Stockholm School of Economics

The Interaction of Retail Investors with Financial Markets – Cross-Country Evidence from Google Search Data

Leicht, Simon* and Pütz, Henning†

Abstract

This paper analyzes retail investors' interaction with financial markets and provides empirical findings confirming noise trader models. We use stock market prime indices in a cross-country analysis and large US stocks in a single-stock analysis. Moreover, we apply an innovative measure of retail investors' attention - the Search Volume Index of Google - and find bi-directional Granger-causality between retail investors' attention on the one hand and volatility, trading volume and liquidity on the other hand. Implied volatility predictively explains retail investors' attention to a great extent. In turn, retail investors' behavior - after their attention was roused - explains a substantial part of future realized volatility. Additionally, retail investors are more active in the markets during periods of strong market fluctuations. Structural characteristics and behavioral biases like stock market participation rates, impatience and uncertainty avoidance are potentially able to explain cross-country variation.

Tutor: Michael Halling

Keywords: investor behavior, investor attention, search engine data, noise trader, retail investors, cross-country, risk aversion, volatility, liquidity

JEL classification: D03, D14, G2, G10

Acknowledgements: We thank our supervisor Michael Halling for the great guidance, continuous support and helpful suggestions. Furthermore we would like to thank all friends and relatives for their valuable input. Special thanks goes to Alina Roth.

* Stockholm School of Economics, 40501@student.hhs.se

† Stockholm School of Economics, 40386@student.hhs.se

Table of Contents

Table of Contents **2**

List of Figures **4**

List of Tables **5**

List of Abbreviations **6**

1. Introduction **7**

2. Literature **9**

3. Data and Methodology **12**

3.1 Data 12

3.2 Methodology 18

4. Results **24**

4.1 Indices 25

4.1.1 Realized Volatility – Complete Sample Period 25

4.1.2 Realized Volatility – Subsamples 29

4.1.3 Implied Volatility 33

4.1.4 Illiquidity 36

4.1.5 Trading Volume 38

4.2 Stocks 41

4.2.1 Attention Distribution across Stocks 42

4.2.2 Investors’ Attention Triggers 42

4.2.3 Investors’ Actions on the Stock Market 43

5. Discussion **45**

5.1 Investors’ behavior in General 45

5.1.1 Time Varying Behavior 45

5.1.2 Trigger Events 49

5.1.3 Attention Distribution 51

5.2 Investors’ behavior in the Cross-Country Setting 53

5.2.1 Structural Characteristics 54

5.2.2 Behavioral Characteristics 57

6. Robustness, Limitations and Further Research **60**

6.1 Robustness 60

6.2 Limitations 61

6.3 Further Research 63

7. Conclusion 65

References 66

Appendix 70

A. Figures.. 70

B. Tables.. 76

List of Figures

Figure 1: Overall search volume stocks 70
Figure 2: Realized Volatility: Impulse response functions CAC 40 and FTSE MIB (complete period) 71
Figure 3: Impulse response functions realized volatility CAC 40 (all periods) 72
Figure 4: Implied Volatility: Impulse response functions Dow Jones and CAC 40 (complete period) 73
Figure 5: Illiquidity: Impulse response functions Dow Jones and CAC 40 (complete period) 74
Figure 6: Trading Volume: Impulse response functions Dow Jones and CAC 40 (complete period) 75

List of Tables

Table 1: Realized volatility: Granger causality test statistics 26
Table 2: Realized volatility: Variance decomposition 27
Table 3: Implied volatility: Granger causality test statistics 34
Table 4: Implied volatility: Variance decomposition... 35
Table 5: Illiquidity: Granger causality tests and variance decomposition 37
Table 6: Trading Volume: Granger causality tests and variance decomposition............ 40
Table 7: Cross-Country Structural Characteristics... 54
Table 8: Cross-Country Behavioral Characteristics... 58
Table 9: Google search queries used for indices and stocks 76
Table 10: Summary Statistics indices: Search volume index and realized volatility...... 77
Table 11: Summary Statistics indices: Implied volatility, illiquidity and volume.......... 78
Table 12: Realized volatility: VAR-estimation results (Complete period)..................... 79
Table 13: Realized volatility: VAR-estimation results (Pre-Crisis period) 80
Table 14: Realized volatility: VAR-estimation results (Crisis period) 81
Table 15: Realized volatility: VAR-estimation results (Post-Crisis period)................... 82
Table 16: Implied volatility: VAR-estimation results (Complete and pre-crisis period) 83
Table 17: Implied volatility: VAR-estimation results (Crisis and post-crisis period) 84
Table 18: Illiquidity: VAR-estimation results (Complete and pre-crisis period)............ 85
Table 19: Illiquidity: VAR-estimation results (Crisis and post-crisis period) 86
Table 20: Trading Volume: VAR-estimation results (Complete and pre-crisis period) . 87
Table 21: Trading Volume: VAR-estimation results (Crisis and post-crisis period)...... 88
Table 22: Summary Statistics Stocks: Realized volatility (RV) 89
Table 23: Summary Statistics Stocks: Google search volume index (SVI).................... 90
Table 24: Summary Statistics Stocks: Trading volume (TV) 91
Table 25: Summary Statistics Stocks: Illiquidity (ILLQ) 92
Table 26: Variance decomposition test statistics of stock sample 93
Table 27: Granger causality test statistics of stock sample 94

List of Abbreviations

ADF	Augmented Dickey-Fuller
AEX	Amsterdam Exchange Index
a.m.	ante meridiem
CAC	Cotation Assistée en Continu
Corp.	Corporation
DAX	Deutscher Aktien Index
DJIA	Dow Jones Industrial Average
FEVD	Forecast Error Variance Decomposition
FTSE	Financial Times and the London Stock Exchange
IBEX	Índice Bursatil Español
Inc	Incorporation
IRF	Impulse Response Function
IV	Implied Volatility
JB	Jarque-Bera
LB	Ljung-Box
L.P.	Limited Partnership
MAIC	Akaike's information criterion
MIB	Milano Italia Borsa
ML	Maximum Likelihood
MSBIC	Multivariate Schwarz's Bayesian Information Criterion
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
p.m.	post meridiem
RV	Realized Volatility
SBIC	Schwartz's Bayesian information criterion
SEC	US Securities and Exchange Commission
SVI	Search Volume Index
UTC	Coordinated Universal Time
U.S.	United States
USA	United States of America

1. Introduction

Empirical evidence on retail investors' behavior suggests that these market participants exhibit various cognitive biases and do not act as rational agents (Barber and Odean, 2011). Interaction with financial markets as a consequence of their behavioral characteristics is controversially discussed in the academic world. The present work aims to shed further light on retail investors' interaction with financial markets. The bi-directional relationship between retail investors' attention on the one hand and stock market volatility, trading volume and liquidity on the other hand is analyzed.

The influence of noise traders, defined as non-rational market participants without superior information[1], on financial markets has been recognized since the models of Kyle (1985) and Black (1986). Often the behavior of noise traders and retail investors, characterized as non-sophisticated individual investors, is linked (DeLong et al, 1990). Empirical evidence supports the assumption that both behave similar[2] and trade for non-informational reasons - including "misperception of future returns, shifts in risk aversion, or hedging needs"[3] (Foucault et al, 2011, p.1370). Following this line of argument, we expect to empirically observe that retail investors' interaction with financial markets follows patterns captured by noise trader models. These models suggest that noise trader activity increases volatility, liquidity and trading volume in the markets and affects asset prices (e.g. Black, 1986; Campbell and Kyle, 1993; Wang, 2010). Retail investors' behavior in financial markets has also been linked to attention in the sense that investors tend to get active after their attention has been roused. Specifically, Barber and Odean (2008) show that many investors tend to only purchase stocks which have caught their attention in the first place. As humans do not have unlimited attention, which is actually a scarce cognitive resource (Kahneman, 1973), the following question arises: What directs retail investors' attention towards financial markets? Among others, Lux and Marchesi (1999) find that noise traders' attention is caught by volatility in financial markets and that their subsequent action causes additional volatility. Based on the above, we set up vector autoregressive (VAR)-models capable of analyzing the bi-directional causal relationship between investors' attention and measures of financial market activity.

We make use of a recently emerged measure of investors' attention originating from individuals' search volume on Google. Google's Search Volume Index (SVI), freely available on Google Trends, offers various advantages formerly used measures like news headlines, press

[1] See e.g. Kyle (1985), Black, (1986), DeLong et al (1990), Foucault et al (2011)

[2] As defined by Shleifer and Summers (1990)

[3] For detailed evidence of non-informational trading behavior see e.g. Barber and Odean (2002)

indices or extreme stock market returns did not exhibit. Firstly, it measures actual investors' attention (or information demand) rather than information supply. Secondly, it is a measure that quantifies attention effectively and standardized over time compared to other measures. Thirdly, SVI data captures investors' attention in a very timely manner, as it records the actual moment when information demand is satisfied. In contrast, press indices merely capture the presence of information supply - but not the point in time when individuals actually make use of it. Lastly, it provides the opportunity to predict future events. The academic literature provides more and more cases where Google search data have been used to either "nowcast" or predict future events. One example: By looking at the search volume of search queries related to influenza, flu outbreaks could be successfully predicted more than a week before the Center for Disease and Control was able to report them (see Ginsberg et. al, 2009). Being able to predict future events can be extremely helpful in Finance. The process to form an investment decision, especially in the context of retail investors, consists of various steps which can elapse over a longer period of time. Namely, investors' attention is aroused, information demand is satisfied, investment decisions are made and lastly actions on financial markets take place. As we intend to capture the initial step of this process, our data is potentially predictive for future financial market activities. Consequently, we can analyze how today's investors' attention influences the stock markets tomorrow - and vice versa.

With our study, we contribute to the existing literature in various aspects. Firstly, we analyze the statistical bi-directional causality between investors' attention and various financial market related measures in a cross-country environment. Secondly, we do not only focus our analysis on Granger-causality, but also on the predictive power of investors' attention on various measures and vice versa. Thirdly, we relate our cross-country findings to various structural and behavioral measures. Thereby, we offer insights into the underlying patterns of country differences in investors' interactions with financial markets. Fourthly, the large sample allows us to investigate the time varying underlying factors of retail investors' behavior. Finally, we examine the relationship between investors' attention and stock market volatility for 22 US stocks with regard to differences in attention attributed to those stocks.

The remainder of the thesis is structured as follows: Section 2 provides an overview on related literature. Section 3 describes data sources, characteristics and transformation as well as the methodology applied. Subsequently, the results of the cross-country analysis on the index level as well as the results for 22 US titles on the stock level are described in the fourth Section. Section 5 discusses the results and relates them to existing literature. Furthermore, robustness and limitations of the model are presented in Section 6. Finally, Section 7 concludes.

2. Literature

Since our study investigates the interaction between retail investors and financial markets using an innovative measure of investors' attention, we focus on two main strains of literature. Firstly, we describe studies on retail investors' behavior and their influence in financial markets, which constitutes the academic foundation of the present work. We begin with outlining important research on the presence and influence of noise traders in market models. Then, we present academic evidence of linkages between noise traders and retail investors, as well as insights into characteristics and behavior of retail investors with regard to financial markets. Secondly, we illustrate how Google Search Volume has been applied to various research questions, specifically as a measure of investor attention.

The role of noise traders in financial markets has been investigated since the early 1980s. Black (1986) characterizes noise traders as market participants who do not possess superior information, but trade on what they perceive as superior or new information. Kyle (1985) sets up a market auction model that shows how noise traders lead to a camouflage effect for insiders that enables the latter to make profits on the former's expense. On this basis, De Long et al (1990) conclude that noise trading may lead to a high deviation of prices from their fundamental values. However, noise traders are able to realize higher returns due to the additional risk they create and therefore survive in financial markets. The presence of noise traders in a financial market has also implications for market liquidity, volatility and trading volume. The models of Black (1986), Kyle (1985), DeLong et al (1990), Campbell and Kyle (1993) and Wang (2010) indicate a highly positive relation between the presence of noise traders in a market and the market's volatility. Further, in their stochastic multi-agent model of financial markets, Lux and Marchesi (1999) find that volatility causes investors' attention which in turn causes additional volatility[4]. In the models of Black (1986), Merton (1987) and Wang (2010), a positive relation between noise traders and market liquidity can be concluded as well. Additionally, in an experimental market simulation, Bloomfield et al (2009) find that bid-ask spreads are considerably lower when uninformed traders are participating in a market. Moreover, Bloomfield et al (2009) and Wang (2010) find that uninformed noise traders lead to significantly higher trading volume in financial markets.

To relate the present work to academic literature of noise trading, we make the assumption that retail investors to a certain extent behave like the noise traders do in the various models. In fact, several studies link retail investors' behavior to noise traders' behavior (Kumar and

---

[4] For empirical confirmation see Dimpfl and Jank (2011)

Lee, 2006; Foucault et al, 2011; Kelley and Tetlock, 2013). For example Foucault et al (2011) summarize previous studies regarding individual investors' behavior in the sense that non-informational reasons are the foundation of trading for retail investors. Other studies show that behavior among retail investors is highly homogeneous (Kumar and Lee, 2006; Barber et al. 2009a). As a consequence, they can drive asset prices away from their fundamental values and influence stock returns. For example Kumar and Lee (2006) show that retail investor's behavior is "systemically correlated"[5] and that retail investors act on a financial market similarly. Barber et al (2009) relate mispricing based on retail investors' behavior to stocks with high idiosyncratic volatility and Hvidkjaer (2008) relates retail investors' behavior to return patterns. Literature empirically investigating the link between retail investors and stock market volatility indicates a positive effect. For example Foucault et al (2011) show that retail investors' behavior has a positive effect on stock market volatility, which suggests that retail investors act like noise traders.

Barber and Odean (2011) document various abnormal behavioral patterns of retail investors. They find that individual investors "underperform standard benchmarks, sell winning investments while holding losing investments, are heavily influenced by limited attention and past return performance in their purchase decisions, engage in naïve reinforcement learning by repeating past behaviors that coincided with pleasure while avoiding past behaviors that generated pain, and tend to hold undiversified stock portfolios."[6] In their study they identify various cognitive biases including overconfidence, familiarity bias, disposition effect and attention bias. Again Barber and Odean (2008) show that retail investors, contrary to institutional investors, tend to only buy shares which caught their attention in the first place. This indicates a general pattern, such that the attention of retail investors to financial markets in general seems to be a very important factor for their future actions. Therefore, being able to measure investors' attention accurately and categorize it correctly is crucial for investigating the interaction between retail investors and financial markets. Among others, Da et al (2012) discuss the Search Volume Index as measure for investors' attention and attribute several positive characteristics to it. By comparing retail order executions from SEC Rule 11Ac1-5 they confirm that the SVI actually measures retail investors' attention and exhibits forecasting abilities.

Google's Search Volume Index has become increasingly popular among scientific researchers in recent years. In 2009, Hyunyoung Choi and Hal Varian, two Google employees,

[5] See Kumar and Lee (2006), page 2452

[6] See Barber and Odean (2011), page 2

published a working paper with the intention to encourage the use of search volume data provided by Google for academic studies. They exemplary show that forecast models for retail sales as well as macroeconomic indicators can be improved by taking into account Google Trends data (Choi and Varian, 2009a, 2009b, 2012). Since then, many studies followed their example.

In Finance academia, Google's SVI has mainly been used as a measure of investors' attention and information demand. For instance, Da et al (2011) link Google search queries to retail investors' attention as sophisticated investors likely use other sources of information. By examining search queries of terms related to finance and economics, several studies find a positive relationship between Google search volume and trading volume on stock markets (Bank et al, 2011; Preis et al, 2010; Vlastakis and Markellos, 2012). Moreover, a relationship is discovered between search volume and equity market liquidity (Bank et al, 2011), as well as realized historical (Dimpfl and Jank, 2011, 2012) and implied volatility (Vlastakis and Markellos, 2012). Analyzing the expected variance risk premium, Vlastakis and Markellos (2012) find evidence for the hypothesis that information demand increases with investors' level of risk aversion. An interrelation between search volume and volatility can also be found in non-equity financial markets (Kita and Wang, 2012). Other studies relate Google search volume of particular search terms to increased returns and construct trading strategies to exploit findings that inhibit historical positive alpha (Bank et al, 2011; Da et al, 2011, 2013; Preis et al 2013).

Our study is most closely related to this of Dimpfl and Jank (2011, 2012[7]). In line with our methodology, they use Google search volume to analyze causal relationships between investors' attention and realized volatility of prime equity indices in different countries. Their study not only concludes a high correlation between search volume and realized volatility of prime equity indices, but also a causal relationship. However the focus of the authors is to improve realized volatility forecasts by including information on search volume in volatility forecast models, whereas our focus is to shed light into cross-country variations and to give reasonable explanations for them. Furthermore, our focus is to provide empirical evidence for noise trader models' predictions about investors' interaction with financial markets regarding not only realized volatility, but also implied volatility, illiquidity and trading volume.

[7] Dimpfl and Jank published two different versions of their working paper (2011, 2012). In the following, when referring to their study, we will quote the former, more comprehensive version (2011).

3. Data and Methodology

3.1 Data

The indices in our sample include the Dow Jones Industrial Average (DJIA) for the United States, DAX 30 for Germany, FTSE 100 for the United Kingdom, CAC 40 for France, AEX for the Netherlands, Nikkei 225 for Japan, Bovespa for Brazil, IBEX 35 for Spain and FTSE MIB for Italy. The time period ranges from January 2006 to September 2013 and is further split up into three subperiods. Additionally, we take a closer look at the stock level and conduct the same analysis on 22 selected stocks from the NYSE and NASDAQ stock exchanges in the time horizon ranging from January 2008 until September 2013. The selection of indices and stocks analyzed in this paper, as well as the time horizon, is determined by data availability and the quality of data available.

Google Search Volume Index (SVI)

We use Google search volume as a measure of retail investors' attention and information demand. Google Trends provides the "Search Volume Index" (SVI) - a daily updated index measuring the popularity of search queries on the Google search engine. However, the SVI cannot directly be linked to the absolute number of search queries for specific terms. Rather, the measure is constructed such that for a specific search term, a value of 100 is assigned to the point in time when the number of searches for this term was largest compared to the total number of overall searches on Google in the requested time period. Then, all other data points are distributed between 1 and 99, based on the number of searches relative to overall search volume, although in relation to the peak of searches. A value of 0 is assigned to time points when search volume does not exceed a certain minimum threshold defined by Google[8].

The SVI time series are available on a daily basis beginning in January 2004 and can be split up regionally on a country, region and metropolitan level. However, for search terms with a low search volume, the SVI might only be provided on a weekly or even monthly basis. It is worth noting that Google defines one day as the time period ranging from 12:00:00 a.m. until 11:59:59 p.m. Pacific Standard Time[9]. This fact biases our results in the way that daily search volume in Central Europe and the United Kingdom is defined as submitted search queries between 09:00:00 a.m. until 08:59:59 a.m. and 08:00:00 a.m. until 07:59:59

[8] For further details on how Google's Search Volume Index is constructed see https://support.google.com/trends

[9] Pacific Standard Time is defined as UTC-08:00 and applied at the United States and Canadian West Coast, for example in Mountain View, California, where Google Inc. is headquartered.

a.m., respectively[10]. We believe that this time shift does not bias the validity of our results on a wide margin, as stock exchanges in Europe are closed before 8:00 a.m., which implies that no new information transmitted through stock prices will arrive in advance on a given day. A bigger problem, however, is the time shift in Japan. As daily search volume is defined as the number of searches between 05:00:00 p.m. and 04:59:59 p.m., it comprises searches submitted partly on two consecutive days. Therefore, we have to be careful when analyzing and interpreting results obtained for Japan. For Brazil, the effect again is negligible as time shift is even less than for the European countries.

Furthermore, daily SVI time series can only be extracted for a maximum of 90 consecutive days. Hence, to construct longer daily time series, we have to download weekly data, which is available for the whole sample period, and daily data for each 90 day interval within. We then merge the datasets such that the search volume of each day relative to the average of the corresponding 7 days period is multiplied with the weekly search volume. By following this procedure, we are able to construct a daily measure reflecting all variations within a week, but also the relative value of the week with regard to the whole sample period. By construction, the SVI data provided by Google already account for a general time trend of the rising popularity of Google and the Internet between 2004 and today. However, only if relative search volume of stock market related queries to overall search volume remains constant, this construction issue is able to solve a potential data bias.

To extract SVI time series for the nine indices, we use their most commonly used abbreviations, for example "DAX" for the DAX 30 of Germany[11] (see Table 9). To avoid ambiguous search queries and other possible noise in our data, we filter the SVI by the corresponding country of each index. For instance, for the DAX 30, we only use search volume originated in Germany. This assumption does not undermine our results as several studies have provided evidence of a significant home country bias for equity investors. Further, the home bias is particularly severe for retail and unsophisticated investors[12]. In addition, we control for ambiguous search terms with Google Correlate[13], which provides the terms with the highest correlation in search queries. By using Google Correlate, we ensure that the search terms we choose are not correlated with queries for terms not related to the specific stock, index or equity market. Past Google SVI data can be requested up to January 2004. However, as the first

[10] Due to the difference of dates when time is shifted to summer time, there are six weeks per year with further time differences between the United States and Central Europe

[11] Single exceptions to this rule is Italy, where the index name is ambiguous and might be confused with a popular movie title

[12] See Obsfeld and Rogoff (2000); French and Poterba (1991) among others, or Jacobs and Weber (2012) for relation to Google search volume

[13] See http://www.google.com/trends/correlate

periods contain many missing values, we start our analysis in January 2007 for the Nikkei 225 of Japan and in January 2006 for all other indices.

To obtain the SVI for the stocks in our sample, we proceed analogously to the indices. Former studies have already focused on the SVI of U.S. stocks, analyzing search queries related to companies' names or stock ticker symbols. In our opinion, both approaches might lead to suboptimal results in our case as we do not expect retail investors to know all stock ticker symbols. In addition, search volume for a company's plain name might be influenced by information demand for its products and services[14]. Consequently, we extract Google SVI for the combination of "stock" and the respective company's most common name (see Table 9). However, for many search terms only weekly SVI data is available, so we proceed with our analysis on a weekly basis for stocks. The selection of the 22 stocks in our sample is determined by availability and validity of SVI data. In total, we analyze search volume for more than 50 US stocks, but for many of them the Search Volume Index exhibits either many missing values or an extreme time trend.

Realized Volatility (RV)

For the realized volatility time series of the nine equity indices under investigation, we rely on publicly available data provided by the Oxford-Man Institute of Quantitative Finance at the University of Oxford[15]. The daily realized volatility time series are calculated using intraday stock market data (computed every minute) aggregated on a ten-minute basis. However, the Oxford-Man Institute does not provide those data for our U.S. stock sample. Here, we apply a naïve measure of volatility and simply sum up the squared daily log returns of each stock over one trading week and take the square root:

$$\sigma_{w,i} = \sqrt{\sum_{d=1}^{w} {r^2}_{i,d}}$$

where $r_{i,d}$ is the log return of stock i on trading day d within week w for which the standard deviation is calculated. To obtain stock returns, we use daily closing prices of the particular stocks provided by ThomsonReuters Datastream. We are aware of the fact that our naïve measure of volatility might be very noisy. Yet, as we do not have access to intraday trading data of the stocks chosen and for reasons of consistency with the index analysis, we see it as an appropriate measure of realized volatility.

[14] For example the search query "Apple" is heavily biased by the demand for product and service information for Apple Inc.'s products; see also Da et al (2012)

[15] http://www.oxford-man.ox.ac.uk/

Implied Volatility (IV)

With the intention to analyze implied volatility, we rely on publicly quoted implied volatility indices, the VXJ for the Dow Jones, the VDAX-NEW for the DAX 30, the VFTSE for the FTSE 100, the VCAC for the CAC 40, the VAEX for the AEX and the VJX for the Nikkei 225. These indices calculate the annualized implied volatility (as standard deviation) from prices of out-of-the-money index put and call options over the next 30 days using the risk-free interest rate. The data series are extracted from Datastream, matching the daily observations for the SVI time series. Unfortunately, publicly quoted implied volatility time series are not available for the Bovespa, IBEX 35 and FTSE MIB, as well as the majority of stocks in our sample, so we constrain the analysis of interrelations between search volume and implied volatility on the six stock indices outlined above.

Trading Volume (TV)

Another measure of market activity we include in our study is daily trading volume, quoted in the respective country's currency. We take the difference in absolute share prices into account and calculate the daily trading volume for one stock i as the number of shares S traded on that particular day d times the stock's closing price P on the same day:

$$TV_{i,d} = S_{i,d}\, P_{i,d}$$

However, daily trading volume for all equities included in the particular index is only available for four indices on Datastream. Therefore, our analysis of retail investors' information demand and trading volume is limited on the DAX 30, Dow Jones, CAC 40 and AEX. For these indices, we sum up all stocks' daily trading volumes as defined above to construct an aggregated daily measure of trading volume for each index N:

$$TV_{n,d} = \sum_{i=1}^{I} TV_{i,d}$$

To match the weekly observations of search volume for the 22 stocks, we construct weekly measures of trading volume by simply summing up daily trading volume over one week w:

$$TV_{i,w} = \sum_{d=1}^{w} TV_{i,d}$$

We are aware of the fact that not all equity trading is conducted via stock market exchanges. However, the number of stocks traded on the main stock exchange of each country should be representative for the relative trading volume, especially over longer periods. Moreover, we are interested in the behavior of retail investors and expect those to use stock exchanges for buying and selling stocks as over-the-counter transactions are merely executed by institutional investors. If more than one exchange trading platform is available, we use the one with the highest daily trading volume[16].

Liquidity (ILLQ)

To examine the interrelation between search volume and market liquidity, we construct Amihud-Illiquidity measures for the equity indices and selected U.S. stocks (see Amihud, 2002). Originally, Amihud proposed this measure as a yearly indicator for illiquidity, taking the simple average over all daily illiquidity values of one year. In our study, we slightly modify the measure to yield daily (for indices) and weekly (for stocks) measures of illiquidity. Further, we interpret the results obtained in a context of liquidity, with low values of the constructed measure (i.e. low illiquidity) indicating a higher liquidity. We are aware of the fact that low illiquidity is not necessarily equal to high liquidity. However, as we interpret all of our results with regard to a specific and limited time period, high values of illiquidity indicate times of low market liquidity within the specific time period.

For measuring daily liquidity, other indicators as for example the bid-ask spread or the price response to signed order flows would probably give us more reliable estimates. However, as we do not have access to intraday and microstructure data, we have to rely on our rather naïvely constructed daily measure of illiquidity. Nevertheless, it is a good indicator for the daily price impact of the order flow as it measures the absolute percentage price change per dollar of trading volume for one specific day (see Amihud, 2002). For the 22 stocks, we construct the Amihud-measure as follows:

$$ILLIQ_{i,d} = \frac{|R_{i,d}|}{TV_{i,d}}$$

where $|R_{i,d}|$ is the absolute percentage return of stock i on day d calculated from closing prices. $TV_{i,d}$ is the daily trading volume for the respective stock i on day d, derived as shown in the section above. Again, all inputs are obtained from Datastream. For the equity indices, we calculate the illiquidity measure for each single stock and day. To obtain an aggregated measure of illiquidity on day d over all stocks I in index n, we take the equally-weighted average of all stocks' daily illiquidity measures:

$$ILLIQ_{n,d} = \frac{1}{I}\sum_{i=1}^{I}\frac{|R_{i,d}|}{TV_{i,d}}$$

[16] In order to obtain closing prices and trading volume of the shares in the DAX 30 index, we rely on data from the XETRA trading system

However, as already mentioned in the section above, trading volume is only available for four indices, so we are limited in constructing illiquidity measures to the DAX 30, Dow Jones, CAC 40 and AEX.

Sample Selection

To match the search volume data with the time series of volatility, illiquidity and trading volume for each index, we exclude all non-trading days and all days with missing trading volume or volatility data. As a consequence, each of the time series related to the same index exhibits the same length. However, as public holidays vary between the selected countries, time series corresponding to different indices may vary in their length. Ranging from January 2006 (January 2007 for the Nikkei 225) until the end of September 2013, we have extracted around 1250 daily observations for each index (950 for the Nikkei 225).

In addition to analyzing the whole time period, we split our sample in three roughly equal time periods to analyze how the interrelation between investors' attention and market activity evolves over time. The first subsample ranges from the beginning of January 2006 (January 2007 for the Nikkei 225) until the end of December 2007. The second subsample ranges from the beginning of January 2008 until the end of December 2010 and the third subsample comprises the period between the beginning of January 2011 until the end of September 2013. This segmentation of time periods enables us to analyze the impact of the financial crisis (the second subsample contains the peak period of the financial crisis) and of the Eurozone crisis ("post-crisis period", third subsample). When looking at the realized volatility time series, we observe that the crisis period exhibits the highest average volatility of the three subperiods, followed be the post-crisis period. However, we do not divide the time series data of stocks into different subsamples. Due to the fact that only weekly data can be used, we are left with no more than 300 observations for each stock. An even smaller subsample would severely constrain the significance and validity of the results.

3.2 Methodology

Data Transformation and Summary Statistics

Our main goal is to shed light on the interrelation between investors' attention measured by Google's SVI on the one hand and realized and implied volatility, as well as liquidity and trading volume on the other hand. As these measures are scaled differently, we normalize each of the time series by dividing every single value within by the series' simple mean, thereby setting the long term average of the whole sample period equal to one:

$$Indices\text{: } Y_{i,d}(norm) = \frac{Y_{i,d}}{\frac{1}{D}\sum_{d=1}^{D} Y_{i,d}} \quad for\ measure\ Y_i\ on\ day\ d\ for\ d = 1, \ldots, D$$

$$Stocks\text{: } Y_{i,w}(norm) = \frac{Y_{i,w}}{\frac{1}{W}\sum_{w=1}^{W} Y_{i,w}} \quad for\ measure\ Y_i\ in\ week\ w\ for\ w = 1, \ldots, W$$

Each normalized value can now be interpreted as the multiplier of the corresponding original value with the long term average of the original series, i.e. a value of 2.5 in the normalized series corresponds to an original value that is 2.5 times as high as the original sample's mean.

