An action – identifying noise traders entering the market with Google and Twitter

Carolin Hartmann*, Hans-Peter Burghof* and Marc Mehlhorn**

ABSTRACT

In this paper we measure and compare the impact of changes in Google search volume (GSV) and Twitter volume (TV) on financial markets. We find that information investors access via Google and Twitter have an impact on financial markets and predictive power. The impact of TV is more important. First, changes in GSV and TV have a positive impact on changes in turnover. Traders enter the market (Easley et al. 1996). Second, we show that changes in TV increase the share of noise traders on the market (De Long et al. 1990), while changes in GSV have no significant impact. Our results suggest that we are able to measure investor sentiment with TV. Further, we expect that the different impact of GSV and TV lies in the different use of Google and Twitter by investors.

Key words: forecasting, investor behavior, noise trader, search engine data, social media, investor sentiment

JEL: G10, G14, G17

1 Introduction

Google and Twitter have users worldwide[1]. They are indicators for the public mood (Bollen et al. 2011). In July 2018 Twitter had 335 million monthly users and 76 percent of search engine users choose Google. Especially for financial markets Twitter becomes more and more important. U.S. president Donald Trump uses Twitter as his direct medium of communication[2]. Elon Musk's tweet on 7 August 2018 about taking TESLA private led to an increase of the share by 11 percent. Short sellers lost approximately 1.3 bn USD.

Current research on financial markets shows that Google search volume (GSV) and Twitter volume (TV) have an impact on financial markets (Da et al. 2011; Dimpfl & Jank 2016; Hamid & Heiden 2015; Bollen et al. 2011; Mao et al. 2015). We contribute to the literature by quantifying the impact of TV and GSV on financial markets. We use GSV and TV as proxies for new information which is taken into account by investors. Our data set is unique and combines data on Google and Twitter on a daily level. We apply a two-step procedure. First, we measure the market entry of traders using GSV and TV. Following the idea of Easley et al. (1996) we find that changes in GSV and TV have a positive impact on turnover on the same day and the next day. This indicates that more traders enter the market. Second, we investigate with the DSSW model (De Long et al. 1990) the effect of GSV and TV on volatility. We

[1] Google is a search engine like an encyclopedia established in 1997. Twitter is a microblogging service established in 2006. Twitter captures active participation, tweets but users can also follow other users, obtain information or forward it.

[2] http://www.trumptwitterarchive.com/

The authors would like to thank Sowa Labs (http://www.sowalabs.com) for providing the sentiment labeled Twitter data about the 30 DJIA companies and the DALAHO of the University of Hohenheim for the data provision (https://wiso.uni-hohenheim.de/en/109608). Further, we want to thank the seminar participants of the Augustin Cournot Doctoral days for their comments. Special thanks go to Maxime Merli and Carola Mueller for their support.

* University of Hohenheim, Institute of Financial Management, Mailing address: Schwerzstraße 38, 70599 Stuttgart, GERMANY, Tel: +49 (0) 7 11 45 92 29 00, Email: carolin.koch@uni-hohenheim.de

** Vereinigung Baden-Württembergische Wertpapierbörse e.V. Stuttgart Financial, Mailing address: Börsenstraße 4, 70174Stuttgart, GERMANY, Tel: +49 (0) 7 11 22 29 85 72 6, Email: mehlhorn@stuttgart-financial.de

find that an increase in TV leads to an increase in volatility. Thus, an increase in the share of noise traders on the market on the same day and the next day. Changes in GSV have no measurable impact on the share of noise traders on the market. Therefore changes in TV are better suited to assess movements on financial markets.

According to the Efficient Market Hypothesis (Fama, 1981) stock prices should on average be driven by fundamentals and only move if new information about fundamentals enter the market and are priced correctly. The idea of behavioral finance is that investors do not behave rational (Kahneman & Tversky 1979; Shiller 2003). As a consequence, we observe investor sentiment and limits-to-arbitrage on financial markets (De Long et al. 1990; Shleifer & Vishny 1997; Baker & Wurgler 2006; Baker & Wurgler 2007). According to Mao et al. (2015) Google and Twitter are possible online data sources of investor sentiment, as they capture the behavior of investors on the market. Arguably, Google and Twitter offer a huge amount of information but it is questionable if (a) this information is useful for financial markets, (b) or --useful or not-- used by participants of financial markets, and (c) if Google and Twitter data is reliable in case it is used.

Before the use of online data sources, Barber and Odean (2008) measured investor behavior using unusual trading volume, extreme returns and news coverage. They find that abnormal trading volume is the best indicator of investor attention. But they use an indirect measure as (1) they cannot differentiate between investor types and (2) they do not know to which available information investors pay attention. Da et al. (2011) measure the interest of investors in stocks using the Google search volume index (GSVI). They find that GSVI is a direct proxy of retail investor attention and that it is a predictor of stock price movements. Their measure outperforms the measures of Barber and Odean (2008). These results are confirmed by Fink and Johann (2014), Dimpfl and Jank (2016) and Dimpfl and Kleinman (2017). Dimpfl and Jank (2016) show that Google has predictive power. An increase in Google search volume increases the volatility of the DJIA[3] on the next day. Hamid and Heiden (2015) use the same set up as Dimpfl and Jank (2016) to forecast weekly volatility. They find a significantly better in-sample and out-of-sample predictions including GSV, especially in times of high volatility. These findings are in line with the outcome of further studies (Vlastakis & Markellos 2012; Vozlyublennaia 2014; Andrei & Hasler 2015; Choi & Varian 2012).

Information measured by Twitter is not related to a special group of investors in the literature. Mao et al. (2015) uses Google search queries and Twitter updates as an indicator of online investor sentiment. They find that the bullishness of Twitter updates on a daily level is a suitable indicator for investor sentiment. Their results indicate that a high Twitter bullishness leads to an increase in returns. They show that there is a positive correlation between Twitter bullishness and Google. Changes in the bullishness of Twitter are followed by changes in Google. Bollen et al. (2011) find that moods extracted from Twitter tweets are able to improve DJIA predictions. Tafti et al. (2016) find a real-time relationship between the activity on Twitter and the trading volume of Nasdaq100 firms. Alexander and Gentry (2014) show that 77 percent of the Fortune 500 companies in America tweeted in 2013. They point out that companies use social media platform to republish company information and to have a channel to keep investors updated e.g. for live tweeting during special events such as annual general meeting SEC (SEC 2013). A tweet by Carl Ichan to his 90 thousand followers on the company Apple led to an increase of the stock by 18 bn USD (Carr 2013).

According to the existing literature and based on our findings, we expect that the difference between the impact of GSV and TV on turnover is justified in the way individual investors use Google and Twitter.

[3] As a measure of volatility they use realized volatility which they calculate following Andersen et al. (2003)

Google as a search engine, is used to get information on a certain topic. It can lead to trades but this is not true for all searches. Twitter seems to observe the financial markets in a better way. If someone tweets or retweets a statement concerning a share of a DJIA index company there is a better chance that it translate into a trade. Moreover some people use Google to verify the Twitter news which can explain why the influence of Google gets lower as soon as Twitter is integrated in the search. We do not expect that people use Google before they are active on Twitter.

In this paper, we test if investors actually use Google or Twitter as source of information by estimating their impact on (1) stock turnover and (2) volatility. According to the model by Easley et al. (1996) new traders arrive on the market when there is new information. Furthermore, we conjecture that traders who use Google or Twitter as their source of information are not institutional. They would rather classify as retail investors, uninformed traders, or noise traders (Black 1986; De Long et al. 1990; Baker & Wurgler 2007). The model of DeLong, Shleifer, Summers, Waldmann (DSSW) provides a theoretical approach to identify the impact of noise traders on prices (De Long et al. 1990). We reformulated the DSSW model equation to estimate the impact of GSV (TV) on stock volatility. The positive correlation between changes in GSV (TV) and volatility implies that traders using these online data sources can produce noise on the market. We find that TV increases the share of noise traders on the same day and the next, GSV does not. The purpose and use of Google is to verify information while Twitter is potentially more prone to manipulation. Google is a medium for retail investors which has no measurable impact on the share of noise traders on the market. The impact of Twitter is different as users can not only search for information but also tweet and follow information. The reaction to this information leads to more market entries and leads to an increase of noise traders on financial markets. The recent Elon Musk case on taking Tesla private (Bain & Mott 2018) and the study of Mao et al. (2015) support our findings.

The procedure of this paper is as follows. We discuss the theoretical models in Section 2. We describe our dataset in Section 3. We present evidence that we measure market relevant information, market entries and noise traders with changes in GSV (TV) in Section 4. We test our results in Section 5. We critical reflect the database in Section 6 and conclude in Section 7.

