CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Tomáš Vakrman

Google searches and financial markets: IPOs and uncertainty

Master Thesis

Prague 2014

Author: Bc. Tomáš Vakrman
Supervisor: PhDr. Ladislav Krištoufek Ph.D.

Academic year: 2013/2014

Declaration of Authorship

1. Hereby I declare that I have compiled this master thesis independently, using only the listed literature and sources.
2. I declare that the thesis has not been used for obtaining another title.
3. I agree on making this thesis accessible for study and research purposes.

Prague, May 16, 2014 Tomáš Vakrman

Acknowledgments

I would like to express my deepest gratitude to the thesis supervisor, PhDr. Ladislav Krištoufek Ph.D., for his patience, support and priceless guidance. I would also like to thank to Petr Beňas for his help with programming, Mr. Patton for provision of the IPO database and both Mrs. Love and Mr. Decker for their Stata plug-in for PVARs. Finally, my love goes to my family and Hana P. for their everlasting support and inspiration in life.

Bibliographic entry

VAKRMAN, Tomáš. Google searches and financial markets: IPOs and uncertainty. Prague, 2014. 173 pp. Master Thesis (Mgr.). Charles University in Prague, Faculty of Social Sciences, Institute of Economic Studies. Supervisor: PhDr. Ladislav Krištoufek Ph.D.

Bibliografický záznam

VAKRMAN, Tomáš. Google searches and financial markets: IPOs and uncertainty. Praha, 2014. 173 str. Magisterská práce (Mgr.). Univerzita Karlova v Praze, Fakulta sociálních věd, Institut ekonomických studií. Vedoucí práce: PhDr. Ladislav Krištoufek Ph.D.

Abstract

This thesis studies how the investor attention proxied by Google search volume affects different aspects of market behavior. My results show that a surge in online attention is associated with an increase in trading activity and stock price volatility, but no effect is detected for daily returns. Yet, if market sentiment is taken into account, the relationship comes to the surface for returns as well. The returns tend to decrease with attention hikes in negative sentiment periods and the opposite is observed for periods of positive sentiment, suggesting that Google web search captures predominately attention of sentiment investors. Moreover, I demonstrate that with the outburst of financial crisis, the interdependence between attention and trading activity was intensified. Lastly, I provide evidence that web search may shed some light on IPO-related puzzles. The initial returns seem to be higher for IPOs that receive above average attention, and are likely to be reversed in long-term. In addition, it is ascertained that web search volume may act as a proxy for market overreaction to the offerings.

JEL Classification	D83, G02, G10, G12
Keywords	attention, Google, internet search, individual investor, IPO, sentiment, web search

Author's e-mail	tomas.vakrman@gmail.com
Supervisor's e-mail	kristoufek@ies-prague.org

Abstrakt

Tato práce zkoumá, jaký vliv má pozornost investorů, zastoupená objemem vyhledávání na Google, na různé aspekty tržního chování. Ve výsledcích studie ukazuji, že růst vyhledávání spjatého s firmou je spojen s nárůstem obchodovaného množství a vyšší volatilitou cen akcií, nicméně dopad na denní výnosy nebyl potvrzen. To ovšem neplatí v případě, bereme-li v potaz sentiment na akciovém trhu. Výnosy mají tendenci klesat s růstem on-line pozornosti v dobách negativního sentimentu a naopak stoupat v dobách pozitivního sentimentu, což naznačuje, že vyhledávání na Google zachycuje převážně pozornost těch investorů, kteří podléhají náladě na trhu. Dále ukazuji, že s vypuknutím finanční krize došlo k zintenzivnění závislosti mezi pozorností investorů a obchodní aktivitou na trhu. V neposlední řadě tato práce přichází s poznatkem, že vyhledávání na internetu může vnést trochu světla do problematiky prvotních veřejných nabídek akcií. Počáteční výnosy při emisích akcií bývají vyšší u firem, jež zaznamenaly nadprůměrnou pozornost investorů, přičemž tyto firmy mají naopak nižší výnosnost v dlouhém období. Závěr práce přináší zjištění, že pomocí objemu vyhledávání lze předpovědět přehnanou reakci investorů na prvotní nabídky akcií.

Klasifikace JEL	D83, G02, G10, G12
Klíčová slova	Google, individuální investor, IPO, pozornost, sentiment, vyhledávání na internetu, vyhledávání na webu

E-mail autora	tomas.vakrman@gmail.com
E-mail vedoucího práce	kristoufek@ies-prague.org

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The Internet, a revolutionary invention from 1965 with more than two billion users by 2014, has undoubtedly changed the world we live in. It allowed its users to access an unprecedented amount of information in very short time. Due to the abundance of available information, attention has become a scarce resource that needs to be correctly allocated in order to acquire the information of interest. For vast majority of Internet users, search engines serve as the gateway to all that information; and Google, with its 69% market share and more than one billion unique visitors every month, is the uncrowned king among them. Since economics focuses on studying the behavior of individuals, it is worth noticing that such online behavior leaves a digital trace. All individual search queries typed into search bar are saved by Google and the processed statistics on search are made publicly available by the company via its online facility Google Trends. Thus, the Google search volume data produces a direct measure of people's attention that is freely available, timely and representative to whole population of internet users.

When Lehman Brothers declared bankruptcy in September 2008, the Google search volume for the keyword "*Lehman Brothers*" exceeded the historical average 38 times. Three years later, the Czech ex-president Václav Klaus visited Chile and experienced the highest increase in popularity in his career when he moved a ceremonial pen from the table into his pocket during the signing ceremony. Demand for information on Google about Mr. Klaus grew 33 times. In the recent history, the political crisis in Ukraine caused 16 fold increase in Google search volume. The above mentioned examples nicely show the wide range of topics people pay attention to.

Recently, researchers realized the extreme potential of internet search data and provided evidence that they can be used to track or even anticipate several social phenomena. The utilization stems from influenza tracking (Eysenbach, 2005; Ginsberg et al., 2008; Dugas et al., 2012), consumer interest and its impact on product sales (Choi and Varian, 2009b; Goel et al., 2010a; Kulkarni, 2012) to macroeconomic indicators (Cooper et al., 2005; Choi and Varian, 2009a; D'Amuri and Marcucci, 2009).

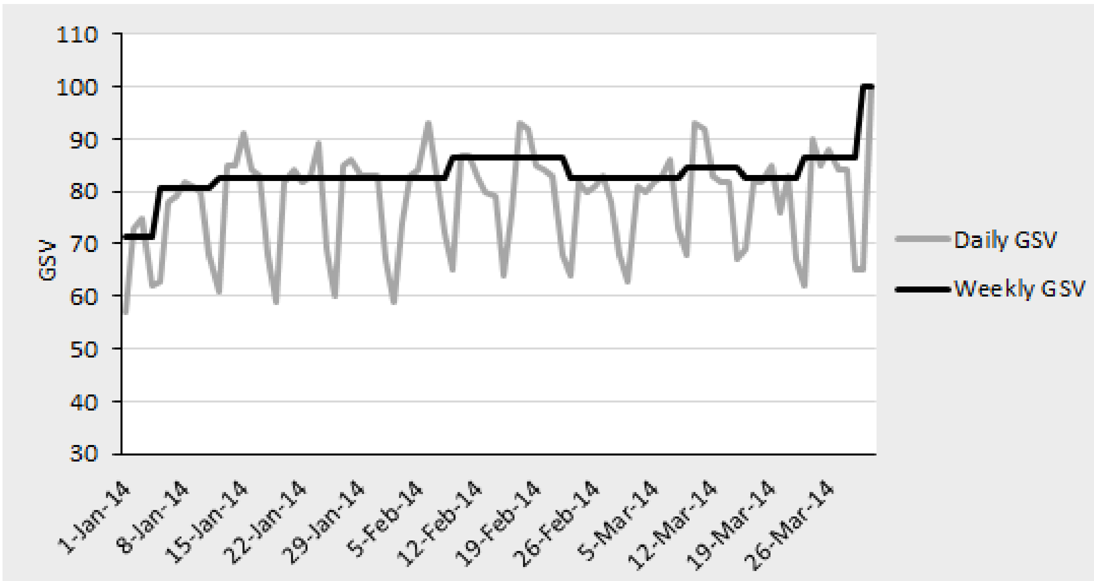
The work of Merton (1987) suggests that attention may be also relevant for complex reality of financial markets and Preis et al. (2008) were the first ones, as far as I am aware, to prove this hypothesis right using attention proxied by web search data. Since then, many researchers used online attention to either track or forecast various financial indicators. I build up on their work and provide a complex review of Google data ability to predict financial data; namely the trading volume, stock price volatility and stock returns.

Most of the authors use weekly or monthly frequency Google search volume to proxy for investor attention (for example Da et al., 2011; Bank et al., 2011), yet as Figure 1.1 demonstrates, considerable amount of information is lost if weekly data are used. Therefore, in in line with Dimpfl and Jank (2011); van Themaat (2012), I employ Google search volume on daily frequency as a measure of attention. As for the search term, I follow Vlastakis and Markellos (2012); Bank et al. (2011) and use name-like unique search term that is unequivocally associated with the firm, rather than stock ticker as Da et al. (2011); van Themaat (2012). I believe the specified search term is more capable of capturing attention of noise traders, who I am mostly interested in, as they are more likely to exhibit herd behavior that should nicely show up in Google search volume.

Contrary to other authors, who mostly focused on one or two different financial indicators and used different methods for modeling the relationship with online attention, I aim to apply comparable methodology for the three indicators and employ the same data set for each. Thus, I am not only able to describe the impact of investor attention on different aspects of market behavior, but also capable of performing pair-wise comparison between the impacts, in terms of strength and nature. The bilateral relationship with web search is examined both in one-by-one time-series setting and panel-data setting. In addition, I try

Figure 1.1: Daily and weekly frequency GSV for "DuPont"

This figure compares daily and weekly Google search volume (GSV) for keyword "DuPont" from January 2014 to March 2014. The gray line represents daily frequency GSV and the black line weekly frequency GSV.



to answer a question whether the relationship between financial indicators and investor attention differs in periods of high uncertainty. Finally, I question the conclusion of Da et al. (2011) that retail investor attention and retail investor sentiment are positively related. I show that the attention in different sentiment levels can yield to opposing effects on returns.