To further evaluate the issue of non-normality, we run Jarque-Bera (JB) tests (see Jarque and Bera, 1987) on all measures, testing normality of the time series. The test's null hypothesis is that skewness and kurtosis of the series resemble those of a normal distribution. The test statistic follows a χ^2-distribution and is defined as:

$$JB_{stat} = t_{S}^{2} + t_{K}^{2} \sim X^2(2) \quad where\ t_S = \frac{\hat{S}}{\sqrt{\frac{6}{T}}} \quad and\ \ t_K = \frac{\hat{K} - 3}{\sqrt{\frac{24}{T}}}$$

with sample size T, skewness S and kurtosis K. Another necessary condition for the vector autoregression we want to implement is stationarity of the input time series. Non-stationarity could severely bias the significance of the parameters obtained (see Brooks, 2008). A time series is defined to be stationary if its first and second moments are time-invariant. Therefore, we test all time series for stationarity and run an augmented Dickey-Fuller (ADF) test (see Elliott et al, 1996) on unit root non-stationarity. For the ADF-test, we have to run the regression

$$\Delta y_t = \alpha_0 + \rho y_{t-1} + \gamma t + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p} + \varepsilon_t$$

with a constant α representing a drift, a deterministic time trend t and the lag order of the autoregressive process p. Then, we are able to estimate the regression parameters with ordinary least squares (OLS) or maximum likelihood (ML) and finally derive the test statistic:

$$ADF_{stat} = \frac{\hat{\rho}}{\sqrt{\widehat{Var}(\hat{\rho})}}$$

The optimal lag length is estimated by Schwartz's Bayesian information criterion (SBIC, see Schwarz, 1987) with the restriction on maximum lag length of 100 for time series associated with the nine indices and 15 for time series associated with the 22 stocks. However, this statistic follows a special tabulated distribution (see Said and Dickey, 1984). The null hypothesis is $\rho = 0$ and is rejected if the test statistic is smaller than the tabulated value at a specific significance level. The third test we run is the Ljung-Box (LB) test (see Ljung and Box, 1978) on autocorrelations of the time series' lags. The LB-test tests a specified number of lags on their autocorrelations being zero. The test statistics looks as follows, with k lags, a maximum lag length of m and τ being the autocorrelation coefficients:

$$LB_{stat} = T(T+2)\sum_{k=1}^{m}\frac{\hat{\tau}_k^2}{T-k} \sim\chi_m^2$$

Again, we use a maximum lag length of 100 for time series corresponding to indices and 15 for those corresponding to stocks. As the LB-test returns p-values for every lag, we evaluate the sum of all p-values.

When looking at the summary statistics, we can see that all but four time series corresponding to indices and six corresponding to stocks are stationary at the 5% significance level (see appendix: Table 10-11 for indices and Table 22-25 for stocks). However, the time series in the sample are not normally distributed and exhibit excessive skewness and kurtosis. Therefore, we proceed with taking simple logarithms of all the time series, as proposed by former studies working with Google SVI (Vlastakis and Markellos, 2012; Da et al, 2011; Dimpfl and Jank, 2011). As we can see in the summary statistics, the log-series are still not close to a normal distribution, but excess skewness and kurtosis is reduced significantly. Moreover, we can reduce the number of non-stationary time series. Particularly, as we primarily focus on volatility series, taking logs leads to significantly improved stationarity for implied volatility time series and reduces the non-stationary of realized and implied volatility. In the following sections of this paper, when talking about Google search volume index, realized and implied volatility, as well as illiquidity and trading volume, we are always referring to the logarithmic series of those measures.

Vector Autoregressive Model (VAR)

We continue by setting up the vector autoregressive models (VAR-models) to examine the interrelations between the logarithmic time series. VAR-models are multidimensional generalizations of autoregressive model and allow variables to depend on their own previous values (or "lags"), as well as on other variables' lags and error terms (see Brooks, 2008). We choose to run a bivariate vector autoregression with two variables, the Google SVI on the one side and volatility, illiquidity or trading volume on the other side. We rely on the VAR-methodology because of the increased flexibility compared to the simple regression model. Following Lux and Marchesi (1999), we expect both series to be interrelated and not linearly dependent on each other. The VAR-methodology allows us to run Granger-causality tests in order to test statistical causality. As it is our main goal to comment on the portion of retail investors' influence on equity markets, the VAR-methodology also allows us to conduct forecast error variance decompositions and to construct impulse response functions. For every measure in each country and stock, we set up a VAR-model of the following structure:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \phi_{10} \\ \phi_{20} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \alpha_{11} \\ \beta_{21} & \alpha_{21} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \cdots + \begin{pmatrix} \beta_{1k} & \alpha_{1k} \\ \beta_{2k} & \alpha_{2k} \end{pmatrix} \begin{pmatrix} y_{1,t-k} \\ y_{2,t-k} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix}$$

$$with\ Var(u_{z,1}) \neq Var(u_{z,2}) \neq \cdots \neq Var(u_{z,i}) \quad and\ Cov(u_{z,t},\ u_{z,t'}) \neq 0\ for\ t \neq t'$$

where $y_1$represents either one of the logarithmic time series of volatility, illiquidity or trading volume and y_2 always represents the log SVI. We also include constants (ϕ_{10}, ϕ_{20}) and error terms $(u_{1,t}, u_{2,t})$ in our model. We further define our model to allow for heteroscedastic and correlated error terms (see Sheppard, 2009). The general bivariate VAR-model is capable of including up to k lags of each variable. In contrast to affiliated studies, we do not predetermine the lag length of our model. Rather, we rely on information criteria to define the optimal lag length k for each model individually (see Brooks, 2008). One advantage of using information criteria over other estimation methods is that they do not require the error terms to be normally distributed, an assumption that would be unlikely in our case. We choose to rely on the multivariate version of Schwarz's Bayesian information criterion (MSBIC, see Schwarz, 1987) and allow for a maximum lag length of ten. One advantage of the MSBIC is that it asymptotically delivers the correct model order and generally proposes smaller lag length than Akaike's information criterion (MAIC) (see Schwarz, 1987; Brooks, 2008). The MSBIC is defined as:

$$MSBIC = ln|\hat{\Sigma}| + \frac{k' \ln(T)}{T}$$

with $\hat{\Sigma}$ being the variance-covariance matrix of residuals of the estimated VAR-models. We determine the MSBIC for VAR-models up to a lag length of 10 and proceed with the model that gives us the smallest value.

After estimating the optimal lag length and fitting the corresponding VAR-models, we proceed with analyzing the interrelations between search volume and market activity. In order to conclude about the significance of parameters obtained by fitting the VAR-models, we conduct student's t-tests (see Brooks, 2008):

$$t_{stat}(i,j) = \frac{\gamma_{ij}}{SE(\gamma_{ij})} \sim t_{T-p}$$

The p-values obtained by conducting t-tests with the null hypothesis of the coefficient being equal to 0 indicate the significance with which the coefficients enter the VAR-equations. In combination with the signs of the estimated coefficients, we can draw conclusions about the significance and direction of a correlation between the dependent variable and the corresponding lags.

Granger-Causality Test

Although p-values of t-tests can give us insights into which parameters are significant in the VAR-model, their results might be misleading when estimating the overall effect of the explanatory variables and their lags on the dependent variable. Particularly for models with multiple lags and changing signs of the corresponding parameters, we are not able to draw reliable conclusions about the overall influential effect of the lags of one variable on the contemporaneous term of another variable within the VAR-model. Therefore, we conduct Granger-causality tests (see Granger, 1969) on the parameters of our VAR-models. A Granger-causality test attempts to answer the question whether lagged values of y_2 cause statistically significant changes in $y_{1,t}$ (and vice versa) in the presence of the lagged values of y_1. In this case, we conclude that y_2 Granger-causes y_1 (and vice versa). Regular Granger-causality tests with the assumption of homoscedastic and uncorrelated error terms are conducted using the likelihood ratio and a test statistic that follows an F-distribution (see Brooks, 2008) However, as we assume heteroscedasticity and correlation of the error terms, we have to use a Wald test where the parameter covariance matrix is estimated under the assumptions of heteroscedasticity and correlation of the residuals to compute the Granger-causality test statistics (see Lütkepohl, 2005; Sheppard, 2009). This test statistic follows an asymptotic χ^2-distribution. As the null hypothesis of the Granger-causality test is no causality, lower p-values provide evidence of a larger causal relationship. However, it has to be noted that Granger-causality does not imply that a stringent economic causality is existent. Rather, a pure statistical causality is present as Granger-causality simply describes a significant correlation between the current value of the dependent variable and past values of the explanatory variables.

Impulse Response Function (IRF)

By analyzing the Granger-causality test statistics and comparing the p-values, we can draw conclusions about which variables have statistically significant influence on others. On the contrary, Granger tests will not reveal the sign of this relationship, i.e. if there is a positive or negative Granger-causal relationship between the variables. The test also does not reveal the persistence of the effect, i.e. how long it lasts over time. To shed further light on these issues, we construct Impulse Response Functions (IRFs) to test how the variables respond to shocks in the error terms ($u_{1,t}$ and $u_{2,t}$). IRFs trace out the responsiveness of the respective dependent variable in every single VAR equation to shocks in the error terms under the assumption that everything else is held constant (see Brooks, 2008). First, a unit shock is applied to one of the error terms. Then the effect on the dependent variables is noted over a predefined time period. As a shock in the error term will be transmitted to all dependent variables through the dynamic VAR structure, the effect can be observed for every single time step (or "lead") $n = 1, 2, \dots k$ over the specified time horizon k. To construct the IRFs, we use Cholesky decomposition and define our measures of market activity (volatility, illiquidity and trading volume in the specific model) as contemporaneously exogenous (see Lütkepohl, 2005). Technically, this means that a shock in volatility, illiquidity or trading volume (i.e. the error term corresponding to these measures) can affect search volume instantaneously when the shock is applied. Contrarily, search volume cannot affect volatility contemporaneously, so an effect of the SVI on market activity can be observed the earliest at lead one. Economically, the intuition is that retail investors' attention and thus search volume is triggered by increased market activity, which is in line with the noise trader models of Lux and Marchesi (1999) as well as with the applied methodology of Dimpfl and Jank (2011). When interpreting the IRFs, we have to keep in mind that a unit shock is applied under the assumption that all other terms are held constant, which gives us valuable insight into the behavior of our models although this assumption is not necessarily realistic in a practical application.

Forecast Error Variance Decomposition (FEVD)

To answer the question how much of the variation of equity market volatility, illiquidity and trading volume can be attributed to variations in search volume and vice versa, we construct Forecast Error Variance Decompositions (FEVD, see Lütkepohl, 2005; Brooks, 2008). FEVD tells us what proportion of the forecast error variance of the dependent variable for

$n = 1, 2, \dots k$ forecast steps (“leads”) can be explained by shocks in each explanatory variable. Again, we order the variables such that increased market activity triggers search volume, what in turn might trigger more market activity. In the bivariate case applied in this study, the FEVD splits up the percentage of the variance in each dependent variable that is caused by shocks on its own error term versus the percentage of variance that is caused by shocks on the other variable’s error term (see Lütkepohl, 2005; Brooks, 2008). Technically, the result tells us how much of the movement in volatility, illiquidity or trading volume are due to movements in the SVI (or vice versa) over a specific time period. Economically, the results can be interpreted as the quantity of information each of the two variables contributes to search volume and market activity separately. In our analysis, we will focus on the long-term forecast error variance decomposition and define “long term” as 100 forecast steps (corresponding to 100 days) for indices and 20 forecast steps (corresponding to 20 weeks) for stocks.

4. Results

To investigate the interaction between retail investors' behavior proxied by their information demand measure (SVI) and financial markets, we examine two main environments (stock market indices and individual stocks), four different financial market related measures (realized volatility, implied volatility, trading volume and illiquidity) and three different time periods (pre-crisis, crisis and post-crisis). This section is organized as follows: We briefly summarize all results of our analysis before we report the results in more detail. Firstly, the results of the VARs (Granger-causality tests and variance decompositions) for SVI and stock market indices are presented. Thereby, we show the results of the interaction between SVI and realized volatility, followed by implied volatility, trading volume and illiquidity – each for the complete sample and for the different time periods. Secondly, we present the results on the individual stock level for 22 US stocks. Thereby, we will focus on the Granger-causality tests for realized volatility followed by an analysis of the interaction between SVI and the other measures (trading volume and illiquidity) and differences regarding search volume.

Overall, our results suggest a bi-directional statistical causality between SVI and realized volatility for all indices but the IBEX 35 and the FTSE MIB. This means that realized volatility in previous lags is statistically correlated with today's SVI and SVI in previous lags is correlated with realized volatility today. Realized volatility for a given stock market index induces increased information demand of retail investors which in turn causes increased volatility for this particular stock market index. These results are in line with a study by Dimpfl and Jank (2011) for a shorter period of time and a subset of indices. Furthermore, the interpretation of these results goes hand in hand with the conclusions of several noise trader models including Lux and Marchesi (1999). For different sample periods, we find further evidence that market fluctuations (i.e. high volatility) seem to play a crucial role for both information demand triggered by volatility and volatility triggered by investors' attention. Granger-causality results are most pronounced in the crisis period, less in the post-crisis period and least in the pre-crisis period. The variance decomposition over a forecast period of 100 days yields differing results for the indices, ranging from about 17% variance in realized volatility explained by variation in the SVI for the CAC 40 to 1% in the case of the FTSE MIB. Economically speaking, 17% of realized volatility in a forecast period of 100 days is explained by retail investors' behavior. The relationship in the other direction, i.e. how much of investors' attention can be attributed to past volatility, shows different patterns for the each country. These observations indicate that there must be differing underlying patterns across countries determining the respective retail investors' behavior. We will present potential underlying causes in Section 5.

Using implied volatility as the relevant market measure gives a slightly different picture. Effects in both directions are evident for all indices but the DJIA and the NIKKEI 225. Here, volatility causes an increase in search volume but retail investor attention does not feedback on implied volatility. Furthermore, implied volatility explains a much greater share in investors' attention than realized volatility. We will also give potential explanations for this observation in Section 5. The results for trading volume and illiquidity show similar patterns: Bi-directional causality can be found for most indices over the whole sample period. Results for illiquidity are most significant during the crisis whereas for trading volume, the pattern of significant Granger-causal relations changes over the three subsample periods and is rather inconclusive.

Results for stocks show that realized volatility causes investor attention for a subsample of stocks. These are mainly the stocks with most overall investor attention, indicating that retail investors attribute most of their attention to a limited number of stocks. Interestingly, investor attention does not significantly Granger-cause realized volatility on the stock level.

4.1 Indices

4.1.1 Realized Volatility – Complete Sample Period

VAR Parameters and Granger-Causality

In the complete sample period, bi-directional causality can be inferred from the p-values of the Granger-Causality tests for almost all indices: Past SVI helps to explain present realized volatility (RV) and past RV helps to explain present SVI. P-Values indicate significance in both directions at the 1% level for the DAX 30, DJIA, FTSE 100, CAC 40, AEX and Bovespa and at the 10% level for the NIKKEI 225 (see Table 1). Results for the IBEX 35 and the FTSE MIB differ somewhat to the extent that RV also Granger-causes SVI (p-values of 1% and 5% significance level respectively), but no statistical significant relationship can be found in the other direction. VAR coefficients in both directions are greatest in value for the first lag of all indices but the FTSE MIB and significant for almost all indices. Coefficients for the other lags of higher order are smaller in value and only partly significant (see Tables 12-15 for VAR-parameter estimations in different lags). The results indicate that retail investors react quickly to increased volatility with increased information demand and that volatility in turn increases quickly after retail investors' information demand increases. Overall, the results suggest that retail investors from all countries in the sample satisfy their increased information demand caused by market movements. However, in contrast to other countries, in Spain and Italy retail investors do not get active in the markets or at least do not produce additional volatility. In other words, realized volatility in Spain and Italy helps to explain the information demand from investors but their information demand is not predictive for future volatility.

Table 1: Realized volatility: Granger causality test statistics

The table presents the results obtained by conducting Granger causality tests on log Google search volume index (SVI) and log realized volatility series (RV) for the nine equity indices. Granger-Statistics are given for the significance of the explanatory variable statistically causing the response variable. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding test statistic.

		DAX 30		Dow Jones		FTSE 100		CAC 40		AEX	
Response variable		RV	SVI	RV	SVI	RV	SVI	RV	SVI	RV	SVI
Explanatory variable		SVI	RV	SVI	RV	SVI	RV	SVI	RV	SVI	RV
Complete period	Granger-Statistic	18.54	16.08	18.37	25.87	40.76	25.42	38.63	33.91	24.12	24.74
	p-value	***.002****	***.007****	***.001****	***.000****	***.000****	***.000****	***.000****	***.000****	***.000****	***.000****
Pre-crisis	Granger-Statistic	2.89	6.64	1.62	5.72	9.19	6.22	4.47	13.75	3.23	4.83
	p-value	*.408*	***.084****	*.446*	***.057****	***.056****	*.184*	*.107*	***.001****	*.199*	***.089****
Crisis	Granger-Statistic	12.29	15.87	8.04	5.42	27.66	21.06	21.60	8.93	23.29	13.36
	p-value	***.006****	***.001****	***.018****	***.066****	***.000****	***.000****	***.000****	***.030****	***.000****	***.004****
Post-crisis	Granger-Statistic	8.02	7.27	9.36	15.82	12.18	3.55	19.19	14.58	8.31	11.79
	p-value	***.046****	***.064****	***.009****	***.000****	***.002****	*.169*	***.000****	***.002****	***.040****	***.008****

		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
Response variable		RV	SVI	RV	SVI	RV	SVI	RV	SVI
Explanatory variable		SVI	RV	SVI	RV	SVI	RV	SVI	RV
Complete period	Granger-Statistic	10.66	16.31	22.27	24.09	5.34	44.43	7.58	12.32
	p-value	***.059****	***.006****	***.004****	***.002****	*.376*	***.000****	*.181*	***.031****
Pre-crisis	Granger-Statistic	0.04	1.45	12.13	9.95	5.30	14.62	0.93	4.08
	p-value	*.841*	*.228*	***.007****	***.019****	*.258*	***.006****	*.627*	*.130*
Crisis	Granger-Statistic	3.40	40.62	15.92	5.78	12.11	29.93	5.72	4.43
	p-value	*.183*	***.000****	***.001****	*.123*	***.017****	***.000****	*.126*	*.218*
Post-crisis	Granger-Statistic	3.84	1.52	6.48	9.74	4.90	8.62	2.91	3.14
	p-value	*.279*	*.677*	*.166*	***.045****	*.298*	***.071****	*.573*	*.535*

Variance Decomposition and Impulse Response Functions

The above results naturally raise the question of how much of future realized volatility can be attributed to retail investors and how this manifests over time. To answer this question, we look at the long term variance decomposition of our VAR models and the impulse response functions. Table 2 reports the values for the variance decomposition in both directions over a forecast period of up to 100 days.

Table 2: Realized volatility: Variance decomposition

The table presents the results obtained by conducting forecast error variance decomposition on log Google search volume index (SVI) and log realized volatility series (RV) for the nine equity indices. The values indicate how much the vertical variable (left axis) explains in the variance of the horizontal variable (above axis) at forecast step 100.

		DAX 30		Dow Jones		FTSE 100		CAC 40		AEX	
		RV	SVI	RV	SVI	RV	SVI	RV	SVI	RV	SVI
Complete period	RV	0.927	**0.189**	0.888	**0.415**	0.908	**0.243**	0.831	**0.230**	0.911	**0.276**
	SVI	**0.073**	0.811	**0.112**	0.585	**0.092**	0.757	**0.169**	0.770	**0.089**	0.724
Pre-crisis	RV	0.975	**0.034**	0.995	**0.219**	0.975	**0.197**	0.982	**0.099**	0.977	**0.067**
	SVI	**0.025**	0.966	**0.005**	0.781	**0.025**	0.803	**0.018**	0.901	**0.023**	0.933
Crisis	RV	0.865	**0.295**	0.906	**0.420**	0.809	**0.295**	0.758	**0.193**	0.840	**0.318**
	SVI	**0.135**	0.705	**0.094**	0.580	**0.191**	0.705	**0.242**	0.807	**0.160**	0.682
Post-crisis	RV	0.910	**0.258**	0.932	**0.364**	0.911	**0.138**	0.852	**0.231**	0.892	**0.215**
	SVI	**0.090**	0.742	**0.068**	0.636	**0.089**	0.862	**0.148**	0.769	**0.108**	0.785

		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
		RV	SVI	RV	SVI	RV	SVI	RV	SVI
Complete period	RV	0.935	**0.091**	0.922	**0.018**	0.975	**0.225**	0.992	**0.018**
	SVI	**0.065**	0.909	**0.078**	0.982	**0.025**	0.775	**0.008**	0.982
Pre-crisis	RV	1.000	**0.081**	0.952	**0.034**	0.997	**0.079**	0.996	**0.011**
	SVI	**0.000**	0.919	**0.048**	0.966	**0.003**	0.921	**0.004**	0.989
Crisis	RV	0.988	**0.141**	0.801	**0.188**	0.916	**0.298**	0.990	**0.007**
	SVI	**0.012**	0.859	**0.199**	0.812	**0.084**	0.702	**0.010**	0.993
Post-crisis	RV	0.959	**0.019**	0.988	**0.048**	0.975	**0.154**	0.996	**0.049**
	SVI	**0.041**	0.981	**0.012**	0.952	**0.025**	0.846	**0.004**	0.951

Three interesting observations can be made by looking at the variance decomposition. Firstly, the amount of realized volatility that can be attributed to retail investors (through their information demand and consecutive market behavior) varies severely among the countries analyzed. Rather high shares can be found for the CAC 40 (17%) and the DJIA (11%) followed by a "middle group" consisting of the FTSE 100 (9%), AEX (9%), the DAX 30 (7%), Bovespa (8%) and the NIKKEI 225 (6.5%) as well as a "low" group including the IBEX 35 (3%) and the FTSE MIB (1%). These results are in line with the non-significant Granger-causality tests for the IBEX 35 and the FTSE MIB: Spanish and Italian retail investors' information demand seems not to feed through to realized volatility in the markets.

Secondly, the amount of retail investors' information demand, which can be attributed to past realizations of volatility, varies dramatically between the countries. We can observe that for the DJIA RV explains about 42% of the variance in SVI, followed by a "middle" group consisting of the AEX (28%), FTSE 100 (24%), CAC 40 (23%), DAX 30 (19%) and the IBEX 35 (23%), as well as a "low" group including the NIKKEI 225 (9%), FTSE MIB (2%) and Bovespa (2%). These results are also mostly supported by the noise trader model (Lux and Marchesi, 1999) because they indicate that market movements induce noise traders' information demand. However, especially for Italian and Brazilian investors this seems not to be the case.

Thirdly there seem to be countries in which retail investors' information demand is triggered relatively less by market movements, but where their influence on market volatility is relatively high (see Table 2). For example, in the Netherlands, realized volatility seems to induce investors' information demand (27.6%), but retail investors' attention seems not to contribute to the explanation of future volatility (8.9%). French retail investors' attention is triggered to a great extent by RV (23.0%) but in contrast to Dutch retail investors their attention also seems to affect volatility severely (16.9%). This indicates that French retail investors either get active to a greater extent or move the markets to a greater extent than Dutch retail investors. Another striking observation is that Italian investors' activities are neither related to information demand nor to volatility in the markets.

To summarize, various noise trader models indicating that retail investors increase volatility are supported in Germany, France, the USA, the UK, the Netherlands, Brazil and Japan. However, behavior of Spanish and Italian investors differs. Italian investors' behavior cannot be linked to the model at all. In contrast, Spanish investors seem to satisfy their information demand when volatility is extreme, but do not cause additional volatility. Generally, the results of the variance decomposition shed further light on how and to what extent information about financial markets is processed and translated into action by retail investors in different countries. One potential reason explaining these patterns could be the limited impact of retail investors on their respective markets due to the low propensity of individuals to hold stocks. Further, the aforementioned patterns could also be influenced by the ratio of stocks that are privately held and therefore directly tradable by individuals as well as stocks held in mutual or pension funds. Moreover, differing behavioral characteristics across the analyzed countries might influence the way retail investors react to volatility in the markets. In this case, bounded rationality, risk aversion and uncertainty avoidance could be related to the creation of volatility as well. We will try to give plausible explanations of these patterns in Section 5 where we qualitatively relate the results to structural and behavioral country-specific characteristics.

Figure 2 exemplarily shows the impulse response functions (forecast period of 100 days) for the CAC 40, one of the indices where Granger-causality and variance decompositions are significant and among the highest in the sample, and for the FTSE MIB. In contrast to the CAC 40, the FTSE MIB is one of the indices where Granger-causality of SVI on RV is not statistically significant. Also, SVI explains very little of the variance in RV (and vice versa). Two main observations can be made with respect to the impulse response functions.

Firstly, the results of the variance decomposition can be confirmed. In particular, we can see that there is a response of realized volatility to a unit shock in SVI: First, volatility rises to a higher level, followed by a slight drop and another increase, which fades away over the forecast period. Regarding the response of SVI to a unit shock in realized volatility for the CAC 40, one can observe that SVI reacts strongly positive and reaches a higher level compared to inception which slowly fades away over the forecast period of 100 days. In contrast, the responses of search volume and realized volatility to shocks in the error terms for the FTSE MIB are very slow and exhibit little impact, even with changing signs for the first days for the reaction of search volume.

Secondly, the impulse response analysis gives insights in the persistence of the effects described. We can observe that the effect of a shock in either variable on the other variable is persistent for quite a long period and fades away after 40-50 days. Comparing the impulse response functions of the CAC 40 and the FTSE MIB further confirms that interaction between retail investors and realized volatility is less strong in Italy (see Figure 2).

4.1.2 Realized Volatility – Subsamples

Granger-Causality

The results for the different subsample periods (pre-crisis, crisis, post-crisis) offer some interesting insights about the interaction of retail investors with financial markets when market fluctuations exhibit different levels (see Table 1 for granger causality tests and Tables 12-15 for VAR-parameter estimations in different sample periods). While coefficients for Granger-causality are significant for almost all indices in the whole sample period, significance seems to partly vanish in sub-samples for the different indices and directions. In general, the effects in both directions are the most pronounced in the crisis period, slightly less in the post-crisis period and the least in the pre-crisis period. This is especially true for the

Granger-causality of investors' attention on realized volatility, meaning that for some countries retail investors' activities seem to cause additional volatility only in times of already volatile stock markets.

Before the crisis, SVI is Granger-causal for volatility only in one out of nine countries at the 1% level. Volatility causing investors' attention before the crisis is only significant at the 1% level for two out of nine indices. In the post-crisis and especially the crisis period, results show significant bi-directional Granger-causality for more indices. These findings are in line with previous studies (e.g. Dimpfl and Jank, 2011) which show that using the SVI to improve volatility forecasts is most successful when volatility is high. Furthermore, these findings are in line with the noise trader model of Lux and Marchesi (1999), where noise traders cause additional volatility following a volatility shock.

Non-significance of Granger-causal relationships in the different sub-periods (especially pre-crisis) as opposed to significance for the complete sample period can have various reasons. Firstly, interaction patterns between retail investors and financial markets during particular periods might simply not exist. Secondly, the smaller sample periods with fewer observations make it statistically less likely to detect an existing effect. Thirdly, the usage of Google as a source of information by investors could have been too low before 2008 to accurately reflect investors' attention, but has reached a critical value since then. Overall, we think that lower (at lower confidence levels) or no significance of Granger-causality in sub-periods implies that the effect in both directions is less pronounced or might be non-existing after all. Furthermore, the third explanation can be ruled out for the post-crisis sample period. This makes us confident that the Granger-causal relationship between search volume and volatility is less evident in non-crisis periods.

Crisis Period: Looking at the results for the crisis period in more detail, we can observe significance regarding the relation of RV on SVI for all indices but the Bovespa and the FTSE MIB. Compared to the other sample periods, Japanese investors' information demand seems to be triggered by volatility only in the crisis. This finding further strengthens our conclusion that besides of structural underlying patterns, especially current market fluctuations seem to influence the interaction of retail investors and financial markets. The fact that we cannot find significant results for Brazilian investors' attention being caused by volatility in the crisis period could simply be explained by the fact that Brazil was not affected by the financial and succeeding economic crisis as much as Europe and the USA.

For all indices besides the Nikkei 225 and the FTSE MIB, we can observe that investors' attention Granger-causes realized volatility. Remarkably, in the crisis period the IBEX 35 is significant in both directions. The findings indicate that Spanish retail investors seem to be influenced by financial markets and influence financial markets mostly in times of high volatility. This further strengthens the above inference about crisis activity. Italian investors, as already indicated in the complete sample period, are not influenced by volatility and in turn their information demand / attention measure does not explain volatility in the future. The results of bi-directional Granger-causality for all other countries exhibit significance, yet on lower levels than in the complete sample period (e.g. 10% level for the DJIA for RV on SVI).

Pre-crisis period: Pre-crisis results show no significance for most indices regarding investors' attention causing volatility. In particular, significance can only be observed for the FTSE 100 (10% level) and the Bovespa (1% level). Looking at the other direction, Granger-causality of volatility on investors' attention is only significant at lower levels and for fewer indices. Specifically, significance cannot be observed for the FTSE 100, the Nikkei 225 and the FTSE MIB. While we are not surprised to see no effects for Italian investors, it is interesting that British investors' attention seems not to be triggered by volatility during the pre-crisis period, but still explains future volatility partly. Furthermore, it is remarkable that Brazilian investors (Bovespa) are the only ones in which the interactions are pronounced in both directions. On the one hand, one would expect to see less conclusive results for emerging market economies where internet penetration has been dramatically lower in the early years of our sample than in the later parts. On the other hand, one could also make the argument that especially in emerging market economies, information sources are less accessible and therefore Google search plays a larger role in satisfying retail investors' information demand.

