Google Search Volume: *Influence and Indication for the Dutch Stock Market*

S. Chen*

Bachelor Thesis

Erasmus University Rotterdam

2011

*StudentID: 319847

Abstract

This paper studies the relationship between stock specific and market level internet search volume on stocks and the Dutch stock market, using the listed stocks in the AEX index. Internet search volume is obtained weekly from the Google Insights database for the period between January 2004 and April 2011. As introduced by earlier studies, internet searching activity is an adequate proxy for investor recognition and should therefore be relevant for modeling trading activity and stock activity. The results obtained show that Google search volume is significantly influential not only for the traded volume, but also the historical stock volatility. This significance is proven to be stable by means of a Quandt-Andrews breakpoint test.

Acknowledgements

I am heartily thankful to my supervisor, Prof. dr. Dick van Dijk, for his guidance and support during the process of producing this thesis.

Contents

1 Introduction

The work of Merton (1987) suggests that investor attention may be relevant for stock pricing and stock liquidity. An increase in the attention of investors may indicate an increase in trading activity. However, measuring investor attention is a rather difficult task in practice, because there is no indicator that can solely represent the amount of attention for a certain stock. In empirical studies investigating the relationship between investor attention and stock returns several proxies for attention have been employed. Grullon et al. (2004) employ product advertisement activity of companies in order to explore the consequences for the visibility of the company in the stock market. They find that advertisement expenditure is positively related to the number of investors and to stock liquidity. Fang and Peress (2009) investigate the cross-sectional link between media coverage and expected stock returns. They have shown that companies with less media coverage have higher returns, acknowledging that the relationship between media coverage and liquidity may play an important role. Likewise, Frieder and Subrahmanyam (2005) employ the perception of a company's brand, and Mitchell and Mulherin (1994) and Berry and Howe (1994) derive a general measure of information flow that incorporates both stock specific and market wide information supply, in the form of the number of published news announcements. All these factors fall under information supply in stock market. However, in financial markets, information resources are one the of the most valuable matters for investors. Therefore, demand for information is an obvious estimator of the level of attention from investors. After all, a rational investor needs to acquire sufficient information before handling. However, measuring information demand seems to be an impossible task, because information can be acquired through infinite resources.

Amongst others, the work of Da et al. (2009) suggests the use of Google search volume to measure the attention of the individual investor. Google, being the world's largest internet search engine, provides the weekly search volume of a stocks underlying company's name, which is a rather innovative measure of the attention for a certain firm. Using this search volume as a proxy for investor attention seems to be adequate for two reasons. First of all, because the internet has become a rather popular and obvious way to seek for information on a certain stock for individual investors, not only due to the enormous

increase of internet usage all over the world during the last few decades, but also due to the great broadening of internet information supply. Using internet search activity recognizes the fact that the internet nowadays is used to revolutionize the consumption of information in the financial world (see Barber and Odean (2001), Antweiler and Frank (2004), Rubin and Rubin (2010)). And second, the search activity on Google should adequately be able to represent the attention of the investor, because an individual will only search a certain term, if he or she is demanding information about the object underlying the term. This adequacy is confirmed amongst others by the work of Vlastakis and Markellos (2010) and Bank et al. (2010) for respectively the 30 largest stocks traded in the NYSE and NASDAQ (representing the U.S. market) and the listed stocks in the Xetra index (representing the German market). One of the findings of Vlastakis and Markellos (2010) is that the internet search volume is positively related to trading volume and historical and implied measures of return volatility, even after controlling for variations in the market return and internet information supply, in terms of news coverage. Similarly, Bank et al. (2010) find that search volume is a powerful measure of investor recognition and that an increase in the internet search volume relates to higher trading activity, improved stock liquidity and leads to higher future returns in the short run.

For this thesis I perform an empirical application focusing on the Dutch stock market. Unlike the United States and Germany (countries considered in earlier studies, both possessing a very strong economical position in the world), the Netherlands is a country with an open, but relatively small and dependent economy. This raises the question whether the search volume on Google, a very "global" indicator of attention, can also reach the same level of influence and indication for the Dutch stock market. I am well aware of possible influence of the variations in the market return and stock activity, which are commonly found strong significant variables in a model describing stock movements. Therefore, I will perform my analysis while controlling for these variables. I choose to investigate the stocks listed in the Amsterdam Exchange (AEX) index. This index contains securities of the largest Dutch firms, which is a good representation of the Dutch stock market.

The purpose of this research is to examine the indication that Google search volume is able to provide for movements in the trading activity and stock market activity. The remainder of this paper is organized as follows. Section 2 describes the data resources and

sample construction, and presents a preliminary descriptive analysis of the data. Section 3 studies the relationship between Google search volume and trading activity. Section 4 studies the relationship between Google search volume and market activity, from the perspective of return volatility. Finally, section 5 concludes.