2 Theoretical models

We use a two-step procedure to measure (1) the impact of new information on trading and (2) the change in the share of noise traders on the market due to changes in GSV and TV.
To measure the arrival of new traders on the market we adapt an approach by Easley et al. (1996). Their market microstructure model measures the probability of informed trading. According to Easley et al. (1996) the probability of informed trading is lower (higher) for high (low) volume stocks. They find that the probability of information-based trading of informed and uninformed traders, depends on new information on the market. More traders arrive on the market if new information exits. The probability of informed trading PI is defined as:

$$PI(t) = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon}\,. \tag{1}$$

If new information exists, informed traders arrive on the market with the arrival rate of μ. They are risk neutral and competitive. The arrival rate of uninformed traders is ε. The trading of uninformed traders is noise. Information occurs on the market with probability α. Normally both trader types are on the market. We simplify the model approach to find out if new information leads to a higher arrival rate of traders on the market. First, we use GSV and TV as a proxy for new investor relevant information. Second, we measure if new traders arrive on the market, without distinguishing between μ and ε. Third,

we look at the effect of information on turnover. We want to measure the impact of new information on trading. Thus, we do not differentiate between high and low volume stocks, as all stocks are part of the DJIA which consist solely of blue chip stocks.[4] Furthermore, we look at turnover instead of volume. We measure turnover as the ratio between the amount of shares traded and the number of shares outstanding (Lo & Wang 2000). This allows us to qualify for different trading volumes and number of shares outstanding of stocks.
We expect to find an impact of changes in GSV and TV on trading, if they include new information. Our findings confirm the expectations, as we find that GSV and TV have an impact on turnover. This means that new traders[5] enter the market.

Further, we asses if GSV and TV increase the share of noise traders on the market. According to Black (1986) noise is the source of inefficiency on financial markets but is also the reason why trading in financial markets is possible. Kumar and Lee (2006) show that trades conducted by individual investors contain a systematic component, meaning that individual investors can move stock prices. The DSSW model of DeLong et al. (1990) distinguish between sophisticated investors, who are rational traders and noise traders, who are irrational traders. In contrast to sophisticated investors, noise traders are subject to sentiment, meaning that their beliefs about future prices deviates from the fundamental value (Long et al. 1990; Baker & Wurgler 2007). The unpredictable behavior of noise traders increases the risk of trading for rational traders (De Long et al. 1990). Rational investors cannot foresee the reaction of the noise traders. Thus, it takes some time until rational investors and arbitrageurs force prices back to their fundamental value (De Long et al. 1990; Shleifer & Vishny 1997; Baker & Wurgler 2007). Shleifer and Vishny (1997) qualify this risk as limits-to-arbitrage. The higher the share of noise traders on the market, the higher is the deviation of the stock prices from the fundamental value (De Long et al. 1990).

Black (1986) connects information, investor types and noise trading to volatility. He finds that noise trading can create volatility in the future. The more noise traders on the market, the higher is volatility. Antweiler and Frank (2004) asses internet stock message on the DJIA and find that they help to predict market volatility. If individual investors behave as noise traders this can have a positive effect on volatility, contributing to idiosyncratic volatility above and beyond cash-flow news (Foucault et al. 2011). Foucault et al. (2011) find that changes in volatility are not entirely explained by changes of the fundamental value or changes in news. They look at the effect of a French stock market reform which increases costs for speculative trading for individual investors. Applying the DSSW model (De Long et al. 1990), they find that individual investors who behave as noise traders increase volatility.

The DSSW model is a two-period overlapping generations model. The aim of the agents is to maximize their utility. Two types of agents exist, the sophisticated investors i and the noise traders n. The share of noise traders is equal to μ; the share of sophisticated investors is equal to $1-\mu$. In the first period t they can decide between an investment in a safe asset s and in a risky asset u. The price of the safe asset is equal to one. The price of the risky asset is equal to p_t in period t. The price p_{t+1} in the second period is unknown. It depends on the agents' expectations about p_{t+1} built in the first period ${}_tp_{t+1}$. The expectations of the sophisticated investors are based on fundamentals. Noise traders are subject to sentiment as their beliefs about the price development are misperceived. The misperception is measured by ρ_t[6].The higher ρ_t the more bullish is the noise trader. The risk aversion of the agents is taken into

[4] In the initially model they Easley et al. (1997) look at over 1000 stocks here the differences between the stocks and the trading volumes are more pronounced.
[5] We have no personal data so we cannot say if really more single trader enter the market or if the traders on the market trade more.
[6] The variable is independent and identically normally distributed random variable with a mean of $\rho^*$and a variance of σ_ρ^2.

account in the coefficient γ. The price of the risky asset today p_t is a discounted function of the expected price of the risky asset in $t+1$:

$$p_t = \frac{1}{1+r}\left[r + {}_t p_{t+1} + \mu\rho_t - 2\gamma\ {}_t\sigma^2_{p_{t+1}}\right]. \tag{2}$$

It consists of the return the agents will earn in the second period r; the expected price of the risky asset in the second period ${}_t p_{t+1}$, the misspecification of the price by the noise trader times the share of the noise traders on the market $\mu\rho_t$, the behavior of the agents according to risk γ and the variance of the risky asset ${}_t\sigma^2_{p_{t+1}}$. Assuming a steady state equilibrium[7] the price of the risky asset p_t depends on known parameters. The unknown variance of the future price becomes ${\sigma^2}_{p_{t+1}}$, with:

$${\sigma^2}_{p_{t+1}} = \frac{\mu^2\sigma^2_\rho}{(1+r)^2}. \tag{3}$$

To determine the influence of noise traders on the variance of the price we take the partial differential of the variance $\sigma^2_{p_{t+1}}$ with respect to the share of the noise traders μ:

$$\frac{\partial\sigma^2_{p_{t+1}}}{\partial\mu} = \frac{1}{(1+r)^2}\ 2\mu\sigma^2_\rho > 0. \tag{4}$$

The share of noise trader has a positive effect on the variance of the price of the risky asset tomorrow. If the share μ of noise traders on the market increases, the variance of the risky price in the next period increases as well. The size of the effect is determined by the share of noise traders on the market, the variance of the bullishness σ_ρ of the noise traders and the quadratic discount factor $(1+r)^2$. From the effect of noise traders on the variance we can deduce the effect on the volatility. The volatility is the standard deviation of the variation of the price in the second period $\sigma_{p_{t+1}}$. It can be obtained by taking the square root of the variance:

$$\sigma_{p_{t+1}} = \sqrt{\sigma^2_{p_{t+1}}} = \sqrt{\frac{\mu^2\sigma^2_\rho}{(1+r)^2}} = \left(\frac{\mu^2\sigma^2_\rho}{(1+r)^2}\right)^{\frac{1}{2}} = \frac{\mu\,\sigma_\rho}{1+r}\,. \tag{5}$$

Taking the partial derivative of the standard deviation of the price in $t+1$ with respect to the share of noise traders μ on the market shows the effect of noise traders on volatility:

$$\frac{\partial\sigma_{p_{t+1}}}{\partial\mu} = \frac{\sigma_\rho(1+r)-0}{(1+r)^2} = \frac{\sigma_\rho}{1+r}\,. \tag{6}$$

The share of noise traders μ has an influence on the volatility of the future stock price but the direction of the effect cannot be clearly determined at first sight. The variable ρ is an i.i.d. normal random variable, it can be positive or negative. The standard deviation $\sigma_\rho$of ρ is positive. The return of the assets r is not bounded by definition and can be positive or negative[8]. If an increase of the share of noise traders leads to an increase in the variance, it also has a positive effective on the standard deviation. The more noise traders are on the market the higher the volatility on the market and vice versa.

---

[7] Consider steady state equilibria and assuming the same distribution for p_t and p_{t+1}

[8] Even if investors hold only positive amounts of both assets, the fact that returns are unbounded gives each investor a chance of having negative final wealth

In our empirical study, we use changes in GSV and TV as proxies for changes in investor behavior due to new information. Our expectation is that changes in GSV and TV lead to changes in volatility, representing an increase in the share of noise traders on the market. Our findings show that this holds for TV but not for GSV. We can show that Twitter is a measure of investor sentiment.

3 Data and descriptive statistics

The sample covers a timespan from 6 June 2013 until 31 December 2016. The balanced panel contains market data for 29 stocks of the Dow Jones industrial average index (DJIA). One observation stands for one stock per day. We include only weekdays from Monday to Friday to avoid unclear weekend effects (Andersen et al. 2003). All variables are expressed as percentage changes, e.g. for volatility:

$$\Delta volatility_t = ln\left(\frac{v_t}{v_{t-1}}\right). \tag{7}$$

The use of logarithmic variables leads to better behaved variables and reduces the correlation between the independent variables. The data sources are Datastream, Google Trends and Sowa Labs[9]. A list of the stocks can be found in the appendix.

Google

The data on GSV is obtained via Google Trends as shown in earlier research (Fink & Johann 2014; Hamid & Heiden 2015; Dimpfl & Jank 2016). This specific Google website gives the possibility to obtain information about the search volume for various search terms. As a result, the Google search volume Index (GSV) is generated. The GSV is a time series which depicts the relative search volume of a specific search term on Google during a certain time span or point in time. Depending on the length of the time span the time series ranges from eight minutes intervals up to monthly data. It is further possible to create filters for example for regions or special topics such as “all categories” or “finance”. Moreover there is a Google trend feature which allows to specify the search term. Besides the option “search term” it is possible to choose a “topic”. As an example Google trend states that if someone googles “London” also “Capital of Great Britain” is integrated in the search if the search is conducted by topic. This means that choosing the option “topic” covers a broader range for a specific search term than the search term alone. For the search of the constituents of the DJIA index the finance filter in combination with the topic filter was chosen on a daily level. The data is obtained for each stock being part of the DJIA in this time span.