Post-crisis period: Results in the post-crisis period show significance in both directions for a subsample of indices. While RV Granger-causes investors' attention in 6 out of 9 countries, SVI Granger-causes realized volatility in a subset of 5 out of 9 countries. Interestingly, we see that Spanish and Italian investors' behavior seems not to follow the pattern of other indices. As we will discuss in Section 5, stock market participation might be a crucial factor.

Variance Decomposition

Summarizing the results, we can conclude that the variance decomposition for different sample-periods offers three main insights. Firstly, we observe that during the crisis period a greater share of volatility can be explained by investors' attention. The same applies to the share of investors' attention explained by volatility. In other words, during non-financial-crisis periods both volatility and investors' attention are explained by their own lags to a greater extent and in the crisis period the respective other variable explains a greater share. Secondly, in the post-crisis period, volatility explains a greater share of investors' attention and investors' attention explains a greater share of volatility compared to the pre-crisis period. Both insights strengthen the conclusion of time-varying behavior of retail investors. Thirdly, the differences in country specific characteristics found in the complete sample period are mostly confirmed. While Spanish and Italian retail investors' interaction with their respective stock markets seems to be quite low (especially for Italy), other investors exhibit a much higher level of interaction.

The share of investors' attention explained by realized volatility ranges from 42% (DJIA) to 0.8% (FTSE MIB) in the crisis period. It is remarkable how much of investors' attention is triggered by realized volatility in a crisis period, indicating that retail investors are much more interested in financial markets during those times. Especially Spanish investors' attention triggered by realized volatility rises to levels similar to those of investors from other European countries during the crisis. However, they don't seem to cause additional volatility to the same extent other investors do. Italian investors' attention does not explain volatility and the same is true vice versa. This further indicates different underlying factors for Spain and Italy. We can observe that six out of nine indices show higher explained levels of investors' attention in the post-crisis period than in the pre-crisis period. Spanish investors' attention, however, follows a contrary pattern. Here, contrary to the other indices, realized volatility explains a greater share of investors' attention during the pre-crisis period than during the post-crisis period. A potential explanation could be the magnitude by which the Spanish economy was hit during the crisis period. This could have forced Spanish retail investors to sell privately held stock which in turn might have induced lower stock market participation of Spanish investors in the post-crisis period.

Regarding retail investors' attention explaining realized volatility in the crisis period, we observe values ranging from 24% (CAC) to 1% (FTSE MIB). Compared to the complete sample period, this value increases by about 40% during the crisis period. This pattern is even more striking for other indices where the percentage of variance explained increases by 100% or more (DAX 30, FTSE 100, AEX, Bovespa, IBEX 35). Interestingly, for the DJIA we cannot observe the same pattern. Variance in volatility explained by investors' attention actually decreases slightly during the crisis period (11% vs 9%). Moreover, we can observe that retail investors' attention seems to contribute more to realized volatility during the post-crisis period where values are in the range between 0.4% and 19.9%, than during the pre-crisis period where values range between 0% and 4.8%. However, this observation is not valid for the IBEX 35, the Bovespa and FTSE MIB. Results for these indices do not reveal a specific pattern.

The impulse response functions further indicate different behavior of retail investors in the analyzed time periods. During the crisis period, impulse response functions show that effects are persistent for about 40-60 days. In the post-crisis period, they are only persistent for about 30-40 days and mostly below 20 days in the pre-crisis period. The impulse response functions for the CAC40 for different sample periods can be found in Figure 3 in the appendix[17].

4.1.3 Implied Volatility

Unfortunately, analyzing the interaction between retail investors and financial markets measured by implied volatility has some limitations due to the non-existence (or non-accessibility) of implied volatility indices for Bovespa, IBEX 35 and FTSE MIB. Especially the lack of data for Spain and Italy is unfortunate since results for realized volatility indicate different underlying patterns in those countries. Nevertheless, implementing our VAR-model methodology including Granger-causality, variance decomposition and impulse response function analysis gives us three main insights (see Table 3 and Table 4 for Granger-causality tests and Tables 16-17 for VAR-parameter estimations in different time periods). Firstly, implied volatility seems to explain future investors' attention to a much higher degree than realized volatility. Secondly, investors' attention explains future implied volatility to a much lesser extent than it does explain future realized volatility. Thirdly, results regarding significance and explanation power of effects in both directions are evident for Continental Europe's countries but not at all or to a much lower extent for the USA, the UK and Japan. Below, we will highlight the most interesting insights in more detail for the different countries, the different time periods and our different analysis tools. Furthermore, we will compare the results with those the realized volatility analysis yields and will discuss the differences in Section 5.

[17] IRFs of other indices are available upon request, but are not reported within this thesis because of their mere number and the lack of new information provided by them.

Table 3: Implied volatility: Granger causality test statistics

The table presents the results obtained by conducting Granger-causality tests on log Google search volume index (SVI) and log implied volatility series (IV) for six equity indices. Granger-Statistics are given for the significance of the explanatory variable statistically causing the response variable. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding test statistic.

		DAX 30		Dow Jones		FTSE 100	
	Response variable	IV	SVI	IV	SVI	IV	SVI
	Explanatory variable	SVI	IV	SVI	IV	SVI	IV
Complete period	Granger-Statistic	11.17	51.05	4.04	69.58	7.42	52.51
	p-value	***.011*****	***.000******	*.258*	***.000******	***.060****	***.000******
Pre-crisis	Granger-Statistic	4.99	1.32	5.02	15.61	0.09	16.27
	p-value	***.083****	*.516*	*.170*	***.001******	*.760*	***.000******
Crisis	Granger-Statistic	3.55	51.39	0.79	31.45	5.31	24.52
	p-value	*.169*	***.000******	*.672*	***.000******	*.070**	***.000******
Post-crisis	Granger-Statistic	6.85	10.98	0.45	18.56	4.78	13.05
	p-value	***.009******	***.001******	*.798*	***.000******	***.092****	***.001******

		CAC 40		AEX		Nikkei 225	
	Response variable	IV	SVI	IV	SVI	IV	SVI
	Explanatory variable	SVI	IV	SVI	IV	SVI	IV
Complete period	Granger-Statistic	14.07	39.81	18.69	40.96	4.09	31.30
	p-value	***.007******	***.000******	***.001******	***.000******	*.252*	***.000******
Pre-crisis	Granger-Statistic	2.20	14.23	4.04	5.88	0.01	13.41
	p-value	*.333*	***.001******	*.133*	***.053****	*.936*	***.000******
Crisis	Granger-Statistic	8.53	21.50	6.74	19.48	0.05	43.20
	p-value	***.014*****	***.000******	*.034***	***.000******	*.824*	***.000******
Post-crisis	Granger-Statistic	5.91	10.42	7.20	24.25	4.21	7.36
	p-value	*.052**	***.005******	***.027*****	***.000******	*.122*	***.025*****

Effects of IV on SVI: In the complete sample period, we observe significant results regarding implied volatility Granger-causing investors' attention for all indices at the 1% level. These results are in line with those of realized volatility. However, the variance decomposition shows that in a 100 day forecast period implied volatility explains the variance in investors' attention to a much greater extent. Values range from up to 53% (DJIA) to 16% (NIKKEI 225) whereby the European indices exhibit values around 30%. The realized volatility analysis yields results between 42% (DJIA) and 14% (NIKKEI 225) whereby the European indices range between 20% and 30%. The greater explanation power of implied volatility is even more striking when looking at the crisis period. Implied volatility explains up to 62% (DJIA) of variance in investors' attention during this period. Results for other indices are also about 10 to 20 percentage points higher for implied volatility than for realized volatility. Another interesting observation is that the patterns of country differences completely confirm the realized volatility analysis, indicating robustness of the results obtained. Furthermore, in line with the results regarding realized volatility, both number of significant indices and level of significance are higher in the post-crisis period than in the pre-crisis period. Moreover, the share of investors' attention explained by volatility is higher in the post-crisis period than in the pre-crisis period. Both observations, regarding significance and variance decomposition, indicate more pronounced activity in the post-crisis period than in the pre-crisis period.

Table 4: Implied volatility: Variance decomposition

The table presents the results obtained by conducting forecast error variance decomposition on log Google search volume index (SVI) and log implied volatility series (IV) for six equity indices. The values indicate how much the vertical variable (left axis) explains in the variance of the horizontal variable (above axis) at forecast step 100.

		DAX 30		Dow Jones		FTSE 100	
		IV	SVI	IV	SVI	IV	SVI
Complete period	IV	0.946	**0.268**	0.943	**0.525**	0.945	**0.303**
	SVI	**0,054**	0,732	**0,057**	0,475	**0,055**	0,697
Pre-crisis	IV	0.927	**0.030**	0.960	**0.222**	0.999	**0.185**
	SVI	**0.073**	0.970	**0.040**	0.778	**0.001**	0.815
Crisis	IV	0.917	**0.441**	0.978	**0.621**	0.902	**0.390**
	SVI	**0.083**	0.559	**0.022**	0.379	**0.098**	0.610
Post-crisis	IV	0,904	**0,290**	0,977	**0,363**	0.943	**0.171**
	SVI	**0,096**	0,710	**0,023**	0,637	**0.057**	0.829

		CAC 40		AEX		Nikkei 225	
		IV	SVI	IV	SVI	IV	SVI
Complete period	IV	0.852	**0.300**	0.913	**0.307**	0.978	**0.161**
	SVI	**0.148**	0.700	**0.087**	0.693	**0.022**	0.839
Pre-crisis	IV	1.000	**0.085**	0.962	**0.051**	1.000	**0.109**
	SVI	**0.000**	0.915	**0.038**	0.949	**0.000**	0.891
Crisis	IV	0.797	**0.342**	0.934	**0.460**	1.000	**0.155**
	SVI	**0.203**	0.658	**0.066**	0.540	**0.000**	0.845
Post-crisis	IV	0.934	**0.312**	0.889	**0.284**	0.975	**0.118**
	SVI	**0.066**	0.688	**0.111**	0.716	**0.025**	0.882

Effects of SVI on IV: While we can see an exemplified effect of implied volatility (compared to realized volatility) on retail investors' attention, the exact opposite is the case for investors' attention causing implied volatility. In the complete sample period, significance at the 1% level for Granger-causality can only be observed for two out of six countries, while the analysis of realized volatility shows significance in five out of those six. Remarkably, for US investors we cannot confirm significance regarding Granger-causality in the complete period nor in any sub-period. This result indicates that US investors' attention does not influence implied volatility, whereas it does influence realized volatility. The same applies to Japanese investors. Furthermore, the share of implied volatility explained by investors' attention is substantially lower than the share of realized volatility. In fact, values range between 2.2% (NIKKEI 225) and 14.8% (CAC 40) for implied volatility compared to 6.5% (NIKKEI 225) and 16.9% (CAC 40) for realized volatility. Nevertheless, the variance decomposition indicates that effects are most pronounced in the crisis period and that they are higher in the post-crisis period than in the pre-crisis period (except for the DJIA). Furthermore, country specific patterns seem to be confirmed since relative values are similar to those in the analysis of realized volatility. The impulse response functions show characteristics similar to those in the realized volatility analysis, however, as expected, with lower magnitudes (see Figure 4 for the IRFs of the CAC40 and the DJIA)[18].

4.1.4 Illiquidity

We focus our analysis on the indices for which data was available and we were able to construct the daily Amihud-Illiquidity measure. The indices for which this was possible include the DAX 30, DJIA, CAC 40 and AEX. The causal effect of illiquidity on investor attention is economically more difficult to interpret. While volatility can be observed easily by retail investors, illiquidity is rather a proxy for special situations in the market which not necessarily would trigger investors' attention itself. Therefore, this is merely a statistical causality than an economic one. Nevertheless, the results reveal three main insights. Firstly, illiquidity and retail investors' attention interact on a statistically significant level in both directions, but compared to the volatility analyses, coefficients for more distant lags change signs from positive to negative. In other words, retail investors decrease illiquidity in the short term, but increase it in the longer term – they can therefore be regarded as liquidity providers in the short term. Secondly, interaction in both directions appears to be the strongest during the crisis period. Thirdly, French retail investors' attention feeds through to the greatest extent and illiquidity Granger-causes investors' attention for US investors the most. Table 5 reports Granger-causality and variance decomposition stats for the illiquidity measure and Tables 18-19 report the VAR-parameter estimations for different time periods.

[18] IRFs of other indices are available upon request, but are not reported within this thesis because of their mere number and the lack of new information provided by them.

Table 5: Illiquidity: Granger causality tests and variance decomposition

The table presents the results obtained by conducting Granger causality tests and variance decomposition on log Google search volume index (SVI) and log Amihud-illiquidity series (ILLQ) for four equity indices. Granger-Statistics are given for the significance of the explanatory variable statistically causing the response variable. For the variance decomposition, the values indicate how much the vertical variable (left axis) explains in the variance of the horizontal variable (above axis) at forecast step 100. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding test statistic.

Granger causality test

		DAX 30		Dow Jones		CAC 40		AEX	
	Response variable	ILLQ	SVI	ILLQ	SVI	ILLQ	SVI	ILLQ	SVI
	Explanatory variable	SVI	ILLQ	SVI	ILLQ	SVI	ILLQ	SVI	ILLQ
Complete period	Granger-Statistic	13.67	16.82	19.57	32.88	33.02	27.19	5.930	37.76
	p-value	***.018*****	***.005******	***.012*****	***.000******	***.000******	***.000******	*.655*	***.000******
Pre-crisis	Granger-Statistic	7.29	4.13	19.95	12.91	4.67	4.66	2.33	3.11
	p-value	***.063****	*.248*	***.001******	***.012*****	*.198*	*.198*	*.506*	*.375*
Crisis	Granger-Statistic	11.60	5.08	12.12	16.54	14.66	14.83	9.62	11.84
	p-value	***.021*****	*.280*	***.033*****	***.005******	***.012*****	***.011*****	***.087****	***.037*****
Post-crisis	Granger-Statistic	9.430	14.830	19.070	6.830	10.560	10.950	7.140	17.850
	p-value	***.024*****	***.002******	***.000******	***.033*****	***.061****	***.052****	*.211*	***.003******

Variance decomposition

		DAX 30		Dow Jones		CAC 40		AEX	
		ILLQ	SVI	ILLQ	SVI	ILLQ	SVI	ILLQ	SVI
Complete period	ILLQ	0.955	**0.032**	0.881	**0.168**	0.809	**0.156**	0.976	**0.093**
	SVI	**0.045**	0.968	**0.119**	0.832	**0.191**	0.844	**0.024**	0.907
Pre-crisis	ILLQ	0.962	**0.033**	0.951	**0.038**	0.986	**0.014**	0.996	**0.014**
	SVI	**0.038**	0.967	**0.049**	0.962	**0.014**	0.986	**0.004**	0.986
Crisis	ILLQ	0.919	**0.014**	0.800	**0.254**	0.832	**0.075**	0.847	**0.135**
	SVI	**0.081**	0.986	**0.200**	0.746	**0.168**	0.925	**0.153**	0.865
Post-crisis	ILLQ	0.941	**0.130**	0.955	**0.127**	0.914	**0.132**	0.963	**0.209**
	SVI	**0.059**	0.870	**0.045**	0.873	**0.086**	0.868	**0.037**	0.791

In the complete sample period, we observe bi-directional Granger-causality for all indices but the AEX, which is only Granger-causal in the direction of illiquidity on investors' attention. This means that illiquidity in markets today influences investors' attention tomorrow and vice versa. However, compared to the other measures we analyzed, the direction of the effect is not completely clear. Looking at the parameters of our estimated VAR-model and the impulse response functions for a unit shock in one of the variables (see Figure 5 for IRFs of the CAC40 and the DJIA) , we can see that the first lag parameter is negative for all indices, but that the parameters of succeeding lags are partly positive and especially of a similar magnitude . In other words, investors' attention seems to reduce illiquidity in the very short term, but illiquidity rises afterwards again, stays at a higher level and fades away after 40 to 60 days. The other direction, is unambiguous: More illiquidity in markets today seems to Granger-cause more investors' attention tomorrow.

Variance decomposition analysis reveals some insights about the degree to which illiquidity and retail investors interact. Interestingly, over the complete sample period, we can observe that US retail investors seem to react the most to illiquidity, followed by French and Dutch investors, whereas German investors' attention is not triggered by illiquidity to a great extent. The picture turns around when looking at the other direction, namely how much of the variance of illiquidity can be explained in a 100 day forecast by investors' attention. French Investors' attention seems to explain the most (19%), followed by US investors (12%), whereas German and Dutch investors' attention seems not to be able to explain much of future illiquidity (5% and 2% respectively). Generally, for all indices, investors' interaction with financial markets with regard to illiquidity appears to be larger in the crisis than before the crisis. Explained variance decomposition shares are also higher in the crisis than in the post-crisis period.

The analysis regarding the different sample periods reveals significance for most indices in the crisis and post-crisis period. Furthermore, variance decomposition indicates that effects are most pronounced in the crisis period. In fact, retail investors' attention explains up to 20% of variance in illiquidity (DJIA) for a 100 day forecast period in the crisis sample. Furthermore, illiquidity explains up to 25% of variance in investors' attention (DJIA) in the crisis period.

4.1.5 Trading Volume

Analogously to our analysis of illiquidity, we are limited by the non-availability of data regarding trading volume on indices for the FTSE 100, NIKKEI 225, Bovespa, IBEX 35 and FTSE MIB. Thus, unfortunately, our analysis only contains the DJIA, DAX 30, CAC 40 and AEX. Moreover, the effect of volume on retail investors' attention is economically more difficult to interpret than the effect of volatility. While volatility can be directly observed by retail investors, volume is merely a proxy for events in the market that could draw retail investors' attention. Nevertheless, looking at the complete sample period, we can confirm bi-directional Granger-causality of investors' attention and trading volume for almost all analyzed indices at the 1% level (see Table 6 for granger-causality tests and variance decompositions, as well as Tables 20-21 for VAR-parameter estimations in different sample periods). However, the results do not seem to be conclusive when looking at the different subsample periods. Contrary to volatility, effects are not more pronounced in the crisis period than in the other periods. This can potentially be contributed to the above mentioned non-economical relationship between volume and investors' attention, especially in the crisis period.

Effects of Volume on SVI: In fact, we cannot observe any significance of volume Granger-causing investors' attention for any of the indices in the crisis period. In the pre-crisis period however, our results suggest that volume Granger-causes investors' attention in Germany, the US and the Netherlands. In the post-crisis period, volume Granger-causes investors' attention at the 1% significance level for the AEX at the 5% level for the CAC 40 and at the 10% level for the DJIA. Since we observe that volume indeed Granger-causes investors' attention in the complete period, we reason that there is no underlying pattern that creates more attention from retail investors in times of severe market fluctuations. In contrast, our results suggest at least for some indices significance in the pre and post-crisis period, but not in the crisis period.

Effects of SVI on Volume: Evidence for investors' attention Granger-causing volume can hardly be detected in the pre-crisis sample period. However, in the crisis period, attention of investors from Germany and the US seem to Granger-cause volume. While this effect is not observable for German investors in the post-crisis period, it is for US investors and also exists for Dutch investors. Again, as with volume Granger-causing investors' attention, results for the different sample periods do not follow a pattern. We will discuss potential reasons for these findings in Section 5.

Table 6: Trading Volume: Granger causality tests and variance decomposition

The table presents the results obtained by conducting Granger causality tests and variance decomposition on log Google search volume index (SVI) and log trading volume series (TV) for four stock indices. Trading volume is given in the respective country's currency. Granger-Statistics are given for the significance of the explanatory variable statistically causing the response variable. For the variance decomposition, the values indicate how much the vertical variable (left axis) explains in the variance of the horizontal variable (above axis) at forecast step 100. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

Granger causality test

		DAX 30		Dow Jones		CAC 40		AEX	
	Response variable	TV	SVI	TV	SVI	TV	SVI	TV	SVI
	Explanatory variable	SVI	TV	SVI	TV	SVI	TV	SVI	TV
Complete period	Granger-Statistic	31.54	19.34	23.77	12.42	28.60	30.73	16.35	19.59
	p-value	***.000******	***.002******	***.000******	***.029*****	***.000******	***.000******	***.006******	***.001******
Pre-crisis	Granger-Statistic	2.67	44.11	4.23	12.05	5.73	19.34	5.63	8.28
	p-value	*.615*	***.000******	*.238*	***.007******	***.057****	***.000******	*.344*	*.141*
Crisis	Granger-Statistic	19.18	2.68	7.28	0.49	8.25	4.61	10.13	1.81
	p-value	***.002******	*.748*	***.026*****	*.781*	*.143*	*.465*	***.072****	*.875*
Post-crisis	Granger-Statistic	5.260	3.360	13.92	5.600	11.61	22.29	0.380	9.730
	p-value	*.154*	*.340*	***.001******	***.061****	***.041*****	***.000******	*.944*	***.021*****

Variance decomposition

		DAX 30		Dow Jones		CAC 40		AEX	
		TV	SVI	TV	SVI	TV	SVI	TV	SVI
Complete period	TV	0.903	**0.135**	0.934	**0.247**	0.839	**0.055**	0.924	**0.127**
	SVI	**0.097**	0.865	**0.066**	0.753	**0.161**	0.945	**0.076**	0.873
Pre-crisis	TV	0.983	**0.382**	0.996	**0.305**	0.977	**0.213**	0.973	**0.131**
	SVI	**0.017**	0.618	**0.004**	0.695	**0.023**	0.787	**0.027**	0.869
Crisis	TV	0.962	**0.254**	0.960	**0.127**	0.969	**0.118**	0.926	**0.134**
	SVI	**0.038**	0.746	**0.040**	0.873	**0.031**	0.882	**0.074**	0.866
Post-crisis	TV	0.982	**0.120**	0.919	**0.151**	0.971	**0.137**	0.998	**0.087**
	SVI	**0.018**	0.880	**0.081**	0.849	**0.029**	0.863	**0.002**	0.913

Looking at the variance decomposition for this analysis over a 100 day forecast period after a shock in one of the variables, we can see a similar pattern regarding the relative magnitude or explanative power of the variables as we can observe in the volatility analysis. French retail investors have quite a strong influence on stock markets in their home country as SVI contributes to 16% of the variance in volume. France is followed by Germany (10%), the Netherlands (8%) and the United States (7%). The same applies for the variance share of investors' attention explained by trading volume. Here we can also see the same pattern as in the volatility analysis. Namely, US investors seem to be influenced by volume the most (25%) followed by German (14%), Dutch (13%) and French investors (6%). Interestingly and similar to the volatility analysis, for trading volume the effect of the market on investors is much stronger than the effect of the investors on the market. Besides country specific differences, we can observe similar effect channels and persistence patterns for almost all indices when looking at the impulse response functions (see Figure 6 for the exemplary IRF of the CAC40 and the DJIA). However, this is only true for the DAX 30, AEX and CAC 40. The DJIA exhibits a completely different picture for the evolvement of retail investors' effect on trading volume. While for the DJIA, an unexpected shock in investors' attention yields a slightly positive effect on volume which eventually vanishes away after about 60 days, results for the other indices indicate a positive effect in the first days, then turning into a negative effect which is persistent for a longer time period of up to 100 days. In other words, retail investors seem to trade in the first days after their attention has been roused, but trade less in later periods.

4.2 Stocks

To further investigate the interrelation between investors' attention and market activity, we proceed with analyzing internet search volume and volatility, as well as illiquidity and trading volume on the single stock level. Here, our objective is twofold. Firstly, we are interested if the observations and conclusions obtained by analyzing the equity indices can be found for the stock sample as well. Again, we try to validate the theory that enhanced activity on markets affects retail investor attention which in turn leads to increased market activity on the stock level (see Black, 1986; Lux and Marchesi, 1999; Bloomfield et al, 2009 among others). Secondly, we want to validate the conclusions of Barber and Odean (2008), who show that retail investors become active in trading mainly for those stocks that largely attract their attention. Due to timely and cognitive constraints, retail investors limit their choice set of investment opportunities to stocks that recently caught their attention (see Barber and Odean, 2008). Therefore, we are interested in differences between stocks that are more attention-grabbing and stocks that receive less retail investor attention. By differentiating between stocks that exhibit varying search volume, we want to further evaluate to what extent market activity can be attributed to attention aroused for these stocks.

Generally, we have to be careful when interpreting results for the stock sample as the measures we use, in particular for realized volatility and illiquidity, might exhibit some noise. Moreover, the sample of weekly data only consists of 300 observations for each stock. However, an analysis on a weekly basis is in line with former studies using Google's SVI[19].

[19] see Bank et al (2011), Da et al (2011) or Vlastakis and Markellos (2012)

4.2.1 Attention Distribution across Stocks

To begin, we evaluate the overall search volume of each stock under examination in the whole sample period relative to all other stocks. The relative overall search volume varies widely for the stocks in our sample (see Figure 1). Most strikingly, by far the highest share of attention attracted by stocks in our sample is attributable to those of Google and Apple, each accounting for approximately double the search volume than the third stock in this list, General Electric (GE). Moreover, 18 out of the 22 stocks reveal a search volume less than one fourth of Apple's search volume, the last eight stocks even under one tenth. Therefore, we can differentiate between stocks that are remarkably attention-grabbing and those that are relatively less. Although it is obvious that Apple, Google and GE can be classified as attention-grabbing stocks, it is not clear where to draw an exact line to distinguish attention-grabbing stocks from those lacking this characteristic.

According to Barber and Odean (2008), retail investors' attention to specific stocks is aroused through news about the stock, as well as abnormal returns and trading volume. In turn, retail investors become active in financial markets and trade these stocks. As a consequence, we expect that results of Granger-causality should be more pronounced for stocks that exhibit a larger overall search volume. However, as for all stocks in our sample search volume exceeds a certain threshold set by Google, even the stocks with relatively low search volume compared to Apple and Google are still far more attention-attracting than the average stock in the S&P 500. Therefore, our study is limited to U.S. stocks that attract the most overall attention. Nevertheless, it is worth evaluating differences between stocks with the highest search volume and lower search volume, as large differences in investor attention are apparent within our sample.

4.2.2 Investors' Attention Triggers

When analyzing the sample of stock market indices, in particular the Dow Jones, we observed that realized volatility affects search volume of retail investors to a certain extent (see Tables 26-27 for Granger-causality tests and variance decompositions of the stock sample). Looking at the U.S. stocks separately, we can validate this relationship only for a minority of stocks. Out of the 22 stocks, 5 exhibit a significant Granger-causal relationship on a 5% significance level and 5 additional stocks on a 10% level. The percentage of search volume explained by realized volatility varies significantly between stocks. Over a long-term forecast horizon of 20 weeks, GE's stock shows the highest percentage (about 32%), followed by Citigroup and IBM (around 20% and 19% respectively). For abnormal trading volume causing search volume, we find a similar picture, as 4 stocks are significant on a 5% level and 4 additional stocks on a 10% significance level. However, in contrast to volatility, variations in trading volume are able to explain a larger part of changing investor attention with 5 stocks exceeding 25%. For Apple, as much as 60% of varying investor attention is attributable to trading volume shocks. Evaluating illiquidity Granger-causing search volume, we find even less significant stocks. Further, only a very small percentage of changes in search volume can be explained by illiquidity movements over a 20-week forecast horizon with Apple (around 25%) being the only exception. Although results are significant for some stocks, we are not able to conclude a general Granger-causal relationship of increased market activity rousing retail investor attention. These results are in contrast to our findings when evaluating the DJIA on a daily basis, where a bidirectional causal relationship is apparent over the complete sample period. However, it is worth noting that only 15 out of the 22 stocks analyzed were part of the Dow Jones Industrial Average over most of the time period analyzed[20].

When evaluating the results obtained with regard to the differences in overall search volume for the stocks in our sample, the picture becomes clearer. Interestingly, we find that the 5 stocks that exhibit a Granger-causal relationship between realized volatility and search volume on a 5% significance level are part of the 7 stocks with the highest overall search volume. Further, if we account for the stocks significant on a 10% level, all 8 stocks with the largest overall search volume are significant, in contrast to 2 out of the remaining 14 stocks. Hence, we conclude that a sudden increase in volatility causes retail investor attention specifically for the most attention-grabbing stocks in our sample. When looking at liquidity and trading volume causing search volume, we cannot observe any pattern with regard to relative search volume of each stock over the sample period.

4.2.3 Investors' Actions on the Stock Market

Now we want to shed further light on the relationship of increased retail investor attention to stocks in our sample to market activity for those stocks. As Granger-causality can be detected for only 5 stocks on a 5% significance level, in addition to 2 stocks significant on a 10% level, we are not able to conclude a general effect of retail investor attention on stock volatility. In addition, the fraction of realized volatility explainable by changing investor attention is small for all stocks over a 20-week forecast period. Although the majority of stocks do not exhibit an unambiguous Granger-causal relationship for attention causing liquidity or

[20] On June 8, 2009, Citigroup dropped out of the DJIA, reducing the number of stocks analyzed that are part of the index from 16 to 15. On September 20, 2013, Hewlett-Packard was replaced by Nike, both part of our stock sample.

trading volume, a significant relation is apparent for exactly half of all stocks on a 10% significance level. Regarding liquidity, for 9 stocks a 5% significance is indicated with 2 more stocks showing a 10% significance. For all significant stocks except one (Exxon), an increase in search volume leads to a negative response of illiquidity at lead one, indicating a positive relation between search volume and liquidity on a one-week basis. However, the second lag exhibits a positive response of illiquidity to search volume for all but two significant stocks (Google and Disney) and over an even longer horizon of 5-10 weeks the relation stays positive for all significant stocks. Following this and in line with the findings for the index sample, we conclude that investor attention reduces illiquidity of most stocks in the short terms, but a bounce-back of illiquidity is apparent over longer periods. It is worthwhile noting that we see exactly the same pattern in the stock market indices analysis. When looking at the effects of investor attention on trading volume, we observe a Granger-causal relationship for 7 stocks on a 5% significance level and 4 stocks on a 10% level. However, when looking at forecast error variance decompositions, the fraction of illiquidity and trading volume explainable by changes in investor attention is very small for all stocks.