I state following conjectures that I strive to test:

Conjecture 1. *"An increase in web search volume for firm-related query is associated with an increase in trading volume at the same day as well as the day after"*

There are two underlying ideas for the hypothesis. First, one must be aware of the stock's existence in order to get involved in trading with it. Second, the trading activity should react to news regarding the firm and so should the investor attention.

Conjecture 2. *"An increase in web search volume for firm-related query is associated with an increase in volatility at the same day as well as the day after"*

The hypothesis builds on the previous one and the fact that trading volume and stock price volatility are positively correlated. In addition, the positive interdependence is predicted by theory Andrei and Hasler (2011).

Conjecture 3. *"Web search volume for firm-related query CANNOT predict daily stock returns as Google data are NOT able to capture buying/selling intention of traders"*

The hypothesis is based on the findings by Preis et al. (2010), but contradicts the theoretical predictions of Barber and Odean (2008). Yet, I believe that solely the attention cannot affect returns in specific direction and that certain key is necessary to disentangle the intentions of investors.

Conjecture 4. *"High uncertainty makes investors to consult their trading decision with Google more often - that is, it enhances the severity of interdependence between web search and trading volume"*

I rely on the belief that investors pay more attention to trading when the uncertainty is high, as the risk increases. In addition, I think that investors are likely to use more timely source of information when news can change quickly.

I also use the Google search data to assess the impact of attention on initial public offering (IPO) returns. I model both the initial and long-term returns in order to test Google data usability in explaining two IPO stylized facts, the high initial returns and long-term underperformance. Lastly, I test whether the theoretical model of Ma and Tsai (2002) goes well with attention measured by Google search volume.

I state three further conjectures for IPOs and Google data:

Conjecture 5. *"IPOs that experience high attention from investors will exhibit higher initial returns"*

While I believe that Google data are incapable of disentangling the investors' buying/selling intentions, it does not apply to IPOs since stocks are not yet available on the market and short selling is limited.

Conjecture 6. *"IPOs that experience high attention from investors will exhibit lower long-term return"*

In line with attention/sentiment based theories (for example Derrien, 2005) and Conjecture 5, I suppose that the high initial returns induced by high attention of retail investor will revert in long term.

Conjecture 7. *"IPOs that experience high attention from investors will exhibit higher market reaction"*

It follows from the definition of market reaction as defined by Ma and Tsai (2002) , which is supposed to measure the overall investor reaction to IPO, thus, one may suspect market reaction to be positive for high attention IPOs and vice versa.

Conjecture 8. *"Investor attention prior stock emission will not affect the true discount of IPO"*

This stems from the definition of true discount as defined by Ma and Tsai (2002). It is supposed to measure the real underpricing of IPO and one can hardly expect retail investors to be able to correctly estimate the true value of the offering and thus the actual underpricing.

The remainder of this paper is organized as follows. Chapter 2 provides review of related literature. Chapter 3 describes the data resources, variable construction and sample construction. Chapter 4 outlines the methodology used in computational chapters. Chapter 5 studies the relationship between Google search volume and different aspects of market behavior. Chapter 6 examines two IPO stylized facts in relation with Google data. Finally, Chapter 7 concludes.

Chapter 2

Literature review

This chapter is divided into three parts. The first part introduces a summary of the application of Internet related data in research. The second part focuses on Google data employment in finance. Finally, the third part focuses on IPOs; two IPO stylized facts are discussed in the relation with Google Trends data.

2.1 Non-financial application of web search data

To my knowledge, web search data were for the first time used in Cooper et al. (2005). The authors found link between Yahoo! search activity associated with specific cancers and their estimated incidence and mortality. Independently on Cooper et al. (2005), Ettredge et al. (2005) published paper, in which they showed that web search is better predictor of US unemployment than the official unemployment insurance claims data; the authors used occurrence of job-related web search in WordTracker's Top 500 Key-word Report. The two studies started a wave of web search data application in academic research, covering wide spectrum of topics in different fields of study.

Most of the early research covered health related topics; most notably, the web search data were used in influenza outbreaks tracking (Eysenbach, 2005; Ginsberg et al., 2008; Carneiro and Mylonakis, 2009; Dugas et al., 2012; Hulth et al., 2009; Polgreen et al., 2008), giving birth to Google-operated, freely available flu tracking model.[1] Pelat et al. (2009); Zhou et al. (2011) extended the

[1] Accesible at http://www.google.org/flutrends

application on illnesses different from flu, while others used alternative Internet-related data, such as Twitter and blog posts (Corley et al., 2009; Achrekar et al., 2011).

In the field of economics, to my knowledge, it took longer until researchers realized the untapped potential of web search data. The expansion of economics-related research using Internet data is connected with the launch of the publicly available Google web search database *Google Trends* or more precisely with its more sophisticated counterpart *Google Insights for Search* in 2008.[2] The pilot paper using a Google data set was written by two Google-related economists Choi and Varian (2009b), who showed that Google Trends data are successful in nowcasting (i.e. more rapid estimation of present data) different economic quantities; namely US retail, home and automotive sales, and travel destinations. In addition, the authors published a follow-up paper extending the application of Google Trends data on a prediction of the present US unemployment (Choi and Varian, 2009a).

Since then, many authors built on the groundbreaking papers of Choi and Varian. Arguably, the most popular topic in studies using Google Trends has been the unemployment; the Google data were shown to predict the current or future unemployment data for the US (D'Amuri and Marcucci, 2009; Ettredge et al., 2005; Choi and Varian, 2009a), Israel (Suhoy, 2009), Italy (D'Amuri, 2009), Germany (Askitas and Zimmermann, 2009) and the UK (McLaren and Shanbhogue, 2011). The correctness of the Google data use in the unemployment studies was confirmed by Baker and Fradkin (2011), who examined the drivers of job related web search. The authors found web search activity to be higher for those at the onset of an unemployment spell and those nearing the exhaustion of their benefits. Nevertheless, the application in the macroeconomics field is far broader; ranging from the inflation forecasting (Guzman, 2011) to the predictions of GDP (Li et al., 2013).

Another stream of the economic research followed Choi and Varian (2009b) and applied the Google data in the field of microeconomics. Chamberlin (2010) replicated their study for UK, while others applied Google Trends in the prediction of commercial success of various cultural products; namely movies, video

[2]The two versions merged into one in 2012; from now on, I will use the term Google Trends for both version of the database interchangeably.

games, and songs (Goel et al., 2010a,b; Kulkarni, 2012); house sales and prices (Wu and Brynjolfsson, 2009); and house foreclosures (Webb, 2009).

Somewhere between micro- and macroeconomics studies lies Google data application as a proxy for consumer sentiment (Della Penna and Huang, 2009; Vosen and Schmidt, 2011), in which the authors demonstrated that the Google data are a superior predictor of consumer spending to prime consumer sentiment indices.

2.2 Financial application of web search data

Since this thesis strives to deal with Google Trends data application in finance, the following subsection is devoted to similarly aimed studies.

2.2.1 Trading volume

As far as I know, the pilot application of Google Trends data in finance addressed the question, whether changes in web search and weekly transaction volume changes are cross correlated (Preis et al., 2010). Using the S&P 500 stocks data and the search volume for corresponding company names, the authors found significant positive correlation at $\triangle t = 0$ as well as pattern-based complex short-time correlations between the two time series (for the latter, they used method introduced in Preis et al. (2008)).

Bank et al. (2011) chose a different approach to examine the relationship between web search and trading volume; the authors sorted all stocks according to their signed change in search volume and computed equally weighted averages of the respective stock characteristics. Their findings confirmed the existence of positive interdependence between web search and trading volume. In addition, they show that *"large (small) signed change in search volume is associated with a small (large) signed change in illiquidity, as measured by the Amihud (2002) ratio"* (Bank et al., 2011, p. 262). They attributed the web-search-illiquidity relationship to the changes in cost of asymmetric information and claimed that *"Google particularly measures the interest of uninformed investors"* (Bank et al., 2011, p. 263). This view on Google data differs substantially from the one of Preis et al. (2010), who took it only as a proxy for company recognition. The fact that Google Trends data captures the attention of retail investors was also confirmed in more recent research, most notably by Da et al. (2011).

Chen (2011) in his bachelor thesis showed that the investor attention - measured by the Google search volume - may not be merely firm specific, but also general market related. At a Dutch stock market sample, he showed that for most stocks, the trading volume is positively affected by both stock-specific and market-related search volume; nevertheless, the effect was more significant for the market-related search volume. Ramos et al. (2013) examined how the relationship between web search and volume reacts to situations, when either a firm or the market breaks through their 52-week highs or lows. Interestingly, they found that the relationship is intensified when either a firm or the market hits the 52-week highs, but not significantly changed if either a firm or the market hits the 52-week lows. Furthermore, according to the authors, the market information is a more significant determinant in changing the effects of investor attention than the firm specific information. Also web search seem to be transformed in more trading activity in periods of positive returns than in periods of negative returns. Ap Gwilym et al. (2012) provided further support to the hypothesis that the Google search represents investor attention of retail investors rather than institutional investors. They found that the positive interaction between web search and trading volume holds for constituents of Chinese A Shares indices (dominated by retail investors), while it does not for constituents of B Shares indices (dominated by institutional investors).

Lastly, Bordino et al. (2012) examined the web search and trading volume relationship in a bilateral setting. Using NASDAQ-100 index and Yahoo! search for stock tickers, the authors obtained support for web search ability to Granger-cause trading volume as well as for the opposite direction Granger-causality; however, the opposite direction relationship was far weaker. Moreover, they demonstrate that *"adding information about today's query volume reduces the average prediction error (in an auto-regressive model) for tomorrows trading volume by about 5% (...) but the reverse does not hold"* (Bordino et al., 2012, p. 12).

2.2.2 Stock price volatility

Andrei and Hasler (2011) developed a theoretical model for investor attention and stock market volatility. The model features fluctuating investor attention to news, and implies quadratic relationship between the two variables. The authors argue that investor attention affects volatility via two contradicting forces. First, volatility increases quadratically with attention as more information is incorporated in prices; second, volatility decreases linearly with attention since it decreases uncertainty. To test the model empirically, they performed quadratic fit of the one-week ahead S&P 500 volatility on the attention index (constructed using Google search volumes on groups of words with financial or economic content, excluding words with positive or negative connotations). The estimation results are in line with the model predictions, at least as the signs of coefficients are concerned, since the linear term shows a negative sign and the quadratic one a positive sign.