Twitter

The Twitter data is from Sowa Labs and comparable to the data used by Peter et al. (2017). It contains approximately 4.5 million tweets for 29 companies. The data is collected by a Twitter search API. For each company the stock cash-tag is specified (e.g. for Apple it is “$AAPL”). To capture the market movement we use the Twitter Volume (TV) on each stock without separating into positive, negative or neutral sentiment. The tweets are aggregated on a daily level.

Turnover

As a measure of trading activity we follow Lo and Wang (2000) who suggest turnover as the best way to measure trading activity. They calculate turnover as a share between the volume of traded shares and the shares outstanding (Lo & Wang 2000):

[9] Sowa Labs measures Twitter volume and Twitter sentiment by counting the tweets on a certain search term on Twitter

$$Turnover_{it} = \frac{X_{\text{it}}}{N_i}. \quad (8)$$

With X_{it} being the volume of shares traded of one stock i at time t and N_i standing for the total number of shares outstanding of stock i. The variables are obtained from Datastream and calculated for each stock on a daily basis. Brooks (1998) also uses a comparable measure by taking the shares traded per day divided by the number of shares outstanding. Looking at turnover as a proportion between trading volume and number of shares outstanding avoids overweighting of large stocks.

Volatility

The variable volatility allows to capture the movement of the stock markets. It is also considered to be a measure of market sentiment. Volatility can be measured based on historical data or implicit based on options. We use an approach based on historical data. Realized volatility per stock would be the measure suggested in the current research (Andersen et al. 2003; Dimpfl & Jank 2012). We measure volatility as the logarithmic difference between high and low stock prices per day, suggested by Brooks (Brooks 1998; 2014). This simpler approach of realized volatility is necessary due to a lack of data. The daily high and low prices are from Datastream.

$$Volatility_t = ln\left(\frac{Price\ high_t}{Price\ low_t}\right). \quad (9)$$

Squared returns

As an indicator for news on the market we use squared returns to capture larger market movements. According to Barber Odean (2008) extreme returns on stocks, positive or negative, are associated with news on the market. Abhyankar (1995) measures good or bad news by the size of returns (Brooks 1998). Vozlyublennaia (2014) uses past returns as a measure of information. She further suggests a relationship between returns, volatility and investor attention. According to Andersen and Bollerslev (1998), squared daily returns can also be seen as a noisy proxy for the true volatility. We use squared returns as a proxy to measure the impact of news on the market. The continuous returns are calculated based on daily prices for all relevant stocks obtained via Datastream.

The summary statistics are presented in Table 1. The mean value of all observations is close to zero. The standard deviation varies between 0.323 (GSV) and 2.992 (change in volatility). This means that the changes deviate from the mean in both directions but are close to zero over all observations. We find that for changes TV these fat tails on the right are more pronounced than for changes GSV. For changes in GSV (TV) we find a positive skewness which is smaller for GSV (0.080) than for TV (0.283). This means that the tail on the right hand side is longer than on the left. The kurtosis of GSV (TV) is at 6.059 (6.173). The changes in turnover are positively skewed (0.28) and peaked (4.91). This means that we see more extreme changes in turnover than under normal distribution and they are more pronounced on the right, as the skewness is larger than zero. The distribution of changes in volatility and squared returns is close to normal. The number of observations vary as some observations are equal to zero and here we cannot calculate the changes which leads to missing variables. This is especially pronounced for the spreads but also for changes in GSV, TV and volatility.

Table 1. Summary statistics

VARIABLES	Mean	Stan. Dev.	Skewness	Kurtosis	Min	Max	N
ΔGSV_t	0.0001571	0.3230409	0.0803382	6.059349	-2.079442	2.628801	25544
ΔTV_t	0.0000635	0.5802832	0.2832056	6.17311	-4.382027	4.567814	26136
$\Delta Squared\_Return_t$	-0.0006845	0.4739944	0.1554362	3.219353	-2.311268	2.507094	26158

$\Delta Turnover_t$	-0.0005657	0.3484734	0.2756585	4.910202	-1.65539	2.576512	26158
$\Delta Volatility_t$	-0.0103956	2.99265	-0.0186688	3.597933	-13.40264	12.78078	25781

4 Empirical results

We find that changes in GSV and TV have a positive impact on the market entry of traders. Moreover, we find that an increase in changes in TV increases the share of noise traders on the market. Further, lagged changes in GSV and TV have predictive power on changes in turnover. Lagged changes in TV has also a predictive power on changes in volatility.

4.1 Correlation analysis

Table 2 shows the correlation between the different variables. The correlation for almost all variables is below 0.3. The correlation between changes in turnover and changes in volatility is higher (0.4967). This means that an increase in volatility goes in line with an increase in turnover. The positive relation is known and described in the literature. Brooks et al. (2001) confirm this positive relationship. The correlation between changes in squared returns and volatility is at 0.4358. Both variables refer to the changes in the underlying price of the stock but there is no linear dependency. Extreme returns and volatility go hand in hand. Changes in GSV (TV) are positively correlated with changes in turnover and volatility. The correlation is higher for changes in TV. Thus we expect the effect of changes in TV to be more pronounced than the effect of changes in GSV on turnover and volatility. This is in line with Mao et al. (2015) who find that Twitter is a better predictor of investor sentiment than Google. Moreover, changes in GSV and TV are positively correlated (0.1166). To assess collinearity we calculate the condition number. For our sample it is at 1.68. Taking 20 as the critical value, there is no multicollinearity in the sample (Belsley et al. 1980, p.58; Belsley 1991).

Table 2. Correlation

*, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

	$\Delta Squared\_Return_t$	$\Delta volatility_t$	ΔGSV_t	ΔTV_t	$\Delta Turnover_t$
$\Delta Squared\_Return_t$	1.0000				
$\Delta Volatility_t$	0.4358***	1.0000			
ΔGSV_t	0.0441***	0.0674***	1.0000		
ΔTV_t	0.1355***	0.2034***	0.1166***	1.0000	
$\Delta Turnover_t$	0.3301***	0.4967***	0.1165***	0.2891***	1.0000

4.2 Regressions

To analyze the impact of changes in GSV and TV on changes in turnover and volatility we implement the models of Easley et al. (1996) and DeLong et al. (1990). Therefore we apply a panel fixed effects model with firm and time fixed effects. The estimation equation follows the general form (Brooks 2014):

$$y_{it} = \alpha + \beta x_{it} + \mu_i + \lambda_t + v_{it} \quad . \tag{10}$$

Where y_{it} is the dependent variable at time t for company i. The variable α stands for the time invariant intercept. The coefficient β is a $kx1$ vector and the independent variable x_{it} is a $1xk$ vector. The time fixed effect λ_t varies over time but is constant over the different stocks. The cross-sectional fixed effect

μ_i does not vary over time but varies over the different companies. Time $t = 1, \ldots T$ is a daily measure and $i = 1, \ldots, N$ represents each company which is part of the sample.

The procedure is twofold. First, we measure the impact of changes in GSV and TV on trading activity which we measure by turnover. The variable stands for the trading activities of the 29 DJIA stocks in the sample. The dependent variable in equation (11) is the change in turnover. If the turnover changes, we expect that the number of traders who arrive on the market changes. Following the idea of Easley et al. (1996) that new information leads to new market entries of traders. To measure the effect of new information on investors, we use GSV and TV as proxy variables. We expect that an increase (decrease) in GSV and TV leads to an increase (decrease) in turnover. We quantify this effect by looking at the percentage change in turnover. The control variables are lagged changes in turnover, changes in squared returns and lagged squared returns. Step by step we include changes in GSV (TV) and lagged changes in GSV (TV) to obtain the final equation:

$$\Delta Turnover_{it} = \alpha + \beta_1 \Delta Turnover_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta GSV_{it} + \beta_5 \Delta GSV_{it-1} + \beta_6 \Delta TV_{it} + \beta_7 \Delta TV_{i,t-1} + \mu_i + \lambda_t + v_{it}. \quad (11)$$

Second, we apply the DSSW model (De Long et al. 1990) to measure the impact of changes in GSV and TV on the amount of noise traders on the market. The dependent variable is changes in volatility. A positive (negative) coefficient would lead to an increase (decrease) in volatility. In the sense of the DSSW model (De Long et al. 1990) this would mean that the share of noise traders on the market increases (decreases). We control for changes in lagged volatility, squared returns and lagged squared returns. Step by step we include changes in GSV (TV) and lagged changes in GSV (TV).

$$\Delta Volatility_{it} = \alpha + \beta_1 \Delta Volatility_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta BidAsk_{it} + \beta_5 \Delta BidAsk_{it-1} + \beta_6 \Delta Turnover_{it} + \beta_7 \Delta Turnover_{it-1} + \beta_8 \Delta GSV_{it} + \beta_9 \Delta GSV_{it-1} + \beta_{10} \Delta TV_{it} + \beta_{11} \Delta TV_{it-1} + \mu_i + \lambda_t + v_{it}. \quad (12)$$

Table 3 shows the results of the fixed effects regression (11) on changes in turnover. All coefficients are highly significant at the 1 % level. In column (1) a baseline of control variables is included (lagged changes in turnover, changes in squared returns and lagged changes in squared returns). Column (2) includes changes in GSV, column (3) changes in lagged GSV. Column (4) includes changes in TV, column (5) includes changes in lagged TV. R^2 increases from 49% in column (1) to 56% in column (5). In all five equations we implement time and company fixed effects. To control for cross sectional dependence, heteroscedasticity and autocorrelation we use Driscoll and Kraay standard errors (Driscoll & Kraay 1998).