2 Data and sample construction

2.1 Google search volume

The public web facility "Google Insights for Search"[1] by Google provides information on how often a particular search term is searched, relative to the total search volume across various regions of the world, dating back to January 2004. For this thesis, I download the weekly search volume indices concerning the AEX index, limited between the period January 5th 2004 to April 30th 2011. The search volume provided by Google Insights is given as a relative value to the total number of searches in a chosen time interval. For each keyword, the number of searches is normalized so that the given search volume always varies between 0 (i.e., a period in which there were too little searches to pass a designated threshold) and 100 (i.e. a period in which the most searches found place).

Following Vlastakis and Markellos (2010) and Bank et al. (2010), I focus on the naive search volume of names of companies as search terms for a broad measure of attention from the search engine users. Though Da et al. (2009) argue it is preferable to use the stock ticker as search term in order to avoid including irrelevant components of the search volume index (e.g. people interested in purchasing beer searching the term "Heineken"), I assume that this component can be seen as random noise which will be eliminated by the normalisation by Google and which should not severely influence the index. In order to maximize accuracy of the sample estimates and also to keep the sample efficient, I expel stocks from the sample which a) have been listed in the AEX index for a period shorter than three years; b) take on the value 0 for a period longer than two consequetive months within the sample period. Due to these restrictions the stocks Aperam, Corio, Unibail-Rodamco and Wereldhave have been removed from the sample. In addition to stock specific search volumes, I also employ a measure using the keyword "AEX" accounting for the general market related search volume of the Dutch stock market. Table 1 presents the list of the 21 stocks remaining after purification of the sample, along with the corresponding stock tickers and applied search queries.

Table 2 contains the skewness, kurtosis and Jarque-Bera statistic of the stocks search volume. The majority of the stocks search volume series is positively skewed and normality

[1] See http://www.google.com/insights/search/.

can be rejected in all but 5 cases with 95% confidence level. Therefore, the search volume series referred to hereafter are logarithmically transformed. Table 3 contains the results of Augmented Dickey-Fuller tests (ADF, Dickey and Fuller (1979)) for stationarity of the search volume series. On the basis of information criteria, an intercept and a linear deterministic trend is included in the specification of the test regressions. The results indicate that the search volume of all stocks is stationary around a deterministic trend.

2.2 Stock activity

In my analysis, I first study the relationship between trading activity and search volume in order to examine whether the number of searches on Google can serve as a proxy for investors recognition. Consequently, I study the relationship between historical stock volatility and search volume to estimate the impact of search volume on market activity.

2.2.1 Trading activity

An obvious measure of trading activity is the traded stock volume, which was employed amongst others by Chordia et al. (2001) and Chordia et al. (2007). For the sample of stocks listed in the AEX index, I use end of the week closing stock prices ($P_{k,t}$) and the number of shares traded per week ($TS_{k,t}$) to compute weekly traded stock volumes:

$$TV_{k,t} = log(P_{k,t}TS_{k,t}) \tag{1}$$

where $TV_{k,t}$ corresponds to the traded volume in Euros of stock k in week t. Table 4 contains the skewness, kurtosis and Jarque-Bera statistic of the traded stock volumes and table 5 presents the ADF test results for stationarity. Again on the basis of information criteria, an intercept and a linear deterministic trend is included in the specification of the test regressions. Normality can be rejected for the majority of the stocks and the ADF statistics indicate that the majority of the traded volumes is stationary around a deterministic trend.

2.2.2 Realized volatility

I choose to study the association between investors attention and market activity from the perspective of historical return volatility. One of the most popular measures of historical volatility in the literature is realized volatility, due to its accuracy (see Barndorff-Nielsen and Shephard (2002), Andersen et al. (2001a), Andersen et al. (2001b)). To compute the realized volatility, I first compute the daily returns using end of the day stock prices ($P_{k,i}$) as follows:

$$r_{k,i} = log\left(\frac{P_{k,i}}{P_{k,i-1}}\right) \tag{2}$$

where $r_{k,i}$ corresponds to return of stock k on day i. Subsequently, the realized volatility is given by the summation of the squared returns over N days:

$$RV_{k,t} = \sum_{i=1}^{N} r_{k,t,i}^2 \tag{3}$$

where $r_{k,t,i}^2$ corresponds to the squared return of day i of stock k in week t. It is an obvious choice to compute weekly realized volatility to match the rest of the data, thus I will implement $N = 5$. Table 6 presents the skewness, kurtosis and Jarque-Bera statistic of the realized volatility. All realized volatility series are positively skewed and none of these are normally distributed. Therefore, the realized volatility referred to hereafter are logarithmically transformed. The ADF test results for stationarity are presented in table 7 and show that the majority of the realized volatility series is stationary around a deterministic trend.

3 Google search volume and trading activity

I start by investigating the relationship between Google search volume and trading activity. The employed measurement for the latter is the traded stock volume (as introduced by section 2.2). This relationship provides insight into the adequacy of Google search volume as a proxy for investor attention, and the level of influence it carries for the traded stock volume.