As for empirical findings, Dimpfl and Jank (2011) addressed the web search data's ability to describe and predict market volatility. According to the authors, past surge in web search Granger-causes future volatility and the effect is concentrated in the first lag with a positive sign. Furthermore, they also showed that the effect prevails in long term. On the other hand, Chen (2011) came to less convincing results, as some of the stocks in his sample show negative and some positive interdependence between the firm specific search volume and volatility; and circa 40% of his sample does not show any significant relation at all. Conversely, he found the market related search to be positive and significant predictor of volatility for all stocks in his sample.

Vlastakis and Markellos (2012) took the web search as a proxy for the information demand from investors, where the supply is represented by financial information in news. Similarly to Chen (2011), they showed that the market-level information demand is a stronger and more unambiguous predictor (strictly positive relationship) of volatility than its firm specific counterpart. In addition, they found that the strength of the relationship escalates in high return market states. Ramos et al. (2013) found contradicting results. They demonstrated that if market hits the 52-week highs (or lows), there is no significant change in the strength of the relation; nonetheless, if a firm price breaks through the 52-week lows, the relationship between web search and volatility intensifies, while the opposite is truth for the 52-week highs breaks. Furthermore, they provided an evidence for the asymmetric nature of the relationship (positive change has a positive impact, while negative one does not impact volatility).

2.2.3 Prices and returns

While the empirical predictions on the interdependence between attention and volume are rather straightforward, and the similar may be told about the interdependence between attention and volatility (albeit in case of volatility it is slightly more complicated due to convexity issues), the empirical findings about investor attention and (abnormal) returns are somewhat contradicting. And so is the theory.

Merton (1987) presented a theoretical model with incomplete information in which he argues that stocks with less investor attention - less-widely known firms with smaller investor base - yield higher returns so that that idiosyncratic risk is compensated to the investors. The model predictions are in line with the empirical theory on "neglected" stocks (Arbel and Strebel, 1982; Arbel et al., 1983). On the other hand, Barber and Odean (2008) offered different point of view on the issue. They argue that (retail) investors can choose from a broad selection of stocks when buying but face limited choice when selling, as they can solely sell the stock they already own if not short-selling (which is not frequent for retail investors). Therefore, they buy stocks that recently caught their attention. By contrast, when selling, they pay attention to how the stocks in their portfolio have behaved. The asymmetric approach to selling and buying does not hold for institutional investors, since they 1) face search problem also when selling (due to holding a broad portfolio of stocks) and 2) do not confront the problem of attention being a scarce resource. The authors found a support for their hypothesis, as the individual investors appear to be net buyers on high volume days, after high previous-day returns and when the stock is extensively covered in news, i.e., the retail investors are net buyers of attention-grabbing stocks. This finding is not in line with the predictions of Merton (1987), as his model does not assume any bias towards the high attention stocks. Additionally, the authors stated predictions of their model on the return dynamics with respect to changes in investor attention - short-term up-rise in prices as the attention of investors rises, followed by medium/long-term reversal.

Preis et al. (2010) was the first one to empirically test return predictability on web search, finding no significant correlation between the two time series. It should be noted, that the authors found significant correlation for web search and volume in their data set, which indicates that neither selling nor buying is preferred when one looks for company name on Google and gets involved in trading subsequently; inconsistent with Barber and Odean (2008) predictions.

Bank et al. (2011) chose a distinct method to assess the validity of Barber and Odean (2008) model predictions. The authors divided their sample into three quantiles according to the change in Google search volume and form a zero investment strategy (that goes long in portfolio with largest change in web search and short in the portfolio with lowest change in web search). They found the next month's return of the long portfolio to be on average 0.347% higher, than the return of the short one; nevertheless, the effect does not prevail if controlled for four factors of Carhart (1997). Thus, as the authors claim, the support for attention-induced risk premium of Barber and Odean (2008) seems weak, while negative effect of investor attention on next month's returns as predicted by Merton (1987) is not present in their setting at all. Yet, one might object that the Merton model predictions may become apparent in long-term only. In accordance with the objection, Kristoufek (2013) demonstrated that portfolios with higher weighs of peripheral stocks and lower weights of popular stocks dominate uniformly weighted portfolios. Therefore, it seems that a short-term increase in web search drives the returns up, while overall high level of attention the opposite way.

Ap Gwilym et al. (2012) showed, using weekly Google search data, that weekly returns tend to be driven up by attention; and that the effect is significantly higher for constituents of Chinese A Shares indices than constituents of B Shares indices. Additionally, they found that the impact of attention on returns is mostly short lived and that the lagged search has negative effect on returns, suggesting that the surge in current web search generates price pressure which is subsequently corrected by lower near-term future returns. All in all, their results give support to Barber and Odean (2008) model predictions on the retail investors' bias towards the attention-grabbing stocks as well as to the predictions on the return reaction to changes in attention. Shi et al. (2012) reached similar results using Baidu search data; their results provide support to the short term price pressure for the attention grabbing stocks with subsequent long-term reversal. Also Ramos et al. (2013); Kita and Wang (2012) (the later used FX market data) showed the negative relation between longer-term cumulative returns a Google searches. In addition Ramos et al. (2013) demonstrated that the short-term price pressure is increased if either a firm or the market prices hit the 52-week highs, while the long-term effect is exacerbated if market sinks to 52-week low. The most notable validation of Barber and Odean (2008) model was, however, presented by Da et al. (2011) - arguably the most famous paper using the Google data - who vindicated model's predictions with an interesting precision. They found strong evidence for the price pressure hypothesis in short-term, as well as for the long-term abnormal return reversal. Furthermore they argue the price impact is stronger for small firms' stocks and stocks mainly traded by retail investors.

The dynamics and magnitude of the predictive power was, more thoroughly, examined by Zhang et al. (2013). They showed that predictive power is declining convexly with increasing lag of the web search, using daily abnormal returns of the Chinese stock market and Baidu web search data. Interesting point to the discussion was brought by Mondria and Wu (2011), who argue that the effect of investor attention on returns depends on whether the company is local from investor's point of view. They showed that increase in abnormal asymmetric attention, a relative web search volume of local versus non-local investors, increases the next-month abnormal returns (stemming from buying pressure from the local investors).

2.3 Google data, investor sentiment and IPOs

A substantial part of my thesis is devoted to the Google data application, as a proxy of investor attention, in the setting of Initial public offerings; or, more precisely, to their application in explaining two IPO stylized fact - the long-term underperformance and the high initial returns also known as the IPO underpricing. Therefore, I present a brief literature review covering these IPO phenomena, especially in connection with the investor attention.

The long-term underperformance (i.e. inferior performance to non-issuing firms) is arguably the most attractive area of IPO academic research. It was, to my knowledge, Stern and Bornstein (1985) who first pointed their fingers on IPO long-term performance, as they showed that issuing firms underperform S&P 500 by 22% in the long-term. The underperformance was confirmed by several studies (see for example Ritter, 1991; Spiess and Affleck-Graves, 1995),[3] most notably by Loughran and Ritter (1995), who called the long-term performance of newly issued stock a puzzle. The existence of the puzzle was questioned by several studies; for example Brav et al. (2000) reported that the underperformance disappears if the benchmarks are matched on firm size and book-to-market ratios. Conversely, Eckbo and Norli (2000) attributed the potential underperformance to a lower risk of IPO stocks, providing evidence that the issuers have a lower leverage ratios and a higher liquidity than the matched firms in years following IPO. After controlling for additional risk of peer companies they could not reject the hypothesis of zero abnormal returns of IPO stocks. Ritter and Welch (2002), in their well-known comprehensive review of IPO related literature, argue that the benchmarking of long-term performance of IPOs is highly sensitive to employed methodology, as well as to the choice of sample period. In addition, they note that despite the similar (unappealing) performance of issuers and their peers with comparable characteristics, the equally weighted post-IPO returns still underperform market indices.

The existence of the second IPO stylized fact, underpricing, is rather indisputable. Ritter and Welch (2002) reported that from 1980 to 2001, the average difference between the offer price and the first day closing price had been 18.8% for US issuers. Furthermore, there had been a positive price change for 70% of issuing firms, while negative initial return had been exhibited only by 14% of the IPOs.

What is unclear about underpricing is why would firms voluntarily leave money on the table. Ritter and Welch (2002) offered wide variety of explanations based on both symmetric and asymmetric information; however, most of them have smaller or bigger shortcomings. The most promising stream of literature struggling to explain the underpricing seems to be focused on behavioral side of investors. Ritter (1991) shed some light on the problematic by pointing out that investors tend to be periodically overoptimistic about the potential of issuing firms; and that the firms take advantage of it by timing the issues so they

[3](Spiess and Affleck-Graves, 1995) performed the analysis for SEOs instead of IPOs

correspond with these "*windows of opportunity*". Loughran and Ritter (1995) provided a support to the hypothesis, as they showed that the first day returns are significantly higher following periods when the market has risen. In line with the investor sentiment theory, it was shown that the underpricing is positively associated with news and non-lead analyst research coverage of the IPOs (Demers and Lewellen, 2003; Aggarwal et al., 2002).

Ljungqvist et al. (2006) and Derrien (2005) offered theoretical models for IPO pricing and initial returns in presence of investor sentiment. Ljungqvist et al. (2006) built the model on the assumption that sentiment investors are budget-constrained and cannot buy the entire IPO. Thus, in order to induce rational investors to participate, firm must set the offer price below the price noise traders are willing to pay. Derrien (2005), on the other hand, stressed out the assumption that "*aftermarket price support is costly for the underwriter*". (Derrien, 2005, p. 490). While the models are different in construction, their predictions are rather similar. They all predict a high underpricing in presence of a high investor sentiment and consequently the poor long-term performance. Derrien (2005, p. 490) aptly noted that it is not the firms who leave the money on table, but rather "*the overoptimistic noise traders who pay excessive prices for IPO shares on the aftermarket*".

The empirical evidence favors these models. Cook et al. (2006) revealed that underwriters promote IPOs in order to induce the sentiment investors into the market for it. It was also reported that the sentiment influences initial pricing and that underwriters do not solely base the valuation on fundamentals and comparable-valuation (Colaco et al., 2013). The most notable empirical validation of the sentiment theories are the higher initial returns of IPOs that exhibited an above average abnormal attention (measured by Google search volume), and subsequent return reversal of such stocks in the long-term (Da et al., 2011).