Our results show that days with high trading activity are followed by days with a decrease in trading activity. A positive change in turnover by 10 % in $t-1$, leads to a decrease in turnover in t by -2.96 %. This is in line with the summary statistics in table 1, which shows that overall the mean of changes in turnover is close to zero. We find that news lead to an increase in turnover. On the same day, a 10 % change in squared returns (lagged squared returns) would lead to an increase of changes in turnover around 0.3 % (0.2%). Despite the fact that we control for news with squared returns, we find that changes in GSV and TV have a positive effect on the trading activity. This means that GSV and TV capture new information which increases the share of traders for the market. An increase in changes of GSV (TV) by 10 % would lead to an increase in turnover on the same day by 0.51 % (1.73 %). An increase in lagged changes of GSV (TV) by 10% would lead to an increase in turnover today by 0.27 % (0.92%). The effect of TV is higher than the effect of GSV which can be interpreted as an indicator for the lead-lag structure (Mao et al. 2015). This is in line with the correlation we find in Table 2 and the study by Mao et al. (2015). We conclude that TV is of high relevance for the trading activity of investors.

Table 3. Market entry

Column (1) to (5) report the result on market entry: $\Delta Turnover_{it} = \alpha + \beta_1 \Delta Turnover_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta GSV_{it} + \beta_5 \Delta GSV_{it-1} + \beta_6 \Delta TV_{it} + \beta_7 \Delta TV_{i,t-1} + \mu_i + \lambda_t + v_{it}$. Column (1) includes the control variables changes in lagged turnover, changes in squared returns and changes in lagged squared returns. In column (2) the variable GSV is added. Column (3) adds the lagged change in GSV. Column (4) adds the change in TV. Column (5) adds the lagged change in TV. All regressions are Panel fixed effects regression, including time and company fixed effects. The regressions are estimated with Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

VARIABLES	(1) $\Delta Turnover_t$	(2) $\Delta Turnover_t$	(3) $\Delta Turnover_t$	(4) $\Delta Turnover_t$	(5) $\Delta Turnover_t$
$\Delta Turnover_{t-1}$	-0.296***	-0.294***	-0.299***	-0.287***	-0.341***
	(-36.874)	(-36.253)	(-38.043)	(-35.482)	(-50.660)
$\Delta Squared\_Return_t$	0.038***	0.038***	0.038***	0.033***	0.031***
	(30.039)	(30.508)	(30.712)	(32.701)	(32.837)
$\Delta Squared\_Return_{t-1}$	0.020***	0.020***	0.019***	0.018***	0.016***
	(22.976)	(22.933)	(22.802)	(23.756)	(22.916)
ΔGSV_t		0.071***	0.095***	0.066***	0.051***
		(8.467)	(9.152)	(8.011)	(6.627)
ΔGSV_{t-1}			0.049***	0.045***	0.027***
			(6.458)	(6.316)	(4.068)
ΔTV_t				0.147***	0.173***
				(23.162)	(24.356)
ΔTV_{t-1}					0.092***
					(18.783)
Constant	0.060***	0.049***	0.047***	-0.051***	0.023***
	(38.549)	(21.977)	(18.724)	(-10.252)	(6.232)
Time Fixed Effects	YES	YES	YES	YES	YES
Company Fixed Effects	YES	YES	YES	YES	YES
R-squared	0.492	0.495	0.497	0.545	0.560
N	25,566	24,970	24,833	24,811	24,801

Table 4 represents the result of the panel fixed effect regression on changes in volatility (12). Column (1) includes the control variables changes in lagged volatility, turnover, lagged turnover, squared returns and lagged squared returns. Column (2) includes the changes in GSV and column (3) the lagged changes in GSV. Column (4) includes the changes in TV and column (5) the changes in lagged TV. We find that an increase in yesterday's volatility leads to a decrease in volatility today. Comparable to the results for changes in turnover in table 3, we observe a mean reversion trend for volatility. As expected by the correlation, the coefficients of changes in turnover and lagged turnover are always positive and highly significant at the 1 % level. This means that an increase in trading activity leads to higher volatility on the market. The impact on the same day is higher than the impact from the previous day. The same holds for news measured by squared returns. An increase in news on the market leads to higher changes in volatility. The impact on the same day (0.047) is higher than the impact from the previous day (0.022). If we include the changes in GSV and TV as proxy variables of new information in column (2) to (5), we find that only the effect of TV is statistically significant. The coefficients of changes in GSV and lagged GSV are not significant. Thus, there is no measurable impact on changes in volatility. Like Mao et al. (2015) we find that TV is more important for financial markets than Twitter. Changes in TV have a positive effect on changes in volatility on the same day (0.04) and the day before (0.02), allowing for one day ahead volatility forecasts. With TV we measure new information from investors that comes on the market. The share of noise traders increase (decreases) if TV increases (decreases). TV allows us to measure investor sentiment. We expect that this deviation from the mean is only temporary (Mao et al. 2015). Arbitrageurs and rational investors bring the changes in volatility back to zero but as they face the risk to trade against noise traders it takes a while (De Long et al. 1990; Shleifer & Vishny 1997).

Table 4. Share of noise trader

Column (1) to (5) report the result on market entry: $\Delta Volatility_{it} = \alpha + \beta_1 \Delta Volatility_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta Turnover_{it} + \beta_5 \Delta Turnover_{it-1} + \beta_6 \Delta GSV_{it} + \beta_7 \Delta GSV_{it-1} + \beta_8 \Delta TV_{it} + \beta_9 \Delta TV_{it-1} + \mu_i + \lambda_t + \upsilon_{it}$. . Column (1) includes the control variables changes in lagged turnover, changes in squared returns and changes in lagged squared returns. In Column (2) the variable GSV is added. Column (3) adds the lagged change in GSV. Column (4) adds the change in TV. Column (5) adds the lagged changes in TV. All regressions are Panel fixed effects regression, including time and company fixed effects. The regressions are estimated with Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

VARIABLES	(1) ΔVolatility$_t$	(2) ΔVolatility$_t$	(3) ΔVolatility$_t$	(4) ΔVolatility$_t$	(5) ΔVolatility$_t$
ΔVolatility$_{t-1}$	-0.474***	-0.474***	-0.474***	-0.473***	-0.475***
	(-84.207)	(-82.124)	(-82.114)	(-82.277)	(-82.730)
ΔTurnover$_t$	0.571***	0.574***	0.574***	0.550***	0.543***
	(44.500)	(44.194)	(44.198)	(41.829)	(40.157)
ΔTurnover$_{t-1}$	0.250***	0.250***	0.250***	0.245***	0.232***
	(26.312)	(25.718)	(25.435)	(25.257)	(21.979)
ΔSquared_Return$_t$	0.047***	0.047***	0.047***	0.047***	0.047***
	(35.853)	(35.259)	(35.622)	(35.424)	(35.283)
ΔSquared_Return$_{t-1}$	0.022***	0.023***	0.022***	0.022***	0.022***
	(22.541)	(22.030)	(22.047)	(22.186)	(22.020)
ΔGSV$_t$		-0.004	-0.002	-0.007	-0.010
		(-0.616)	(-0.243)	(-0.862)	(-1.213)
ΔGSV$_{t-1}$			0.004	0.004	0.001
			(0.548)	(0.525)	(0.069)
ΔTV$_t$				0.036***	0.043***
				(7.560)	(8.386)
ΔTV$_{t-1}$					0.020***
					(3.720)
Constant	0.106***	0.106***	0.106***	-0.136***	-0.024***
	(53.016)	(45.415)	(44.074)	(-42.267)	(-4.053)
Time Fixed Effects	YES	YES	YES	YES	YES
Company Fixed Effects	YES	YES	YES	YES	YES
R-squared	0.593	0.593	0.594	0.595	0.596
N	25,566	24,970	24,833	24,811	24,801

Our results are unique, as we are the first to compare the influence of changes in GSV and TV in a panel data setting for the DJIA index stocks. We would have expected that both, GSV and TV, have a positive influence on changes in turnover and on changes in volatility. The expectations hold for turnover. We find an impact of changes in GSV and TV on changes in turnover. In the sense of Easley et al. (1996) we find that information in GSV and TV lead to market entries. The effect is more pronounced for TV than for GSV. Furthermore, we find that both have predictive power on turnover. The coefficients are smaller but still significant from one day to the next. For volatility we find a significant impact of changes in TV on volatility but no impact of GSV. In the sense of De Long et al. (1990) we find that TV leads to an increase of the share of noise traders on the market. This effect is observable for the same day and the next, indicating the predictive power of TV on volatility.

5 Robustness tests

We modify our standard approach to test if our results are consistent. First, we confirm the choice of a panel data model with fixed effects. Second, we include bid-ask spreads as liquidity indicator for the market. Third, we look at the impact of GSV and TV on high and low trading volume. Fourth, we consider the different industry sectors in more detail. Fifth, we run our regressions on changes in turnover and volatility using the procedure by Arellano and Bond (1991). Overall we find that our results are consistent.