No pattern with regard to differences in overall search volume of the sample stocks can be detected for stocks Granger-causing realized volatility and trading volume. However, when looking at illiquidity, we observe that 6 out of the 7 most attention-grabbing stocks exhibit a significant Granger-causal coherence. Therefore, it is tempting to conclude that search volume Granger-causes illiquidity only for stocks that naturally attract high attention from retail investors. However, as 4 out of the 10 stocks with the least overall search volume exhibit a significant causal relation from search volume to illiquidity, and as differences in total search volume for those stocks are relatively small, we need to be careful in drawing valid conclusions here.

5. Discussion

To better understand the drivers, impact channels and prediction likelihood of retail investors' interaction with financial markets, we focus on the interpretation of observed patterns of retail investor activity, how they are related to financial and behavioral theory and how we can explain differences in countries through country-specific characteristics of retail investors. We will first present interpretations of our results relating to overall patterns, followed by interpretations on the country level.

5.1 Investors' behavior in General

Generally, we observe three interesting patterns with respect to investors' behavior, which we will discuss below. Firstly, behavior is time varying. Secondly, implied volatility explains investors' attention to a much greater extent. Lastly, attention of retail investors is distributed mainly on few stocks.

5.1.1 Time Varying Behavior

Two main insights can be gained with respect to time varying behavior. Firstly, results for volatility and liquidity are most pronounced in the crisis period. Secondly, post-crisis results for volatility and liquidity are more pronounced than pre-crisis results. This could indicate a general positive trend. Below we will discuss these findings and provide reasoning of potential underlying factors. Moreover, we will comment on the ambiguous results regarding crisis patterns of the trading volume measure.

Realized Volatility in the crisis period

Our results suggest that retail investors interact with financial markets more extensively in times of high market fluctuations. The main results of the realized volatility analysis show that for all countries included in the sample, the interaction of retail investors is most pronounced when only looking at the crisis period itself or when the financial crisis is included in the sample period. Some retail investors, in particular the Spanish, actually do only exhibit interaction with financial markets during this period. To interpret these results, we have to further elaborate on their meaning and how we can relate them to financial theory.

Effect of volatility on SVI: The fact that volatility Granger-causes investors' attention shows that stock market volatility induces retail investors to attribute to the stock markets one of the scarcest cognitive resources of humans – attention (Kahneman, 1973). Our results suggest that realized volatility explains investors' attention during crises better than in other periods. One of the reasons for the increased attention of retail investors could be increased risk aversion during a crisis. Various studies show that risk aversion rose during the financial crisis (e.g. Guiso et al, 2011). Eckhoudt and Godfroid (2000) discuss that information value and demand can increase with risk aversion. Therefore, information demand about financial markets induced by realized volatility should be higher during a crisis, when risk aversion is high. Our results are in line with this inference.

Effect of SVI on volatility: The general link between investors and financial markets is investigated in various noise trader models (Lux and Marchesi, 1999; Alfarno and Lux 2007) and empirically supported (Dimpfl and Jank, 2011; Foucault et al, 2011). All of them either predict or find that noise traders increase the volatility in the markets. Our results contribute to these findings by confirming these models in a longer sample period and for various stock market indices. Furthermore, the results shed light on the link by identifying a crisis as the period with most interaction. We have clear indication that retail investors' attention explains substantially more of volatility in financial markets when volatility is already high. Moreover, our results confirm that the raised attention of investors during a crisis period actually impacts the volatility on financial markets - most likely through their trading behavior. In addition, the impulse response functions indicate that effects are much more persistent during times of high volatility.

Realized Volatility in the pre- and post-crisis period

The variance decomposition analysis indicates that realized volatility is more predictive and explains a greater share of investors' attention in the post-crisis than in the pre-crisis sample. The same is true for retail investors' attention explaining realized volatility. We find two plausible explanations for this observation: Increased internet affinity as well as persistent risk aversion after a crisis period.

Higher post- than pre-crisis variance decomposition shares for realized volatility (explaining future investors' attention and vice versa) indicate rising retail investors' activity. In between 2006 and 2012, average internet penetration in our sample has increased from 58.4% to 76.3% (Worldbank, 2013). Furthermore, the amount of people using online banking has increased by about 20 % (Eurostat, 2013). We reckon that both effects play a role in increased investors' interaction with financial markets. Studies show that households that are connected to the internet are more likely to participate in the stock markets (e.g. Bogan, 2006). Moreover, higher internet penetration might also have lowered transaction costs through online brokerage and might have decreased information asymmetries due to easier and faster accessibility of relevant information. Thus, internet penetration might have increased the attention of retail investors to financial markets' movements, but also their propensity to get active after their attention is roused. The trend in online banking penetration shows the increased tendency of retail investors to pursue financial transactions online. This could also have affected the trading behavior of retail investors, since the barriers for trading in the stock markets deceased severely through the possibility to use online brokers fast and cheap. This might be an explanation why investors' attention explains a much greater share of realized volatility after the crisis than before. Unfortunately, we cannot neglect the fact that an increased usage of Google as source of information is the main driver behind this pattern. However, if we assume that Google search queries are also representative for retail investors who satisfy their information demand through other sources, the explained variance in volatility and attention is representative for all retail investors (see Da et al, 2012 for a comparison of Google Trends data and SEC trading filings to confirm this assumption). Therefore, we can conclude that at least to a certain degree, retail investor activity in financial markets has increased after their attention has been roused. Hence, it explains a greater share of market volatility post-crisis than pre-crisis.

As mentioned in section above, volatility in stock markets seems to be the main driver for retail investors' interaction with financial markets. For European countries, the crisis period did not end in 2009/2010. Instead, as a consequence of the financial and banking crisis in Europe, the EURO-crisis emerged. Equity markets were quite volatile - especially in 2011. However, volatility also spiked several times when EURO-crisis related news emerged later on. Therefore, we do not only attribute the pattern of increased interaction of retail investors with financial markets over time to technical developments, but also to market fluctuations. Higher risk aversion in times of an economic and financial crisis correlates with higher investors' attention succeeding volatility shocks. Furthermore, academic literature also relates negative returns to increased investors' attention. Hacamo and Reyes (2011) empirically prove the hypothesis of a negativity bias in attention allocation for retail investors. They conclude that negative returns entice more attention of retail investors than positive returns. Especially during the second half of 2011 (part of the post-crisis period), returns for European stock markets were negative. This explanation perfectly matches the higher share of explained investors' attention by realized volatility during the crisis and post-crisis period compared to the pre-crisis period. The noise trader theory developed by Lux and Marchesi (1999), in which increased investors' attention due to a shock in volatility induces investors to get active in the market, thereby creating volatility, helps to explain the increased share of volatility caused by retail investors in the crisis and post-crisis period.

Related to above findings regarding the increase in risk aversion, our results further suggest that effects for realized volatility Granger-causing investors' attention (and vice versa) are much more persistent when volatility is high, i.e. in the crisis period. However, we also find that effects are more persistent in the post- than in the pre-crisis periods. Interestingly, our results seem to confirm findings of Guiso et al (2011) who show that risk aversion is higher after a crisis than before. The reason: Risk aversion does not fade away instantaneously. For example, retail investors' attention could be roused by less severe shocks in volatility because recent strong market fluctuations are still in memory. If risk aversion stays high succeeding a severe crisis and higher levels of risk aversion cause higher information demand by retail investors, this would lead to more persistent information demand following the crisis period. Our impulse response functions for the different periods confirm this inference by indicating the highest persistence of effects during the crisis period of about 40-60 days, followed by 30-40 days in the post-crisis period and mostly below 20 days in the pre-crisis period. Apparently, retail investors attribute a much larger share of their scarce resource attention to stock markets over a longer time horizon when volatility and risk aversion are both high.

Illiquidity in the crisis period:

We do not only find interesting insights about retail investors' interaction with financial markets regarding realized volatility, but also regarding liquidity. While Bank et. al, (2011) show that retail investors' attention increases liquidity of stocks in Germany, we conducted our analysis on a macro level for different countries. In our sample of Germany, the USA, France and the Netherlands, we can confirm that retail investors reduce illiquidity in the short run. However, the impulse response functions clearly indicate that this effect is only persistent over a period of some days before it turns into the opposite direction. Retail investors' attention then corresponds to a higher level of illiquidity. A potential economic explanation of these results is that retail investors do usually not trade for a longer period of time, but rather transform their information demand into trading stocks in the short run (at least if they hold stocks directly). Their holdings do not require a sell-off or stretched buying over a longer period of time due to their small size. Therefore, their potential to sell/buy stocks is limited and they will only sell/buy for a short period of time. This reasoning would be in line with an experimental study conducted by Bloomfield et al (2009). According to them, noise traders provide positive effects on market liquidity measured by volume and spreads. They also link noise traders' liquidity enhancing trading strategies to a temporary price impact of market orders rather than to a reduction of the permanent price impact of trades. This explanation fits very well to the results indicated by the impulse response functions presented in this empirical work.

Trading Volume:

Firstly, we can confirm that retail investors' attention is positively linked to trading volume (see Bank et. al, 2011; Preis et. al, 2010) for different countries including Germany, France, the USA and the Netherlands. We find a Granger-causal effect of retail investors on trading volume. However, in contrast to volatility, the effect of trading volume is not stronger during the crisis period. The effect on trading volume by retail investors could depend on their share traded in the stock markets at a particular time. Retail investors generally trade with a time lag. While volume is likely to peak when institutional investors trade, volume could already be lower again when retail investors get active. Therefore, the additional volume effect induced by retail investors' activities might be covered up by less active trading behavior from institutional investors. Further research with separated volume data for retail and institutional investors can potentially shed light on this. Interestingly, our results suggest that retail investors' attention explains a lot of the higher volatility during the crisis period, but explains little of the trading volume. As a consequence, retail investors' activity during a crisis period might be even more irrational than in non-crisis periods leading to increased volatility but not volume.

Regarding the effect of trading volume on retail investors' attention, we cannot find any pattern during the crisis period either. Trading volume can hardly directly be linked economically to retail investor attention, but is merely a proxy for events on the financial market. Instead, our other findings suggest that volatility, which is directly observable by retail investors, triggers retail investors' attention. If volume is not perfectly correlated with volatility, which should be the case in reality, not detecting an effect of volume on investors' attention would make sense.

5.1.2 Trigger Events

Another interesting finding is the difference between implied and realized volatility as a measure of retail investors' interaction with financial markets. We observe that implied volatility explains a much larger share of future investors' attention than realized volatility does. More surprisingly, retail investors' attention, in turn, explains much less of future implied volatility than of realized volatility. By analyzing the differences between realized and implied volatility, we try to shed further light on the way retail investors interact with financial markets.

Our previous analyses and interpretations might induce a critical question: Does volatility truly cause retail investors' attention or are there other factors influencing volatility itself (e.g a news event)? Realized volatility measures the actual volatility occurring during a specific time period - in our case one day. Implied volatility, on the other hand, measures markets' expectations about future volatility over a certain period - in our case 30 days. If we relate this difference to news events , we can distinguish the major underlying sources for both measures. Realized volatility is more likely to capture unexpected one-time or idiosyncratic news events, which influence stock markets at a specific day. During those events, implied volatility would not rise if the market does not expect the news event to have a structural impact which will be persistent over a certain period of time. On the other hand, implied volatility would then be more likely to capture those news events, which inherit structural changes that persist over a longer time period. Relating this argument to our findings about investors' attention, we could infer that retail investors seem to get more active when structural news events occur, that alter the markets' expectations about future volatility. For example, the collapse of Lehman Brothers in 2008 probably made market participants aware of the risk inherent in the financial system and changed expectations about future volatility quite dramatically. On the other hand, there are other idiosyncratic news events like worse than expected earnings announcements. These might rather induce high volatility on the occurrence date of this specific event, but not alter long term volatility expectations. Following this line of argument, our results suggest that retail investors' attention is more likely to be triggered by events which have implications about expected future volatility. Putting it differently, it is rather the news event with long-term implications that triggers retail investors' attention than the actual fluctuation of prices on the stock market on a particular day. Of course, news events are not easily distinguishable and often overlap regarding their instant and long-term effect on volatility. Therefore, it is plausible that we find an effect of both, realized and implied volatility, on retail investors' attention.

The results for retail investors' attention Granger-causing implied volatility reveal that retail investors' attention can help to explain tomorrow's implied volatility for some indices. However, the explanation power is much smaller than for realized volatility (smaller share in variance decomposition) and the effect is not detectable for USA and Japan (no statistical significance). This difference among countries might indicate that US and Japanese markets are more efficient than their European counterparts because of the following:

Among others, Christensen and Prabhala (1998) empirically show that implied volatility is an unbiased and efficient forecast for realized volatility, subsuming all information in past and current realized volatility. Thus, implied volatility tomorrow contains information about today's and tomorrow's realized volatility. If we observe an increase in implied volatility tomorrow granger-caused by an increase in investors' attention today, today's implied volatility did not account for the additional realized volatility retail investors cause succeeding their information demand. In other words, if the implied volatility in the future is influenced by the retail investors' attention today, it must have adjusted for the additional realized volatility retail investors bring to the market tomorrow. Therefore, an influence of retail investors' attention on implied volatility tomorrow indicates that the market did not efficiently take into account all information about retail investors' behavior. This would imply that professional investors form their expectations about the future without taking retail investors' behavior into account. Interestingly, investors' attention does not Granger-cause implied volatility in the USA and Japan, indicating that US markets might be more efficient than their European counterparts – or that institutional investors adapted to retail investors' behavior.

To sum up, implied volatility and its underlying pattern influence investors' attention. Yet, retail investors influence realized rather than implied volatility, indicating a disconnection between the different acting groups in financial markets. Nevertheless, we are aware of the fact that this conclusion is based on various assumptions. Therefore, our study rather provides some ideas for future research questions regarding this topic.

5.1.3 Attention Distribution

Analyzing the dynamics between investors' information demand and financial markets on the stock level helps us to further investigate the origin of attention as well as the magnitude and dynamics on the financial markets. In contrast to the sample of stock market indices, we are not able to generally confirm noise trader models for the stock sample as we cannot conclude an overall bidirectional Granger-causal relationship between retail investors' attention and market activity. In general, retail investors do not seem to get active in trading stocks after a shock in market activity for the specific stock. Moreover, retail investors' market activity does not necessarily increase after their attention towards specific stocks is roused. However, the results suggest differences in individual investors' attention on the 22 stocks under examination. Some stocks, in particular Apple, Google and GE attract far more attention than others. Barber and Odean (2008) argue that individual investors distribute their attention carefully and only on a few stocks that grab their attention through certain events. This argumentation is supported by our results: Search volume for stocks that attract the highest attention from retail investors is Granger-causally driven by increases in volatility. Further, investor attention is attracted contemporaneously, as the impact is generally highest for forecast step 0 in the impulse response function. However, by contemporaneously we mean that the shock increases attention within the same week. This should not be surprising given the amount of media coverage and the information and communication channels from companies and financial markets to individual investors. We infer that investors distribute their attention to stocks that exhibit unexpected information arrival, extreme returns, increased media coverage or corporate announcements because these events increase stock volatility (see Drake et al 2012; Engle et al, 2011; Barber and Odean, 2008; Tetlock, 2007; Greene and Smart, 1999). However, similar to our argument in 5.1.2, inferring from the lower magnitude and significance in the stock sample, it seems that rather systemic change in volatility (expectations) triggers attention by retail investors than idiosyncratic stock specific events.

Our data does not suggest any difference for increased attention Granger-causing volatility and trading volume for attention-grabbing stocks compared to stocks that attract less attention. For investor attention causing liquidity, we find that mainly attention-grabbing stocks exhibit increasing liquidity in the first week after attention rises. Further, and in line with the findings in the previous section, illiquidity bounces back over longer time horizons. Hence, we can conclude that retail investors, as suggested by noise trader models, are liquidity providers also on the stock level. A potential reason why we are not able to detect a clear causal link between increased investors' attention and both, volatility and trading volume, could be that individual investors do not trade at all on the stock market following their information acquisition. It is likely that a lot of individual "unsophisticated" investors, who research information regarding a specific stock, do not manage their stock holdings directly. Instead, they own those stocks indirectly via pension plans, mutual funds or company stock participation programs. Moreover, defined contribution plans, where individuals have influence on their stock holdings, have become more and more popular in recent decades (see Thaler and Benartzi, 2001). This suggests that – on the one hand - investors know which stocks they own and therefore are interested in the stock's performance as it directly contributes to their current wealth. Further, they are involved in investment decisions and therefore interested in new opportunities to invest. On the other hand, these investors are not able to trade instantaneously on their information as their stock holdings are managed externally. However, this situation is not able to explain the Granger-causal relation from search volume to liquidity we find in our sample for attention-attracting stocks. Although increased liquidity might be exogenous and related to the event that triggered the initial volatility (see Lee et al, 1993), we assume that retail investors do trade on noise or events that caught their attention. This behavior is supported by the indices results and the country analysis (as well as by Da et al, 2011; Barber and Odean, 2008 and Black, 1986).

Additionally, we find significant results not related to overall retail investors' attention. For stocks exhibiting lower search volume, we assume that no market, corporate or stock event was impacting enough to attract attention from retail investors. Therefore, the causal relation between volatility and search volume is missing. However, for some of these stocks, a causal link from increased investors' attention to increased volatility, liquidity or trading volume is given. Although we cannot find a pattern linked to overall investors' attention in our sample, it has to be noted that the 22 stocks we examine are not a random sample but instead the 22 stocks which exhibit the highest or most valid search volume when being researched on Google. It is also not utterly out of question that events not related to increased market activity trigger attention of retail investors for these stocks. Furthermore, some anticipated events, for example earning announcements and company events like annual meetings, product presentations or important fairs, might be preceded by increased investors' attention before financial markets react (see Drake et al, 2012 and Da et al, 2011). In particular, for announcement events, Drake et al (2012) suggests that retail investors' attention measured by Google search volume is spread over a period surrounding the announcement. Moreover, we find differences in behavior of specific stocks in our sample that are not straightforward to explain. Exxon for example exhibits a positive causal relationship from search volume to illiquidity, indicating that an increase in investors' attention is followed by a decrease in liquidity. Further, a negative causal relationship between investors' attention and trading volume is detected for Hewlett-Packard. However, we expect these differences to be explained by individual firm or stock characteristics or specific events affecting our results. Although it would be interesting to take a closer look at those stocks and their specific characteristics that might explain the unexpected behavior, we leave them to further research. Kumar et al (2006) for example give some explanations - consistent with noise trader models - which specific stocks might be preferred by retail investors.

In general, we have to keep in mind that due to data availability, our analysis was conducted on a weekly basis, thus possibly hiding certain contrasting dynamics that happen consecutively but within the same week and revealing only the overall effect. Moreover, Kumar and Lee (2006) find that stocks affected by retail investors' activity on financial markets mainly are small stocks. As most of the stocks in our sample are those exhibiting the largest overall market capitalization, influence of retail investors might be too small to detect.

5.2 Investors' behavior in the Cross-Country Setting

As described in section 4, we find varying results for the different countries regarding bidirectional Granger-causality and explanation power between investors' attention on the one hand and volatility, volume and liquidity on the other hand. Below, we will discuss several potential underlying reasons for cross-country variance and relate them to existing research. Specifically, we will focus on structural and behavioral explanations. Both will help us to explain the necessary conditions for retail investors' interaction with financial markets.

5.2.1 Structural Characteristics

Several structural characteristics of countries have to be considered when trying to explain the varying results of retail investors' interaction with stock markets in different countries. These include stock market affinity, retail investors' portfolio allocation, how stocks are owned and traded (mutual funds, direct ownership, ETFs), internet penetration, willingness to pursue financial transactions online, financial literacy, share of the market capitalization held by retail investors, wealth and income distribution, trading volume share of retail investors and many more. Unfortunately, there are data constraints on several measures. We chose to focus on stock market affinity, internet penetration and online banking usage.

Table 7: Cross-Country Structural Characteristics

This table presents structural characteristics of countries and the correlation of these measures with the explained share of variance in RV by SVI on the one hand, and with the variance share of SVI explained by RV.

	Stock Market Participation Rate (%)	Internet Penetration (%)	Online Banking (%)	a) Share of variance in RV explained by SVI	b) Share of variance in SVI explained by RV
Germany	25.4	79.0	45.0	7.3%	18.9%
USA	49.7	74.5	51.0	11.2%	41.5%
England	39.4	80.7	52.0	9.2%	24.3%
France	43.0	71.1	54.0	16.9%	23.0%
Netherlands	24.9	88.9	80.0	8.9%	27.6%
Japan		76.1		6.5%	9.1%
Brazil		38.2		7.8%	1.8%
Spain	12.8	61.8	32.0	2.5%	22.5%
Italy	10.4	48.7	21.0	0.8%	1.8%
Correlation with a)	**0.87**	0.39	0.63	1.00	1.00
Correlation with b)	**0.74**	0.68	0.61		1.00

In order to require information demand about the stock markets and to get active in the market, the first logical condition is to own shares or to be willing to participate in the stock markets. The number of retail investors active in a specific country therefore is both a good indicator for volatility causing investors' attention, and investors' attention funneling through to future realized volatility. Table 7 shows individuals' stock market participation rate for different countries. Interestingly, these figures reveal a pattern which is very consistent with our results for bi-directional Granger-causality of retail investors' attention and realized volatility. As described in section 4, Italian and Spanish investors are the only ones whose attention does not Granger-cause future realized volatility. The fact that stock market participation rates are extremely low in these countries indicates that stock market participation might play a major role in the way retail investors affect financial markets. Compared to the average of the other countries (36.5%) their values are about 20 to 25 percentage points lower (Italy: 10.4; Spain: 12.8). Since realized volatility seems to Granger-cause future investors' attention, it appears that retail investors in Italy and Spain are satisfying their information demand. Yet, their succeeding actions are not funneling through to increased future realized volatility. Low stock market participation rates could also explain this pattern, as investors' actions might not reach a critical mass to influence stock market volatility.

When relating stock market participation to the variance decomposition results, we would expect to see a greater share of future realized volatility explained by investors' attention in countries where stock market participation is high. Table 7 shows stock market participation rates and the 100 day forecast variance decomposition values for each country and in each direction - complemented by their correlation. Results are as expected: Stock market participation is highest in France and the USA and lowest in Italy and Spain, mirroring the explained share of future realized volatility by investors' attention (correlation of 0.87). If individuals in a certain country do not own/trade stocks, their impact on volatility as explained in noise trader models is simply not existent.

Logically, this argument is also valid for volatility Granger-causing investors' attention. We can assume that most individuals satisfy their information demand succeeding increased levels of realized volatility only if they are indeed stockholders. Even if this was not true, it is highly likely that non-stockholders search for information about stock markets in a similar manner than stockholders do. Therefore, the more stockholders live in a country, the more does the information demand correlate with trigger events like news causing realized volatility. Stock market participation rates correlate with the share of investors' attention explained by realized volatility quite strongly (0.74), but not as much as for the opposite direction (0.87). Nevertheless, the above observations indicate that the correlation between retail investors' activities and volatility in financial markets rises with the stock market participation rate. In other words, stock market participation rates are able to explain the share of investors' attention caused by realized volatility (and vice versa) extremely well. We observe very similar correlation patterns for illiquidity and investors' attention. However, due to limited data availability we were only able to analyze four indices. Therefore, we should be conservative with interpreting these results in combination with stock market participation. Nevertheless, investors' attention explains most of illiquidity in the USA and France, which also exhibit the highest stock market participation rates. The correlation between the share of illiquidity explained by investors' attention and the stock market participation rates is 81% and 90% for the other direction.

Furthermore, internet affinity of a country could play a major role in explaining retail investors' interaction with financial markets. Clearly, the construction of our study will most likely produce results which correlate with country specific usage statistics. However, internet penetration also proxies stock market participation, financial literacy, decreased transaction costs and decreased information asymmetries. Therefore, it might contain information about investment behavior in different countries (see Bogan, 2006 as well as Glaser and Klos, 2013). Additionally, important for retail investors to get active in the market, there must be an affinity to carry out financial transactions online. We proxy the propensity to carry out transactions online by the percentage of online banking users in the different countries. Table 7 shows internet penetration for different countries, online banking usage as well as shares of investors' attention explained by realized volatility and vice versa in a 100 day forecast period . The pattern of retail investors' interaction with financial markets in Italy and Spain is confirmed by internet penetration and online banking usage. Especially, Spain and Italy exhibit very low values of both measures. This could - on the one hand - help to explain the limited interaction with financial markets. On the other hand, it might point to the fact that for those countries retail investors' behavior is not accurately measured by the Google SVI. To overcome this potential limitation, we would have to make the assumption that Google search volume proxies retail investors' behavior in the whole population, which is not unlikely (see Da et al, 2011). Furthermore, stock market participation rates indicate similar results for Italy and Spain. Therefore, we can conclude with reasonable confidence that in those countries retail investors' interaction with financial markets is more limited than in other countries analyzed. Including Spain and Italy in the sample, correlation between the share of explained investors' attention by volatility is 0.68. Vice versa we observe a correlation coefficient of 0.39. Excluding Italy and Spain, we cannot find a strong correlation of investors' behavior with internet affinity in the other countries. We can observe high internet penetration rates for all countries besides Spain and Italy and the penetration has risen enormously during our sample period. As described in section 5.1.1 raising internet penetration correlates with higher shares of explained retail investors' activity. This in turn might indicate that it is rather a certain threshold of internet penetration than the level itself which helps to explain retail investors' interaction with financial markets and that the countries besides Spain and Italy have already reached this threshold.

5.2.2 Behavioral Characteristics

Our results raise the question whether behavioral characteristics of the retail investors differ and whether these characteristics can help to explain the cross-country variation. A vast amount of literature has engaged in investigating cultural differences in different regional settings. For financial markets related investigations, especially differences in the behavior of individuals dealing with risk and uncertainty are important. Since we link retail investors to noise traders, we are interested in non-rational behavior and cognitive biases which can be attributed to both noise traders (e.g. DeLong et.al, 1990) and retail investors (e.g. Barber and Odean, 2011).

Specifically, we want to relate our results to measures of loss aversion, uncertainty avoidance and impatience. The University of Zurich conducted the cross-country study "International test for risk attitudes" (INTRA), in which economics students from almost 60 countries answered a standardized test containing well-known behavioral finance questions. Among others, loss aversion and impatience measures were constructed. Hens (2013) uses these measures to analyze the connection between mutual fund flows and returns in a cross-country study. He discovers higher volatility in fund flows for countries in which loss aversion and impatience are high, indicating that mutual fund investors' behavioral characteristics can explain the additional volatility. Uncertainty avoidance is an indicator for the propensity of individuals to feel uncomfortable with uncertainty or ambiguity. The measure was originally developed by Hofstede (1991) to measure cultural differences and has been widely used in economic and finance literature since then (e.g. Rieger et. al, 2011).

Table 8: Cross-Country Behavioral Characteristics

This table presents behavioral characteristics of countries and the correlation of these measures with the explained share of variance in RV by SVI on the one hand, and with the variance share of SVI explained by RV. Moreover, correlations with the ratio of both aforementioned measures is reported. Data for loss aversion and patience is obtained from Hens and Caliskan (2013). Data for uncertainty avoidance is obtained from http://geert-hofstede.com/, which summarizes several empirical studies.

	Loss Aversion	Patience	Uncertainty Avoidance	a) Share of variance in RV explained by SVI	b) Share of variance in SVI explained by RV	c) Ratio of a) and b)
Germany	31.1	8.9	65.0	7.3%	18.9%	38.5%
USA	21.2	6.6	46.0	11.2%	41.5%	26.9%
UK	31.8	7.1	35.0	9.2%	24.3%	37.9%
France	21.4	6.5	86.0	16.9%	23.0%	73.7%
Netherlands	17.6	8.5	53.0	8.9%	27.6%	32.3%
Japan	21.7	7.4	92.0	6.5%	9.1%	72.0%
Spain	25.7	4.6	86.0	2.5%	22.5%	
Italy	21.8	4.4	75.0	0.8%	1.8%	
Correlation with a)	-0.38	**-0.73**	0.37	1.00		
Correlation with b)	-0.05	0.28	-0.58		1.00	
Correlation with c)	-0.18	-0.33	**0.89**			1.00

Table 8 reports loss aversion and impatience indices for the countries analyzed in this thesis, as well as correlation coefficients for the data series with variance decomposition shares in a 100 day forecast period . Impatience appears to be a factor which correlates with the share of realized volatility explained by investors' attention remarkably well. The USA and France are the countries which exhibit the largest explained share in realized volatility by investors' attention. Residents from these countries also appear to be among the most impatient in the sample. Hens (2013) points out that impatience coincide with irrational and spontaneous trading behavior. In our setting, this could explain that retail investors' attention in France and the USA funnels through to actual trading action to causing additional volatility. Following the noise trader model of Lux and Marchesi (1999), realized volatility causes investors' attention in the first place and investors' attention then causes additional realized volatility caused by their irrational trading behavior. If impatience is a characteristic, which measures the likelihood of irrational trading behavior, we would observe that more impatience goes hand in hand with a higher share of explained realized volatility through investors' attention. In other words, the less patient retail investors in France and the USA seem to react more strongly and directly to shocks in volatility. Overall, the correlation between the share of explained volatility by investors' attention and the patience measure is -0.73.

Interestingly, none of the measures we identified to characterize cultural differences between countries correlates with the amount of investors' attention explained by realized volatility to a degree higher than +/-0.6. Since the sample size is very small, we refrain from interpreting these correlations. Although there might be some underlying pattern, the behavioral characteristics we focus on are not able to fully explain when and how retail investors satisfy their information demand.

