	sell
0.667	0.208	0.665	0.667	0.662	0.466	0.798	0.675	Weighted Avg.

=== Confusion Matrix ===

a	b	c	<-- classified as
1067	117	177	a = buy
129	377	84	b = hold
338	118	485	c = sell

Supplementary Appendix References

Blair, D.C. and Maron, M.E. (1990) Full-text information retrieval: Further analysis and clarification, *Information Processing & Management* **26**(3), 437-447.

Cawley, G.C. and Talbot, N.L. (2003) Efficient leave-one-out cross-validation of kernel Fisher discriminant classifiers. *Pattern Recognition* **36**(11), 2585-2592.

Porter, M.F. (1980) An algorithm for suffix stripping, *Program* **14**(3), 130-137.

Quinlan, J.R. (1993) *C4.5: Programs for Machine Learning*, Morgan Kaufmann. San Francisco, CA, USA.

Witten, I.H. and Frank, E. (2005) *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann. San Francisco, CA, USA.