5.1 Fitting of the model

We use a panel data model with time and company fixed effects with Dirscoll and Karay standard errors. To validate our model we conduct several tests. The Hausman's specification test confirms[10] that a fixed effect model with company and time fixed effects is more appropriate for our data than a random effects model. The necessity of time fixed effects is confirmed by a significant Wald test. Moreover, we test for cross-sectional dependence. Therefore we conduct the Breusch-Pagan Lagrange Multiplier test of independence and the Parsan test for cross-sectional dependence. Our results show that we have cross sectional dependence, meaning that the residuals are correlated across the different companies. To overcome this problem we apply Driscoll and Kraay standard errors (Driscoll & Kraay 1998)[11]. Further, we find that we have heteroscedasticity, meaning that our variance is not constant. We can solve this problem by using robust standard errors[12]. Here we can use Huber and White robust standard errors, or the Driscoll and Kraay standard errors. We run our estimations with Driscoll and Kraay, Huber and White and clustered standard errors. Overall the coefficients and the significance levels stay the same, only the constant changes.[13]

In Table 5 we compare different ways to estimate the standard equation (11) for turnover including all control variables. In column (1) we use Huber and white standard errors. In column (2) we use Driscoll and Kraay standard errors. We find that our coefficients do not change. The standard errors in column for Huber and White in column (1) are smaller than for Driscoll and Kraay. The main difference is the constant, it is larger for the robust standard errors in in column (1).

Table 5. Market entry with different standard errors

Column (1) and (2) report the result on market entry: $\Delta Turnover_{it} = \alpha + \beta_1 \Delta Turnover_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta GSV_{it} + \beta_5 \Delta GSV_{it-1} + \beta_6 \Delta TV_{it} + \beta_7 \Delta TV_{i,t-1} + \mu_i + \lambda_t + v_{it}$. In column (1) we use Huber and White robust standard errors. In column (2) we use Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

	(1)	(2)
VARIABLES	ΔTurnover$_t$	ΔTurnover$_t$

[10] (Hausman, J. A. 1978. Specification tests in econometrics. Econometrica 46: 1251–1271.)
[11] Hoechle, Daniel, "Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence
[12] Huber and White
[13] We do not report the results of the clustered standard errors on the company level, as they lead to the same results as the Huber and White standard errors.

ΔTurnover$_{t-1}$	-0.341***	-0.341***
	(-58.173)	(-50.660)
ΔSquared_Return$_t$	0.031***	0.031***
	(27.549)	(32.837)
ΔSquared_Return$_{t-1}$	0.016***	0.016***
	(18.389)	(22.916)
ΔGSV$_t$	0.051***	0.051***
	(3.998)	(6.627)
ΔGSV$_{t-1}$	0.027**	0.027***
	(3.231)	(4.068)
ΔTV$_t$	0.173***	0.173***
	(20.317)	(24.356)
ΔTV$_{t-1}$	0.092***	0.092***
	(11.838)	(18.783)
Constant	0.144**	0.023***
	(3.445)	(6.232)
Time Fixed effects	YES	YES
Company Fixed effects	YES	YES
R-squared	0.560	0.560
N	24,801	24,801

In Table 6 we look at regression (12) on changes in volatility including all control variables. In column (1) we use Huber and White standard errors. In column (2) we use Driscoll and Kraay standard errors. Again the coefficients do not change. For lagged changes the significance level differs, as the standard error is higher in column (2) than in column (1). The constant is positive for the Huber and White standard error and close to zero but negative for the Driscoll and Kraay standard error.

Table 6. Share of noise trader with different standard errors

Column (1) and (2) report the result on market entry: $\Delta Volatility_{it} = \alpha + \beta_1 \Delta Volatility_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta BidAsk_{it} + \beta_5 \Delta BidAsk_{it-1} + \beta_6 \Delta Turnover_{it} + \beta_7 \Delta Turnover_{it-1} + \beta_8 \Delta GSV_{it} + \beta_9 \Delta GSV_{it-1} + \beta_{10} \Delta TV_{it} + \beta_{11} \Delta TV_{it-1} + \mu_i + \lambda_t + v_{it}$. In column (1) we use Huber and White robust standard errors. In column (2) we use Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

	(1)	(2)
VARIABLES	ΔVolatility$_t$	ΔVolatility$_t$
ΔVolatility$_{t-1}$	-0.475***	-0.475***
	(-88.948)	(-82.730)
ΔTurnover$_t$	0.543***	0.543***
	(31.395)	(40.157)
ΔTurnover$_{t-1}$	0.232***	0.232***
	(20.263)	(21.979)
ΔSquared_Return$_t$	0.047***	0.047***
	(29.859)	(35.283)
ΔSquared_Return$_{t-1}$	0.022***	0.022***
	(23.252)	(22.020)
ΔGSV$_t$	-0.010	-0.010
	(-1.093)	(-1.213)
ΔGSV$_{t-1}$	0.001	0.001
	(0.066)	(0.069)

ΔTV_t	0.043***	0.043***
	(7.180)	(8.386)
ΔTV_{t-1}	0.020**	0.020***
	(3.511)	(3.720)
Constant	0.241***	-0.024***
	(5.156)	(-4.053)
Time Fixed effects	YES	YES
Company Fixed effects	YES	YES
R-squared	0.596	0.596
N	24,801	24,801

5.2 Liquidity on the market

Another indicator for movements on stock markets is the market liquidity. Originally the Corwin and Schulz (2012) approach is a bid-ask spread estimator from daily high and low prices.[14] The idea behind that estimator is that the high-low ratio reflects both the stock's variance and its bid-ask spread. The market is more (less) liquid on days with low (high) spreads. We calculate the spread measure following Corwin and Schulz (2012) and compute the percentage change of the spread.[15]

In Table 7 the summary statistic on spread in Panel A shows that the mean change is close to zero. In Panel B we report the correlation between changes in spreads and other variables. An increase in spreads is negatively connected with changes in squared returns, volatility, GSV, TV and turnover. For turnover this means that an increase in the spread, thus a wider bid-ask spread, is correlated with a decrease in turnover. Wider spreads lead to less liquid markets which leads to a decrease in turnover and volatility.

Table 7 Descriptive statistic on spreads

*, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

Panel A							
VARIABLES	Mean	Stan. Dev.	Skewness	Kurtosis	Min	Max	N
$\Delta Spread_t$	0.0258852	1.26125	0.0011681	5.800737	-8.502703	9.658206	16076
Panel B							
VARIABLES	$\Delta Squared\_Return_t$	$\Delta Volatility_t$	ΔGSV_t	ΔTV_t	$\Delta Turnover_t$	$Spread_t$	
$\Delta Spread_t$	-0.0748***	-0.2536***	-0.0171	-0.0298***	-0.0218***	1.000	

In table 8 we include the changes in spreads in regression (12) on volatility. We find that the impact of spreads, as a proxy for market liquidity, on volatility is negative and significant at the 1% level. This holds for the same day (-0.057) and the day before (-0.055). On less liquid markets, we observe that people trade less and that this leads to a decrease in volatility on the market. The number of observations decreases to around 10,000 as we have more missing observations for spreads. The R^2 is at 63 %.

Table 8. Share of noise trader and market liquidity

Column (1) to (5) report the result on market entry: $\Delta Volatility_{it} = \alpha + \beta_1 \Delta Volatility_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta BidAsk_{it} + \beta_5 \Delta BidAsk_{it-1} + \beta_6 \Delta Turnover_{it} + \beta_7 \Delta Turnover_{it-1} + \beta_8 \Delta GSV_{it} + \beta_9 \Delta GSV_{it-1} + \beta_{10} \Delta TV_{it} + \beta_{11} \Delta TV_{it-1} + \mu_i + \lambda_t + v_{it}$.
Column (1) includes the control variables changes in lagged turnover, changes in squared returns, changes in lagged squared returns, as well as changes in spreads and lagged spreads. In Column (2) the variable GSV is added.

[14] An example to calculate the bid ask spread is provided on Corwin's homepage https://www3.nd.edu/~scorwin/

[15] We also compute the zero spread but due to less observation we do not consider it here.