A correlation analysis of Google search volume and traded stock volume, as shown in table 8, supports the existence of a significant association between the two variables. The evidence is strong for both stock specific and market related search volume. Specifically, all but 3 correlation coefficients of stock specific search volumes are significant at 99% confidence level, and all but 6 correlation coefficients of market related search volume are significant at 95% confidence level. Although it seems that the association between stock specific search volume and traded stock volume is stronger, it is remarkable that for each stock at least one of the two variants of search volume is significant positively correlated.

3.1 Regression model

The multivariate regression in equation (4) is used in order to study the relationship between traded volume and Google search volume. A frequently researched relationship is the one between trading activity and stock volatility. This relationship is commonly found to be significantly positive (e.g., Bjursell et al. (2010), Chordia et al. (2001)). Thus I include weekly realized volatility as a regressor in order to account for the explaining power realized variance entails. This is useful due to the significance in correlation found between realized volatility and search volume (section 4). In order to control for the effect of market returns, the returns are employed as a regressor. The first lag of traded volume is also included as a regressor in order to control for autoregressive patterns in the model.

$$TV_{k,t} = \alpha + \beta_1 SV_{k,t} + \beta_2 SV_{M,t} + \beta_3 RV_{k,t} + \beta_4 MR_{k,t} + \beta_5 TV_{k,t-1} + \epsilon_{k,t} \tag{4}$$

Here $TV_{k,t}$ is the traded volume of stock k in week t, α is the constant, $SV_{k,t}$ is the Google search volume of stock k in week t, $SV_{M,t}$ is the market related Google search volume in

week t, $RV_{k,t}$ is the realized volatility of stock k in week t, $MR_{k,t}$ is the market return of stock k in week t and $\epsilon_{k,t}$ are the errors. For all regressions, a Breusch-Pagan test for heteroscedasticity (Breusch and Pagan (1979)) is applied. White standard errors (White (1980)) are implemented whenever the null hypothesis of homoscedasticity is rejected.

3.2 Results

The results presented in table 9 are in line with the correlation analysis and the findings of earlier studies and indicate a factual relationship between traded stock volume and Google search volume. Specifically, at least one search volume variable is significant at 95% level for all but 4 of the stocks, even after controlling for effects of realized volatility, market returns and the first auto-lag. Stock specific search volume is significant in 12 cases, whereas market related search volume is significant in 7 cases. The impact of market return appears to be limited in most of the cases, while, as expected, the persistency with the first lag of traded stock volume is strongly positive for all stocks. Note that the historically proven association between traded stock volume and realized volatility can also be concluded from these results. The adjusted R^2 values range between 40.73% for AEGON and 91.02% for Boskalis Westminster, which indicates a good fit of the models.

In order to assess the stability of the model, I perform a Quandt-Andrews test (QA, Andrews (1993)). This test examines one or more structural breakpoints in the sample, with a chosen trimming region of 15% of the sample. Under the null hypothesis of "no structural breakpoints", I obtain the maximum of Likelihood Ratio F-statistics for every regression. These are presented in the last column of table 9. The significance of these statistics is based on the probability values calculated using the Hansen method (Hansen (1997)). The results show that the model is stable in all but 5 cases. This means that the interdependence between traded volume, and Google search volume and the other regressors has been stable over the sample period for the majority of the stocks, which is remarkable since the sample includes the financial crisis from 2007 to 2009. This result is in line with the assumption that Google search volume is an adequate proxy for investor recognition, which is a factor that is found to have persistent relevance for the trading activity of stocks (Merton (1987)).

4 Google search volume and market activity

Consequently, I study the association between Google search volume and market activity. As introduced by section 2.2.2, I do this from the perspective of realized volatility. I start by a correlation analysis of search volume and realized volatility. The results presented by table 10 report the existence of a strong significant association between Google search volume and realized volatility. In particular, the correlation coefficients for market related search volume are all significantly positive at 99% confidence level. Stock specific search volume also seems to be linked to realized volatility, but the signs of the coefficients are mixed and strength of significance vary.

4.1 Regression model

The multivariate regression in equation (5) is used in order to study the relationship between realized volatility and Google search volume. As stated in section 3, it is needed to include traded stock volume as a regressor to account for the explaining power it entails, due to the commonly found positively significant relationship between trading activity and realized volatility. To control for the effect of market returns, the returns are added as regressor. Likewise, the first lag of weekly realized variance is included as a regressor in order to control for autoregressive patterns in the model.

$$RV_{k,t} = \delta + \gamma_1 SV_{k,t} + \gamma_2 SV_{M,t} + \gamma_3 TV_{k,t} + \gamma_4 MR_{k,t} + \gamma_5 RV_{k,t-1} + \eta_{k,t} \tag{5}$$

Here $RV_{k,t}$ is the realized volatility of stock k in week t, δ is the constant, $SV_{k,t}$ is the Google search volume of stock k in week t, $SV_{M,t}$ is the market related Google search volume in week t, $TV_{k,t}$ is the traded volume of stock k in week t, $MR_{k,t}$ is the market return of stock k in week t and $\epsilon_{k,t}$ are the errors. For all regressions, a Breusch-Pagan test for heteroscedasticity is applied and White standard errors are implemented whenever the null hypothesis of homoscedasticity is rejected.