Finally, it should be noted that the terms initial return and underpricing need not to be necessarily interchangeable. Ma and Tsai (2002) pointed out that under the sentiment theory, the initial return may actually have two parts: true discount and market reaction[4]. A high sentiment is associated with market over-

[4]Ma and Tsai (2002) measure the true discount and market overreaction as $(fundamental\ price - offer\ price)$ and $(first\ day\ closing\ price) - (fundamental\ price)$, respectively.

reaction, whether it may not effect the true discount of IPO.

Chapter 3

Data

This chapter is divided into two parts. The first discusses how different variables used throughout the paper are constructed. The second focuses on sample construction for the two empirical parts of the paper.

3.1 Variable construction

This subsection discusses the nature of Google Trends data and provides an overview of the web search variables construction.

3.1.1 Google search volume

Google Trends data

Google Trends is a public web facility of Google Inc. that provides information on the web search volume for distinct queries realized by the Google search engine users. It is not possible to obtain the absolute search volume for a keyword of interest, since Google publishes an indexed number instead. The number, which is often referred to as *GSV* (= Google search volume; I will follow the convention and use the term *GSV* in this thesis for now on, whenever referring to the "*raw*" Google index value), shows "*...how many searches have been done for the terms you've entered, relative to the total number of searches done on Google over time. This analysis indicates the likelihood of a random user to search for a particular* *search term from a certain location at a certain time*".[1] The data are available on daily, weekly and monthly frequency from 2004 to the present,[2] in form of an online graph or a downloadable *.csv* file.

The process that Google performs to compute *GSV* can be formally expressed as follows. Let's denote $ASV_{keyword}^{t,g}$ the Absolute Search Volume for given *keyword* in time t and geographic region g. *GSV* is obtained from *ASV* in following steps. First, the Absolute Search Volume for the keyword is divided by the total number of searches ASV_{total} , to obtain the Relative Search Volume ($RSV_{keyword}$):

$$RSV_{keyword}^{t,g} = \frac{ASV_{keyword}^{t,g}}{ASV_{total}^{t,g}}. \tag{3.1}$$

Second, *GSV* is obtained by scaling *RSV* in a way that the maximum $RSV_{keyword}^{t,g}$ over time t gets *GSV* value equal to 100:

$$GSV_{keyword}^{t,g} = \frac{RSV_{keyword}^{t,g}}{MAX(RSV_{keywordl}^{t_0,g}, \ldots, RSV_{keyword}^{t_T,g})} \cdot 100, \tag{3.2}$$

where t_0 and t_T represent the lower and upper bound of the specified time interval.

Google also imposes some limitations on the data availability:

1. Only the data of weekly frequency can be downloaded in *.csv* format for the entire available interval. As far as I am concerned, monthly data are only available in online graphic form. The daily data, on the other hand, are only accessible for the intervals between one to three months, irrespective to selected form. It causes a problem as *GSV* in each of the one- to three-month intervals is scaled differently (according to the maximum *GSV* value in the interval), and therefore the *GSV* time series is not continuous if one wants to use the entire sample from 2004. To overcome this problem, I downloaded the daily *GSV* in monthly intervals and weighted the values

[1]https://support.google.com/trends/?hl=en#
[2]Actually, there is a lag of two days til the data are putted online by Google.

by the monthly *GSV* indices available in online form,[3] that is:

$$GSV^d_{adjusted} = GSV^d_{original} \cdot GSV^m_{original}, \tag{3.3}$$

here d is day in month m, and $GSV^d_{original}$and $GSV^m_{original}$ represent daily and monthly *GSV* values available from Google Trends website, respectively. The resulting $GSV^d_{adjusted}$ series has daily frequency and one maximum value equal to 100, to which all other $GSV^d_{adjusted}$ values are scaled. All calculations from now on are done using $GSV^d_{adjusted}$, if not stated otherwise.

2. The data on all frequencies are available only if certain threshold is surpassed, i.e. if $ASV^{t,g}_{keyword} > T_i;\ i \in \{d, w, m\}$, where *d, w and m* represent daily, weekly and monthly frequency, respectively; it also holds that $T_d > T_w > T_m$ as Google Trends may return weekly or monthly data if daily data are requested (or monthly data if weekly data are requested). Nevertheless, the exact values of T_i are not released by Google.

Search variable construction

Most researchers dealing with Google Trends data usually prefer to use the adjusted search volume indices rather than the raw *GSV*. It is a convenient approach as it allows to deal with non-stationarity issues as well as with different "normal" levels of search volumes for distinct queries. Da et al. (2011), for example, use Abnormal Search Volume Index (ASVI):

$$ASVI^k_{i,t} = \ln(GSV_{i,t}) - \ln\left(\text{median}(GSV_{i,t-1}, \ldots, GSV_{i,t-k})\right), \tag{3.4}$$

where $i = 1, \ldots N$ represents firms in the sample and *k* integer defines the length of the time window, over which the "normal" level of attention is depicted. The authors use $k = 8$, however, the selection criteria for *k* may differ according to researcher's preferences and distinct data frequencies.

I decided to construct several different measures of search volume and compare their behavior with respect to the financial variables. Firstly, I followed Da

[3] One might also consider averaging the weekly values in each month, to get the estimates of the monthly *GSV*, as it requires far less work.

et al. (2011) and constructed $ASVI^k$ for $k = 1, \ldots, 7$. Since Da et al. (2011) used weekly data, they did not face the problem of non-trading days. Nevertheless, if one uses daily data instead, she[4] necessarily has to decide whether to include non-trading day *GSV*s to the analysis, or not; since financial data are only available for trading days while *GSV* is available for weekends and holidays as well. I overcame this issue by computing $ASVI^k$ with non-trading days both included and excluded from the list of *GSV*s, denoted as $ASVI^{k,inc}$ and $ASVI^{k,exc}$, respectively.

In addition, one might also question the seasonality of weekly data. Therefore, I also computed *ASVI* measured against the median of *GSV*s from the same days in previous weeks, $ASVI^{k,week}$, for $k = 1, \ldots, 4$:

$$ASVI_{i,t}^{k,week} = \ln(GSV_{i,t}) - \ln(\text{median}(GSV_{i,t-1week}, \ldots, GSV_{i,t-kweeks}). \quad (3.5)$$

Lastly, I consider whole new variable for web search that counts with the possibility that investor attention might be spread between more than one day. Therefore, I created five-day weighted average of *GSV* in following way:

$$GSV_{i,t}^{weighted} = \frac{\sum_{j=0}^{4}(5-j) \cdot GSV_{i,t-j}}{15} \quad (3.6)$$

Afterward, I calculated the abnormal values of $GSV_{i,t}^{weighted}$:

$$ASVI_{i,t}^{k,weighted} = \ln(GSV_{i,t}^{weighted}) - \ln(\text{median}(GSV_{i,t-1}^{weighted}, \ldots, GSV_{i,t-k}^{weighted}), \quad (3.7)$$

for $k = 1, \ldots, 7$.

In addition, it should be noted that some *GSV* values for given date and keyword are not available due to not reaching the limit value. Several cures to the problem might be considered. Fist, one may insert a zero value instead, but I see this procedure as incorrect, since the missing values only suggest that the search volume lies in the interval $\langle 0, T_d)$, not that it truly equals zero. Thus, zero is a downward biased estimate of the missing value. Alternatively, a random

[4]If the gender of an individual referred to in a sentence is unknown, "she" would be used as the generic pronoun to provide respect to the opposite gender.

number from $(0, T_d)$ interval can be plugged in, but it would introduce a synthetic volatility and trend to the *GSV* time series that may be different from the actual volatility and trend in it.[5] Third option is to drop the missing values, which only results in smaller sample size, but does not bring synthetic information to the data. The only shortcoming is that it selectively drops the lowest values of *GSV* from the sample, thus, certain information might be missed this way.

In Chapter 5, if I refer to ASVI without further specification , I always mean $ASVI^{2,exc}$ since it is the overall best performing indicator. Missing values of *GSV* are dealt with by the third option, that is, they are excluded from the sample. In Chapter 6, I use $ASVI_{i,t-i}^{26-i}$ as the measure of attention, where *i* is the number of days to IPO.[6] Conversely to Chapter 5, I do not drop missing values of *GSV* from the sample, but I fill in a number from an interval $(0, T_d)$ instead. The objection against this procedure raised in previous paragraph is not valid in the cross-sectional setting, as I am interested in the actual size of *GSV* in one point of time rather than in the relative size of various *GSVs* in different points of time.

3.1.2 Financial variables

In this subsection, I present a construction of financial variables used in the thesis.

Trading volume

First, I would like to assess whether web searches influence the trading activity. I follow Chordia et al. (2001), Chordia et al. (2007) and Bank et al. (2011), among others, who used very obvious measure of the trading activity - traded volume of firm *i*'s stock in day t in US dollars, or more precisely its natural logarithm:

$$VOLUME_{i,t} = \ln TV_{i,t} = \ln(VO_{i,t} \cdot P_{i,t}), \tag{3.8}$$

where VO is the number of shares traded and P the respective price.

[5] Imagine entering three follow-up values such $T_d > GSV_t^{synthetic} > GSV_{t+1}^{synthetic} > GSV_{t+2}^{synthetic} \geq 0$, while the actual values were $T_d > GSV_{t+2}^{actual} > GSV_{t+1}^{actual} > GSV_t^{actual} \geq 0$.

[6] The different size of the window against which the abnormality of attention is measure is caused by the data availability, as I was able to obtain GSV only up to 26 days prior IPO. Nevertheless, it should not influence the results in any way.