Column (3) adds the lagged change in GSV. Column (4) adds the change in TV. Column (5) adds the lagged changes in TV. All regressions are panel fixed effects regression, including time and company fixed effects. The regressions are estimated with Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

	(1) $\Delta Volatility_t$	(2) $\Delta Volatility_t$	(3) $\Delta Volatility_t$	(4) $\Delta Volatility_t$	(5) $\Delta Volatility_t$
$\Delta Volatility_{t-1}$	-0.440***	-0.440***	-0.440***	-0.439***	-0.442***
	(-36.924)	(-36.270)	(-36.039)	(-35.792)	(-36.626)
$\Delta Turnover_t$	0.538***	0.541***	0.543***	0.518***	0.506***
	(27.613)	(27.060)	(26.933)	(24.774)	(23.741)
$\Delta Turnover_{t-1}$	0.229***	0.231***	0.233***	0.228***	0.208***
	(14.500)	(14.551)	(14.521)	(14.155)	(12.001)
$\Delta Squared\_Return_t$	0.045***	0.045***	0.045***	0.045***	0.044***
	(22.043)	(21.470)	(21.541)	(21.529)	(21.487)
$\Delta Squared\_Return_{t-1}$	0.023***	0.023***	0.023***	0.023***	0.023***
	(15.745)	(15.465)	(15.368)	(15.293)	(15.242)
$\Delta Spread_t$	-0.058***	-0.058***	-0.058***	-0.057***	-0.057***
	(-10.178)	(-10.116)	(-10.027)	(-9.908)	(-9.918)
$\Delta Spread_{t-1}$	-0.056***	-0.056***	-0.056***	-0.055***	-0.055***
	(-15.534)	(-15.394)	(-15.226)	(-15.157)	(-15.174)
ΔGSV_t		0.002	0.002	-0.003	-0.007
		(0.177)	(0.158)	(-0.212)	(-0.480)
ΔGSV_{t-1}			-0.003	-0.002	-0.007
			(-0.237)	(-0.189)	(-0.631)
ΔTV_t				0.041***	0.052***
				(4.955)	(5.863)
ΔTV_{t-1}					0.033***
					(4.105)
Constant	-0.500***	0.333***	-0.497***	0.340***	-0.492***
	(-20.458)	(9.615)	(-20.013)	(9.772)	(-20.077)
Time Fixed Effects	YES	YES	YES	YES	YES
Company Fixed Effects	YES	YES	YES	YES	YES
R-squared	0.633	0.633	0.633	0.634	0.635
N	10,297	10,068	10,008	9,995	9,993

5.3 High and low trading volume

We assume that GSV and TV are proxies for new information on the market taken into account by investors. We find that both affect trading activities on the market and at least TV has an impact on the share of noise traders on the market which we cannot proof for GSV. Thus, we assume to measure investor sentiment. We expect that especially in times of high trading volume Google and Twitter are important[16]. Assuming that lower (higher) trading volume stands for less (more) public attention. To see the impact of GSV and TV on the trading activity, we create 5 groups out of the 29 DJIA stocks. We perform our standard regression on changes in turnover (11) for the five different trading volume groups. In table 9 we report the results for the top 20% and lowest 20% of observations according to trading volume.

Our results show that the impact of changes in GSV and TV on changes in turnover is higher (lower) for the high (low) volume stocks. The impact of changes in GSV and lagged GSV is only significant for the high volume stocks. Thus, only when stocks have a certain trading volume GSV has an impact. TV has a significant impact at the 1% level on low volume and high volume stocks on the same day and the

[16] Dimpfl z.B. und weitere Studien

day after. However, the impact is higher for the top 20% (0.246) than for the lowest 20% (0.062). We can confirm that Google and Twitter are different and that the impact of Twitter on financial markets is more pronounced than the impact of Google.

According to Easley et al. (1996) the probability of informed trading is higher (lower) among low (high) volume stocks. Based on this idea and knowing that we measure noise traders with changes in TV, we assume that more informed traders trade the low volume stocks, as we see smaller or no impact of TV and GSV. For high volume stocks we assume to observe less informed traders but more noise traders.

Table 9 Market entry with different standard errors

Column (1) and (2) report the result on market entry: $\Delta Turnover_{it} = \alpha + \beta_1 \Delta Turnover_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta GSV_{it} + \beta_5 \Delta GSV_{it-1} + \beta_6 \Delta TV_{it} + \beta_7 \Delta TV_{i,t-1} + \mu_i + \lambda_t + v_{it}$. In column (1) we use Huber and White robust standard errors. In column (2) we use Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

VARIABLES	(1) 20% lowest volume $\Delta Turnover_t$	(2) 20% highest volume $\Delta Turnover_t$
$\Delta Turnover_{t-1}$	-0.368***	-0.362***
	(-25.040)	(-21.535)
$\Delta Squared\_Return_t$	0.019***	0.035***
	(9.672)	(15.954)
$\Delta Squared\_Return_{t-1}$	0.011***	0.018***
	(6.34)	(9.378)
ΔGSV_t	0.009	0.097***
	(0.66)	(5.45)
ΔGSV_{t-1}	0.001	0.044**
	(0.067)	(2.618)
ΔTV_t	0.062***	0.246***
	(6.134)	(17.594)
ΔTV_{t-1}	0.042***	0.134***
	(4.999)	(10.819)
Constant	0.096***	0.357***
	(5.775)	(10.632)
Time Fixed Effects	YES	YES
Company Fixed Effects	YES	YES
R-squared	0.66	0.641
N	4,736	5,066

5.4 Industry and sector groups

We classify the different stocks into industry and 9 sector groups according to the profile from Yahoo finance[17][18]. The sectors are Industrials (17%), Technology (17%), Consumer Cyclicals (14%), Financial Services (14%), Consumer Defensive (10%), Healthcare (10%), Communication services (7%), Energy (7%) and Basic Materials (4%). Running our standard panel fixed effects regression on changes in turnover (11) and volatility (12) with time, company fixed effects and sector fixed effects or time and sector fixed effects, does not lead to significant results. Here the sectors have no impact. Running our regressions for all nine sectors separately shows the different impact on changes in turnover and

[17] https://finance.yahoo.com/quote/A/profile?p=A (30.8.2018, example Apple)

[18] We only look at the effects of the nine different sectors as there are 24 industry groups

volatility. We find that the coefficients of the control variables are comparable to what we find in table 3 and 4. For changes in GSV and TV the results are different. First, the results do not turn out significant for all sectors. Second, for some sectors the sign of the coefficient changes. Although, the results are better for changes in TV than for changes in GSV. One explanation is that not all sector have the same media coverage on Google and Twitter. The tables can be found in the appendix.

5.5 Arellano Bond

We look at the changes in turnover and volatility. One control variable we use is the lagged dependent variable. This can lead to first order autocorrelation and endogeneity problems (Roodman 2006). Thus, we estimate our standard regressions on changes in turnover (11) and volatility (12) applying the approach suggested by Arellano and Bond (1991; Roodman 2006). We restrict the set of independent control variables to the lag of the dependent variable, changes in GSV and TV at time t. Our results are presented in table 10. In column (1) we present the results of regression (11) on changes in turnover. Here we find that that the lagged changes in turnover have a negative and significant impact. The impact of changes in GSV and TV remain positive and significant. Without controlling for squared returns, the impact of GSV is higher than of TV. To confirm the validity of the instruments we use the Sargan test and the Hansen test. We perform the overconfidence test by Sargan (0.016) and Hansen (0.250). Both tests confirm that we have valid instruments. In column (2) we present the results of regression (12). We find that an increase in yesterday's volatility leads to a decrease of volatility today. Moreover, we find that the impact of GSV is weaker (0.097) than the impact of TV (0.246) but both remain highly significant.

Table 10. Arellano Bond for market entries and noise trader

Equation of Arellano Bond!!!! For turnover and volatility.

VARIABLES	(1) $\Delta Turnover_t$	(2) $\Delta Volatility_t$
$\Delta Turnover_{t-1}$	-0.1581***	
	(-7.34)	
$\Delta Volatility_{t-1}$		-0.2884***
		(-15.66)
ΔGSV_t	0.8247**	0.097***
	(2.19)	(5.45)
ΔTV_t	0.7160***	0.246***
	(7.53)	(17.594)
Constant	-0.0006**	-0.0008
	(-2.09)	()
Arellano Bond test for AR(1)	0.0000	0.000
Arellano Bond test for AR(2)	0.002	0.049
Sargan test	0.016	0.096
Hansen test	0.250	0.271
N	25,494	25,494

6 Critical reflection

So far we discussed the influence of GSV and TV on turnover and volatility. The rise of this new information source has also weaknesses which cannot be neglected. We take the information we obtain as a sources of information to measure investor attention and find a noise proxy for retail investor behavior but we have to ask ourselves if the information is not manipulated somehow. This emphasizes the necessity to critically address the question of data manipulation.

Generally speaking data manipulation is not a new phenomenon. In their paper on market manipulation Leinweber and Madhavan (2001) describe the history of 300 years of market manipulation. While the attributes for a successful manipulation are still access to media, anonymity, scalability, time and impact the technical opportunities developed. The game is on a new level with new technological opportunities. In earlier days not everyone had access to information and not everyone could publish information. Newspapers and special finance magazines were the medium of interest when it came to spreading rumors. Nowadays the internet can connect everyone and leads to a larger audience. Different trading strategies, such as pump-and-dump or bluffing, attract especially uninformed retail investors while the informed investors leave the game at the peak with high profits. The SEC has even an own office dealing with this kind of market manipulation and insider trading. They inform investors in their regular alerts on social media about fraud and stock rumors (SEC 2015; SEC 2014)

It is not clear to what extend Google and Twitter data is affected. In 2016 the most popular social media scams were dating and romance. The numbers show further that the reported loss due to scams is increasing (Commission 2017). According to Gerth (2017) market manipulation with Twitter often miss the mark. There are more than one million stock relevant tweets on twitter per day and manipulation is easy as a fake account is sufficient to commit fraud. Instead of human beings bots can manage accounts as well. Looking at market manipulation, especially short selling attacks are a problem. There are examples but no proof that the whole market is affected and to which extent. A recent report by the Federal Financial Supervisory Authority (Gillert 2017) discussed the topic of short selling attacks or too positive mailings. They pointed out that wrong information no matter if positive or negative can lead to reputational damages. Retail investors have to verify information to protect themselves.