4.2 Results

The results of the regression in (5) are presented in table 11. Stock specific search volume appears to be a significant regressor in 13 cases, while in all cases market related search volume is proven to be a significant regressor at 95% confidence level. Also, where stock specific search volume relates to realized volatility with mixed strength and direction, all estimated coefficients for market related search volume are strong significantly positive. The effect of market returns appears to be limited and, as expected, realized volatility seems to be highly persistent with most coefficients on the first lag being significantly positive. The association between trading activity and realized volatility is also once again strongly confirmed. The adjusted R^2 values range between 22.91% for ASML and 64.60% for Fugro, which suggests a fairly good fit of the models.

To assess the stability of the model, I perform another Quandt-Andrews test examining possible structural breakpoints in the sample, with a chosen trimming region of 15% of the sample. In the last column of table 11 contains the maximum of Likelihood Ratio F-statistics for every regression. The corresponding null hypothesis is that there are no structural breakpoints within the sample. The significance of the F-statistics is based on the probability values calculated using the Hansen method. The results show that the model is stable in all but 3 cases. This means that the interdependence between realized volatility, and Google search volume and the other regressors has been stable over the sample period for the majority of the stocks. Because the sample period includes the financial crisis from 2007 to 2009, this result is remarkable, but however in line with the findings in the analysis of the relationship between trading activity and Google search volume.

5 Conclusion

This thesis studies the relationship between investor attention and market activity by employing a novel proxy for investor attention, namely Google search volume. The adequacy of using search volume to proxy investor attention is shown by correlation analyses between search volume, and traded stock volume and realized volatility. These results are also in line with earlier studies (e.g., Merton (1987), Vlastakis and Markellos (2010), Bank et al. (2010)). For the listed stocks in the AEX index, I have shown that movement in the number of searches on Google for these stocks on both stock specific and market level significantly leads to movement in trading activity, in terms of traded stock volume, and market activity, in terms of realized volatility. This effect is even significant after controlling for effects of market returns, autoregressive patterns and also the commonly found significant association between trading activity and stock volatility. Due to my lack of resources, I was not able to collect variables for information supply (e.g., news announcements and other media coverage), which is historically proven to be of influence for stock activity. It is an interesting possibility for elaboration of this topic to add these variables when available, in order to test for the persistency of the significance of internet search activity for stock activity, when controlling for information supply variables.

6 Appendix

Table 1: **List of stocks in the sample and search queries.**

Stock	**Ticker**	**Search Query**
AEGON	AGN	"aegon"
Kon. Ahold	AH	"ahold"
Air France-KLM	AF	"air france klm"
Akzo Nobel	AKZA	"akzo nobel"
ArcelorMittal	MT	"arcelor mittal"
ASML Holding	ASML	"asml"
Kon. Boskalis Westminster	BOKA	"boskalis"
Kon. DSM	DSM	"dsm"
Fugro	FUR	"fugro"
Heineken Holding	HEIA	"heineken"
ING Groep	INGA	"ing"
Kon. KPN	KPN	"kpn"
Kon. Philips	PHIA	"philips"
Randstad	RAND	"randstad"
Reed Elsevier	REN	"reed elsevier"
Royal Dutch Shell	RSDA1	"shell"
SBM Offshore	SBMO	"sbm offshore"
TNT	PNL	"tnt"
TomTom	TTM	"tomtom"
Unilever Certificate	UNA	"unilever"
Wolters Kluwer	WKL	"wolters kluwer"
AEX	AEX	"aex"

Table 2: **Normality statistics of search volume.**

This table presents the test statistics for normality of the search volumes at stock specific and market level. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Skewness**	**Kurtosis**	**JB-Statistic**
AEGON	3.6758	30.6997	13072,67 ***
Ahold	1.0385	3.6408	75,19952 ***
Air France-KLM	0.1597	3.9747	16,74549 ***
AkzoNobel	0.5117	2.7476	17,68471 ***
Arcelor Mittal	0.2506	3.1914	4.583
ASML	0.4997	3.0333	15,91396 ***
Boskalis	1.7287	9.1980	801,7121 ***
DSM	0.1888	2.6871	3.829
Fugro	-0.0194	3.9336	13,89557 ***
Heineken	1.4535	8.8686	682,6863 ***
ING	-0.0033	1.7108	26,45342 ***
KPN	0.1717	3.3233	3.541
Philips	0.5135	2.7182	18,05121 ***
Randstad	0.0523	2.8760	0.419
Reed Elsevier	1.3807	6.9718	372,4658 ***
Royal Dutch Shell	2.8838	20.2603	5271,302 ***
SBM Offshore	1.7275	9.6702	735,9266 ***
TNT	0.0028	4.6183	41,68237 ***
TomTom	0.1107	3.0716	0.862
Unilever	5.3820	42.1667	26260,83 ***
Wolters Kluwer	-0.5845	2.5377	25,14973 ***
AEX	5.4261	43.5265	28015,99 ***