Stock price volatility

Second, I would like to examine how web search sets prices in motion, i.e. the price volatility. The most natural measure of volatility would have been its realized values. Nevertheless, it would require high frequency data and demanding modeling, which is both beyond the scope of this thesis. Instead, I decided to use price range estimator of volatility proposed by Garman and Klass (1980), which provides reasonable trade-off between efficiency/precision and estimation complexity (Shu and Zhang, 2006; Chou et al., 2010). The major shortcoming is its inability to capture the overnight shocks; the bias, however, decreases with number of transactions (Garman and Klass, 1980) and thus it is not a significant problem for the data set of highly traded stocks. Formally, the estimator is defined:

$$\begin{aligned}\hat{\sigma}^2_{GK} = {} & 0.511[\ln(P_t^{High}/P_t^{Low})]^2 - 0.19\{\ln(P_t^{Close}/P_t^{Open})[\ln(P_t^{High}) \\ & + \ln(P_t^{Low}) - 2\ln(P_t^{Open}) - 2[\ln(P_t^{High}/P_t^{Open})\ln(P_t^{Low})/P_t^{Open})]\} \\ & - 0.383[\ln(P_t^{Close}/P_t^{Open})]^2, \end{aligned} \tag{3.9}$$

where P_t^{High} and P_t^{Low} are daily maximum and minimum prices, and P_t^{Open} and P_t^{Close} are daily opening and closing price. As mentioned in Garman and Klass (1980), their estimator can be presented practically as:

$$\hat{\sigma}^2_{GK_t} = 0.5\left[\ln(P_t^{High}/P_t^{Low}\right]^2 - [2\ln(2) - 1]\left[\ln(P_t^{Close}/P_t^{Open})\right]^2. \tag{3.10}$$

I take the natural logarithm[7] of Garman-Klass estimator as the measure of volatility throughout the thesis:

$$VOLATILITY_{i,t} = \ln(\hat{\sigma}^2_{GK_{i,t}}). \tag{3.11}$$

[7]For standard deviation and variance, the logarithmic specification differs in scale only. Thus, the results are not affected, in terms of significance, by the application of logarithmic transformation. For the sake of convenience, I use the term *"volatility"* for the logarithmic transformation of Garman-Klass volatility for the remainder of the thesis.

Daily returns

Lastly, I am keen on examining the response of returns to changes in web search volume. Most of the researches use abnormal returns (van Themaat, 2012; Da et al., 2011), nonetheless, it should be noted that such approach is unequivocally linked with a need for correct specification of underlying model for returns.[8] To avoid the problem, I use *normal* returns (as for example Preis et al. (2010)), in logarithmic form, instead of their abnormal counterpart:

$$RETURNS_{i,t} = r_{i,t} = \ln(P_{i,t}^{adjClose}) - \ln(P_{i,t-1}^{adjClose}), \tag{3.12}$$

where $P_t^{adjClose}$ and $P_{t-1}^{adjClose}$ are adjusted close prices provided by Yahoo! Finance.

3.1.3 IPO variables

For the case study chapter on IPOs, I need to define the initial and long-term cumulative returns. Firstly, I define (logarithmic) initial return (which I refer to as *"initial return"*, *"first day return"* or *"IR"* for now on for the sake of convenience) as:

$$IR_i = \ln(P_{i,t}^{Close}) - \ln(P_i^{Offer}), \tag{3.13}$$

where P_t^{Close} and P^{Offer} refer to the closing price on the first day of trading (t) and the offering price, respectively. The long-term cumulative log return is defined as:

$$CLR_i = \ln(P_{i,t+k}^{Close}) - \ln(P_{i,t}^{Close}), \tag{3.14}$$

where t either refers to closing price on the first day of trading or the closing price one month after IPO; and k is either 91, 183 or 366 days. The two starting dates are considered to control for potential immediate drop in price after the first day of trading.

[8] To be correct, I note that this also applies to the volatility measure as specified in this thesis.

3.2 Sample construction

This section describes the sample construction process. Two different data sets are used in this thesis: the first one, which constitutes of Dow Jones companies, is employed in Section 5; the second one is a sample of IPO firms, which is used in Section 6.

3.2.1 Dow Jones Industrial Average data set

The data set consists predominantly of financial and web search data from years 2004 to 2013. I use companies that were constituents of Dow Jones Industrial Average (DJIA) index continuously between 2004 and 2013; which totals 19 stocks. Dow Jones companies are convenient to use for few reasons. Firstly, the index encompasses a very limited number of firms.[9] Secondly, DJIA is the most well-known and most frequently used indexes in the world and represents about one quarter of the total U.S. equity market capitalization.[10] Third, the well known constituents ensures that firm-related web search is available at Google Trends.

My next empirical choice concerns the identification of a stock in Google. Ramos et al. (2013) list three options, what may be considered as the *keyword*: a researcher can either (1) look for complete name of the firm ("E. I. du Pont de Nemours and Company", for example), (2) like Da et al. (2011), use the ticker of the stock ("DD" or "NYSE: DD"), or (3) like Bank et al. (2011) and Vlastakis and Markellos (2012), consider a unique search term that is unequivocally associated with the firm ("DuPont", or firm's name without the legal label). I am not aware of any paper using the first option and the provided example demonstrates why; hardly anyone uses the complete name of the firm, which is confirmed by the unavailability of Google Trends data for several of the 19

[9] The reason why I decided for the limited size sample is the Google Trends data availability. Since I use daily search volumes, which are accessible only in 93 day intervals (see Section 3.1.1 for details on data availability), it requires enormous number of downloads. To make the downloading task manageable, I use a web crawling program that downloads results of Google Trends queries in form of *.csv* file (for which I kindly thanks to my friend and IT-specialist, Petr Beňas). The use of API does not, however, assure the mass extraction of web search data as Google also imposes restrictions over the number of searches that can be performed from the same IP address and user account. Thus, I use only small sample of firms.

[10] http://www.investopedia.com/articles/analyst/102501.asp

stocks (see Section 3.1.1 for details on the search volume thresholds). The second option is very appealing, since, as Da et al. (2011) correctly pointed out, Google is not able to distinguish between two types of search; the term "Home Depot" might be entered by someone who looks for investment opportunity, but it is far more likely that it is entered by a person who simply wants to buy a new lawn mower. Despite the appealing attributes, it is inappropriate for the DJIA sample, since some of the tickers have generic meaning and the corresponding *GSV* is therefore likely to be very noisy and hardly capable of capturing any investor interest. Excluding those would further reduce the already limited sample size. Therefore, I decided to use the third option. Although the search for general company related term does include some irrelevant component, as Vlastakis and Markellos (2012) wisely point out, this component is probably either random noise or purely deterministic. Moreover, Bordino et al. (2012) showed that the search volume time series for a ticker and a variation to company's name are highly correlated. The third option, in addition, has some advantages over the other measures. Vlastakis and Markellos (2012, p. 1811) stated that it is a broad measure that captures "*information demand by investors which is related to the firm in general rather than only to the stock*" and Bank et al. (2011, p. 240) appended that it allows to receive information from a *"much broader, and potentially relevant audience,"* as ISIN, WKN or other tickers are only used by the professional market participants (Fink and Johann, 2013).

To specify the exact search term for each stock I follow similar procedure to Vlastakis and Markellos (2012). I inserted the full company name, and all the variations known to me to Google Trends and searched for the keywords with the highest search volume. Afterward, I entered the most widely used terms to Google and Wordtracker,[11] and identified additional variations that I unintentionally omitted in the first step. The last step consisted of identification of words with generic meanings. If no generic meaning was found, I used the search term with the highest search volume; in opposite case, the second most popular option is employed. Table A.1 lists the companies in DJIA sample along with the corresponding stock tickers, search queries[12] and number of available

[11] http://www.wordtracker.com/

[12] I will refer to those term as to "name" for now on, even if it is not exactly correct.

GSV values.[13]

In addition to Google Trends data, I employ time-varying aggregate market sentiment in several regression, for which I use sentiment measure developed by Baker and Wurgler (2006). The sentiment data are available both on yearly and monthly bases, in form of levels and changes; and are orthogonalized with respect to a set of macroeconomic conditions. I obtained the data from Jeffrey Wurgler's website.[14]

Finally, the corresponding financial data come from Yahoo! Finance database, namely the daily prices P_t^{High}, P_t^{Low}, P_t^{Open}, P_t^{Close} and $P_t^{adjClose}$ (where $P_t^{adjClose}$ is the close price adjusted for dividends and splits), and the daily trading volume - all values are in US dollars.

Table A.2 lists and describes all variables used in the computational sections for DJIA data-set.

3.2.2 IPO data set

I use the firm database of emerging growth IPOs (Kenney and Patton, 2013),[15] for the identification of firms going public between years 2004 and 2010. It contains a complete list of emerging growth firms going public at the US exchanges from 1990 to 2010; the database also contains various variables that pertain either to the firms going public or the offerings themselves. The complete list of variables can be found in respective guide written by the authors.[16] The database excludes following types of firms and filings from the Thomson Financial Venture Expert, SDC data and other comprehensive lists of IPOs: mutual funds, real estate investment trusts (REITs), asset acquisition or blank check companies, foreign F-1 filers, and all spin-offs and other firms that are not true emerging growth firms.[17]

I use all the companies included in Kenney-Patton database that went public between years 2004 and 2010, with the exception of the unit offerings and one firm that went public on Over-the-Counter Market; totally, it encompasses 547

[13]It should be noted that Vlastakis and Markellos (2012)also used DJIA companies for their analysis, and the search query they used largely corresponds to the one in Table A.1.

[14]http://pages.stern.nyu.edu/~jwurgler

[15]I would like to thank a lot to Mr. Patton that he provided me an access to the database.

[16]http://hcd.ucdavis.edu/faculty/webpages/kenney/misc/Firm_IPO_Database_Guide.pdf

[17]Similar types of firms were excluded also by Da et al. (2011) in their analysis.

companies. The exact same procedure, as described in Section 3.2.1, was used for the identification of search queries.[18] When the identification was completed, I downloaded the daily search volumes for all firms, for an interval starting three months before IPO and ending on the first day of trading.[19] Out of the 547 companies, the daily data were available only for 75 of them; which reveals arguably the biggest shortcoming of Google Trends data - their seldom availability at the daily frequency for infrequently used keywords. It should be, however, noted that the data availability increases in time; only 5% search queries are available for 2004 IPOs in the list, while for 2010, the availability reaches 23%. Therefore, if also IPOs from 2011 to 2013 were included, the data set would have been most probably far larger. Nevertheless, the construction of a comprehensive IPO database between years 2011 and 2013 is beyond the scope of this thesis.[20] See Table A.3 for more details on data availability.

The Kenney and Patton (2013) database, unfortunately, does not contain data on post-IPO performance. Therefore, the financial data on the first day closing prices came from *SCOOP Track Record from 2000 to Present* IPO database,[21] and were controlled against data from Yahoo! Finance, Google Finance, NASDAQ web site database and IPO news coverage. The reason why I did not use solely Yahoo! Finance data, as I did for the DJIA stocks, was the often missing information on stock prices in the first days of trading in the database.

For the long-term performance, the data availability is also poor, as some of the companies were already acquired, merged or delisted; and therefore do not anymore appear in the freely available databases of financial data. Thus, I used Quantshare Trading Software,[22] or more specifically the *Historical EOD data Downloader for Delisted/Bankrupt Stocks* plug-in,[23] to download the stock prices for such stocks. Unfortunately, Quantshare often returns wrong results, so I only used data from Quantshare that matched information available from

[18]The complete list of search terms is available from the author upon request .