As already briefly discussed in section 4, we observe that for some countries volatility explains a high share in investors' attention, but investors' attention in turn does not explain much of realized volatility. For example in France and Germany, we can see that the share of explained volatility (by investors' attention) is rather small compared to the share of explained investors' attention (by volatility) (see Table 8). A potential interpretation of this measure is the degree to which retail investors get active after their attention has been drawn towards the stock market. Volatility normally coincides with uncertainty. Therefore, the degree to which individuals in a country are willing to accept uncertainty is crucial for the propensity of investors to get active. Interestingly, France and Germany exhibit among the highest uncertainty avoidance measures as introduced by Hofstede (1991), while the USA exhibits one of the lowest (see Table 8). At the same time, France and Germany are the countries in our sample where the aforementioned ratio is the lowest and the USA is the country for which it is the highest. Qualitatively, US investors can better handle uncertainty and decide more often not to get active in the financial markets after they have satisfied their information demand in uncertain times. On the other hand, French and German investors might not be willing to accept the uncertainty and rather sell stocks. The correlation coefficient for this relationship of 0.89 indicates a quite strong relationship . However, since the sample only consists of four indices, we don't consider our analysis to be statistically solid, but rather indicative for a potential existing relationship.

The same pattern becomes obvious when looking at volume instead of realized volatility. Uncertainty avoidance seems to strongly correlate with the share of variance in volume explained by investors' attention and vice versa. This correlation might indicate a similar behavior as described above in which investors with high uncertainty avoidance measures get active in financial markets to a greater extent in times of high volatility.

6. Robustness, Limitations and Further Research

6.1 Robustness

Considering the methodology we implement in this study, the inputs and specification of our VAR-models arguably are most critical to the results obtained[21]. To address concerns relating these issues, we evaluate the robustness of the results obtained quantitatively and qualitatively.

In order to estimate the lag length of the VAR- models, we chose to rely on Schwartz's Bayesian information criterion (MSBIC). However, other related studies have used predefined lag length of order two or three[22]. To test the robustness of results obtained and evaluate their sensitivity to the specification of lag length for the VAR-models used, we specify new VAR-models for search volume and our measures for market activity (realized and implied volatility, as well as illiquidity and trading volume) for the indices over the whole sample and all sub-periods. This time, we do not use information criteria but predefine the lag length for all VAR-models to two lags. The resulting Granger-test statistics and variance decompositions are very similar to those derived by the original VAR-models using MSBIC[23]. Differences include changes in significance levels for a small number of measures. More important, all main conclusions of our study are still supported. Even with uniform lag length of two for the VAR-models, we conclude a bidirectional causal relationship between Google's SVI and realized volatility for almost all indices over the complete sample period. Interestingly, the Granger test statistics of the NIKKEI 225 and IBEX 35 have become more significant whereas those for the FTSE MIB have become less significant. Further, the results for realized and implied volatility are still time-variant and most significant during the crisis period, followed by the post-crisis period. Although variance in realized volatility explained by variation in the SVI have increased for most indices over the complete period, realized volatility of the CAC 40 still exhibits the highest portion attributable to individual investors' attention. Another finding supported by the new models is the greater share of investors' attention that is explained by implied volatility in relation to realized volatility over the complete sample period. Finally, illiquidity and trading volume still exhibit a significant bi-directional Granger-causal relation over the entire sample period.

[21] Braun and Mittnik (1993) find severe inconsistencies for impulse response functions and variance decomposition by misspecifications of the VAR-models used, for example incorrectly specified lag length

[22] Kita and Wang (2012) use lag length of 2 for their VAR-models; Dimpfl and Jank (2011) use lag length of 3 for their VAR-models

[23] Results for the VAR-models with uniform lag length of two are not reported here, but are available from the authors on request.

Furthermore, our results arguably might depend critically on the specification of input time series. To reduce the number of non-stationary series and decrease excess skewness and kurtosis, we transformed our data by taking natural logarithms. However, there is a tradeoff between working with original data time series and transforming the series to improve distributional characteristics. On the one hand, using data transformation reduces the probability of obtaining ambiguous VAR-parameters, as non-stationarity of input variables can bias the statistical significance of the VAR coefficients[24]. On the other hand, the transformation process ignores certain information contained in the time series, leading to the possibility of manipulated conclusions about the interrelations between two data series. One major concern for using non-stationary series is that the Wald-test statistic does not follow an exact chi-squared distribution anymore when implemented on non-stationary series (see Lütkepohl, 2005). To account for this issue, and to improve the distributional characteristics of our input series, we opted to stay in line with previous studies working with Google Trends and volatility data. Hence, we transform our observed data series by taking natural logarithms[25]. By taking logs, we could transform some of the stock sample data series to come close to a normal distribution (see JB-test statistics in Tables 10-11). However, we have to note that working with non-transformed normalized data series might lead to different results.

6.2 Limitations

Various limitations to specific interpretations have been pointed out in the present work. In this section we want to complement those with general limitations of our study with regard to a broader context. To clearly infer that retail investors' activity is caused by financial markets and that their action in turn is causal for developments in these financial markets, we have to make the critical assumption that the same individuals of whom we measure information demand or attention are those getting active in the markets (Dimpfl and Jank, 2011). We are aware of the fact that this is quite a strong assumption. However, to draw valid conclusions, it would be sufficient that those who use Google as an information source for making investment decisions are representative for the majority of retail investors, even if other sources of information are used as well. Google is probably more extensively used by younger individuals who are not necessarily those owning financial assets and being active on financial markets. However, given Google's market share in the countries analyzed, as well as the results of different studies (see literature review, most notably Da et al, 2012), Google Trends seems to be very effective in capturing the behavior of individuals and groups. More-

[24] See Brooks (2008)

[25] See methodology used by Vlastakis and Markellos (2012); Da et al (2011); Dimfl und Jank (2011)

over, as the most active retail investors on financial markets are those owning online (discount) brokerage accounts (see Barber and Odean, 2002, 2009), we can assume that Google's market share is representative for the portion of this group's information demand that is captured by the SVI. However, we have to be careful when interpreting SVI data for Japan, as in contrast to the other countries, Google is not the market leading search engine in this country.

Unfortunately, for the stock sample analyzed, Google search volume measured on a daily basis cannot be used before 2010 due to periodically missing values. In addition, we are not able to utilize daily realized volatility data for our stock sample. As results obtained by the analysis of the index sample clearly indicate that the strongest effects normally occur during the first days, results on the individual stock level might be biased due to the (forced) reliance on weekly data. In particular, weekly results might hide certain contrasting dynamics that happen consecutively but within the same week and only reveal the overall effect. Looking at the results of the indices analysis, this seems quite likely as a large portion of the effects take place within the first days here – i.e. investors process information quickly. However, results obtained on the index level also show that bi-directional interaction patterns become less pronounced in the course of a week, indicating that results for the stock sample are rather under than overestimated.

Another critical assumption is that investors who satisfy their information demand by using Google are retail investors and further can be characterized as noise traders. This assumption is crucial for relating our empirical results to academic theory and noise trader models. Since professional and institutional investors, in particular in stock markets, usually have access to more sophisticated information sources like Bloomberg, ThomsonReuters etc., we are confident that stock market related search volume on Google mainly captures retail investors' information demand. Several characteristics are attributed to noise traders, which makes it impossible to generally relate retail investors' behavior to all noise trader models. However, noise traders are mainly characterized as investors without superior information who act irrationally (see Black, 1986; DeLong et al, 1990). We deem it as rather unlikely that investors using the Google Search engine as source for financial information in contrast to more sophisticated information sources possess superior or inside information. In addition, non-rational behavior seems to be confirmed for retail investors by academic literature. (see discussion by Da et. al, 2012 and section 2 for linking retail investors and noise traders). However, we cannot fully neglect the possibility that the behavior of investors captured by the SVI deviates from noise traders' characteristics in academic models.

Concerning the interpretation of results obtained in the present study, certain limitations arise as well. Firstly, the cross-sectional analysis of underlying patterns on a country (index) level is limited by the sample size of countries. Due to data availability, our sample comprises between four and nine countries, depending on the measure under examination. Therefore, relating cross-country differences to country specific structural and behavioral measures is merely qualitative than quantitative. However, we still believe that these interpretations help to relate our statistical findings to interesting economic and behavioral differences across countries. As a consequence, we provide an outlook on potential further research topics and relate our results to a greater economic context.

Finally, we want to address potential limitations concerning our main findings about time varying behavior of retail investors and in relation to the sample period. Our results indicate that retail investors' interaction with financial markets is most pronounced during the financial crisis period and more pronounced in the post-crisis than the pre-crisis period. Primarily, we relate these findings to technological development and different behavior in periods of high market fluctuations. Unfortunately, Google's search volume data can only be used appropriately from year 2006 on. Therefore, the whole sample period ranging from 2006 to 2013 is relatively biased towards times of high market fluctuations compared to previous time periods. Specifically, our sample period includes the global financial crisis, the succeeding economic downturn, the EURO-crisis as well as the Fukushima nuclear disaster. As a consequence, our study might over-interpret retail investors' influence on financial markets in general. On the one hand, the finding of increased interaction during crises periods is very interesting in itself. On the other hand, practical importance, e.g. for improving volatility forecast models, might be severely reduced in times of lower market volatility.

6.3 Further Research

Future research can be conducted in various directions. Firstly, the bi-directional relationships could be analyzed in different financial market environments. For example, one would expect that retail investors are less active on bond markets and commodity markets than on stock markets. Moreover, investigating the individual stock level in more detail might yield further insights of retail investors' and noise traders' behavior on equity markets. Since our results on the stock level are inconclusive regarding the influence of investors' attention on volatility, using a more sophisticated daily measure for realized volatility might add interesting insights. Furthermore, results regarding the difference between realized and implied volatility promise to be interesting when analyzed on the stock level. Further analysis could also be conducted to identify the underlying drivers of the high predictive power of implied volatility on investors' attention. Secondly, our research could be expanded by matching investors' attention measured by Google with actual trading data of retail investors received from online brokers. Thereby, the link between investors' attention and actual trading actions could be analyzed on a more detailed level. This would not only link the study to those investigating retail investors' behavior (e.g. Barber and Odean, 2011), but also address one of the structural limitations of using the SVI to analyze retail investor interaction with financial markets. Thirdly, the cross-country analysis could be extended to a greater number of countries. Thereby, our limitation of a limited sample size could be addressed. As already pointed out in section 5.2.1, more structural measures including ownership type (mutual fund, directly, ETF) could then be tested with regard to explanatory power of retail investors' activity. Furthermore, a cultural comparison along other dimensions not used in this thesis could be of interest and might contribute to behavioral finance academia.

Regarding further analysis on the single stock level, Kumar and Lee (2006) find that stocks with low institutional ownership are preferred by retail investors. Therefore, it would be highly interesting to evaluate our results in the background of non-institutional holding fractions for the 22 stocks over the sample period in order to shed further light on the influence of retail investors on those stocks.

Lastly, we would like to point out the possibilities of practically applying our findings. Dimpfl and Jank (2011) already showed that forecast models for realized volatility can be improved significantly by using Google's SVI. Due to the multi-step investment decision process of retail investors, this finding can be built on the inherent forecast abilities of a measure capturing investors' attention. Further investigating our findings to forecast implied volatility might be of great interest for option pricing in both a practical and academic environment. Moreover, since implied volatility is often used as risk-aversion indicator (for example the VIX), an improved estimation of future implied volatility might help to develop a better understanding of risk-aversion in financial markets.

7. Conclusion

In the present thesis, we apply a new method to measure the attention of individuals – the Google Search Volume Index – to investigate the interaction of retail investors with financial markets in a cross-country setting. We provide empirical evidence for existing noise trader models and shed further light on the underlying factors of retail investor activity.

On the country level, retail investors' attention is caused by volatility and this attention funnels through to additional volatility in the future. Moreover, it seems like retail investors' attention is triggered to a greater extent by persistent volatility shocks (implied volatility) than by idiosyncratic volatility shocks (realized volatility). Retail investors' activity, however, helps to explain future realized volatility rather than future implied volatility to a greater extent, implying that some market participants actually take future retail investors' behavior into account. Additionally, retail investors create liquidity in the short-term and increase trading volume. Further, retail investors attribute a larger share of their attention to stock markets in times of high market fluctuations and in more recent times. During these times, their activity also has the greatest explanation power for volatility and illiquidity. Potential reasons include an increased and persisting risk aversion and the increased propensity to become active on financial markets due to technological developments. On the stock level, retail investors are likely to attribute their limited cognitive resource attention to the most popular stocks.

Not surprisingly, stock market participation rates in the different countries correlate with retail investors' financial markets interaction measures. This explains the low stock market interaction of Spanish and Italian retail investors compared to US and French retail investors. Additionally, country-specific behavioral patterns like individuals' impatience and tendency to avoid uncertainty seem to contribute to the magnitude and likelihood of interaction with financial markets. French and American investors are among the most impatient and their information demand therefore explains a large share of future volatility. French and German investors are among the most uncertainty avoiding and therefore seem to be the most likely to funnel through their information demand to actual trading.

Our interpretations relate the findings to behavioral finance and thereby indicate interesting areas of future research in academia. Moreover, the possibility to measure retail investors' attention precisely and in a timely manner in combination with our results showing how, when and why retail investors are active in the markets, might be of practical importance. Professionals could apply our findings to improve volatility forecasts or build trading strategies to exploit the predictive power of retail investors' attention.

References

Amihud, Y., 2002, Illiquidity and stock returns: cross-section and time-series effects, Journal of financial markets, 5(1), 31-56.

Alfarano, S. and Lux, T, 2007, A noise trader model as a generator of apparent financial power laws and long memory, Macroeconomic Dynamics 11(Supplement S1), 80-101.

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

Bank, M., Larch, M. and Peter, G., 2011, Google search volume and its influence onliquidity and returns of German stocks, Financial Markets and Portfolio Management 25, 239–264.

Barber, B.M., Lee, Y.T., Lee, Y.J. and Odean, T., 2009, Just How Much Do Individual Investors Lose by Trading?, The Review of Financial Studies, Vol. 22, No. 2 (Feb., 2009), pp. 609-632.

Barber, B.M., and Odean, T., 2002, Online investors: Do the slow die first? Review of Financial Studies 15, 455–487.

Barber, B.M. and Odean, T., 2008, All that Glitters: The Effect of Attention on the Buying Behavior of Individual and Institutional Investors, Review of Financial Studies, 21: 785-818.

Barber, B.M., and Odean, T., 2011, The behavior of individual investors, available at SSRN 1872211.

Barber, B.M., Odean, T., and Zhu, N., 2009a, Do Retail Trades Move Markets?, Review of Financial Studies 22, 151-186.

Barber, B.M., Odean, T., and Zhu, N., 2009b, Systematic Noise, Journal of Financial Markets, 12, 547-469.

Benartzi, S. and Thaler, R. H., 2001, Naive diversification strategies in defined contribution saving plans, American economic review, 79-98.

Black, F., 1986, Noise, Journal of Finance 41, 529–543.

Blankespoor, E., Miller, G., and White, H., 2011, The impact of managerial dissemination of firm disclosure, Working Paper, University of Michigan.

Bloomfield, Robert, Maureen O'Hara, and Gideon Saar, 2009, How noise trading affects markets: An experimental analysis, Review of Financial Studies 22, 2275–2302.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., and Weber, I., 2012, Web search queries can predict stock market volumes, *PloS one*, *7*(7), e40014.

Brooks, C., 2008, Introductory econometrics for finance, Cambridge university press.

Campbell, J. Y., and Kyle, A. S., 1993, Smart money, noise trading, and stock price behaviour, Quarterly Journal of Economics 108, 905–939.

Choi, H. and Varian, H, 2009b, Predicting the present with Google trends, Working Paper pp. 1–23.

Choi, H., and Varian, H., 2012, Predicting the present with google trends, Economic Record, 88(s1), 2-9.

Choi, H. and Varian, H, 2009b, Predicting the present with Google trends, Working Paper pp. 1–23.

Christelis, D., Georgarakos, D., and Haliassos, M., 2013, Differences in portfolios across countries: Economic environment versus household characteristics, Review of Economics and Statistics, 95(1), 220-236.

Christensen, B. J., and Prabhala, N. R., 1998, The relation between implied and realized volatility, Journal of Financial Economics, 50(2), 125-150.

Da, Z., Engelberg, J. and Gao, P, 2010a, In search of earnings predictability, Working Paper.

Da, Z., Engelberg, J. and Gao, P., 2010b, The Sum of All FEARS: Investor Sentiment and Asset Prices, Working Paper.

Da, Z., Engelberg, J. and Gao, P., 2011, In Search of Attention, The Journal of Finance 66(5), 1461–1499.

Da, Z., Liu, Q., and Schaumburg, E., 2013, A Closer Look at the Short-Term Return Reversal, Management Science.

DeLong, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, 1990, Noise trader risk in financial markets, Journal of Political Economy 98, 703–738.

Dimpfl, T., and Jank, S., 2011, Can internet search queries help to predict stock market volatility? (No. 11-15), CFR working paper.

Dimpfl, T., and Jank, S., 2012, Can Internet Search Queries Help to Predict Stock Market Volatility?, Paris December 2012 Finance Meeting EUROFIDAI-AFFI Paper.

Drake, M., Roulstone, D. and Thornock, J., 2011, Investor Information Demand: Evidence from Google Searches around Earnings Announcements, Working Paper .

Eeckhoudt, L., and Godfroid, P., 2000, Risk aversion and the value of information, The Journal of Economic Education, 31(4), 382-388.

Engle, R., 2011, What Is Happening With Financial Market Volatility and Why? In Volatility (pp. 29-45), Springer US.

Eurostat, 2013, Individuals using the Internet for Internet banking, Retrieved on November 13, 2013 from http://epp.eurostat.ec.europa.eu/tgm/table.do?tab=table&plugin=1&language=en&pcode=tin00099.

Foucault, T., Sraer, D. and Thesmar, D. J, 2011, Individual Investors and Volatility, The Journal of Finance 66(4), 1369–1406.

French, K.R. and J.M. Poterba, 1991, International Diversification and International Equity Markets, American Economic Review 81:222-226.

Gawlik, E., Kabaria, H., and Kaur, S., 2011, Predicting tourism trends with Google Insights.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L., 2009, Detecting influenza epidemics using search engine query data, Nature 457(7232), 1012–1014.

Goyenko, R., Holden, C., Trzcinka, C., 2008, Do measures of liquidity measure liquidity?, Available at SSRN 1108553.

Greene, J., and Smart, S., 1999, Liquidity provision and noise trading: evidence from the "investment dartboard" column. the Journal of Finance, 54(5), 1885-1899.

Guiso, L., Sapienza, P., and Zingales, L., 2011, Time varying risk aversion, Working paper.

Hacamo, I., and Reyes, T., Negativity Bias in Attention Allocation: Retail Investors' Reaction to Stock Returns.

Hens, T. and Nilufer Caliskan, 2013, Behavioural Finance and Mutual Fund Flows: An International Study. Retrieved on November 13, 2013 from https://www.dgfi.com/DGFI/White-Papers/Behavioural-Finance-and-Mutual-Fund-Flows

Hofstede, G., 1991, The confucius connection: From cultural roots to economic growth, Organization Dynamics 16(4), 418.

Hu, G.X., Pan, J., and Wang, J., 2013, Noise as Information for Illiquidity, The Journal of Finance 68(6), 2341-2382.

Hvidkjaer, S., 2008, Small trades and the cross-section of stock returns, Review of Financial Studies 21, 1123–1151..

Jacobs, H., and Weber, M., 2012, The trading volume impact of local bias: Evidence from a natural experiment, Review of Finance, 16(4), 867-901.

Jarque, C. M., and Bera, A. K., 1987, A test for normality of observations and regression residuals, International Statistical Review/Revue Internationale de Statistique, 163-172.

Kahneman, Daniel., 1973, Attention and Effort (Prentice-Hall, Englewood Cliffs, NJ).

Kita, A., and Wang, Q., 2012, Investor attention and fx market volatility, Bangor Business School, 3, 33.

Kogan, L., Ross, S. A., Wang, J., and Westerfield, M. M., 2006, The price impact and survival of irrational traders, The Journal of Finance, 61(1), 195-229.

Kumar, Alok, and Charles M.C. Lee, 2006, Retail investor sentiment and return comovements, Journal of Finance 61, 2451–2486.

Kyle, A.S. (1985), “Continuous Auctions and Insider Trading,” Econometrica 53:1315-1336.

Lee, C. (1993). Market Integration and Price Execution for NYSE-Listed Securities. The Journal of Finance, 48(3), 1009-1038.

Ljung, G. M., and Box, G. E. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297-303.

Lux, T. and Marchesi, M.: 1999, Scaling and criticality in a stochastic multi-agent model of a financial market, Nature 397(6719), 498–500.

Lütkepohl, H., 2005, New Introduction to Multiple Time Series Analysis, Springer Verlag Berlin

Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, Journal of Finance 42, 483–510.

Mondria, J., Wu, T., and Zhang, Z., 2010, The determinants of international investment and attention allocation: using internet search query data, Journal of International Economics 82, 85–95.

Obsfeld, M., and Rogoff, K., 2000, The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?, NBER Macroeconomics Annual 2000, 15 MIT Press, 339–390.

Preis, T., Reith, D., and Stanley, H. E., 2010, Complex dynamics of our economic life on different scales: insights from search engine query data, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368(1933), 5707-5719.

Preis, T., Moat, H. S., and Stanley, H. E., 2013, Quantifying trading behavior in financial markets using Google Trends. Scientific reports, 3.

Rieger, M., Wang, M., and Hens, T., 2011, Prospect theory around the world, NHH Dept. of Finance and Management Science Discussion Paper, (2011/19).

Rubin, A., and Rubin, E., 2010, Informed investors and the Internet. Journal of Business Finance and Accounting, 37(7-8), 841-865.

Said, S. E., and Dickey, D. A., 1984, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika, 71(3), 599-607.

Schwarz, G., 1978, Estimating the dimension of a model. The annals of statistics, 6(2), 461-464.

Sheppard, K., 2009, MFE MATLAB Function Reference Financial Econometrics, Oxford-Man Institute of Quantitative Finance, Retrieved on December 4, 2013 from http://www.kevinsheppard.com/images/9/95/MFE_Toolbox_Documentation.pdf

Shleifer, A., and Summers, L. H., 1990, The noise trader approach to finance, The Journal of Economic Perspectives, 4(2), 19-33.

Sprenger, T. O., Tumasjan, A., Sandner, P. G., and Welpe, I. M., 2013, Tweets and trades: The information content of stock microblogs, European Financial Management.

Stock, J. H., 1996, VAR, error correction and pretest forecasts at long horizons. Oxford Bulletin of Economics and Statistics, 58(4), 685-701.

Tetlock, Paul C., 2007, Giving content to investor sentiment: The role of media in the stock market, Journal of Finance 62, 1139–1168.

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

The World Bank, 2013, Internet users per 100 people, Retrieved on November 13, 2013 from http://data.worldbank.org/indicator/IT.NET.USER.P2.

Wang, F. A., 2010, Informed arbitrage with speculative noise trading, Journal of Banking & Finance 34 (2010) 304–313.

Appendix

A. Figures

Figure 1: Overall search volume stocks

The diagram shows the overall search volume for the 22 stocks in the sample period from January 2008 until September 2013 relative to search volume for Microsoft (indexed at 100).

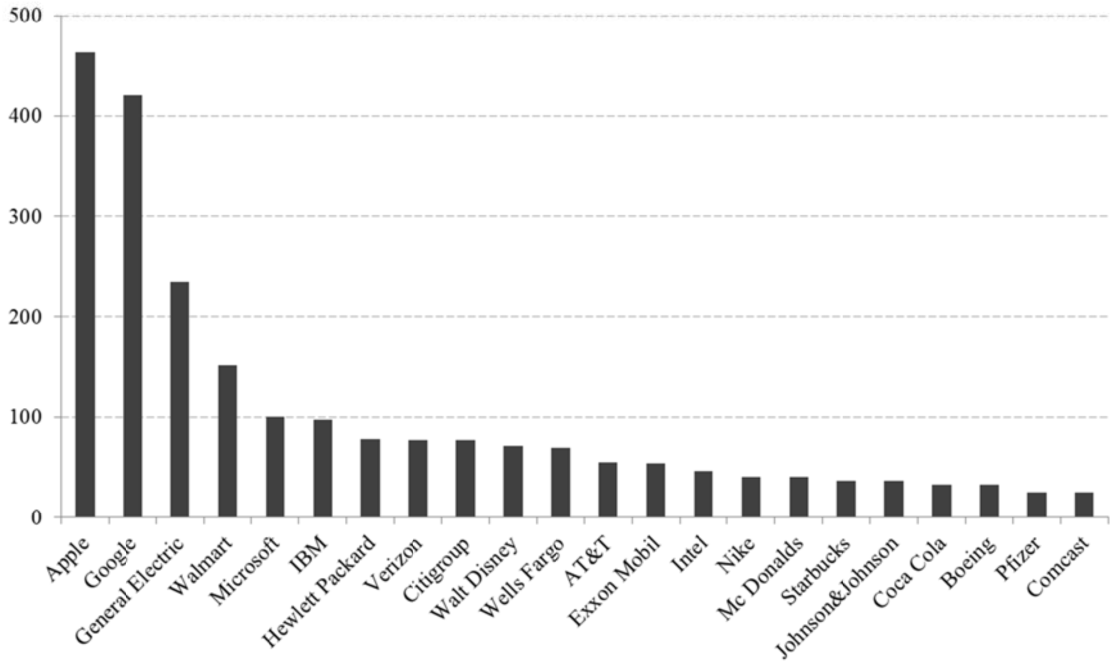
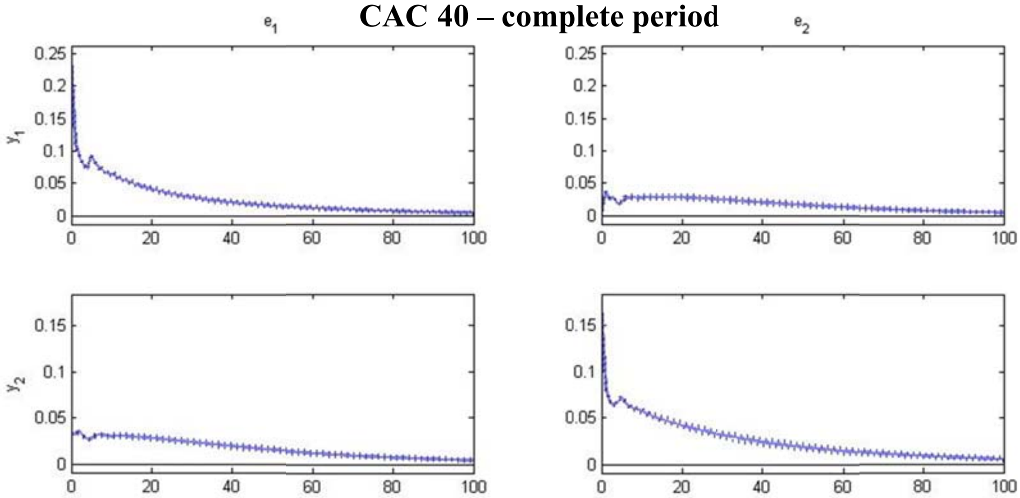


Figure 2: Realized Volatility: Impulse response functions CAC 40 and FTSE MIB

The diagram shows impulse response functions (IRFs) to shocks in the error terms for realized volatility (RV) in the complete time period (January 2006 until September 2013) for the CAC 40 and FTSE MIB. The x-axis of each diagram shows forecast steps, i.e. one step corresponds to one day. The y-axis indicates the magnitude of the impulse response. y1 corresponds to log realized volatility as the dependent variable (i.e. the first VAR-equation), and y2 to the (log) search volume index (i.e. the second VAR-equation) as the dependent variable. e1 is the first error term corresponding to realized volatility, e2 is the

CAC 40 – complete period



FTSE MIB – complete period

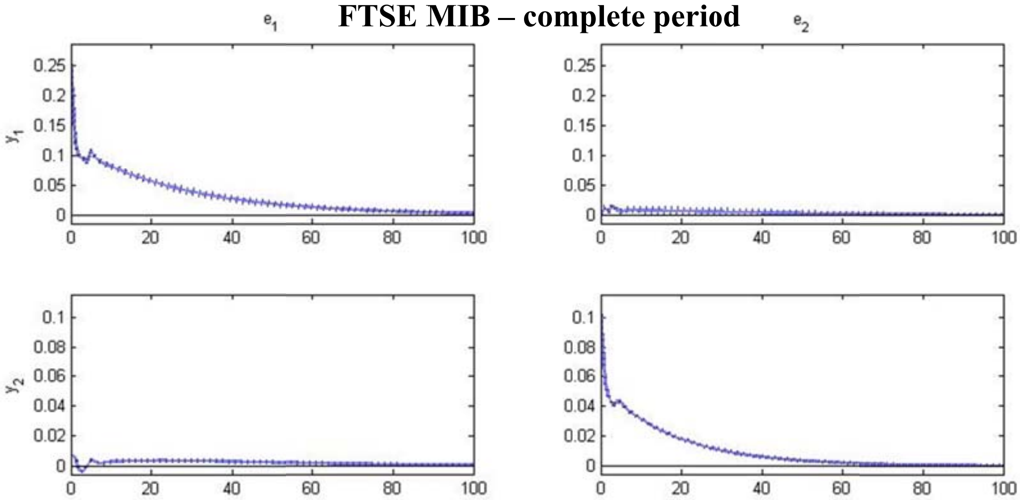
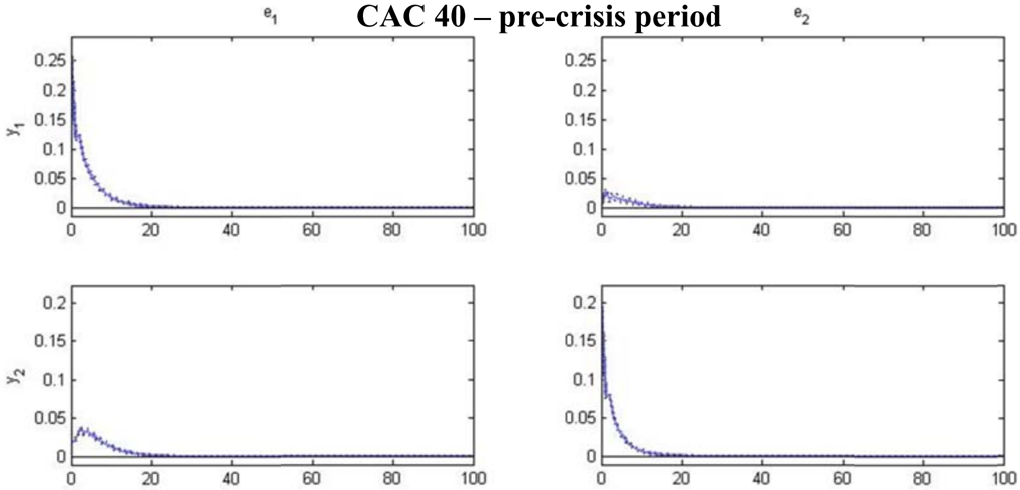