Table 3: **Augmented Dickey-Fuller test results of search volume.**

This table presents the Dickey-Fuller test statistics for stationarity of the natural logarithm of the search volume. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Augmented Dickey Fuller**
AEGON	-9.771 ***
Ahold	-3.622 **
Air France-KLM	-4.810 ***
AkzoNobel	-12.183 ***
Arcelor Mittal	-3.358 *
ASML	-3.709 **
Boskalis	-12.091 ***
DSM	-4.929 ***
Fugro	-1.987
Heineken	-5.273 ***
ING	-3.650 **
KPN	-4.514 ***
Philips	-4.229 ***
Randstad	-5.056 ***
Reed Elsevier	-10.734 ***
Royal Dutch Shell	-9.265 ***
SBM Offshore	-4.342 ***
TNT	-3.431 **
TomTom	-3.395 **
Unilever	-8.723 ***
Wolters Kluwer	-8.642 ***
AEX	-5.530 ***

Table 4: **Normality statistics of traded volume.**

This table presents the test statistics for normality of the traded volumes at stock specific level. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Skewness**	**Kurtosis**	**JB-Statistic**
AEGON	-0.3856	4.2143	32.9341 ***
Ahold	-0.2223	4.9436	31.9656***
Air France-KLM	0.1272	2.2274	10.0631***
AkzoNobel	0.2124	3.0797	2.9721
Arcelor Mittal	-2.1339	11.1622	639.8052 ***
ASML	-0.5004	3.7588	25.1033 ***
Boskalis	-0.6807	2.2457	38.5598 ***
Corio	-0.0456	2.6658	2.8099
DSM	-0.4697	3.1742	14.5292 ***
Fugro	-1.0293	3.0852	67.5665 ***
Heineken	-0.3667	3.5760	13.8408 ***
ING	-0.1825	3.4702	5.6392 *
KPN	0.2732	3.4870	8.5269 **
Philips	-0.4280	3.6643	18.6845 ***
Randstad	-0.9040	3.4215	54.8582 ***
Reed Elsevier	-0.4386	4.9763	74.4130 ***
Royal Dutch Shell	-0.3140	3.3581	6.5770 **
SBM Offshore	-0.2753	3.0263	4.8377 *
TNT	0.0192	3.0936	0.1631
TomTom	0.1872	2.7568	2.5754
Unibail-Rodamco	0.3127	3.0155	3.2774
Unilever	0.0737	3.7051	5.5784 *
Wereldhave	0.1433	2.8702	1.5755
Wolters Kluwer	-0.2496	3.5936	6.4411 **

Table 5: **Augmented Dickey-Fuller test results of traded volume.**

This table presents the Dickey-Fuller test statistics for stationarity of traded volume. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Augmented Dickey Fuller**
AEGON	-9.6631 ***
Ahold	-8.2413 ***
Air France-KLM	-3.3441 *
AkzoNobel	-3.1478 *
Arcelor Mittal	6.7445 ***
ASML	-10.4333 ***
Boskalis	-4.1152 ***
Corio	-2.4950
DSM	-4.8994 ***
Fugro	-2.0126
Heineken	-5.2265 ***
ING	-3.6323 **
KPN	-4.4884 ***
Philips	-4.2173 ***
Randstad	-5.0896 ***
Reed Elsevier	-10.6522 ***
Royal Dutch Shell	-9.2678 ***
SBM Offshore	-4.2884 ***
TNT	-3.4016 *
TomTom	-3.3767 *
Unibail-Rodamco	-1.9231
Unilever	-8.7278 ***
Wereldhave	-2.5785
Wolters Kluwer	-8.6103 ***

Table 6: **Normality statistics of realized volatility.**

This table presents the test statistics for normality of the realized volatility. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Skewness**	**Kurtosis**	**JB-Statistic**
AEGON	4.5663	27.5376	10910.85 ***
Ahold	3.6245	19.9894	2729.503 ***
Air France-KLM	8.8863	109.1755	130377.5 ***
AkzoNobel	15.1017	254.6866	1022780 ***
Arcelor Mittal	4.4530	26.7865	4784.612 ***
ASML	7.7508	95.7364	140708.8 ***
Boskalis	3.5254	18.6554	4667.782 ***
DSM	5.0557	40.6682	24021.23 ***
Fugro	9.3823	118.5830	218242.1 ***
Heineken	11.2909	168.0491	441705.8 ***
ING	5.3286	40.0071	23606.08 ***
KPN	13.0511	216.3939	735641.3 ***
Philips	4.8111	34.7776	17546.56 ***
Randstad	4.3326	27.3455	10628.95 ***
Reed Elsevier	8.4945	100.2109	155005.7 ***
Royal Dutch Shell	12.7920	193.7363	464479 ***
SBM Offshore	8.7751	103.0999	164387.3 ***
TNT	8.4778	90.3383	125987.9 ***
TomTom	4.0809	23.4867	6261.39 ***
Unilever	5.3062	41.0398	24824.48 ***
Wolters Kluwer	2.6252	10.8065	944.0892 ***