[19]For some companies, data for the entire interval were not avaiable. Nevertheless, I included all companies for which at least 75% of *GSV* in the interval $\langle t-26, t\rangle$ was available.

[20]Even though there are IPO databases, such as the previously mentioned SDC, one would still have to go thorough the companies one by one and exclude the types of firms not listed in 2004 to 2010 database.

[21]Available at https://www.iposcoop.com/index.php?option=com_trackrecord&Itemid=200.

[22]Available at http://www.quantshare.com/.

[23]Available at http://www.quantshare.com/item-1270-historical-eod-data-downloader-for-delisted-bankrupt-stocks.

other sources.[24] The final IPO data set contains search volumes and stock prices for 75 firms, albeit long-term cumulative returns are included only for 62 firms.

Table A.4 lists and describes all variables used in the computational sections for IPO data set.

[24] I use data from SCOOP Track Record database, Yahoo! Finance, Google Finance, NASDAQ web site and news coverage for comparison; data available from those sources were usually incomplete so it was not possible to use them directly.

Chapter 4

Methodology

This chapter provides a preview of methodology used in Sections 5 and 6.

4.1 Dow Jones Industrial Average data set

4.1.1 Correlation

Firstly, to examine the existence of the interdependence between the web search data and financial variables, I follow Bordino et al. (2012) and Preis et al. (2010), and use Pearson's cross correlation coefficient ρ between two time series Q_t (search variable) and T_t (financial variable), for lag of $\triangle t = 0$:

$$\hat{\rho} = \frac{\sum_{t=1}^{n}(Q_{t,n} - \overline{Q}_n)(T_{t,m} - \overline{T}_m)}{\sqrt{\sum_{t=1}^{n}(Q_{t,n} - \overline{Q}_n)^2}\sqrt{\sum_{t=1}^{n}(T_{t,m} - \overline{T}_m)}}, \quad (4.1)$$

where $\overline{Q}$ and $\overline{T}$ are the sample averages of the two time series, and $m, n \in \{1, \ldots, 19\}$ refer to stocks in DJIA sample. Firstly I assess the interdependence for $m = n$, that is, how firm's financial data are interdependent with the search volume for its name. Afterward I perform robustness check by computing the correlations for $m \neq n$, i.e., how firm's financial data are interdependent with search volume for the name of other companies in the sample; and compare the initial correlation with median correlation for reshuffled data.

4.1.2 Time series analysis

Stationarity issues

I proceed with regression analysis in the time series setting. Firstly, since most econometric models that I employ require the time series to be stationary, I employ two test for (non-)stationarity. At first, I perform the augmented Dickey–Fuller (ADF) test (Dickey and Fuller, 1979) that a variable follows a unit-root process. The true model of the test is assumed to be:

$$y_t = \alpha + y_{t-1} + u_t, \tag{4.2}$$

where u_t is independently and identically distributed zero-mean error term. The Dickey-Fuller test involves fitting:

$$y_t = \alpha + \rho y_{t-1} + \delta t + u_t \tag{4.3}$$

by OLS and testing $H_0 : \rho = 1$, i.e., y_t follows a unit root process, against the alternative $H_A : \rho < 1$, with a regression restriction $\delta = 0$. Since the above specified model is likely to be plagued by serial correlation, the augmented version of the test fits:

$$\Delta y_t = \alpha + \beta y_{t-1} + \delta t + \varsigma_1 \Delta y_{t-1} + \cdots + \varsigma_k \Delta y_{t-k} + \epsilon_t \tag{4.4}$$

and tests whether $H_0 : \beta = 0$ against the alternative $H_a : \beta < 0$. Table B.1 presents the results for *ASVI, VOLUME, VOLATILITY* and *RETURNS*. Results are discussed in empirical chapter.

If the ADF test rejects unit root, I also employ Robinson's log-periodogram regression estimator (Robinson, 1995) for the long memory diagnostics in a time series. As suggested by Murphy and Izzeldin (2009), Robinson's test is the best performing long memory test for samples with $T \geq 250$, out of 6 commonly used tests (the 5 alternatives are Lo's modified rescaled range or R/S statistic (Lo, 1989), the KPSS statistic (Kwiatkowski et al., 1992), the rescaled variance or V/S statistic (Giraitis et al., 2003), the GPH statistic (Geweke and Porter-Hudak, 1983), and, the $\hat{s}_k$ statistic (Harris et al., 2008). Robinson (1995) proposed a procedure to obtain the estimate of memory parameter *d* of a fractionally integrated process x_t in a model:

$$(1-L)^d x_t = u_t, \tag{4.5}$$

where $d > 0$, u_t is covariance stationary process with zero mean and spectral density function that is both positive and finite at any frequency. Resulting value of $d = 0$ suggests the series is $I(0)$; $d \in (0, 0.5)$ means the series has long memory, but is still covariance stationary; in case of$d \in \langle 0.5, 1)$, the series is no longer covariance stationary, but it is still mean reverting in long term; finally if $d \geq 1$ the series is nonstationary and non-mean-reverting.

The Robinson's estimator is implicitly defined by:

$$\hat{d} = arg\ min_d\ (\ln \overline{C(d)} - 2d\frac{1}{m}\sum_{s=1}^{m} \ln \lambda_s \tag{4.6}$$

$$\overline{C(d)} = \frac{1}{m}\sum_{s=1}^{m} I(\lambda_s)\lambda_s^{2d},\ \lambda_s = \frac{2\pi s}{T},\ \frac{m}{T} \to 0, \tag{4.7}$$

where $I(\lambda_s)$ is the periodogram of the raw time series x_t, given by:

$$I(\lambda_s) = \frac{1}{2\pi T}\left|\sum_{t=1}^{T} x_t e^{i\lambda_s t}\right|^2, \tag{4.8}$$

and $d \in (-0.5, 0.5)$. Under certain mild conditions and finiteness of the fourth moment, Robinson (1995) proved that:

$$\sqrt{m}(\hat{d} - d_0) \to_d\ N(0, \frac{1}{4})\ as\ T \to \infty, \tag{4.9}$$

where d_0 is the true value of *d* (Caporale and Gil-Alana, 2010). A choice must be made on the number of ordinates entering the log-periodogram regression, $nord = N^p$. I follow Robinson (1995) and choose power *p* equal to 0.9.

Table B.2 present the results of Robinson's tests for *ASVI, VOLUME, VOLATILITY* and *RETURNS*.

OLS regression analysis

None of the series exhibit nonstationarity, but the trading volume and volatility clearly has long memory. Arguably, the most correct procedure to model the long memory processes and their relationship with web search would be via ARFIMA(X) models (Watanabe and Ubukata, 2009; Degiannakis, 2008); however, ARFIMA(X) models are rather inconvenient for interpretation. Therefore, I follow the practice employed in most papers dealing with the web search and financial data (for example Da et al., 2011; Vlastakis and Markellos, 2012; Andrei and Hasler, 2011) and use the Ordinary Least Square (OLS) regression with lagged dependent variable (LDV), with Newey-West heteroskedasticity and autocorrelation consistent standard errors (HAC) (Newey and West, 1986). The inclusion of LDV should filter out most of the residual autocorrelation and the Newey-West procedure ensures that the reported standard errors are robust to the leftover (if present) correlation in error terms. For the validity of the employment of LDV with HAC errors see Wooldridge (2009, p. 430 - 431).

Formally, the model is specified as follows:

$$y = X\beta + \epsilon, \tag{4.10}$$

where X is a set containing constant term, lags of dependent variable and independent variables. Newey-West calculates the estimates:

$$\begin{aligned} \hat{\beta}_{OLS} =& (X'X)^{-1}X'y \\ \widehat{Var}(\hat{\beta}_{OLS}) =& (X'X)^{-1}X'\hat{\Omega}X(X'X)^{-1}, \end{aligned} \tag{4.11}$$

that is the coefficients are simply those of standard OLS linear regression. For $lag(0)$, the Newey-West variance estimate equals White formulation (White, 1980):

$$X'\hat{\Omega}X = X'\hat{\Omega}_0X = \frac{n}{n-k}\sum_i \hat{e}_i^2 x_i' x_i, \tag{4.12}$$

where $\hat{e}_i = y_i - x_i\hat{\beta}_{OLS}$ with x_i being the ith row of the X matrix, n is the number of observations and k is the number of regressors in the model 4.10 including constant term and LDV. For $lag(m),\ m > 0$, the Newey-West variance estimates extends to the following formulation:

$$X'\hat{\Omega}X = X'\hat{\Omega}_0X = X'\hat{\Omega}_0X + \frac{n}{n-k}\sum_{l=1}^{m}(1-\frac{l}{m+1})\sum_{t=l+1}^{n}\hat{e}_t\hat{e}_{t-1}(x_t'x_{t-l}+x_{t-l}'x_t), \tag{4.13}$$

where x_t is a row of matrix X observed at time *t*. For the lag specification I use the automatic bandwidth selection procedure of Newey and West (1994).

Vector auto-regressive model

To assess the dynamics of the relationships I use bi-variate VAR(p) model:

$$y_t = \sum_{i=1}^{p} A_i y_{t-p} + \varepsilon_t, \tag{4.14}$$

where $A_i = \begin{pmatrix} a_{11}^i & a_{12}^i \\ a_{21}^i & a_{22}^i \end{pmatrix}$ is 2×2 matrix. A choice must be made on the number of lags, *p*, included in the model. I use Lütkepohl's version of SBIC (Schwarz et al., 1978) criterion (Lütkepohl, 2007). For the sake of convenience, I use the same number of lags for all stocks, based on the mode of lags recommended by criterion. To test the significance of lagged interdependence I use Grange-causality test (Granger, 1969), that is I test whether $a_{12}^1 = a_{12}^2 = \ldots = a_{12}^p = 0$ (or whether $a_{21}^1 = a_{21}^2 = \ldots = a_{21}^p = 0$ for opposite direction Granger-causality) by Wald test.

4.1.3 Panel analysis

Fixed effects with lagged dependent variable

To test the overall performance of the web search in predicting financial variables, I employ panel data analysis on top of the one-by-one time series approach. Since the data dimension is 19×2516, that is $T >> N$, the data are not a *typical* panel but rather comparable time series data observed on a variety of units, which is often referred to as the Time-series-cross-section (TSCS) data.[1]

[1] In following chapters, I use the term *"panel"* when referring to time-series-cross-section for the sake of brevity.