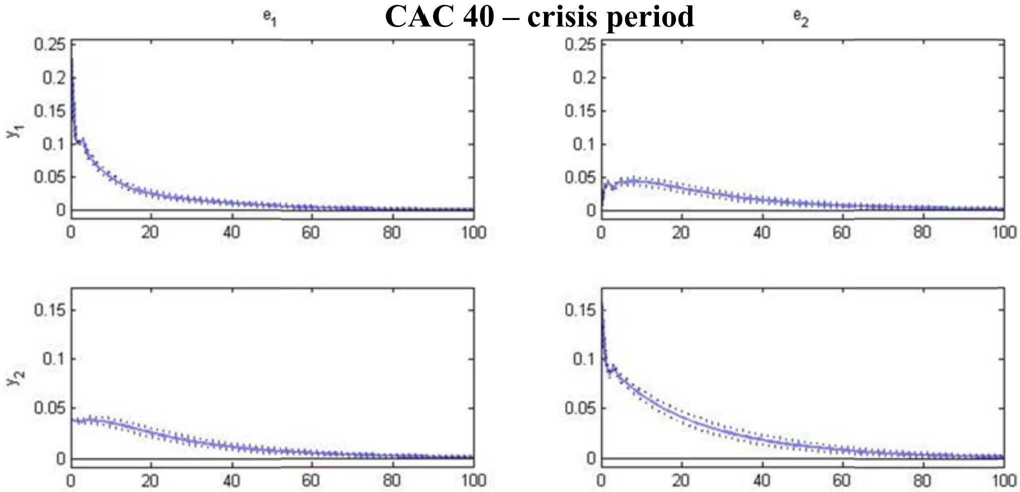
Figure 3: Realized Volatility: Impulse response functions CAC40

The diagram shows impulse response functions (IRFs) to shocks in the error terms for realized volatility (RV) in the three sub periods for the CAC 40. The x-axis of each diagram shows forecast steps, i.e. one step corresponds to one day. The y-axis indicates the magnitude of the impulse response. y1 corresponds to log realized volatility as the dependent variable (i.e. the first VAR-equation), and y2 to the (log) search volume index (i.e. the second VAR-equation) as the dependent variable. e1 is the first error term corresponding to realized volatility, e2 is the second error term corresponding to search volume. The impulse response function plots the effect of a unit shock in each of the error terms to the dependent variable. So the upper right diagram in both graphs shows the effect of a unit shock in search volume to real-

CAC 40 – pre-crisis period



CAC 40 – crisis period



CAC 40 – post-crisis period

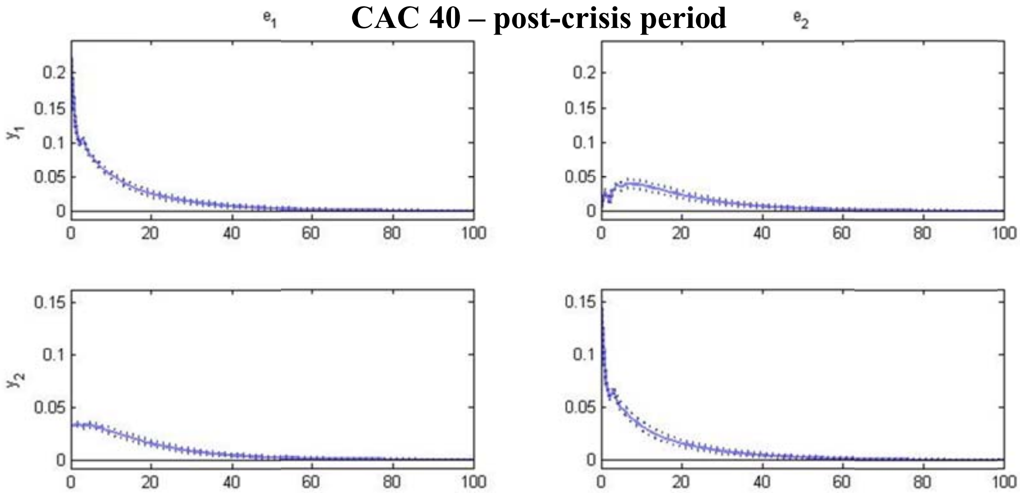
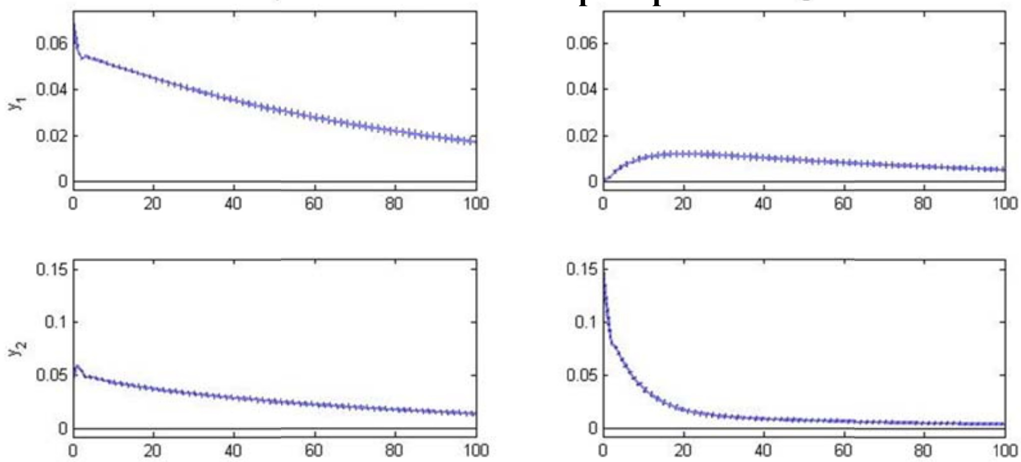


Figure 4: Implied Volatility: Impulse response functions DJIA and CAC 40

The diagram shows impulse response functions (IRFs) to shocks in the error terms for implied volatility (IV) in the complete time period (January 2006 until September 2013) for the Dow Jones Industrial Average and CAC 40. The x-axis of each diagram shows forecast steps, i.e. one step corresponds to one day. The y-axis indicates the magnitude of the impulse response. y1 corresponds to log implied volatility as the dependent variable (i.e. the first VAR-equation), and y2 to the (log) search volume index (i.e. the second VAR-equation) as the dependent variable. e1 is the first error term corresponding to realized volatility, e2 is the second error term corresponding to search volume. The impulse response function plots the effect of a unit shock in each of the error terms to the dependent variable. So the upper right diagram

Dow Jones – complete period



CAC 40 – complete period

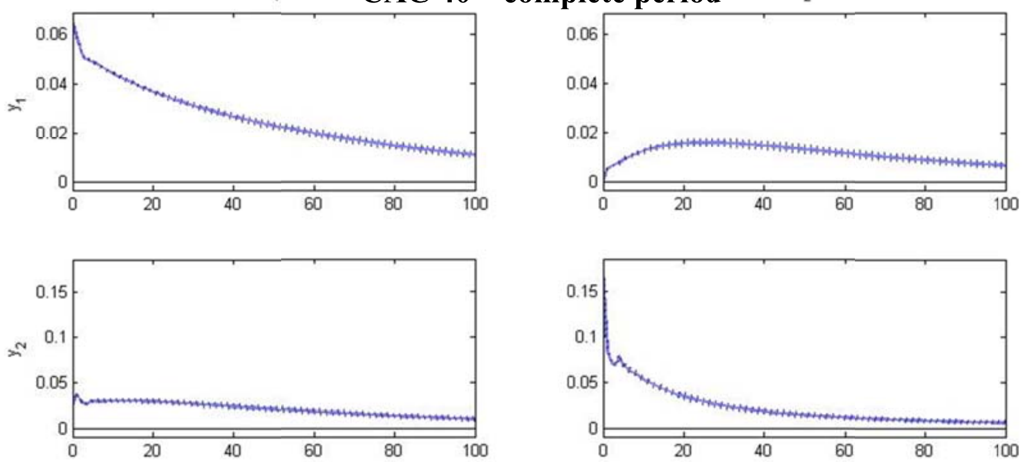
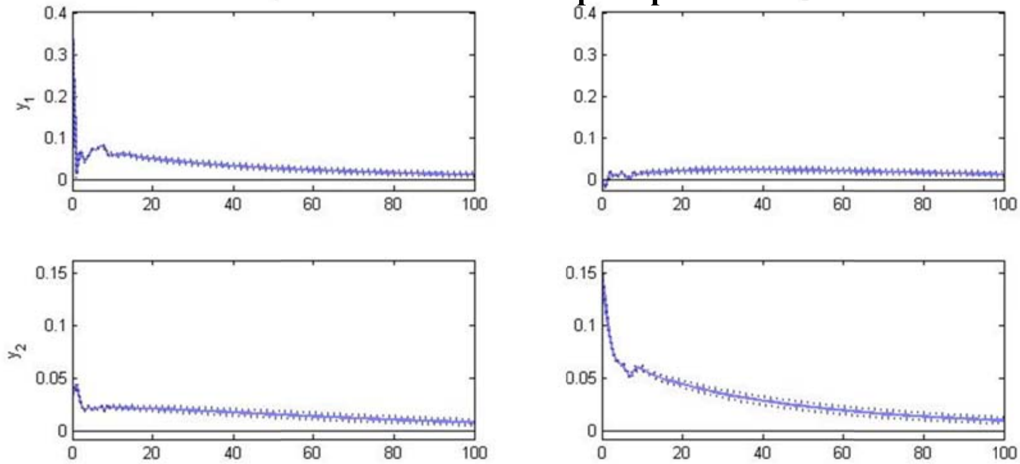


Figure 5: Illiquidity: Impulse response functions Dow Jones and CAC 40

The diagram shows impulse response functions (IRFs) to shocks in the error terms for our Amihud-illiquidity measure (ILLQ) in the complete time period (January 2006 until September 2013) for the Dow Jones Industrial Average and CAC 40. The x-axis of each diagram shows forecast steps, i.e. one step corresponds to one day. The y-axis indicates the magnitude of the impulse response. y1 corresponds to log illiquidity as the dependent variable (i.e. the first VAR-equation), and y2 to the (log) search volume index (i.e. the second VAR-equation) as the dependent variable. e1 is the first error term corresponding to realized volatility, e2 is the second error term corresponding to search volume. The impulse response function plots the effect of a unit shock in each of the error terms to the dependent variable. So the upper

Dow Jones – complete period



CAC 40 – complete period

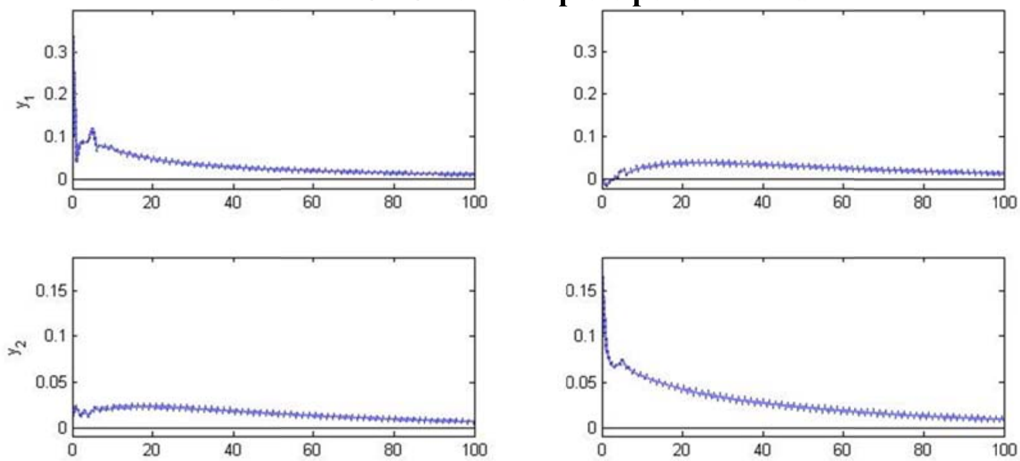
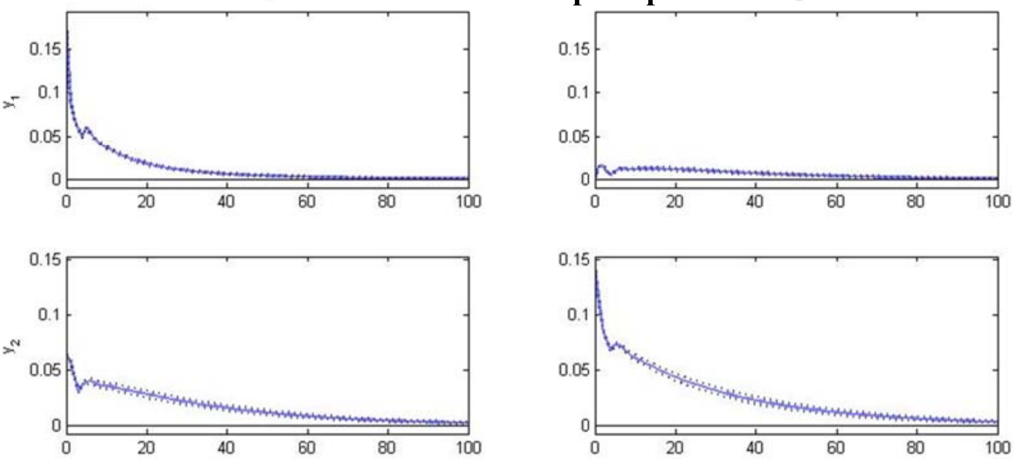


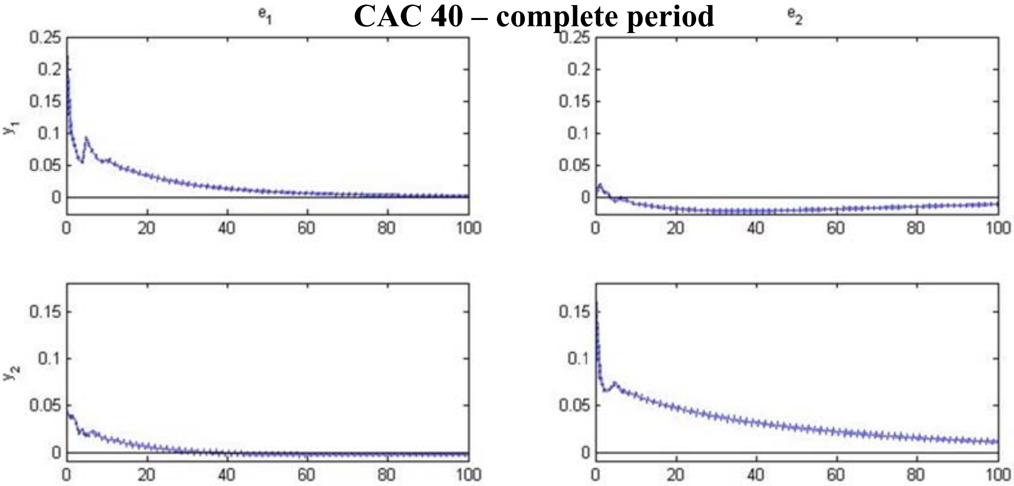
Figure 6: Trading Volume: Impulse response functions Dow Jones and CAC 40

The diagram shows impulse response functions (IRFs) to shocks in the error terms for trading volume (IV) in the complete time period (January 2006 until September 2013) for the Dow Jones Industrial Average and CAC 40. The x-axis of each diagram shows forecast steps, i.e. one step corresponds to one day. The y-axis indicates the magnitude of the impulse response. y1 corresponds to log trading volume as the dependent variable (i.e. the first VAR-equation), and y2 to the (log) search volume index (i.e. the second VAR-equation) as the dependent variable. e1 is the first error term corresponding to realized volatility, e2 is the second error term corresponding to search volume. The impulse response function plots the effect of a unit shock in each of the error terms to the dependent variable. So the upper right diagram

Dow Jones – complete period



CAC 40 – complete period



B. Tables

Table 9: Google search queries used for indices and stocks

The table shows the search queries used for extracting Google search volume indices (SVI) from Google Trends. We use the search term with the highest overall volume and control with Google correlate for ambiguous search terms related to the specific search query.

Prime equity index sample

Country	Index	Search Query Used
Germany	DAX 30	"dax"
United States	Dow Jones Industrial Average	"dow"
United Kingdom	FTSE 100	"ftse"
France	CAC 40	"cac"
Netherlands	AEX	"aex"
Japan	Nikkei 225	"225"
Brazil	Indice Bovespa	"bovespa"
Spain	IBEX 35	"ibex"
Italy	FTSE MIB	"bolsa"

U.S. stock sample

Company	Search Query	Company	Search Query Used
Apple Inc.	"apple stock"	Intel Corporation	"intel stock"
AT&T Inc.	"at&t stock"	Johnson & Johnson	"johnson johnson stock"
Boeing	"boeing stock"	McDonald’s Corporation	"mcdonalds stock"
Citigroup Inc.	"citigroup stock"	Microsoft Corporation	"microsoft stock"
The Coca-Cola Company	"coca cola stock"	Nike Inc.	"nike stock"
Comcast Corporation	"comcast stock"	Pfizer Inc.	"pfizer stock"
ExxonMobil Corp.	"exxon stock"	Starbucks Corp.	"starbucks stock"
General Electric	"ge stock"	Verizon Communications .	"verizon stock"
Google Inc.	"google stock"	Wal-Mart Stores Inc.	"walmart stock"
Hewlett-Packard L.P.	"hp stock"	Walt Disney Company	"disney stock"
IBM Corporation	"ibm stock"	Wells Fargo	"wells fargo stock"

Table 10: Summary Statistics indices: Search volume index and realized volatility

The table presents the summary statistics of the Google search volume index (SVI) and realized volatility (RV) for the nine prime equity indices. Further, p-values of test statistics are given for the Jarque-Bera (JB)-test on normality, the augmented Dickey-Fuller (ADF)-test on the presence of a unit-root (non-stationarity) and the Ljung-Box (LB)-test on independent distribution (no autocorrelations) of the time series. For the ADF-test and LB-test, the time series are tested on 100 lags, whereas the stated p-value of the LB-test gives the sum of all p-values over 100 lags tested. Summary statistics are given for the normalized data series ("norm"), as well as the natural logarithm time series ("log"). Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

Search volume index

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log
Mean	1.000	-0.077	1.001	-0.118	1.001	-0.071	1.009	-0.096	1.000	-0.111	1.000	-0.051	1.018	-0.158	1.000	-0.061	1.000	-0.021
Std. Deviation	0.587	0.343	0.733	0.420	0.553	0.333	0.675	0.413	0.700	0.414	0.337	0.314	0.635	0.596	0.367	0.348	0.242	0.194
Skewness	6.590	1.862	5.982	1.837	6.261	1.962	6.404	1.075	5.745	1.569	1.574	0.194	1.702	0.042	1.537	0.035	4.328	1.475
Kurtosis	63.72	10.377	60.17	7.817	62.10	9.973	63.99	7.412	48.77	7.635	8.647	3.357	8.224	2.247	8.723	3.438	37.65	9.543
Minimum	0.375	-0.981	0.417	-0.874	0.433	-0.837	0.259	-1.353	0.288	-1.244	0.351	-1.046	0.235	-1.447	0.227	-1.481	0.537	-0.622
Maximum	9.254	2.225	12.679	2.540	9.445	2.245	11.062	2.403	9.362	2.237	3.727	1.316	5.205	1.650	4.110	1.413	3.587	1.277
JB-test *(p-value)*	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***
ADF-test *(p-value)*	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***
sum LB-test *(p-values)*	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***

Realized volatility

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log
Mean	1.000	-0.131	1.000	-0.172	1.000	-0.134	1.000	-0.119	1.000	-0.138	1.000	-0.131	1.000	-0.111	1.000	-0.123	1.000	-0.136
Std. Deviation	0.603	0.492	0.758	0.543	0.612	0.495	0.566	0.470	0.621	0.501	0.660	0.474	0.604	0.436	0.540	0.495	0.582	0.513
Skewness	2.816	0.466	3.738	0.797	2.797	0.553	2.794	0.422	2.691	0.548	3.663	0.893	3.709	0.945	2.278	-0.014	2.227	0.223
Kurtosis	16.55	3.441	25.82	3.847	16.81	3.290	17.62	3.436	14.66	3.330	22.93	4.486	24.30	4.892	13.79	3.125	12.23	2.940
Minimum	0.213	-1.544	0.229	-1.475	0.253	-1.373	0.204	-1.590	0.186	-1.683	0.290	-1.238	0.231	-1.467	0.171	-1.769	0.226	-1.489
Maximum	6.483	1.869	10.073	2.310	7.118	1.963	6.819	1.920	6.562	1.881	7.014	1.948	7.294	1.987	6.041	1.799	6.172	1.820
JB-test *(p-value)*	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	.001***	0.500	.001***	.001***
ADF-test *(p-value)*	.000***	.000***	.003***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***
sum LB-test *(p-values)*	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***	.000***

Table 11: Summary Statistics Indices: Implied volatility, illiquidity and trading volume

The table presents the summary statistics of the implied volatility, illiquidity and trading volume for six (volatility) and four (illiquidity, trading volume) prime equity indices. Further, p-values of test statistics are given for the Jarque-Bera (JB)-test on normality, the augmented Dickey-Fuller (ADF)-test on the presence of a unit-root (non-stationarity) and the Ljung-Box (LB)-test on independent distribution of the time series. For the ADF-test and LB-test, the time series are tested on 100 lags, whereas the stated p-value of the LB-test gives the sum of all p-values over 100 lags tested. Summary statistics are given for the normalized data series ("norm"), as well as the natural logarithm time series ("log"). Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

Implied Volatility

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225	
	norm	log	norm	log	norm	log	norm	log	norm	log	norm	log
Mean	1.000	-0.062	1.000	-0.085	1.000	-0.072	1.000	-0.058	1.000	-0.081	1.000	-0.060
Std. Deviation	0.404	0.333	0.476	0.390	0.430	0.362	0.378	0.326	0.459	0.382	0.404	0.327
Skewness	2.141	0.979	2.152	0.826	2.087	0.762	1.885	0.715	1.994	0.809	2.321	1.055
Kurtosis	9.077	3.863	8.962	3.548	9.143	3.502	8.090	3.494	7.864	3.434	9.702	4.365
Minimum	0.501	-0.691	0.467	-0.762	0.487	-0.719	0.478	-0.738	0.247	-1.399	0.518	-0.659
Maximum	3.440	1.235	3.751	1.322	3.506	1.255	3.230	1.172	3.473	1.245	3.235	1.174
JB-test *(p-value)*	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****
ADF-test *(p-value)*	*.077**	*.037***	*.161*	*.062**	*.008****	*.006****	*.020***	*.010****	*.100*	*.048***	*.003****	*.003****
sum LB-test *(p-values)*	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****

Illiquidity

	DAX 30		Dow Jones		CAC 40		AEX	
	norm	log	norm	log	norm	log	norm	log
Mean	1.000	-0.262	1.000	-0.211	1.000	-0.186	1.000	-0.280
Std. Deviation	0.841	0.704	0.977	0.570	0.703	0.595	0.907	0.720
Skewness	2.601	0.330	5.470	1.245	2.450	0.258	2.856	0.375
Kurtosis	15.051	2.683	48.46	5.124	11.99	3.038	14.42	2.932
Minimum	0.105	-2.250	0.239	-1.433	0.176	-1.739	0.119	-2.132
Maximum	9.917	2.294	12.92	2.559	6.284	1.838	8.203	2.104
JB-test *(p-value)*	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****
ADF-test *(p-value)*	*.027***	*.092**	*.004****	*.011***	*.019***	*.069**	*.171*	*.035***
sum LB-test *(p-values)*	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****

Trading Volume

	DAX 30		Dow Jones		CAC 40		AEX	
	norm	log	norm	log	norm	log	norm	log
Mean	1.000	-0.112	1.000	-0.041	1.000	-0.086	1.000	-0.119
Std. Deviation	0.539	0.457	0.299	0.285	0.438	0.415	0.545	0.475
Skewness	2.400	0.449	1.291	0.072	1.475	-0.047	2.132	0.359
Kurtosis	14.15	3.263	6.351	3.767	6.849	3.883	12.34	3.067
Minimum	0.179	-1.721	0.258	-1.353	0.075	-2.592	0.191	-1.654
Maximum	6.309	1.842	2.884	1.059	3.918	1.366	6.500	1.872
JB-test *(p-value)*	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****	*.001****
ADF-test *(p-value)*	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.002****	*.000****
sum LB-test *(p-values)*	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****

Table 12: Realized volatility: VAR-estimation results (Complete period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log realized volatility (RV) over the complete sample period ranging from January 2006 to September 2013 for all nine equity indices. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. Please note that we only present parameters up to lag 5, Total lags indicates the number of lags included in the specific VAR-model. We also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
	RV(t)	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**
RV(t-1)	0.406	0.047	0.450	0.058	0.406	0.073	0.417	0.069	0.436	0.062	0.442	0.057	0.448	0.061	0.487	0.095	0.449	0.006
(p-value)	*.000******	*.001******	*.000******	*.000******	*.000******	*.000******	*.000******	*.000******	*.000******	*.001******	*.000******	*.006******	*.000******	*.000******	*.000******	*.000******	*.000******	*.521*
SVI (t-1)	0.141	0.630	0.110	0.734	0.246	0.575	0.200	0.471	0.132	0.478	0.026	0.348	0.108	0.402	0.073	0.434	0.113	0.512
(p-value)	*.003******	*.000******	*.019****	*.000******	*.000******	*.000******	*.000******	*.000******	*.000******	*.000******	*.409*	*.000******	*.004******	*.000******	*.057***	*.000******	*.047****	*.000******
RV(t-2)	0.156	-0.018	0.205	-0.021	0.143	-0.011	0.153	0.031	0.150	0.019	0.175	0.017	0.160	-0.027	0.119	0.005	0.156	-0.028
(p-value)	*.000******	*.208*	*.000******	*.197*	*.000******	*.509*	*.000******	*.076***	*.000******	*.354*	*.000******	*.444*	*.000******	*.102*	*.000******	*.776*	*.000******	*.006******
SVI (t-2)	0.032	0.154	-0.028	0.015	-0.081	0.106	-0.029	0.145	0.020	0.176	0.036	0.169	-0.019	0.245	-0.030	0.210	-0.064	0.157
(p-value)	*.502*	*.000******	*.589*	*.672*	*.043****	*.001******	*.494*	*.000******	*.562*	*.000******	*.277*	*.000******	*.612*	*.000******	*.420*	*.000******	*.301*	*.000******
RV(t-3)	0.132	-0.007	0.065	-0.017	0.097	-0.008	0.070	-0.016	0.100	-0.004	0.102	-0.021	0.084	-0.016	0.106	-0.029	0.116	-0.003
(p-value)	*.000******	*.637*	*.014****	*.268*	*.000******	*.666*	*.005******	*.398*	*.000******	*.841*	*.000******	*.346*	*.001******	*.321*	*.000******	*.074***	*.000******	*.727*
SVI (t-3)	-0.066	0.058	-0.047	0.063	-0.034	0.071	0.002	0.098	-0.070	0.071	0.006	0.078	-0.007	0.078	-0.013	0.055	0.093	0.072
(p-value)	*.177*	*.094***	*.373*	*.121*	*.397*	*.019****	*.959*	*.000******	*.047****	*.009******	*.848*	*.004******	*.856*	*.005******	*.731*	*.037****	*.155*	*.010******
RV(t-4)	0.053	-0.005	0.137	0.030	0.137	0.019	0.062	-0.021	0.050	-0.025	0.042	0.010	0.020	0.004	0.100	-0.020	0.054	0.015
(p-value)	*.035****	*.739*	*.000******	*.027****	*.000******	*.319*	*.014****	*.253*	*.044****	*.224*	*.131*	*.617*	*.448*	*.844*	*.000******	*.241*	*.039****	*.141*
SVI (t-4)	0.012	0.001	0.072	0.084	-0.025	0.082	-0.052	0.100	0.042	0.115	0.039	0.057	-0.090	0.033	-0.007	0.136	-0.076	0.105
(p-value)	*.804*	*.972*	*.122*	*.017****	*.529*	*.015****	*.149*	*.000******	*.206*	*.000******	*.224*	*.039****	*.023****	*.278*	*.854*	*.000******	*.217*	*.000******
RV(t-5)	0.134	0.008			0.108	-0.026	0.146	-0.011	0.151	0.009	0.096	-0.013	0.063	0.017	0.097	-0.020	0.143	0.015
(p-value)	*.000******	*.586*			*.000******	*.278*	*.000******	*.496*	*.000******	*.640*	*.000******	*.508*	*.022****	*.289*	*.000******	*.228*	*.000******	*.168*
SVI (t-5)	-0.044	0.047			-0.016	0.015	-0.020	0.092	-0.055	0.036	-0.004	0.128	-0.030	0.050	0.011	0.069	-0.036	0.049
(p-value)	*.308*	*.183*			*.667*	*.690*	*.547*	*.001******	*.084***	*.179*	*.903*	*.000******	*.474*	*.094***	*.758*	*.007******	*.523*	*.037****
Total lags	5		4		5		5		5		5		8		5		5	