Table 7: **Augmented Dickey-Fuller test results of realized volatility.**

This table presents the Dickey-Fuller test statistics for stationarity of the natural logarithm of realized volatility. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Augmented Dickey Fuller**
AEGON	-4.3959 ***
Ahold	-6.4141 ***
Air France-KLM	-6.7670 ***
AkzoNobel	-6.6642 ***
Arcelor Mittal	-3.0234
ASML	-2.2934
Boskalis	-8.6426 ***
DSM	-7.7473 ***
Fugro	-7.5118 ***
Heineken	-5.5660 ***
ING	-4.6860 ***
KPN	-9.8063 ***
Philips	-6.2475 ***
Randstad	-5.3524 ***
Reed Elsevier	-5.9716 ***
Royal Dutch Shell	-3.9102 **
SBM Offshore	-6.4225 ***
TNT	-4.8670 ***
TomTom	-3.4319 **
Unilever	-8.6769 ***
Wolters Kluwer	-4.5959 ***

Table 8: **Correlation between traded volume and Google search volume.**

This table presents the correlation coefficients between traded stock volume and Google search volume at stock specific and market level. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Stock specific**	**Market related**
AEGON	0.057	0.214 ***
Kon. Ahold	0.595 ***	0.451 ***
Air France-KLM	0.191 **	0.303 ***
Akzo Nobel	0.711 ***	0.178 **
ArcelorMittal	0.315 ***	-0.057
ASML Holding	0.736 ***	0.114
Kon. Boskalis Westminster	0.414 ***	-0.042
Kon. DSM	0.544 ***	0.141 *
Fugro	0.506 ***	0.173 **
Heineken Holding	0.188 **	0.300 ***
ING Groep	-0.272 ***	0.122 *
Kon. KPN	-0.383 ***	0.359 ***
Kon. Philips	0.109	0.168 **
Randstad	0.469 ***	0.033
Reed Elsevier	0.676 ***	0.443 ***
Royal Dutch Shell	0.443 ***	0.474 ***
SBM Offshore	0.327 ***	0.279 ***
TNT	-0.463 ***	0.216 ***
TomTom	0.415 ***	0.154 **
Unilever Certificate	0.018	0.443 ***
Wolters Kluwer	0.446 ***	0.448 ***

Table 9: **Regression of traded volume on Google search volume.**

This table presents the results of OLS regressions between traded stock volume, and Google search volume and market information variables. α is the constant, β_k ($k = 1, ..., 5$) is the coefficient for respectively stock specific search volume; market related search volume; realized volatility; stock market return; the first lag of traded volume. The last two columns present the corresponding values of the adjusted R^2 statistic and the Quandt-Andrews test statistic. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	α	β_1	β_2	β_3	β_4	β_5	**Adjusted R^2**	**QA-Statistic**
AEGON	6.8160 ***	0.0099	0.1734 **	0.0350 *	0.6978 ***	0.6082 ***	**0.4073**	**27.6030 *** **
Kon. Ahold	10.3524 ***	0.9552 ***	-0.0492	0.1111 ***	-0.5245	0.3045 ***	**0.6550**	**3.1848**
Air France-KLM	0.8866	0.9530 **	0.1692	0.1461 ***	1.0477	0.6969 ***	**0.6208**	**10.4694**
Akzo Nobel	4.4159 ***	0.2565 ***	-0.0756	0.0708 ***	0.4799	0.7660 ***	**0.6797**	**36.6573 *** **
ArcelorMittal	8.0862 ***	0.5424 **	-0.2946 ***	0.0522	0.1857	0.5380 ***	**0.4344**	**6.8378**
ASML Holding	9.3181 ***	0.8433 ***	0.0717	0.1269 ***	0.2394	0.3406 ***	**0.6499**	**11.0053**
Kon. Boskalis Westminster	3.2635 ***	0.1794	-0.2643 ***	0.2195 ***	0.9388 **	0.8837 ***	**0.9102**	**14.1390**
Kon. DSM	6.7640 ***	0.2893 *	-0.1444 **	0.1618 ***	0.5270	0.6287 ***	**0.5843**	**15.8373**
Fugro	2.0205 **	0.3006 *	-0.0514 ***	0.1265 ***	1.1267 ***	0.8602 ***	**0.8514**	**23.6921 ** **
Heineken Holding	8.1410 ***	0.1311 *	0.0703	0.1232 ***	1.0659 **	0.5495 ***	**0.5013**	**9.2544**
ING Groep	4.9724 ***	-0.0248	0.1223 *	0.0347 *	0.3419	0.7384 ***	**0.5505**	**37.3379 *** **
Kon. KPN	10.3087 ***	0.2332	0.0311	0.1821 ***	0.4145	0.4471 ***	**0.4782**	**7.6946**
Kon. Philips	5.6769 ***	0.2177 ***	0.0004	0.0958 ***	0.3746	0.6831 ***	**0.4923**	**24.5215 *** **
Randstad	2.5486 ***	1.1248 ***	0.0547	0.1312 ***	0.8041 **	0.6043 ***	**0.7867**	**9.0661**
Reed Elsevier	9.4595 ***	0.2024 ***	0.1221 *	0.1232 ***	0.3310	0.4420 ***	**0.4936**	**15.2704**
Royal Dutch Shell	5.4893 ***	0.4168 ***	0.1743 ***	0.0453 ***	-0.0813	0.6307 ***	**0.6282**	**19.0833 ***
SBM Offshore	5.0163 ***	0.1837 **	0.1321	0.0789 ***	0.5816	0.6696 ***	**0.5017**	**19.2463 ***
TNT	7.6882 ***	-0.0997	0.0755	0.1045 ***	0.5585	0.6101 ***	**0.4707**	**12.7607**
TomTom	4.9509 ***	0.2348 **	-0.1509	0.1686 ***	0.8807 ***	0.7374 ***	**0.6608**	**14.3469**
Unilever Certificate	9.5029 ***	0.0735	0.1161 **	0.1095 ***	0.3657	0.5051 ***	**0.4619**	**14.2050**
Wolters Kluwer	4.1246 ***	0.6058 ***	0.1198	0.0726 ***	0.3665	0.6110 ***	**0.5498**	**19.8836 ***