Since the time series analysis reveals differences between the stocks in terms of their financial data interdependence with *GSV*, pooled regression does not seem to be proper option. To deal with heterogeneity or individual effects that may or may not be observed, many panel data studies use fixed or random effects models. The random effects (RE) model assumes that individual effect (heterogeneity) is not correlated with the independent variables, which is hardly a case for my data (I tested for the inconsistency of random effects estimator by Hausman specification test (Hausman, 1978), and it confirmed the initial conjecture); so the fixed effects (FE) model seems to be better option for the DJIA dataset.

Nevertheless, as some of the series exhibit long memory (see paragraph about stationarity in 4.1.2), one may want to include LDV to the regression to account for the long term dynamics of dependent variable and to deal with the arising autocorrelation:

$$y_{i,t} = \phi y_{i,t-1} + \alpha_i + \varepsilon_{i,t}. \tag{4.15}$$

The inclusion of LDV to the fixed effect model (FE-LDV), as demonstrated by Nickell (1981), does not come with zero cost; the fixed effects model with autoregressive term results in biased coefficients if estimated by OLS. Nickell (1981) derived the asymptotic bias (as $N \to \infty$) and showed that it equals $O(T^{-1})$.

Many cures to the problem have been proposed. Perhaps the most common are the Dynamic Panel Data estimators, such as Anderson and Hsiao (1982) (AH), or the alternative estimator of Arellano and Bond (1991). Nevertheless, Beck and Katz (2011) reviewed the dynamics-modeling procedures in *TSCS* setting and showed that the AH estimator should not be used for the *TSCS* data; while it is clearly unbiased, the authors argue that the AH estimator pays a high cost, in terms of sampling variability, to get the unbiasedness.

Alternatively, one might address the problem by estimating the FE-LDV model bias and by using it for correcting the estimate, as proposed by Kiviet (1995). The author derives a formula for the bias of the FE-LDV model which has an $O(N^{-1}T^{-3/2})$ approximation error. Yet, neither the Kiviet (1995) procedure comes without shortcomings. Beck and Katz (2011) highlighted four major drawbacks of the procedure. First, the calculations needed to estimate the bias are complex. Second, the formula requires knowledge of the true parameters in equation 4.15. Third, the approximation assumes the *TSCS* data to be balanced (which is not a case for the DJIA dataset as some *GSV* values are missing). Finally, there is no direct way to calculate standard errors using this correction.

Interestingly, the authors also showed that for large *T* (as the bias is decreasing in *T*), FE-LDV model outperforms the AH instrumental variable procedure and its estimates are very close to those of Kiviet correction estimation. Thus, the authors claim that they "*see little reason, in general, not to prefer FE-LDV over the Kiviet estimator when T is twenty or more" (Beck and Katz, 2011, p. 19)*. Kristensen and Wawro (2003) came to similar conclusion by conducting Monte-Carlo simulations on *TSCS* data noting that in case of correlation between unit effects and explanatory variables, FE-LDV model with robust standard errors performs reasonably well (and outperforms OLS with panel corrected standard errors method, which they used as a benchmark). Therefore, I use the fixed effect model with lagged dependent variable; with standard errors clustered by firm to account for residual intra-group correlation that is not filtered out by the inclusion of lags. One might object that clustered standard errors are downward biased in case of small *N*, but Rogers (1994) notes that clustered standard errors with circa 20 equal-sized clusters would only suffer from a very small bias.[2] Formally, the model can be specified as follows (for a case with one independent variable):

$$y_{i,t} = \phi_1 y_{i,t-1} + \phi_2 y_{i,t-2} + \ldots + \phi_k y_{i,t-k} + \phi_{k+1} x_{t-m} + \alpha_i + \varepsilon_{i,t}, \tag{4.16}$$

where number of lags, *k,* is selected based Lütkepohl's version of SBIC and $m \in \{0, ..., T\}$ specifies the lag of independent variable x. The clustered variance estimator that is used to calculate standard errors in 4.16 is defined as:

$$\widehat{Var}(\hat{\Phi}_{Clustered}) = (X'X)^{-1} \sum_{g=1}^{G} X_g' \sqrt{c}\hat{u}_g \sqrt{c}\hat{u}_g' X_g (X'X)^{-1}, \tag{4.17}$$

where *X* is the set of regressors in 4.16, $\hat{u}_g = y_g - X_g\hat{\Phi}$ is the vector of OLS residuals for the *g*th cluster, and $c = \frac{G}{G-1}\frac{N-1}{N-K} \simeq \frac{G}{G-1}$ is a finite sample modification used by Stata to reduce the downwards bias of $\widehat{Var}(\hat{\Phi}_{Clustered})$ (Cameron and

[2]To test the correct specification of standard errors, I ran bootstrap simulation with 1000 replications and the clustered errors did not differ by more than 3% from the bootstrapped one; whereas without clustering, the reported standard errors were ca. 4.3x and 2.1x lower than the bootstrapped counterpart for volume and volatility respectively.

Miller, 2013).

Since the model 4.16 requires stationary variables, I perform Fisher-type tests implemented in Stata, which combine the *p*-values from the panel-specific ADF unit-root tests (see equation 4.4) using four methods proposed by Choi (2001). Three of the methods differ in whether they use the inverse χ^2 , inverse-normal, or inverse-logit transformation of *p*-values, and the fourth is a modification of the inverse χ^2 transformation that is suitable when N tends to infinity (StataCorp, 2013). See Choi (2001, p. 253) for further details on the test statistics.

Panel vector auto-regressive model

Lastly, to further model the dynamics of the systems in panel setting, I employ the panel-data vector auto-regressive (PVAR) method;[3] which combines the VAR approach taking all the variables in the system as endogenous with panel-data approach that allow for unobserved individual heterogeneity. The PVAR model is specified as follows:

$$y_{i,t} = \mu_0 + \alpha_i + \sum_{i=1}^{p} A_i y_{i,t-p} + \lambda_t + \varepsilon_{i,t},\ i = 1, \ldots, N; t = 1, \ldots, T, \tag{4.18}$$

where $y_{i,t}$ is two-variable vector; A_i are2×2 matrices of estimable coefficients, α_i are firm specific fixed effects, λ_t denotes time-effects, and $\varepsilon_{i,t}$ is 2×1 vector of well behaved disturbances.

As Drakos and Konstantinou (2011) note, the panel data setting imposes restriction on the underlying structure of each cross-sectional unit, i.e., on the equality of the coefficients in A_i for all firms. The natural solution is to introduce the firm-specific fixed effects, α_i, to account for the individual heterogeneity. I already showed, however, that the fixed effects are correlated with regressors in case of inclusion of the LDV; hence, the commonly employed mean differencing procedure to remove the fixed effects would create biased coefficients (Nickell, 1981; Arellano and Bond, 1991). I follow the procedure recommended by Love and Zicchino (2006)and use the forward orthogonal deviations instead

[3] I used the Stata plug-in written by Ryan Decker from University of Maryland, which revises the original program created by Mrs. Inessa Love from World Bank, that was firstly used in Love and Zicchino (2006). Therefore, I would like to thank both to Mr. Decker and Mrs. Love for making their PVAR programs for Stata publicly available.

(Arellano and Bover, 1995). This procedure, also referred to as the *Helmert transformation*, removes only the forward mean, i.e., the mean of all the future observations available for each firm-period. This transformation preserves the orthogonality between transformed variables and lagged regressors, so it allows to use the lagged regressors as instruments and estimate the coefficients by system the generalized method of moments.

Once the unknown PVAR parameters are estimated, the program allows to run dynamic simulations, namely the impulse response functions (IRF) and variance decomposition analysis (VD). The IRFs and VDs allow to examine impacts of innovations or shocks in one variable to the other variable(s) in the system. The IRFs measure the dynamics of the response (the coefficients are the average effects of IRFs and permit recognizing the significance of the overall response), and VD gives information on how much variation in one variable is due to the innovation in the other. The responses correspond to one-time shocks holding the shocks to all other variables zero, i.e. the impulse response is orthogonalized. To obtain orthogonalized impulse response functions, one must decompose the residuals in a way that makes them orthogonal. However, such orthogonalization is not unique and depends on the ordering of the variables in the VAR. The common way to deal with the problem is to choose some causal ordering, i.e. *arbitrarily* select the more exogenous variable; and then apply the Choleski decomposition on residual variance-covariance matrix. I follow this procedure, and allocate any correlation between the residuals to the variable that appears earlier in the ordering, which is always the web search variable in my case.

4.2 IPO data set

The IPO regressions are all estimated by cross-sectional OLS. I perform widely applied methodology to test for OLS assumptions. First, the presence of heteroskedasticity is tested by Breusch-Pagan (Breusch and Pagan, 1979) and White (White, 1980) tests. No severe heteroskedasticity is detected in the sample, however, if any of the tests suggest presence of mild heteroskedasticity, White's heteroskedasticity consistent standard errors are calculated (White, 1980). Second, the existence of multicollinearity is tested by variance inflation factors. Lastly, the normality of residuals is tested by Shapiro-Wilk test (Shapiro and Wilk, 1964).

Regardless the residuals not being normally distributed, the Gauss-Markov Theorem states that the ordinary least squares estimate is still the best linear unbiased estimator (BLUE) of the regression coefficients; yet, the *p*-values associated with the coefficients are unreliable. Therefore, when Shapiro-Wilk test suggest that residuals are non-normally distributed, I use bootstrapping (1000 replications) method to estimate the confidence intervals and *p*-values.

Due to limited sample size, the estimates are prone to be largely affected by outliers and observations with high leverage. The observations exhibiting such behavior are identified, using Stata's dfbeta function (identifies how each coefficient is changed by deleting the observation, I apply criterion $|df\beta| > \frac{2}{\sqrt{n}}$ to identify overly influential observations), Cook's DCook (2000), dfits (Welsch and Kuh, 1977) and visual examination; and then omitted from the sample.

Chapter 5

Empirical results - DJIA data set

This chapter presents and describes all empirical results for DJIA data set. The first section contains a basic analysis of web search and financial data interdependence in one-by-one, time series setting. The second section examines the interdependence from the panel-data perspective and provides deeper insight into the nature of the interdependence. In addition, both the sections are further divided according to the analyzed financial variable. Lastly, I present short discussion on search term specification, data frequency and other issues.