We can conclude that at the moment mail and phone are more affected than social media (Commission 2017). Hence, we do not have to adjust for Twitter or Google manipulation in our data. But individual investors need to be aware as well as the regulation environment for the years to come.

7 Conclusion

This paper proposes a measure for market entry and noise trading. The purpose of this paper is to show that GSV and TV incorporate new investor information to measure investor sentiment. First, we show that changes GSV and TV measure the market entry of traders. Second, changes in TV influence the share of noise traders on the market and measure investor sentiment. For GSV the results is not significant. Moreover, GSV and TV have predictive power on a daily level. Our results are robust to various test. Further, the impact of Twitter on financial markets is more important than the impact of Google.

Appendix

List of the stocks in the sample

Ticker	Historical index consituents	GSV	TW	RV
AAPL				
ARNC		NV	NV	NV
AXP				
BA				
BAC		NV	NV	NV
CAT				
CSCO				
DVX				
DD				
DIS				
GE				
GS				
HD				
HPQ		NV	NV	NV
IBM				
INTC				
JMP				NV
JNJ				
KO				
MCD				
MMM				
MRK				
MSFT				
NKE				
PFE				
PG				
T				
TRV				
UNHP				NV
UTX				
V				
VZ				
WMT				
XOM				
Sum	34	31	31	29

Table 9. Market entry per sector

Column (1) to (9) report the result on market entry: $\Delta Turnover_{it} = \alpha + \beta_1 \Delta Turnover_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta GSV_{it} + \beta_5 \Delta GSV_{it-1} + \beta_6 \Delta TV_{it} + \beta_7 \Delta TV_{i,t-1} + \mu_i + \lambda_t + v_{it}$. Column (1) reports the results for the sector Basic Materials. Column (2) stands for the sector Communication Services. Column (3) stands for the sector Consumer Cyclical. Column (4) stands for the sector Consumer Defensive. Column (5) stands for the sector Energy. Column (6) stands for the sector Financial Services. Column (7) stands for the sector Healthcare. Column (8) stands for the sector Industrials and column (9) for the sector Technology. All regressions are Panel fixed effects regression, including time and company fixed effects. The regressions are estimated with Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

	1	2	3	4	5	6	7	8	9
	Basic Materials[19]	Communication Servics	Consumer Cyclical	Consumer Defensive	Energy	Financial Services	Healthcare	Industrials	Technology
VARIABLES	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$	$\Delta Turnover_t$
$\Delta Turnover_{t-1}$	-0.210***	-0.373***	-0.322***	-0.313***	-0.376***	-0.345***	-0.380***	-0.317***	-0.366***
	(-1.38e+13)	(-8.238)	(-15.558)	(-12.766)	(-9.783)	(-17.249)	(-13.634)	(-16.754)	(-19.965)
$\Delta Squared\_Return_t$	-0.019***	0.028***	0.028***	0.031***	0.014***	0.029***	0.026***	0.031***	0.032***
	(-3.18e+12)	(5.123)	(10.529)	(9.615)	(4.278)	(9.788)	(7.920)	(11.517)	(14.365)
$\Delta Squared\_Return_{t-1}$	0.034***	0.011*	0.012***	0.012***	0.004	0.016***	0.017***	0.019***	0.017***
	(6.18e+12)	(2.266)	(5.326)	(3.897)	(0.972)	(6.109)	(6.024)	(7.742)	(9.132)
ΔGSV_t	-0.245***	0.030	0.127**	-0.012	-0.004	0.038	0.029	0.063**	0.137***
	(-1.07e+13)	(0.587)	(3.272)	(-0.608)	(-0.127)	(1.132)	(1.823)	(3.194)	(6.111)
ΔGSV_{t-1}	-0.240***	0.020	0.086*	-0.001	0.006	0.012	0.019	0.030	0.054*
	(-8.87e+12)	(0.345)	(2.242)	(-0.058)	(0.200)	(0.335)	(1.222)	(1.700)	(2.411)
ΔTV_t	0.643***	0.121***	0.196***	0.140***	0.071***	0.132***	0.129***	0.197***	0.228***
	(1.98e+13)	(4.051)	(13.240)	(7.384)	(3.308)	(7.282)	(8.064)	(10.618)	(14.381)
ΔTV_{t-1}	0.052***	0.089***	0.101***	0.066***	0.040*	0.064***	0.055***	0.088***	0.163***
	(3.62e+12)	(4.203)	(8.811)	(4.392)	(1.987)	(4.393)	(4.318)	(7.979)	(10.743)
Constant	0.241***	-0.093	-0.260***	-0.167***	-0.233***	0.255***	-0.071	-0.010	-0.932***
	(1.14e+13)	(-1.669)	(-27.439)	(-6.335)	(-7.871)	(13.280)	(-1.757)	(-0.501)	(-17.693)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Company Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	1.000	0.866	0.655	0.728	0.876	0.659	0.739	0.671	0.698
N	865	1,760	3,553	2,560	1,772	3,553	2,604	3,757	4,377

[19] Only one company

Table 10. Share of noise trader per sector

Column (1) to (9) report the result on market entry: $\Delta Volatility_{it} = \alpha + \beta_1 \Delta Volatility_{it-1} + \beta_2 \Delta SquaredReturn_{it} + \beta_3 \Delta SquaredReturn_{it-1} + \beta_4 \Delta Turnover_{it} + \beta_5 \Delta Turnover_{it-1} + \beta_6 \Delta GSV_{it} + \beta_7 \Delta GSV_{it-1} + \beta_8 \Delta TV_{it} + \beta_9 \Delta TV_{it-1} + \mu_i + \lambda_t + v_{it}$. Column (1) reports the results for the sector Basic Materials. Column (2) stands for the sector Communication Services. Column (3) stands for the sector Consumer Cyclical. Column (4) stands for the sector Consumer Defensive. Column (5) stands for the sector Energy. Column (6) stands for the sector Financial Services. Column (7) stands for the sector Healthcare. Column (8) stands for the sector Industrials and column (9) for the sector Technology. All regressions are Panel fixed effects regression, including time and company fixed effects. The regressions are estimated with Driscoll and Kraay standard errors. *, **, *** indicate statistical significance at the 10%, 5% or 1% levels respectively.

	1	2	3	4	5	6	7	8	9
	Basic Materials[20]	Communication Servics	Consumer Cyclical	Consumer Defensive	Energy	Financial Services	Healthcare	Industrials	Technology
VARIABLES	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$	$\Delta Volatility_t$
$\Delta Volatility_{t-1}$	-0.169***	-0.482***	-0.451***	-0.489***	-0.445***	-0.482***	-0.478***	-0.501***	-0.443***
	(-1.98e+12)	(-12.745)	(-26.322)	(-20.017)	(-12.512)	(-25.126)	(-23.527)	(-30.545)	(-29.440)
$\Delta Turnover_t$	0.094***	0.531***	0.508***	0.515***	0.511***	0.484***	0.528***	0.572***	0.603***
	(7.93e+11)	(5.903)	(12.531)	(9.482)	(5.522)	(14.953)	(9.321)	(18.290)	(18.730)
$\Delta Turnover_{t-1}$	0.091***	0.275***	0.207***	0.234***	0.126	0.222***	0.233***	0.230***	0.222***
	(6.65e+11)	(3.311)	(5.567)	(4.990)	(1.239)	(7.573)	(4.658)	(7.738)	(6.673)
$\Delta Squared\_Return_t$	-0.029***	0.047***	0.048***	0.045***	0.044***	0.050***	0.055***	0.036***	0.042***
	(-2.41e+12)	(5.892)	(14.627)	(9.324)	(6.420)	(13.658)	(9.943)	(10.435)	(15.913)
$\Delta Squared\_Return_{t-1}$	-0.014***	0.016*	0.020***	0.023***	0.022*	0.022***	0.022***	0.019***	0.023***
	(-1.88e+12)	(2.195)	(6.731)	(4.648)	(2.554)	(6.235)	(4.519)	(5.931)	(7.703)
ΔGSV_t	0.004***	-0.098	-0.103*	0.007	-0.014	0.019	-0.024	0.020	-0.038
	(9.67e+10)	(-1.447)	(-2.289)	(0.271)	(-0.272)	(0.409)	(-1.086)	(1.002)	(-1.379)
ΔGSV_{t-1}	0.218***	-0.013	-0.040	-0.006	-0.002	0.051	0.003	0.003	-0.010
	(3.88e+12)	(-0.162)	(-0.904)	(-0.196)	(-0.038)	(1.179)	(0.126)	(0.178)	(-0.320)
ΔTV_t	0.424***	0.019	0.049**	0.030	0.063	0.052***	0.016	0.070***	0.007
	(5.02e+12)	(0.611)	(2.868)	(1.478)	(1.800)	(3.330)	(0.768)	(4.373)	(0.474)
ΔTV_{t-1}	-0.089***	0.013	0.034	0.021	0.027	0.032*	0.009	0.051***	-0.015
	(-1.62e+12)	(0.364)	(1.885)	(1.013)	(0.888)	(2.163)	(0.468)	(3.583)	(-0.825)
Constant	-0.079***	0.329***	0.218***	-0.164***	-0.135*	0.106***	0.047	-0.032	-0.273***
	(-2.47e+12)	(3.938)	(12.713)	(-3.348)	(-2.271)	(3.892)	(0.851)	(-1.582)	(-4.584)
Time Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
Company Fixed Effects	YES	YES	YES	YES	YES	YES	YES	YES	YES
R-squared	1.000	0.879	0.689	0.739	0.856	0.721	0.762	0.726	0.667
N	865	1760	3553	2560	1772	3553	2604	3757	4377

[20] Only one company

Literature

Abhyankar, A.H., 1995. Return and volatility dynamics in the FT-SE 100 stock index and stock index futures markets. *Journal of Futures Markets*, 15(4), pp.457–488. Available at: http://doi.wiley.com/10.1002/fut.3990150405.