Table 10: **Correlation between stock realized volatility and Google search volume.**

This table presents the correlation coefficients between stock realized volatility and Google search volume at stock specific and market level. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	**Stock specific**	**Market related**
AEGON	0.451 ***	0.750 ***
Kon. Ahold	0.278 ***	0.601 ***
Air France-KLM	0.372 ***	0.552 ***
Akzo Nobel	0.317 ***	0.650 ***
ArcelorMittal	0.721 ***	0.797 ***
ASML Holding	0.225 **	0.564 ***
Kon. Boskalis Westminster	0.581 ***	0.707 ***
Kon. DSM	0.060	0.651 ***
Fugro	0.480 ***	0.804 ***
Heineken Holding	-0.133	0.721 ***
ING Groep	0.180 *	0.782 ***
Kon. KPN	-0.154	0.669 ***
Kon. Philips	0.265 **	0.738 ***
Randstad	-0.139	0.681 ***
Reed Elsevier	0.510 ***	0.688 ***
Royal Dutch Shell	0.170	0.572 ***
SBM Offshore	0.206 *	0.696 ***
TNT	-0.460 ***	0.529 ***
TomTom	0.428 ***	0.607 ***
Unilever Certificate	0.416 ***	0.609 ***
Wolters Kluwer	0.184 *	0.650 ***

Table 11: **Regression of stock realized volatility on Google search volume.**

This table presents the results of OLS regressions between stock realized volatility, and Google search volume and market information variables. α is the constant, γ_k $(k = 1, ..., 5)$ is the coefficient for respectively stock specific search volume; market related search volume; traded stock volume; stock market return; the first lag of realized volatility. The last two columns present the corresponding values of the adjusted R^2 statistic and the Quandt-Andrews test statistic. A star, double star and triple star denote significance at 10%, 5% and 1% level, respectively.