Table 13: Realized volatility: VAR-estimation results (Pre-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log realized volatility (RV) over the pre-crisis period ranging from January 2006 to December 2008 for all nine equity indices. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. Please note that we also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
	RV(t)	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**
RV(t-1)	0.393	0.047	0.467	0.072	0.400	0.092	0.425	0.054	0.419	0.054	0.631	0.064	0.523	0.116	0.422	0.066	0.480	0.023
(p-value)	*.000******	*.086**	*.000******	*.015****	*.000******	*.032****	*.000******	*.113*	*.000******	*.179*	*.000******	*.225*	*.000******	*.003******	*.000******	*.059****	*.000******	*.242*
SVI (t-1)	0.101	0.534	0.106	0.670	0.190	0.456	0.128	0.375	0.076	0.358	0.015	0.264	0.088	0.368	0.005	0.358	0.004	0.370
(p-value)	*.232*	*.000******	*.222*	*.000******	*.004******	*.000******	*.039****	*.000******	*.178*	*.000******	*.840*	*.000******	*.103*	*.000******	*.937*	*.000******	*.970*	*.000******
RV(t-2)	0.170	-0.041	0.257	-0.041	0.130	-0.006	0.252	0.062	0.234	0.025			0.040	-0.085	0.134	0.073	0.229	-0.037
(p-value)	*.001******	*.153*	*.000******	*.176*	*.008******	*.869*	*.000******	*.058**	*.000******	*.536*			*.497*	*.061**	*.003******	*.034****	*.000******	*.038****
SVI (t-2)	-0.036	0.137	-0.056	0.017	-0.020	0.070	-0.037	0.223	0.029	0.252			-0.108	0.341	-0.052	0.248	-0.078	0.279
(p-value)	*.631*	*.009******	*.518*	*.761*	*.754*	*.163*	*.517*	*.000******	*.553*	*.000******			*.025****	*.000******	*.404*	*.000******	*.419*	*.000******
RV(t-3)	0.117	-0.039			0.092	-0.018							0.155	-0.020	0.082	-0.089		
(p-value)	*.009******	*.116*			*.060**	*.637*							*.002******	*.584*	*.085**	*.010****		
SVI (t-3)	0.024	0.177			-0.056	0.080							0.082	0.190	0.114	0.027		
(p-value)	*.745*	*.000******			*.353*	*.114*							*.094**	*.000******	*.036****	*.567*		
RV(t-4)					0.217	-0.010									0.162	-0.013		
(p-value)					*.000******	*.795*									*.000******	*.683*		
SVI (t-4)					-0.056	0.092									-0.069	0.150		
(p-value)					*.341*	*.060****									*.197*	*.002******		

Table 14: Realized volatility: VAR-estimation results (crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log realized volatility (RV) over the crisis period ranging from January 2008 to December 2010 for all nine equity indices. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. Please note that we also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
	RV(t)	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**
RV(t-1)	0.440	0.087	0.512	0.044	0.423	0.109	0.412	0.054	0.437	0.043	0.501	0.112	0.474	0.050	0.474	0.153	0.440	0.013
(p-value)	*.000****	*.000****	*.000****	*.073**	*.000****	*.000****	*.000****	*.023***	*.000****	*.128*	*.000****	*.000****	*.000****	*.012***	*.000****	*.000****	*.000****	*.428*
SVI (t-1)	0.169	0.606	0.140	0.755	0.317	0.615	0.233	0.602	0.200	0.588	0.014	0.297	0.137	0.483	0.126	0.472	0.183	0.615
(p-value)	*.025****	*.000****	*.042***	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.754*	*.000****	*.044***	*.000****	*.034***	*.000****	*.036***	*.000****
RV(t-2)	0.150	-0.009	0.347	0.005	0.162	-0.037	0.183	0.014	0.180	-0.010	0.349	0.028	0.207	-0.019	0.109	-0.049	0.224	-0.037
(p-value)	*.001****	*.688*	*.000****	*.849*	*.000****	*.182*	*.000****	*.573*	*.000****	*.759*	*.000****	*.402*	*.000****	*.385*	*.008****	*.129*	*.000****	*.027***
SVI (t-2)	0.049	0.219	-0.053	0.139	-0.178	0.123	0.009	0.134	0.008	0.166	0.068	0.096	0.065	0.239	0.016	0.150	-0.163	0.129
(p-value)	*.515*	*.003****	*.388*	*.003****	*.006****	*.032***	*.897*	*.009****	*.889*	*.000****	*.124*	*.016***	*.358*	*.000****	*.765*	*.000****	*.089**	*.012***
RV(t-3)	0.214	-0.021			0.201	0.023	0.159	-0.018	0.205	0.051			0.157	-0.008	0.107	0.021	0.195	0.016
(p-value)	*.000****	*.367*			*.000****	*.343*	*.000****	*.473*	*.000****	*.080**			*.000****	*.662*	*.016***	*.419*	*.000****	*.268*
SVI (t-3)	-0.087	0.044			0.029	0.092	-0.079	0.164	-0.103	0.120			-0.064	0.202	-0.106	0.077	-0.056	0.134
(p-value)	*.184*	*.440*			*.629*	*.026***	*.164*	*.000****	*.027***	*.006****			*.334*	*.000****	*.063**	*.038***	*.507*	*.000****
RV(t-4)															0.118	-0.053		
(p-value)															*.002****	*.038***		
SVI (t-4)															0.067	0.163		
(p-value)															*.178*	*.000****		

Table 15: Realized volatility: VAR-estimation results (Post-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log realized volatility (RV) over the post-crisis period ranging from January 2011 to September 2013 for all nine equity indices. The horizontal variables correspond to the dependent variables in each VAR-model, the vertical variables correspond to the respective lags. Please note that we also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225		Bovespa		IBEX 35		FTSE MIB	
	RV(t)	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**	**RV(t)**	**SVI(t)**
RV(t-1)	0.416	0.007	0.503	0.076	0.521	0.014	0.509	0.073	0.511	0.065	0.458	0.036	0.406	0.017	0.524	0.054	0.487	-0.015
(p-value)	*.000******	*.723*	*.000******	*.000******	*.000******	*.486*	*.000******	*.001******	*.000******	*.011****	*.000******	*.173*	*.000******	*.494*	*.000******	*.004******	*.000******	*.316*
SVI (t-1)	0.107	0.741	-0.039	0.748	0.224	0.662	0.186	0.494	0.087	0.541	0.102	0.357	0.154	0.573	0.098	0.515	0.090	0.584
(p-value)	*.201*	*.000******	*.676*	*.000******	*.004******	*.000******	*.008******	*.000******	*.164*	*.000******	*.095**	*.000******	*.019****	*.000******	*.253*	*.000******	*.385*	*.000******
RV(t-2)	0.216	0.001	0.238	-0.017	0.274	0.016	0.124	0.019	0.147	0.031	0.175	-0.007	0.221	0.021	0.094	0.008	0.121	-0.003
(p-value)	*.000******	*.973*	*.000******	*.416*	*.000******	*.411*	*.005******	*.454*	*.000******	*.275*	*.000******	*.759*	*.000******	*.392*	*.034****	*.700*	*.005******	*.862*
SVI (t-2)	0.065	0.058	0.201	0.080	-0.068	0.157	-0.104	0.122	0.018	0.169	0.034	0.242	0.001	0.167	-0.086	0.164	0.003	0.097
(p-value)	*.481*	*.336*	*.025****	*.124*	*.358*	*.002******	*.256*	*.167*	*.777*	*.001******	*.576*	*.000******	*.994*	*.000******	*.352*	*.002******	*.978*	*.071**
RV(t-3)	0.195	0.037					0.145	-0.018	0.162	-0.034	0.129	-0.017	0.095	-0.022	0.131	-0.030	0.133	0.003
(p-value)	*.000******	*.047****					*.000******	*.503*	*.000******	*.151*	*.002******	*.463*	*.029****	*.384*	*.002******	*.139*	*.001******	*.823*
SVI (t-3)	-0.036	0.048					0.119	0.177	0.015	0.135	-0.020	0.178	-0.051	0.017	-0.089	0.121	0.068	0.111
(p-value)	*.666*	*.317*					*.025****	*.001******	*.798*	*.003******	*.731*	*.000******	*.512*	*.738*	*.273*	*.004******	*.594*	*.029****
RV(t-4)													0.094	-0.051	0.079	-0.014	0.121	-0.005
(p-value)													*.019****	*.051**	*.049****	*.532*	*.001******	*.765*
SVI (t-4)													-0.105	0.180	0.138	0.101	-0.178	0.099
(p-value)													*.118*	*.000******	*.111*	*.017****	*.107*	*.038****

Table 16: Implied volatility: VAR-estimation results (Complete and pre-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log implied volatility (IV) over the complete and pre-crisis period ranging from January 2006 to September 2013 and January 2006 to December 2008, respectively for six indices. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. Please note that we also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

Complete Period

	DAX 30		**Dow Jones**		**FTSE 100**		**CAC 40**		**AEX**		**Nikkei 225**	
	IV(t)	**SVI(t)**	**IV(t)**	**SVI(t)**	**IV (t)**	**SVI(t)**	**IV (t)**	**SVI(t)**	**IV (t)**	**SVI(t)**	**IV (t)**	**SVI(t)**
IV (t-1)	1.024	0.470	0.826	0.378	0.908	0.343	0.886	0.388	0.837	0.304	0.944	0.323
(p-value)	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****
SVI (t-1)	0.019	0.609	0.008	0.716	0.021	0.574	0.028	0.480	0.028	0.478	0.001	0.365
(p-value)	*.040***	*.000****	*.578*	*.000****	*.022***	*.000****	*.003****	*.000****	*.001****	*.000****	*.820*	*.000****
IV (t-2)	-0.114	-0.253	0.069	-0.150	0.046	-0.079	0.003	-0.233	0.055	-0.037	-0.002	-0.236
(p-value)	*.002****	*.024***	*.079**	*.049***	*.211*	*.387*	*.927*	*.005****	*.545*	*.685*	*.966*	*.063**
SVI (t-2)	0.012	0.159	0.001	0.006	0.000	0.120	-0.004	0.162	-0.002	0.178	0.010	0.204
(p-value)	*.237*	*.000****	*.974*	*.866*	*.986*	*.000****	*.699*	*.000****	*.812*	*.000****	*.167*	*.000****
IV (t-3)	0.068	-0.165	0.074	-0.114	0.020	-0.185	0.033	-0.059	0.027	-0.140	0.033	0.015
(p-value)	*.011***	*.018***	*.007****	*.077**	*.443*	*.002****	*.369*	*.448*	*.451*	*.191*	*.232*	*.867*
SVI (t-3)	-0.018	0.100	0.011	0.125	-0.003	0.127	-0.001	0.111	0.004	0.077	0.002	0.132
(p-value)	*.065***	*.001****	*.389*	*.000****	*.782*	*.000****	*.923*	*.000****	*.632*	*.004****	*.787*	*.000****
IV (t-4)							0.041	-0.009	0.055	-0.033		
(p-value)							*.150*	*.884*	*.015***	*.647*		
SVI (t-4)							-0.003	0.135	-0.013	0.122		
(p-value)							*.797*	*.000****	*.108*	*.000****		

Pre-crisis Period

	DAX 30		**Dow Jones**		**FTSE 100**		**CAC 40**		**AEX**		**Nikkei 225**	
	IV **(t)**	**SVI(t)**	IV **(t)**	**SVI(t)**	IV **(t)**	**SVI(t)**	IV **(t)**	**SVI(t)**	IV **(t)**	**SVI(t)**	IV **(t)**	**SVI(t)**
IV (t-1)	0.916	0.179	0.849	0.415	0.974	0.162	0.927	0.625	0.917	0.117	0.954	0.252
(p-value)	*.000****	*.342*	*.000****	*.003****	*.000****	*.000****	*.000****	*.000****	*.000****	*.592*	*.000****	*.000****
SVI (t-1)	0.024	0.575	-0.012	0.654	0.003	0.563	0.014	0.390	0.020	0.361	-0.001	0.217
(p-value)	*.076**	*.000****	*.572*	*.000****	*.760*	*.000****	*.206*	*.000****	*.068**	*.000****	*.936*	*.000****
IV (t-2)	0.038	-0.166	-0.024	-0.160			0.034	-0.498	0.037	0.017		
(p-value)	*.393*	*.370*	*.704*	*.314*			*.518*	*.005****	*.468*	*.936*		
SVI (t-2)	-0.004	0.222	0.041	-0.123			-0.015	0.233	-0.001	0.251		
(p-value)	*.752*	*.000****	*.079**	*.023***			*.200*	*.000****	*.952*	*.000****		
IV (t-3)			0.133	-0.171								
(p-value)			*.005****	*.086**								
SVI (t-3)			0.000	0.170								
(p-value)			*.989*	*.000****								

Table 17: Implied volatility: VAR-estimation results (Crisis and post-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log implied volatility (IV) over the crisis and post-crisis period ranging from January 2008 to December 2010 and January 2011 to September 2013, respectively for six equity indices. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. Please note that we also included constants in the VAR-estimation which are not presented here. Significance of p-lags is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

Crisis Period

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225	
	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)
IV (t-1)	1.054	0.748	0.845	0.422	0.852	0.316	0.814	0.264	0.923	0.230	0.986	0.188
(p-value)	*.000****	*.000****	*.000****	*.000****	*.000****	*.003****	*.000****	*.004****	*.000****	*.082**	*.000****	*.000****
SVI (t-1)	-0.011	0.571	0.020	0.702	0.024	0.625	0.041	0.622	0.027	0.602	-0.002	0.336
(p-value)	*.427*	*.000****	*.452*	*.000****	*.072**	*.000****	*.028***	*.000****	*.019***	*.000****	*.823*	*.000****
IV (t-2)	-0.088	-0.643	0.127	-0.267	0.109	-0.186	0.116	-0.135	0.047	-0.071		
(p-value)	*.058***	*.000****	*.017***	*.010****	*.012***	*.077**	*.014***	*.155*	*.220*	*.599*		
SVI (t-2)	0.031	0.273	-0.008	0.137	0.001	0.194	-0.001	0.228	-0.012	0.230		
(p-value)	*.071**	*.000****	*.729*	*.003****	*.926*	*.001****	*.957*	*.000****	*.300*	*.000****		

Post-crisis Period

	DAX 30		Dow Jones		FTSE 100		CAC 40		AEX		Nikkei 225	
	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)	IV (t)	SVI(t)
IV (t-1)	0.965	0.102	0.801	0.274	0.966	0.264	0.958	0.310	0.772	0.320	1.014	0.207
(p-value)	*.000****	*.001****	*.000****	*.001****	*.000****	*.001****	*.000****	*.007****	*.000****	*.002****	*.000****	*.133*
SVI (t-1)	0.028	0.803	0.011	0.761	0.050	0.642	0.026	0.528	0.031	0.571	0.027	0.397
(p-value)	*.009****	*.000****	*.693*	*.000****	*.027***	*.000****	*.092**	*.000****	*.086**	*.000****	*.040***	*.000****
IV (t-2)			0.161	-0.180	-0.005	-0.214	-0.003	-0.168	0.171	-0.203	-0.068	-0.120
(p-value)			*.004****	*.024***	*.909*	*.008****	*.942*	*.119*	*.312*	*.055**	*.508*	*.382*
SVI (t-2)			0.006	0.059	-0.023	0.167	0.004	0.211	0.009	0.230	-0.008	0.303
(p-value)			*.771*	*.240*	*.228*	*.000****	*.787*	*.017***	*.633*	*.000****	*.542*	*.000****

Table 18: Illiquidity: VAR-estimation results (Complete, Pre-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log illiquidity (ILLQ) over the complete sample period ranging from January 2006 to September 2013, as well as for the pre-crisis period ranging from January 2006 to December 2008 for all four equity indices where data is available. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. Please note that we only present parameters up to lag 5, Total lags indicates the number of lags included in the specific VAR-model. We also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	Complete Period								Pre-crisis Period							
	DAX 30		Dow Jones		CAC 40		AEX		DAX 30		Dow Jones		CAC 40		AEX	
	ILLQ(t)	**SVI(t)**	**ILLQ(t)**	**SVI(t)**	**ILLQ(t)**	**SVI(t)**	**ILLQ(t)**	**SVI(t)**	**ILLQ(t)**	**SVI(t)**	**ILLQ(t)**	**SVI(t)**	**ILLQ(t)**	**SVI(t)**	**ILLQ(t)**	**SVI(t)**
ILLQ (t-1)	0.130	0.030	0.045	0.049	0.112	0.049	0.079	0.066	0.152	0.004	0.126	0.050	0.183	0.039	0.173	0.041
(p-value)	*.000******	*.000******	*.086**	*.000******	*.000******	*.000******	*.001******	*.000******	*.001******	*.828*	*.003******	*.022****	*.000******	*.126*	*.000******	*.204*
SVI (t-1)	-0.108	0.646	-0.119	0.735	-0.087	0.491	-0.035	0.488	-0.253	0.547	-0.332	0.692	-0.128	0.378	-0.093	0.347
(p-value)	*.070**	*.000******	*.064**	*.000******	*.079**	*.000******	*.417*	*.000******	*.033****	*.000******	*.000******	*.000******	*.128*	*.000******	*.154*	*.000******
ILLQ (t-2)	0.227	-0.006	0.173	-0.017	0.215	-0.008	0.189	0.002	0.164	-0.009	0.196	-0.037	0.300	-0.041	0.303	0.007
(p-value)	*.000******	*.442*	*.000******	*.059**	*.000******	*.429*	*.000******	*.835*	*.000******	*.589*	*.000******	*.097**	*.000******	*.067**	*.000******	*.820*
SVI (t-2)	0.172	0.160	0.188	0.020	-0.011	0.162	-0.012	0.190	0.011	0.122	0.070	-0.114	-0.058	0.195	0.058	0.221
(p-value)	*.024****	*.000******	*.019****	*.578*	*.846*	*.000******	*.807*	*.000******	*0.937*	*.019****	*.489*	*.046****	*.472*	*.000******	*.376*	*.000******
ILLQ (t-3)	0.181	-0.018	0.085	-0.016	0.174	0.010	0.152	-0.011	.166	-0.030	0.137	0.008	0.282	-0.009	0.236	-0.012
(p-value)	*.000******	*.022****	*.001******	*.099**	*.000******	*.387*	*.00***0*	*.348*	*0.000******	*.066**	*.002******	*.718*	*.000******	*.721*	*.000******	*.700*
SVI (t-3)	-0.015	0.054	0.007	0.036	0.071	0.098	0.047	0.071	0.013	0.152	0.003	0.130	0.108	0.124	0.005	0.118
(p-value)	*.840*	*.117*	*.930*	*.356*	*.191*	*.000******	*.363*	*.010******	*.921*	*.000******	*.974*	*.022****	*.181*	*.005******	*.941*	*.006******
ILLQ (t-4)	0.172	-0.004	0.102	0.012	0.142	-0.014	0.095	0.013			0.233	-0.042				
(p-value)	*.000******	*.653*	*.000******	*.241*	*.000******	*.198*	*.000******	*.293*			*.000******	*.042****				
SVI (t-4)	0.074	-0.004	-0.002	0.036	0.064	0.091	-0.047	0.094			0.136	0.081				
(p-value)	*.313*	*.910*	*.978*	*.303*	*.231*	*.001******	*.374*	*.003******			*.111*	*.148*				
ILLQ (t-5)	0.182	0.004	0.140	-0.004	0.181	0.003	0.150	0.001								
(p-value)	*.000******	*.654*	*.000******	*.694*	*.000******	*.752*	*.000******	*.956*								
SVI (t-5)	-0.043	0.054	0.052	0.049	0.118	0.072	0.073	0.015								
(p-value)	*.458*	*.123*	*.488*	*.231*	*.017****	*.010****	*.149*	*.611*								
Total lags	5		8		5		8									

Table 19: Illiquidity: VAR-estimation results (Crisis period, Post-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log illiquidity (ILLQ) over the crisis period ranging from January 2008 to September 2010, as well as for the pre-crisis period ranging from January 2011 to September 2013 for all four equity indices where data is available. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. We also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

Crisis Period

	DAX 30		Dow Jones		CAC 40		AEX	
	ILLQ(t)	SVI(t)	ILLQ(t)	SVI(t)	ILLQ(t)	SVI(t)	ILLQ(t)	SVI(t)
ILLQ (t-1)	0.135	0.027	0.090	0.036	0.091	0.056	0.118	0.060
(p-value)	*.000****	*.035***	*.017***	*.012***	*.010****	*.000****	*.001****	*.001****
SVI (t-1)	-0.164	0.644	-0.003	0.738	-0.097	0.590	0.012	0.577
(p-value)	*.050**	*.000****	*.974*	*.000****	*.228*	*.000****	*.877*	*.000****
ILLQ (t-2)	0.271	-0.009	0.204	-0.018	0.187	0.002	0.198	-0.001
(p-value)	*.000****	*.455*	*.000****	*.146*	*.000****	*.914*	*.000****	*.933*
SVI (t-2)	0.235	0.223	0.271	0.069	-0.082	0.139	-0.066	0.166
(p-value)	*.034***	*.002****	*.054**	*.235*	*.417*	*.010****	*.456*	*.000****
ILLQ (t-3)	0.235	-0.019	0.152	-0.012	0.196	-0.009	0.201	-0.024
(p-value)	*.000****	*.153*	*.000****	*.367*	*.000****	*.603*	*.000****	*.173*
SVI (t-3)	0.030	0.004	-0.109	0.012	0.060	0.105	0.054	0.083
(p-value)	*.784*	*.955*	*.390*	*.831*	*.558*	*.019***	*.578*	*.072**
ILLQ (t-4)	0.183	-0.005	0.143	0.028	0.153	-0.040	0.134	-0.001
(p-value)	*.000****	*.719*	*.000****	*.039***	*.000****	*.013***	*.001****	*.953*
SVI (t-4)	0.011	0.055	-0.042	0.008	0.211	0.099	0.017	0.091
(p-value)	*.903*	*.292*	*.743*	*.884*	*.031***	*.057**	*.850*	*.073**
ILLQ (t-5)			0.225	0.004	0.185	0.009	0.184	-0.002
(p-value)			*.000****	*.796*	*.000****	*.570*	*.000****	*.894*
SVI (t-5)			0.073	0.087	0.041	0.002	0.109	0.002
(p-value)			*.473*	*.079**	*.635*	*.973*	*.162*	*.963*

Post-crisis Period

	DAX 30		Dow Jones		CAC 40		AEX	
	ILLQ(t)	SVI(t)	ILLQ(t)	SVI(t)	ILLQ(t)	SVI(t)	ILLQ(t)	SVI(t)
ILLQ (t-1)	0.157	0.048	0.099	0.052	0.060	0.033	0.073	0.063
(p-value)	*.000****	*.000****	*.012***	*.013***	*.120*	*.102*	*.041***	*.001****
SVI (t-1)	-0.052	0.729	-0.257	0.787	-0.033	0.496	-0.031	0.532
(p-value)	*.666*	*.000****	*.037***	*.000****	*.711*	*.000****	*.726*	*.000****
ILLQ (t-2)	0.320	-0.008	0.241	-0.020	0.200	-0.012	0.231	-0.018
(p-value)	*.000****	*.520*	*.000****	*.167*	*.000****	*.505*	*.000****	*.266*
SVI (t-2)	0.128	0.078	0.447	0.087	0.087	0.110	-0.023	0.173
(p-value)	*.384*	*.188*	*.000****	*.067**	*.415*	*.178*	*.832*	*.000****
ILLQ (t-3)	0.212	-0.010			0.088	0.026	0.175	-0.011
(p-value)	*.000****	*.418*			*.029***	*.165*	*.000****	*.479*
SVI (t-3)	0.110	0.063			0.025	0.108	0.157	0.076
(p-value)	*.366*	*.167*			*.796*	*.022***	*.167*	*.091**
ILLQ (t-4)					0.120	-0.001	0.162	0.004
(p-value)					*.003****	*.963*	*.000****	*.834*
SVI (t-4)					-0.050	0.032	-0.210	0.050
(p-value)					*.628*	*.415*	*.084**	*.312*
ILLQ (t-5)					0.183	0.011	0.230	-0.002
(p-value)					*.000****	*.423*	*.000****	*.905*
SVI (t-5)					0.202	0.110	0.186	0.052
(p-value)					*.045***	*.013***	*.046***	*.298*

Table 20: Trading Volume: VAR-estimation results (Complete, Pre-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log trading volume (TV) over the complete sample period ranging from January 2006 to September 2013, as well as for the pre-crisis period ranging from January 2006 to December 2008 for all four equity indices where data is available. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. We also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	Complete Period								**Pre-crisis Period**							
	DAX 30		**Dow Jones**		**CAC 40**		**AEX**		**DAX 30**		**Dow Jones**		**CAC 40**		**AEX**	
	TVL(t)	**SVI(t)**	**TVL(t)**	**SVI(t)**	**TVL(t)**	**SVI(t)**	**TVL(t)**	**SVI(t)**	**TVL(t)**	**SVI(t)**	**TVL(t)**	**SVI(t)**	**TVL(t)**	**SVI(t)**	**TVL(t)**	**SVI(t)**
TV (t-1)	0.402	0.045	0.450	0.045	0.411	0.068	0.454	0.077	0.419	0.154	0.453	0.094	0.407	0.123	0.413	0.116
(p-value)	*.000****	*.004****	*.000****	*.026***	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.000****	*.010****	*.000****	*.015***	*.000****	*.011***
SVI (t-1)	0.082	0.629	0.105	0.748	0.109	0.472	0.022	0.477	0.057	0.461	0.108	0.664	0.118	0.337	-0.021	0.281
(p-value)	*.054**	*.000****	*.000****	*.000****	*.002****	*.000****	*.435*	*.000****	*.470*	*.000****	*.068**	*.000****	*.030***	*.000****	*.667*	*.000****
TV (t -2)	0.107	-0.041	0.127	-0.032	0.145	0.012	0.125	-0.010	0.063	-0.046	0.145	-0.033	0.153	0.081	0.107	-0.011
(p-value)	*.000****	*.004****	*.000****	*.159*	*.000****	*.456*	*.000****	*.604*	*.325*	*.114*	*.008****	*.404*	*.006****	*.041***	*.026***	*.782*
SVI (t-2)	0.080	0.176	-0.022	0.010	-0.048	0.156	0.040	0.198	0.021	0.165	-0.105	-0.136	0.003	0.200	0.022	0.185
(p-value)	*.067**	*.000****	*.506*	*.788*	*.239*	*.000****	*.215*	*.000****	*.751*	*.001****	*.107*	*.023***	*.945*	*.000****	*.691*	*.000****
TV (t -3)	0.103	-0.018	0.082	-0.033	0.042	-0.057	0.077	-0.041	0.058	-0.085	0.245	0.031			0.068	-0.039
(p-value)	*.000****	*.236*	*.003****	*.175*	*.075**	*.006****	*.001****	*.040***	*.202*	*.005****	*.000****	*.440*			*.088**	*.380*
SVI (t-3)	-0.096	0.061	-0.051	0.075	-0.010	0.120	-0.075	0.091	-0.039	0.154	-0.006	0.173			-0.094	0.075
(p-value)	*.023***	*.088**	*.160*	*.071**	*.776*	*.000****	*.027***	*.001****	*.606*	*.002****	*.891*	*.000****			*.119*	*.095**
TV (t -4)	0.096	0.015	0.051	0.051	0.075	-0.001	0.043	-0.029	0.209	0.097					0.067	-0.043
(p-value)	*.000****	*.305*	*.069**	*.017***	*.002****	*.978*	*.083**	*.163*	*.000****	*.002****					*.139*	*.360*
SVI (t-4)	-0.081	-0.006	-0.010	0.023	-0.066	0.104	0.045	0.126	0.049	-0.034					0.064	0.146
(p-value)	*.112*	*.883*	*.789*	*.543*	*.051**	*.000****	*.151*	*.000****	*.489*	*.485*					*.171*	*.003****
TV (t -5)	0.213	-0.001	0.099	-0.006	0.225	-0.039	0.244	0.012							0.235	0.045
(p-value)	*.000****	*.946*	*.000****	*.742*	*.000****	*.026***	*.000****	*.593*							*.000****	*.289*
SVI (t-5)	-0.044	0.055	0.023	0.085	-0.041	0.097	-0.067	0.036							-0.037	0.025
(p-value)	*.335*	*.143*	*.422*	*.004****	*.213*	*.001****	*.023***	*.191*							*.433*	*.564*

Table 21: Trading Volume: VAR-estimation results (Crisis period, Post-crisis period)

The table presents the results of the VAR-estimation, i.e. estimated parameters and corresponding p-values, for log Google search volume index (SVI) and log trading volume (TV) over the crisis period ranging from January 2008 to September 2010, as well as for the pre-crisis period ranging from January 2011 to September 2013 for all four equity indices where data is available. The horizontal axis corresponds to the dependent variables in each VAR-model, the vertical axis corresponds to the respective lags. We also included constants in the VAR-estimation which are not presented here. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% significance level and *** for 1% significance level for the corresponding parameter.