Alexander, R.M. & Gentry, J.K., 2014. Using social media to report financial results. *Business Horizons*, 57(2), pp.161–167. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0007681313001808.

Andersen, T.G. et al., 2003. Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2), pp.579–625. Available at: http://doi.wiley.com/10.1111/1468-0262.00418.

Andersen, T.G. & Bollerslev, T., 1998. Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39(4), p.885. Available at: http://www.jstor.org/stable/2527343?origin=crossref.

Andrei, D. & Hasler, M., 2015. Investor attention and stock market volatility. *Review of Financial Studies*, 28(1), pp.33–72.

Antweiler, W. & Frank, M.Z., 2004. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance*, 59(3), pp.1259–1294. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2004.00662.x.

Arellano, M. & Bond, S., 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies*.

Bain, B. & Mott, G., 2018. Can Elon Musk Tweet That? The SEC Is Digging In. Available at: https://www.bloomberg.com/news/articles/2018-08-07/can-elon-musk-tweet-that-the-sec-may-have-an-opinion-quicktake.

Baker, M. & Wurgler, J., 2006. Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, 61(4), pp.1645–1680. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2006.00885.x.

Baker, M. & Wurgler, J., 2007. Investor Sentiment in the Stock Market. *Journal of Economic Perspective*, 21(2), pp.129–151.

Barber, Brad M; Odean, T., 2008. All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), pp.785–818. Available at: https://academic.oup.com/rfs/article/21/2/785/1607197.

Belsley, D.A., 1991. A Guide to using the collinearity diagnostics. *Computer Science in Economics and Management*, 4(1), pp.33–50.

Belsley, D.A., Kuh, E. & Welsch, R.E., 1980. *Regression Diagnostics*, New York.

Black, F., 1986. Noise. *The Journal of Finance*, 41(3), pp.528–543.

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

Brooks, C., 2014. *Introductory Econometrics for Finance 3rd edition* 3rd ed., Available at: www.cambridge.org/9781107661455.

Brooks, C., 1998. Predicting stock index volatility: can market volume help? *Journal of Forecasting*, 17(1), pp.59–80. Available at: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1099-131X(199801)17:1%3C59::AID-FOR676%3E3.0.CO;2-H/epdf.

Brooks, C., Rew, A.G. & Ritson, S., 2001. A trading strategy based on the lead–lag relationship between the spot index and futures contract for the FTSE 100. *International Journal of Forecasting*, 17(1), pp.31–44. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0169207000000625.

Carr, D., 2013. Using Twitter to move the markets. The New York Times. Available at: http://www.nytimes.com/2013/10/07/business/media/using-twitter-to-move-the-markets.html?pagewanted=all&_r=0.

Choi, H. & Varian, H., 2012. Predicting the Present with Google Trends. *Economic Record*, 88(SUPPL.1), pp.2–9.

Commission, A.C. and C., 2017. *Targeting scams - Report of the ACCC on scams activity 2016*, Available at: https://www.accc.gov.au/publications/targeting-scams-report-on-scam-activity/targeting-scams-report-of-the-accc-on-scam-activity-2016.

Corwin, S.A. & Schultz, P., 2012. A Simple Way to Estimate Bid-Ask Spreads from Daily High and Low Prices. *Journal of Finance*, 67(2), pp.719–760.

Da, Z., Engelberg, J. & Gao, P., 2011. In Search of Attention. *The Journal of Finance*, 66(5), pp.1461–1499. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2011.01679.x.

Dimpfl, T. & Jank, S., 2012. Can internet search queries help to predict stock market volatility ? *Journal of Economic Literature*, (G10, G14, G17). Available at: http://www.ssrn.com/abstract=1941680.

Dimpfl, T. & Jank, S., 2016. Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management*, 22(2), pp.171–192.

Dimpfl, T. & Kleiman, V., 2017. Investor Pessimism and the German Stock Market: Exploring Google Search Queries. *German Economic Review*.

Driscoll, J.C. & Kraay, A.C., 1998. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics*, 80(4), pp.549–560. Available at: http://www.mitpressjournals.org/doi/10.1162/003465398557825.

Easley, D. et al., 1996. Liquidity, Information, and Infrequently Traded Stocks. *The Journal of Finance*, 51(4), pp.1405–1436. Available at: http://doi.wiley.com/10.1111/j.1540-6261.1996.tb04074.x.

Fink, C. & Johann, T., 2014. May I Have Your Attention , Please : The Market Microstructure of Investor Attention. *SSRN Working Paper*, (September), p.59. Available at: http://ssrn.com/abstract=2139313.

Gerth, M., 2017. Twittern und zittern. *WirtschaftsWoche*, pp.77–78.

Gillert, M., 2017. Marktmanipulation: Short-Attacken - Wie Anleger und Emittenten ins visier von Manipulatoren geraten. *BaFin Journal*, (5), pp.26–28. Available at: https://www.accc.gov.au/publications/targeting-scams-report-on-scam-activity/targeting-scams-report-of-the-accc-on-scam-activity-2016.

Hamid, A. & Heiden, M., 2015. Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization*, 117, pp.62–81. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0167268115001687.

Kahneman, D. & Tversky, A., 1979. Prospect Theory - An Analysis Of Decision Under Risk. *Econometrica*, 47, pp.263–292. Available at: http://www.jstor.org/stable/1914185.

Kumar, A. & Lee, C.M.C., 2006. Retail Investor Sentiment and Return Comovements. *The Journal of Finance*, 61(5), pp.2451–2486. Available at: http://doi.wiley.com/10.1111/j.1540-6261.2006.01063.x.

Leinweber, D.J. & Madhavan, A.N., 2001. Three Hundred Years of Stock Market Manipulations. *the Journal of Investing*, 10(2), pp.1–10. Available at: http://www.iijournals.com/doi/abs/10.3905/joi.2001.319457.

Lo, A.W. & Wang, J., 2000. Trading Volume: Definitions, Data Analysis, and Implications of Portfolio Theory. *Review of Financial Studies*, 13(2), pp.257–300. Available at: https://academic.oup.com/rfs/article-lookup/doi/10.1093/rfs/13.2.257.

De Long, J.B. et al., 1990. Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), pp.703–738.

Long, J.B. De et al., 1990. Noise Trader Risk in Financial Markets Andrei Shleifer. *Journal of Political Economy*, 98(4), pp.703–738. Available at: http://works.bepress.com/cgi/viewcontent.cgi?article=1008&context=brad_delong.

Mao, H., Counts, S. & Bollen, J., 2015. Quantifying the effects of online bullishness on international financial markets. *ECB workshop on using Big Data for forecasting and statistics*, pp.1–15.

Peter, G. et al., 2017. Twitter sentiment around the earnings announcement events. *PLoS ONE*, 12(2).

Roodman, D., 2006. How to do Xtabond2: An Introduction to Difference and System GMM in Stata. *SSRN Electronic Journal*. Available at: http://www.ssrn.com/abstract=982943.

SEC, 2013. *Report of Investigation Pursuant to Section 21(a) of the Securities Exchange Act of 1934: Netflix, Inc., and Reed Hastings*, Available at: https://www.sec.gov/litigation/investreport/34-69279.htm.

SEC, 2015. Updated Investor Alert: Social Media and Investing -- Stock Rumors. Available at: https://www.sec.gov/oiea/investor-alerts-bulletins/ia_rumors.html [Accessed November 29, 2017].

SEC, 2014. Updated Investor Alert: Social Media and Investing - Avoiding Fraud. Available at: https://www.investor.gov/additional-resources/news-alerts/alerts-bulletins/updated-investor-alert-social-media-investing [Accessed November 29, 2017].

Shiller, R.J., 2003. From Efficient Markets Theory to Behavioral Finance. *Journal of Economic Perspectives*, 17(1), pp.83–104.

Shleifer, A. & Vishny, R., 1997. The Limits to Arbitrage. *Journal of Finance*, 52, pp.35–55.

Tafti, A., Zotti, R. & Jank, W., 2016. Real-time diffusion of information on twitter and the financial markets. *PLoS ONE*, 11(8).

Vlastakis, N. & Markellos, R.N., 2012. Information demand and stock market volatility. *Journal of Banking and Finance*, 36(6), pp.1808–1821.

Vozlyublennaia, N., 2014. Investor attention, index performance, and return predictability. *Journal of Banking and Finance*, 41(1), pp.17–35.