Stock	α	γ_1	γ_2	γ_3	γ_4	γ_5	**Adjusted R^2**	**QA-Statistic**
AEGON	-13.9407 ***	1.3440 ***	0.5533 ***	0.2790 **	-0.9696	0.5575 ***	**0.5711**	**17.7874**
Kon. Ahold	-34.3267 ***	-1.2949 **	1.3256 ***	1.5813 ***	-2.3568	0.1351 *	**0.4708**	**3.9421**
Air France-KLM	-6.7145 ***	-0.6321	0.7849 ***	0.1765 ***	-1.0859	0.3189 ***	**0.2716**	**10.1464**
Akzo Nobel	-32.0144 ***	-0.4696	1.7721 ***	1.0392 ***	-2.6217	0.3183 ***	**0.3169**	**16.6213**
ArcelorMittal	-15.7001 ***	1.4203 ***	1.2705 ***	0.1433	0.2567	0.2539 ***	**0.6460**	**2.7058**
ASML Holding	-17.6433 ***	-0.7456 ***	0.4244 ***	0.8097 ***	-1.1893	0.2886 ***	**0.2291**	**21.3152 ****
Kon. Boskalis Westminster	-16.3800 **	0.8183 ***	1.1831 ***	0.2886 ***	1.0020	0.1319 ***	**0.4117**	**10.4581**
Kon. DSM	-18.4157 ***	-0.8297 ***	0.9015 ***	0.8403 ***	-0.2862	0.3304 ***	**0.4797**	**14.9583**
Fugro	-14.6093 ***	0.2427	1.1293 ***	0.3401 ***	-2.4076 ***	0.2889 ***	**0.4146**	**15.1725**
Heineken Holding	-19.6769 ***	-0.2406	0.7103 ***	0.7551 ***	-1.8597	0.3029 ***	**0.3601**	**8.8700**
ING Groep	-11.3063 ***	1.3477 ***	0.8655 ***	0.0079	-1.2837	0.5139 ***	**0.6040**	**20.4504 ****
Kon. KPN	-23.9939 ***	-1.1757 ***	1.0846 ***	1.0287 ***	0.5724	0.0421	**0.3766**	**6.4816**
Kon. Philips	-8.0794 ***	1.1744 ***	0.8825 ***	0.3407 ***	0.0559	0.4207 ***	**0.4083**	**19.9638 ***
Randstad	-13.3895 ***	-1.0142 **	0.8345 ***	0.6334 ***	-0.8039	0.2750 ***	**0.3044**	**10.8309**
Reed Elsevier	-24.0908 ***	-0.1811	0.6451 ***	1.0183 ***	-2.3909	0.2789 ***	**0.3409**	**25.9171 ****
Royal Dutch Shell	-8.1635 ***	-0.4378	0.9174 ***	0.1489	-1.5111	0.4228 ***	**0.3059**	**12.2750**
SBM Offshore	-14.3904 ***	0.5262 ***	1.1842 ***	0.2522 **	-1.7960 *	0.2767 ***	**0.3224**	**18.0592 ***
TNT	-17.7667 ***	0.8174 ***	0.6976 ***	0.4663 ***	-1.1805	0.3666 ***	**0.3344**	**15.6789**
TomTom	-12.0064 ***	-0.5564 **	1.0471 ***	0.4035 ***	-0.0900	0.2698 ***	**0.2344**	**17.4554**
Unilever Certificate	-23.3110 ***	0.0734	0.4816 **	0.8875 ***	-3.3451 *	0.2929 ***	**0.2997**	**11.4443**
Wolters Kluwer	-14.5261 ***	0.3625	0.8988 ***	0.3156 **	-1.4905	0.3008 ***	**0.3007**	**16.1300**

References

T.G. Andersen, T. Bollerslev, F.X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61:43–76, 2001a.

T.G. Andersen, T. Bollerslev, F.X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96:42–55, 2001b.

D.W.K. Andrews. Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856, 1993.

W. Antweiler and M.Z. Frank. Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294, June 2004.

M. Bank, M. Larch, and G. Peter. Google search volume and its influence on liquidity and returns of german stocks. *Working Paper*, 2010.

B.M. Barber and T. Odean. The internet and the investor. *The Journal of Economic Perspectives*, 15(1):41–54, 2001.

O. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of Royal Statistical Society*, 64: 253–280, 2002.

T.D. Berry and K.M. Howe. Public information arrival. *Journal of Finance*, 49(4):1331–46, 1994.

J. Bjursell, A. Frino, Y. Tse, and G.H.K. Wang. Volatility and trading activity following changes in the size of futures contracts. *Journal of Empirical Finance*, 17(5):967–980, 2010.

T.S. Breusch and A.R. Pagan. Simple test for heteroscedasticity and random coefficient variation. *Econometrica (The Econometric Society)*, 47(5):1287–1294, 1979.

T. Chordia, R. Roll, and A. Subrahmanyam. Market liquidity and trading activity. *Journal of Finance*, 56(2):501–530, 2001.

T. Chordia, H. Sahn-Wook, and A. Subralunanyam. The cross-section of expected trading activity. *Review of Financial Studies*, 20(3):709–740, 2007.

Z. Da, J. Engelberg, and P.J. Gao. In search of attention. *Journal of Finance*, 2009.

D.A. Dickey and W.A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.

L.H. Fang and J. Peress. Media coverage and the cross-section of stock returns. *The Journal of Finance*, Forthcoming, 2009.

L. Frieder and A. Subrahmanyam. Brand perceptions and the market for common stock. *The Journal of Financial and Quantitative Analysis*, 40(1):57–85, 2005.

G. Grullon, G. Kanatas, and J.P. Weston. Advertising, breadth of ownership, and liquidity. *The review of financial studies*, 17(2):439–461, 2004.

B.E. Hansen. Approximate asymptotic p-values for structural-change tests. *Journal of Business and Economic Statistics*, 15(1):60–67, 1997.

R. Merton. A simple model of capital market equilibrium with incomplete information. *The Journal of Finance*, 42(3):483–510, 1987.

M.L. Mitchell and J.H. Mulherin. The impact of public information on the stock market. *Journal of Finance*, 49(3):923–50, 1994.

A. Rubin and E. Rubin. Informed investors and the internet. *Journal of Business, Finance and Accounting*, 37(7/8):841–865, 2010.

N. Vlastakis and R.N. Markellos. Information demand and stock market volatility. *Working Paper*, 2010.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980.