	Crisis Period								Post-crisis Period							
	DAX 30		Dow Jones		CAC 40		AEX		DAX 30		Dow Jones		CAC 40		AEX	
	TVL(t)	SVI(t)	TVL(t)	SVI(t)	TVL(t)	SVI(t)	TVL(t)	SVI(t)	TVL(t)	SVI(t)	TVL(t)	SVI(t)	TVL(t)	SVI(t)	TVL(t)	SVI(t)
TV (t-1)	0.417	0.036	0.498	0.013	0.406	0.047	0.448	0.036	0.337	0.000	0.389	0.023	0.392	0.085	0.426	0.067
(p-value)	*.000******	*.241*	*.000****	*.704*	*.000****	*.054**	*.000****	*.221*	*.000****	*.981*	*.000****	*.465*	*.000****	*.000****	*.000****	*.018***
SVI (t-1)	0.141	0.630	0.078	0.786	0.147	0.583	0.056	0.580	0.110	0.753	0.148	0.802	0.135	0.478	0.016	0.542
(p-value)	*.030***	*.000****	*.056**	*.000****	*.007****	*.000****	*.223*	*.000****	*.185*	*.000****	*.014***	*.000****	*.112*	*.000****	*.775*	*.000*
TV (t -2)	0.178	-0.016	0.211	-0.025	0.204	0.002	0.174	-0.006	0.061	-0.036	0.124	-0.071	0.071	0.014	0.073	-0.016
(p-value)	*.000****	*.566*	*.000****	*.490*	*.000****	*.933*	*.000****	*.862*	*.130*	*.067**	*.003****	*.018***	*.060**	*.524*	*.059**	*.551*
SVI (t-2)	0.107	0.214	-0.041	0.156	0.006	0.124	0.050	0.161	0.064	0.086	-0.026	0.102	-0.080	0.101	0.001	0.197
(p-value)	*.126*	*.003****	*.318*	*.001****	*.921*	*.027***	*.351*	*.001****	*.467*	*.167*	*.664*	*.053**	*.409*	*.226*	*.988*	*.000****
TV (t -3)	0.057	-0.003			0.009	-0.036	0.018	-0.025	0.193	0.014			0.057	-0.089	0.157	-0.062
(p-value)	*.127*	*.894*			*.794*	*.236*	*.658*	*.467*	*.000****	*.525*			*.137*	*.025***	*.000****	*.016***
SVI (t-3)	-0.174	0.000			-0.072	0.102	-0.092	0.086	-0.119	0.061			0.125	0.159	-0.002	0.162
(p-value)	*.007****	*.999*			*.213*	*.030***	*.066**	*.073**	*.115*	*.226*			*.062**	*.001****	*.974*	*.001****
TV (t -4)	0.046	-0.007			0.051	0.011	0.018	0.010					0.045	0.049		
(p-value)	*.289*	*.767*			*.221*	*.684*	*.664*	*.774*					*.284*	*.304*		
SVI (t-4)	-0.051	0.044			-0.013	0.105	0.061	0.099					-0.175	0.021		
(p-value)	*.545*	*.524*			*.840*	*.040***	*.253*	*.056**					*.009****	*.605*		
TV (t -5)	0.228	0.018			0.248	-0.021	0.258	-0.011					0.225	-0.050		
(p-value)	*.000****	*.444*			*.000****	*.405*	*.000****	*.781*					*.000****	*.058**		
SVI (t-5)	-0.048	0.021			-0.084	0.035	-0.105	0.017					-0.044	0.142		
(p-value)	*.529*	*.749*			*.180*	*.516*	*.042*	*.711*					*.437*	*.004****		

Table 22: Summary Statistics Stocks: Realized volatility (RV)

The table presents the summary statistics of normalized (“norm”) and logarithmic (“log”) time series corresponding to realized volatility for the 22 US-stocks. Further, p-values of test statistics are given for the Jarque-Bera (JB)-test on normality, the augmented Dickey-Fuller (ADF)-test on the presence of a unit-root and the Ljung-Box (LB)-test on independent distribution of the time series. For the ADF-test, the time series are tested on 100 lags, whereas the stated p-value of the LB-test gives the sum of all p-values over 100 lags tested.

	Apple		Google		General Electric		Wal-Mart		Microsoft		IBM		Hewlett-Packard		Verizon		Citigroup		Walt Disney		Wells Fargo	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	-0.15	1.00	-0.20	1.00	-0.25	1.00	-0.18	1.00	-0.17	1.00	-0.18	1.00	-0.18	1.00	-0.18	1.00	-0.30	1.00	-0.20	1.00	-0.32
Std. Dev.	0.62	0.55	0.71	0.62	0.80	0.70	0.70	0.59	0.69	0.56	0.67	0.60	0.68	0.59	0.70	0.59	1.02	0.72	0.72	0.61	0.97	0.77
Skewness	2.52	0.02	2.16	0.16	2.20	0.14	2.72	0.22	2.58	0.37	1.89	0.06	2.36	0.11	2.87	0.05	3.19	0.63	2.28	0.22	2.41	0.40
Kurtosis	14.10	3.59	9.29	3.09	9.07	2.84	14.61	3.27	12.73	3.25	7.16	3.31	11.40	3.51	15.55	3.87	14.99	3.57	9.75	3.26	8.95	3.05
Minimum	0.15	-1.92	0.13	-2.01	0.14	-1.96	0.16	-1.86	0.24	-1.43	0.12	-2.13	0.10	-2.34	0.09	-2.36	0.12	-2.15	0.12	-2.11	0.11	-2.18
Maximum	5.48	1.70	4.92	1.59	5.40	1.69	5.95	1.78	5.60	1.72	4.14	1.42	5.46	1.70	6.05	1.80	7.13	1.96	5.06	1.62	5.35	1.68
JB-test *(p-value)*	0.00^{***}	0.10^{*}	0.00^{***}	0.48	0.00^{***}	0.49	0.00^{***}	0.16	0.00^{***}	0.03^{**}	0.00^{***}	0.49	0.00^{***}	0.12	0.00^{***}	0.02^{**}	0.00^{***}	0.00^{***}	0.00^{***}	0.16	0.00^{***}	0.02^{**}
ADF-test *(p-value)*	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.03^{**}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
sum LB-test *(p-values)*	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}

	AT&T		Exxon Mobil		Intel		Nike		McDonald’s		Starbucks		Johnson & Johnson		Coca-Cola		Boeing		Pfizer		Comcast	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	-0.20	1.00	-0.21	1.00	-0.16	1.00	-0.20	1.00	-0.17	1.00	-0.18	1.00	-0.20	1.00	-0.17	1.00	-0.16	1.00	-0.16	1.00	-0.18
Std. Dev.	0.76	0.61	0.84	0.62	0.63	0.58	0.69	0.62	0.64	0.58	0.64	0.60	0.78	0.61	0.72	0.57	0.61	0.57	0.66	0.56	0.72	0.58
Skewness	2.86	0.28	3.96	0.34	2.13	-0.22	1.88	0.13	2.32	-0.11	1.73	0.01	3.45	0.20	3.32	0.21	1.79	-0.01	2.25	0.28	2.87	0.32
Kurtosis	15.01	3.35	26.24	3.76	9.31	3.96	7.11	3.11	11.41	3.84	6.85	2.85	20.40	3.76	20.12	3.72	7.63	3.06	9.53	3.17	15.78	3.37
Minimum	0.13	-2.01	0.15	-1.89	0.12	-2.16	0.13	-2.08	0.10	-2.26	0.13	-2.02	0.14	-1.93	0.14	-1.95	0.14	-1.98	0.22	-1.53	0.14	-1.94
Maximum	6.27	1.84	8.10	2.09	4.02	1.39	4.59	1.52	5.17	1.64	3.87	1.35	7.06	1.95	6.59	1.89	4.14	1.42	4.07	1.40	6.15	1.82
JB-test *(p-value)*	0.00^{***}	0.06	0.00^{***}	0.01	0.00^{***}	0.01^{***}	0.00^{***}	0.50	0.00^{***}	0.02^{**}	0.00^{***}	0.50	0.00^{***}	0.02^{**}	0.00^{***}	0.02^{**}	0.00^{***}	0.50	0.00^{***}	0.10	0.00^{***}	0.03^{**}
ADF-test *(p-value)*	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
sum LB-test *(p-values)*	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}

Table 23: Summary Statistics Stocks: Google search volume index (SVI)

The table presents the summary statistics of normalized ("norm") and logarithmic ("log") time series corresponding to the Google search volume index (SVI) for the 22 US-stocks. Further, p-values of test statistics are given for the Jarque-Bera (JB)-test on normality, the augmented Dickey-Fuller (ADF)-test on the presence of a unit-root and the Ljung-Box (LB)-test on independent distribution of time series. For the ADF-test, time series are tested on 100 lags, whereas the stated p-value of the LB-test gives the sum of all p-values over 100 lags.

	Apple		Google		General Electric		Wal-Mart		Microsoft		IBM		Hewlett-Packard		Verizon		Citigroup		Walt Disney		Wells Fargo	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	2.41	1.00	3.41	1.00	2.63	1.00	3.14	1.00	3.28	1.00	3.25	1.00	1.99	1.00	2.41	1.00	2.11	1.00	3.22	1.00	2.92
Std. Dev.	0.72	0.57	0.31	0.26	0.83	0.57	0.53	0.39	0.39	0.32	0.37	0.31	0.76	0.31	0.56	0.35	1.12	0.93	0.36	0.32	0.64	0.47
Skewness	3.58	0.51	2.83	0.83	2.89	1.23	2.84	1.21	2.70	0.65	2.60	0.66	12.77	2.77	8.07	1.13	3.08	0.28	2.46	0.42	2.92	0.94
Kurtosis	27.10	2.99	17.27	5.54	12.94	3.98	12.45	5.09	15.24	4.57	15.60	4.35	193.5	22.18	100.6	8.20	15.90	2.23	16.34	4.00	12.70	5.12
Minimum	0.30	1.39	0.51	2.77	0.35	1.79	0.40	2.30	0.43	2.48	0.44	2.48	0.51	1.39	0.42	1.61	0.16	0.69	0.46	2.48	0.24	1.61
Maximum	7.57	4.61	3.17	4.61	5.89	4.61	3.95	4.61	3.55	4.61	3.69	4.61	12.69	4.61	8.31	4.61	7.81	4.61	3.81	4.61	4.73	4.61
JB-test *(p-value)*	*0.00****	*0.01****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.01***	*0.00****	*0.00****	*0.00****	*0.00****
ADF-test *(p-value)*	*0.00****	*0.00****	*0.00****	*0.00****	*0.02***	*0.02***	*0.05***	*0.12*	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.02***	*0.00****	*0.00****	*0.00****	*0.01***
sum LB-test *(p-values)*	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****

	AT&T		Exxon Mobil		Intel		Nike		McDonald's		Starbucks		Johnson & Johnson		Coca-Cola		Boeing		Pfizer		Comcast	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	3.20	1.00	3.17	1.00	3.34	1.00	3.73	1.00	3.53	1.00	3.19	1.00	3.35	1.00	3.16	1.00	2.56	1.00	3.26	1.00	3.47
Std. Dev.	0.33	0.30	0.41	0.35	0.34	0.32	0.35	0.37	0.35	0.36	0.42	0.40	0.33	0.29	0.41	0.37	0.62	0.44	0.45	0.38	0.34	0.30
Skewness	2.47	0.24	2.70	0.51	1.81	0.00	0.49	-0.38	0.83	-0.37	1.75	-0.01	2.32	0.67	2.43	0.19	4.58	0.82	2.19	0.68	2.03	0.51
Kurtosis	20.08	3.81	15.82	4.55	11.55	3.82	3.16	2.64	5.22	3.03	9.89	3.18	13.19	4.28	15.18	3.78	36.82	4.90	9.94	3.61	10.26	4.02
Minimum	0.39	2.30	0.27	1.95	0.37	2.40	0.34	2.71	0.36	2.56	0.30	2.08	0.50	2.71	0.36	2.20	0.28	1.39	0.42	2.48	0.45	2.71
Maximum	3.87	4.61	3.93	4.61	3.38	4.61	2.25	4.61	2.75	4.61	3.79	4.61	3.35	4.61	3.95	4.61	6.91	4.61	3.54	4.61	2.98	4.61
JB-test *(p-value)*	*0.00****	*0.01***	*0.00****	*0.00****	*0.00****	*0.02***	*0.01***	*0.02***	*0.00****	*0.03***	*0.00****	*0.50*	*0.00****	*0.00****	*0.00****	*0.02***	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****
ADF-test *(p-value)*	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****
sum LB-test *(p-values)*	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****

Table 24: Summary Statistics Stocks: Trading volume (TV)

The table presents the summary statistics of normalized ("norm") and logarithmic ("log") time series corresponding to dollar trading volume for the 22 US-stocks. Further, p-values of test statistics are given for the Jarque-Bera (JB)-test on normality, the augmented Dickey-Fuller (ADF)-test on the presence of a unit-root and the Ljung-Box (LB)-test on independent distribution of the time series. For the ADF-test, the time series are tested on 100 lags, whereas the stated p-value of the LB-test gives the sum of all p-values over 100 lags tested.

	Apple		**Google**		**General Electric**		**Wal-Mart**		**Microsoft**		**IBM**		**Hewlett-Packard**		**Verizon**		**Citigroup**		**Walt Disney**		**Wells Fargo**	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	-0.14	1.00	-0.08	1.00	-0.09	1.00	-0.09	1.00	-0.06	1.00	-0.04	1.00	-0.10	1.00	-0.04	1.00	-0.13	1.00	-0.05	1.00	-0.12
Std. Dev.	0.54	0.53	0.45	0.39	0.49	0.41	0.44	0.41	0.36	0.34	0.30	0.29	0.50	0.43	0.30	0.29	0.56	0.51	0.34	0.32	0.59	0.46
Skewness	1.29	0.02	1.74	0.45	2.43	0.67	1.28	0.25	1.43	0.16	0.94	0.01	3.31	0.06	0.96	0.00	1.79	0.04	1.78	-0.03	3.43	0.67
Kurtosis	4.81	2.56	7.07	3.16	12.32	3.58	5.01	2.58	5.95	3.67	4.16	3.22	24.56	3.96	4.50	3.24	7.03	3.48	10.81	4.30	24.52	3.91
Minimum	0.21	-1.57	0.38	-0.96	0.38	-0.97	0.32	-1.13	0.33	-1.11	0.36	-1.02	0.25	-1.38	0.40	-0.91	0.14	-1.97	0.31	-1.16	0.31	-1.18
Maximum	3.38	1.22	3.15	1.15	4.02	1.39	2.96	1.08	2.57	0.94	2.20	0.79	5.29	1.67	2.21	0.79	3.62	1.29	3.33	1.20	6.20	1.82
JB-test ***(p-value)***	0.00^{***}	*0.27*	0.00^{***}	0.01^{**}	0.00^{***}	0.00^{***}	0.00^{***}	0.06^{*}	0.00^{***}	0.03^{**}	0.00^{***}	*0.50*	0.00^{***}	0.01^{***}	0.00^{***}	*0.50*	0.00^{***}	*0.20*	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
ADF-test ***(p-value)***	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.01^{***}	0.01^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
sum LB-test ***(p-values)***	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.01^{***}	0.01^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}

	AT&T		**Exxon Mobil**		**Intel**		**Nike**		**McDonald's**		**Starbucks**		**Johnson & Johnson**		**Coca-Cola**		**Boeing**		**Pfizer**		**Comcast**	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	-0.04	1.00	-0.08	1.00	-0.06	1.00	-0.06	1.00	-0.04	1.00	-0.09	1.00	-0.05	1.00	-0.04	1.00	-0.06	1.00	-0.06	1.00	-0.04
Std. Dev.	0.30	0.28	0.45	0.39	0.34	0.34	0.39	0.35	0.30	0.29	0.44	0.43	0.35	0.30	0.30	0.28	0.38	0.36	0.43	0.33	0.30	0.30
Skewness	1.68	0.22	1.69	0.53	1.22	-0.37	1.69	0.39	1.17	-0.05	1.32	-0.27	2.58	0.67	1.57	0.19	1.60	-0.03	4.41	0.75	1.05	-0.44
Kurtosis	8.93	4.10	6.54	3.04	6.72	4.20	7.00	3.62	6.09	3.61	5.42	4.09	13.64	5.29	7.79	4.08	8.09	3.94	35.30	6.27	5.54	4.88
Minimum	0.42	-0.88	0.42	-0.88	0.25	-1.39	0.34	-1.07	0.40	-0.92	0.16	-1.82	0.36	-1.01	0.38	-0.96	0.26	-1.36	0.34	-1.07	0.28	-1.27
Maximum	2.85	1.05	3.13	1.14	2.94	1.08	2.85	1.05	2.48	0.91	3.00	1.10	3.27	1.18	2.71	1.00	3.30	1.19	4.86	1.58	2.43	0.89
JB-test ***(p-value)***	0.00^{***}	0.00^{***}	0.00^{***}	0.01^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.01^{***}	0.00^{***}	0.08^{*}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.01^{**}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
ADF-test ***(p-value)***	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.01^{***}	0.01^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}
sum LB-test ***(p-values)***	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}	0.00^{***}

Table 25: Summary Statistics Stocks: Illiquidity (ILLQ)

The table presents the summary statistics of normalized ("norm") and logarithmic ("log") time series corresponding to the Amihud-illiquidity measure for the 22 US-stocks. Further, p-values of test statistics are given for the Jarque-Bera (JB)-test on normality, the augmented Dickey-Fuller (ADF)-test on the presence of a unit-root and the Ljung-Box (LB)-test on independent distribution of time series. For the ADF-test, time series are tested on 100 lags, whereas the stated p-value of the LB-test gives the sum of all p-values over 100 lags tested.

	Apple		Google		General Electric		Wal-Mart		Microsoft		IBM		Hewlett-Packard		Verizon		Citigroup		Walt Disney		Wells Fargo	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	-0.26	1.00	-0.12	1.00	-0.12	1.00	-0.08	1.00	-0.12	1.00	-0.12	1.00	-0.16	1.00	-0.15	1.00	-0.36	1.00	-0.21	1.00	-0.12
Std. Dev.	0.84	0.69	0.52	0.49	0.51	0.50	0.38	0.41	0.57	0.48	0.54	0.49	0.57	0.59	0.62	0.53	1.21	0.76	0.74	0.62	0.51	0.50
Skewness	2.13	0.46	1.68	-0.03	1.20	-0.16	0.64	-0.77	2.24	0.48	2.03	0.04	1.09	-0.24	2.58	0.08	4.16	0.80	2.18	0.26	1.16	0.02
Kurtosis	8.11	2.92	7.67	3.08	4.52	3.14	4.43	3.84	9.76	3.39	9.23	3.72	4.28	2.65	13.23	3.98	26.09	3.83	8.41	3.40	4.26	2.60
Minimum	0.15	-1.93	0.22	-1.52	0.14	-2.00	0.24	-1.44	0.25	-1.39	0.18	-1.69	0.15	-1.89	0.11	-2.20	0.12	-2.08	0.10	-2.32	0.22	-1.49
Maximum	5.64	1.73	3.78	1.33	2.90	1.06	2.76	1.01	4.27	1.45	4.04	1.40	3.36	1.21	5.07	1.62	10.29	2.33	4.81	1.57	3.11	1.14
JB-test ***(p-value)***	*0.00****	*0.01***	*0.00****	*0.50*	*0.00****	*0.43*	*0.00****	*0.00****	*0.00****	*0.01***	*0.00****	*0.04***	*0.00****	*0.09**	*0.00****	*0.01****	*0.00****	*0.00****	*0.00****	*0.06**	*0.00****	*0.33*
ADF-test ***(p-value)***	*0.18*	*0.08**	*0.01***	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.03***	*0.00****	*0.00****	*0.06**	*0.20*	*0.00****	*0.00****	*0.21*	*0.19*	*0.12*	*0.03***	*0.01***	*0.00****
sum LB-test ***(p-values)***	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****

	AT&T		Exxon Mobil		Intel		Nike		McDonald's		Starbucks		Johnson & Johnson		Coca-Cola		Boeing		Pfizer		Comcast	
	norm	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**	**norm**	**log**
Mean	1.00	-0.14	1.00	-0.08	1.00	-0.15	1.00	-0.17	1.00	-0.12	1.00	-0.33	1.00	-0.12	1.00	-0.12	1.00	-0.16	1.00	-0.12	1.00	-0.18
Std. Dev.	0.59	0.52	0.41	0.41	0.65	0.54	0.66	0.57	0.52	0.52	1.08	0.74	0.51	0.50	0.55	0.49	0.60	0.56	0.54	0.49	0.69	0.60
Skewness	2.29	0.14	1.55	-0.44	2.47	0.20	2.01	0.25	1.56	-0.62	3.10	0.70	1.41	-0.22	1.86	0.16	1.53	0.14	1.82	0.06	2.14	0.01
Kurtosis	11.46	3.18	9.49	4.02	11.23	3.88	7.89	3.28	6.99	5.66	13.73	3.50	6.02	3.38	7.32	3.46	5.65	2.75	8.26	3.27	8.41	4.26
Minimum	0.23	-1.49	0.17	-1.77	0.16	-1.81	0.15	-1.87	0.05	-2.91	0.17	-1.79	0.15	-1.90	0.17	-1.76	0.16	-1.81	0.18	-1.71	0.05	-2.94
Maximum	4.48	1.50	3.40	1.22	5.05	1.62	4.47	1.50	3.68	1.30	6.82	1.92	3.40	1.22	3.32	1.20	3.77	1.33	3.92	1.37	4.19	1.43
JB-test ***(p-value)***	*0.00****	*0.46*	*0.00****	*0.00****	*0.00****	*0.01****	*0.00****	*0.10*	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.09**	*0.00****	*0.12*	*0.00****	*0.38*	*0.00****	*0.50*	*0.00****	*0.00****
ADF-test ***(p-value)***	*0.00****	*0.00****	*0.00****	*0.00****	*0.26*	*0.02***	*0.15*	*0.00****	*0.00****	*0.00****	*0.16*	*0.02***	*0.00****	*0.00****	*0.00****	*0.00****	*0.03****	*0.00****	*0.02***	*0.00****	*0.04***	*0.00****
sum LB-test ***(p-values)***	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****	*0.00****

Table 26: Variance decomposition test statistics of stock sample

The table presents the results obtained by conducting forecast error variance decomposition on log Google search volume index (SVI) and log search volume index (SVI) and measures of market activity for the U.S. stock sample. The values indicate the fraction of how much the vertical variable (left axis) explains in the variance of the horizontal variable (above axis) at forecast step 20.

		Apple		Google		General Electric		Wal-Mart		Microsoft		IBM		Hewlett-Packard		Verizon	
		α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI
α = RV	RV	0.998	0.070	0.992	0.118	0.893	0.319	0.949	0.052	0.989	0.131	0.953	0.185	0.987	0.115	0.998	0.051
	SVI	0.002	0.930	0.008	0.882	0.107	0.681	0.051	0.948	0.011	0.869	0.047	0.815	0.013	0.885	0.002	0.949
α = TV	TV	0.981	0.597	1.000	0.236	0.964	0.329	0.993	0.047	0.960	0.155	0.993	0.264	0.957	0.082	0.969	0.127
	SVI	0.019	0.403	0.000	0.764	0.036	0.671	0.007	0.953	0.040	0.845	0.007	0.736	0.043	0.918	0.031	0.873
α = ILLQ	ILLQ	0.951	0.254	0.978	0.009	0.881	0.092	0.955	0.026	0.967	0.020	0.900	0.025	0.894	0.046	0.998	0.081
	SVI	0.049	0.746	0.022	0.991	0.119	0.908	0.045	0.974	0.033	0.980	0.100	0.975	0.106	0.954	0.002	0.919

		Citigroup		Walt Disney		Wells Fargo		AT&T		Exxon Mobil		Intel		Nike		McDonald's	
		α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI
α = RV	RV	0.996	0.196	0.994	0.023	0.972	0.149	0.989	0.035	0.865	0.151	0.993	0.051	0.993	0.039	0.987	0.044
	SVI	0.004	0.804	0.006	0.977	0.028	0.851	0.011	0.965	0.135	0.849	0.007	0.949	0.007	0.961	0.013	0.956
α = TV	TV	0.992	0.071	0.994	0.126	0.906	0.155	0.952	0.047	0.949	0.260	0.980	0.157	0.997	0.046	0.994	0.100
	SVI	0.008	0.929	0.006	0.874	0.094	0.845	0.048	0.953	0.051	0.740	0.020	0.843	0.003	0.954	0.006	0.900
α = ILLQ	ILLQ	0.995	0.081	0.969	0.013	0.978	0.031	0.997	0.049	0.962	0.006	0.935	0.031	0.992	0.026	0.989	0.014
	SVI	0.005	0.919	0.031	0.987	0.022	0.969	0.003	0.951	0.038	0.994	0.065	0.969	0.008	0.974	0.011	0.986

		Starbucks		Johnson Johnson		Coca-Cola		Boeing		Pfizer		Comcast	
		α	SVI	α	SVI	α	SVI	α	SVI	α	SVI	α	SVI
α = RV	RV	0.998	0.024	0.940	0.088	0.990	0.015	0.997	0.130	0.971	0.111	0.998	0.030
	SVI	0.002	0.976	0.060	0.912	0.010	0.985	0.003	0.870	0.029	0.889	0.002	0.970
α = TV	TV	0.999	0.128	0.940	0.057	0.998	0.035	0.975	0.305	0.949	0.077	0.991	0.056
	SVI	0.001	0.872	0.060	0.943	0.002	0.965	0.025	0.695	0.051	0.923	0.009	0.944
α = ILLQ	ILLQ	0.994	0.024	0.964	0.054	0.989	0.010	0.992	0.002	0.995	0.046	0.993	0.019
	SVI	0.006	0.976	0.036	0.946	0.011	0.990	0.008	0.998	0.005	0.954	0.007	0.981

Table 27: Granger causality test statistics of stock sample

The table presents the results obtained by conducting Granger causality tests on log search volume index (SVI) and measures of market activity for the U.S. stock sample. α corresponds to either log realized volatility (RV), log trading volume (TV) or log Amihud-illiquidity (ILLQ). Granger-Statistics are given for the significance of the explanatory variable statistically causing the response variable. Significance of p-values is indicated as follows: * for 10% significance level, ** for 5% and *** for 1% significance level.

		Apple		Google		General Electric		Wal-Mart		Microsoft		IBM		Hewlett-Packard		Verizon	
	Response variable	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**
	Explanatory variable	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**
α = RV	Granger-Stat	0.60	9.21	1.31	4.63	11.17	4.90	4.69	9.86	1.22	6.71	9.83	6.11	3.15	6.68	0.19	4.89
	p-value	*.740*	*.010******	*.519*	*.099**	*.004******	*.086**	*.096**	*.007******	*.543*	*.035****	*.007******	*.047****	*.207*	*.035****	*.907*	*.087**
α = TV	Granger-Stat	1.17	2.42	0.02	2.38	3.84	0.00	1.65	6.02	9.97	6.71	1.15	1.20	10.32	5.09	6.86	0.20
	p-value	*.557*	*.299*	*.893*	*.123*	*.050**	*.983*	*.647*	*.111*	*.019****	*.082**	*.284*	*.273*	*.001******	*.024****	*.009******	*.657*
α = ILLQ	Granger-Stat	2.96	14.46	8.01	1.57	16.77	3.15	5.47	4.32	4.73	4.72	14.98	3.63	10.44	7.06	1.10	7.58
	p-value	*.227*	*.001******	*.018****	*.455*	*.000******	*.207*	*.065**	*.115*	*.094**	*.094**	*.001******	*.163*	*.034****	*.133*	*.578*	*.023****

		Citigroup		Walt Disney		Wells Fargo		AT&T		Exxon Mobil		Intel		Nike		McDonald's	
	Response variable	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**
	Explanatory variable	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**
α = RV	Granger-Stat	0.60	0.96	1.32	0.58	4.16	0.44	2.70	1.44	12.77	4.34	3.87	0.15	0.47	5.04	6.57	1.91
	p-value	*.741*	*.619*	*.725*	*.900*	*.125*	*.804*	*.441*	*.697*	*.002******	*.114*	*.145*	*.925*	*.791*	*.081**	*.087**	*.592*
α = TV	Granger-Stat	0.51	5.86	1.37	0.83	8.18	6.96	4.76	0.06	7.58	6.65	3.54	6.04	1.04	4.90	2.01	4.71
	p-value	*.476*	*.015****	*.242*	*.362*	*.017****	*.031****	*.029****	*.806*	*.056**	*.084**	*.060**	*.014****	*.596*	*.086**	*.367*	*.095**
α = ILLQ	Granger-Stat	2.13	1.63	7.92	0.42	2.83	1.77	0.70	4.61	6.03	0.45	20.71	4.76	3.91	1.01	6.00	1.42
	p-value	*.344*	*.444*	*.048****	*.937*	*.243*	*.414*	*.704*	*.100**	*.014****	*.503*	*.000******	*.092**	*.142*	*.604*	*.050****	*.492*

		Starbucks		Johnson Johnson		Coca-Cola		Boeing		Pfizer		Comcast	
	Response variable	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**
	Explanatory variable	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**	**SVI**	**α**
α = RV	Granger-Stat	0.36	5.37	9.61	4.27	0.89	0.29	0.43	1.35	6.32	0.62	0.20	1.67
	p-value	*.833*	*.068**	*.008******	*.118*	*.642*	*.865*	*.808*	*.509*	*.042****	*.734*	*.904*	*.433*
α = TV	Granger-Stat	0.19	2.87	9.98	1.71	0.48	1.89	2.83	0.18	7.62	0.96	2.11	0.05
	p-value	*.910*	*.238*	*.007******	*.426*	*.491*	*.169*	*.092**	*.668*	*.022****	*.618*	*.146*	*.825*
α = ILLQ	Granger-Stat	1.05	3.37	7.41	3.73	0.91	1.00	2.43	1.07	1.34	0.83	3.47	2.63
	p-value	*.593*	*.185*	*.025****	*.155*	*.636*	*.607*	*.296*	*.586*	*.511*	*.659*	*.325*	*.452*