5.1 Time series setting

5.1.1 Trading volume

Correlation

Firstly, I asses how trading volume of a stock correlates with web search for its name (*ASVI*), using Pearson's cross correlation coefficient. First look in the Column (1) in Table C.1 reveals an existence of positive correlation between the two series for a vast majority of the stocks; the median correlation reaches 4.30% and only data of four firms exhibit negative correlation. The correlation is significant for twelve stocks at 5% level.

In order to assess the robustness of the results for the DJIA set of stocks, I construct a reshuffled data set in which the query volume time series of a company C_i is randomly paired to the trading volume time series of another company

C_j, as in Bordino et al. (2012). Medians of all resulting correlation coefficients of each firm's trading volume with *ASVI* of all other firms are presented in Column (2). The robustness check discloses that some correlation is present also for reshuffled data, however, it is more than two-times weaker on average.[1] The remaining correlation among the reshuffled data can be explained by general market trends and business associations among companies; for example, IBM's trading volume strongly correlates with the web search for Intel.

Non-lagged interdependence

Before approaching to the regression analysis, I run ADF test on *VOLUME* and *ASVI* to test for nonstationarity; Table B.1 shows that none of the time series exhibit unit-root for any stock. Robinson's test for fractional integration, however, unveils the long memory nature of *VOLUME*, with $\hat{d} \in \langle 0.33, 0.45 \rangle$.

After stationarity testing, I proceed to the regression estimation itself. I first investigate whether a non-lagged relationship between ASVI and VOLUME exists. It should be noted, however, that most of the related research works with lagged values of *ASVI*. It makes arguably more sense, as in non-lagged setting, one cannot perfectly distinguish which of the two variables drives the interdependence. On the other hand, if only the lagged setting is considered, one might miss the information about the immediate reaction of the trading activity to the changes in investor attention. Therefore, I consider both lagged and non-lagged setting. To conserve space, I focus mostly on the non-lagged interdependence in the time series section; while in the panel-data section, both options are given the same amount of space.[2]

I fit model 4.10, which includes five lags of *VOLUME* to address its long-memory nature, and a web search variable. All variables throughout the thesis are standardized prior regressing, so the regression coefficient for a variable can be interpreted as the effect of a one-standard deviation change in that variable;

[1]The only stocks that show a stronger interdependence for the reshuffled data - if measured in absolute values - are Coca-Cola and Wal-Mart.

[2]This applies to volatility and returns estimations as well.

and extreme values of *ASVI* are excluded.[3] Results are listed in Table C.2.

$$VOLUME_t^{st} = \alpha + \sum_{i=1}^{5} \beta_i VOLUME_{t-i}^{st} + \beta_6 ASVI_t^{st} + \epsilon_t \tag{5.1}$$

The results provide moderately strong support to Conjecture 1, stating that *"An increase in web search volume for firm-related query is associated with an increase in trading volume at the same day"*. In order to find an evidence to the conjecture, the regression coefficients of ASVI from model 5.1 should be significant[4] and positive. It applies to fifteen out nineteen stocks; from the rest of the stocks, two have negative *ASVI* coefficients (WalMart and Home Depot) and two show insignificant *ASVI* coefficients (Coca Cola and Walt Disney). It should be noted, that the insignificance of *ASVI* coefficients, in contrast to the negative significance, does not necessarily contradict Conjecture 1. The reason is that the insignificance may just suggest that the selected search query for given firm is not the one used by investors to search for company related information.

Despite the statistical significance, *ASVI* seems to bring only minor improvement to pure auto-regressive model in terms of goodness of fit. The average (median) difference in $adjR^2$ between model 5.1 and a restricted version with $\beta_6 = 0$, is 0.90% (0.64%). On the other hand, Table C.3 shows that the significance of the interdependence prevails even if one controls for current stock volatility and returns. Thus, the information contained in *ASVI* seem to be different from an information readily available from the other two financial variables.

Time variation

One might be interested whether the impact of investor attention on trading volume differs in time; especially whether periods of increased uncertainty, such was the financial crisis, has any implications for the interdependence of those two variables. Therefore I divided the sample into three periods: pre-crisis, crisis and post-crisis. The exact definition of financial crisis interval is rather arbitrary,

[3] I manually checked the outlying values of ASVI and identified which can be assigned to the imperfections in original *GSV* data. Afterward, I excluded the suspect observations from the regression. The same procedure is applied on all regressions in Chapter 5.

[4] If I refer to significance in the text, I always mean statistical significance at 5% level if not stated otherwise.

so it is up to the researcher which dates he considers. I chose the duration of US recession as the crisis time limitation. The recession, according to NBER, begun December 2007 and ended July 2009;[5] thus I apply the dates 12/1/2007 and 6/30/2009 as the bounds of the crisis period.

To formally test for the potential differences in *ASVI-VOLUME* relationship, I construct *ASVI* slope dummy variables for the three periods and test their equality. I also constructed level dummies to control for the difference in constant term. The estimated model has following form:[6]

$$\begin{aligned} VOLUME_t^{st} = {} & \alpha_1 PRE_t + \alpha_2 CRIS_t + \alpha_3 POST_t + \beta_1 ASVI_t^{PRE,st} \\ & + \beta_2 ASVI_t^{CRI,st} + \beta_3 ASVI_t^{POST,st} \\ & + \sum_{i=1}^{5} \beta_{3+i} VOLUME_{t-i}^{st} + \epsilon_t, \end{aligned} \tag{5.2}$$

where *PRE*, *CRI* and *POST* are dummy variables corresponding to pre-crisis, crisis and post-crisis period, respectively.

The results, presented in Table C.4, disclose that most of the stocks does not show significant interdependence throughout all three periods; suggesting that the significant results from previous regression might have been dragged by certain points of time when the interdependence was the strongest, even though for other periods the interdependence might have been insignificant. There is, however, apparent increase in the average coefficient size, associated with the financial crisis outbreak. In crisis, the trading volume seemed to react to changes in investor attention circa two time stronger; in numbers, the ASVI coefficients (in absolute value) for crisis period are 81.5% (133.3%) higher on average (median), than they were prior to the outbreak. When the crisis subsided, the

[5] http://www.nber.org/cycles/sept2010.html

[6] The regression, as stated, has one advantage against alternative form

$$\begin{aligned} VOLUME_t^{st} = {} & \alpha + \alpha_2 CRIS + \alpha_3 POST + \beta_1 ASVI_t^{st} + \beta_2 ASVI_t^{CRI,st} \\ & + \beta_3 ASVI_t^{POST,st} + \sum_{i=1}^{5} \beta_{4+i} logVolume_{t-i} + \epsilon_t \end{aligned}$$

As it allows to address the significance of ASVI in each period, not only the difference between coefficients.

strength of the relationship between attention and trading volume weakened, yet still exceeded the pre-crisis level. The results provide support to the Conjecture 4, that people rely more heavily on web search for firm related queries, if trading in periods of enhanced uncertainty.

An interesting pattern might be observed in the results, if one considers the number of significant coefficients in each period. Clearly, the significance increases in time. I can think of two possible explanations for it: 1) investors tend to consult their trading decision more often then they used to; 2) data quality improves in time as the number of Google users increases, therefore, the data include more information on the retail investors' search behavior as well. As the average size of coefficients in the three periods does not correspond to the increasing significance in time, the second fact seem to be contributing more to the observed pattern.

Granger-causality

So far, I took into account solely the *nowcasting* ability of web search over the trading volume. Now, I move on and examine whether any forecasting interdependence exists between the variables. Therefore, I run bi-variate VAR model with 5 lags for *ASVI* and *VOLUME*; and test for Granger-causality between them. The results for all nineteen stocks are available in Table C.5.

The results show that for 11 stocks, web search Granger-causes trading volume.[7] The predictive power seem to be predominantly concentrated around the first lag, suggesting that traders are most likely to trade the same day they get involved in web search for a firm-related information, or the day after.

The opposite way Granger-causality is also present for several stock, nonetheless, they are lower in count; Wald test rejects the null of zero trading volume coefficients in *ASVI* equation for eight stocks. Interestingly, the highest number of significant coefficients is present at lag five (eight stocks) and all of them are negative. The possible explanation is that investors execute their trade after involving in search, control the stock performance for few days after trading, and then relax their search activity for some time. In addition, the weaker Granger-

[7] It should be, however, noted that Granger causality tests are sensitive to lag selection and the results might change if one specifies different number of lags.

causality from trading volume to web search activity than in the opposite direction is in line with the findings of Bordino et al. (2012).

5.1.2 Stock price volatility

Correlation

Similarly to the trading volume, I examine how web volume correlates with the stock price volatility. The results, presented in Table C.6, show that the average (median) correlation between the two time series is lower than in the case of trading volume and web search, reaching 1.89% (1.88%); and the correlation is predominately positive but mostly insignificant. Three stocks in the sample exhibit negative correlation, namely Home Depot, Wal-Mart and Walt Disney; the stocks that also has negative correlation between web search and trading volume. The correlation for reshuffled data is circa $1.5\times$ lower than the initial correlation; and only three stocks' volatility show lower correlation with their own *ASVI* than with *ASVI* of other companies.

Non-lagged interdependence

Prior approaching to the regression analysis, I test whether the stationarity condition is met for volatility. None of the volatility time series exhibit unit root as ADF test rejects the null hypothesis for all nineteen stocks. The Robinson's test show modest fractional integration, with $\hat{d} \in \langle 0.24, 0.34 \rangle$, suggesting long memory nature of volatility.

To estimate the non-lagged relationship between the web search volume and price volatility, I fit following model:

$$VOLATILITY_t^{st} = \alpha + \sum_{i=1}^{5} \beta_i VOLATILITY_{t-i} + \beta_6 ASVI_t^{st} + \epsilon_t, \qquad (5.3)$$

which corresponds to model 5.1 in previous subsection.

The results are less convincing in terms of validity of Conjecture 2 that *"An increase in web search volume for firm-related query is associ- ated with an increase in volatility at the same day"*, than they were for Conjecture 1. The validity of Conjecture 2 is questionable, as only nine stocks exhibit significant relation between the web search volume for their name and their stock price volatility (plus 2 more on 10% level). On the other hand, all has positive coefficients.

Logically, as volume and volatility are highly correlated, the significance of coefficients in models 5.1 and 5.3 corresponds highly; all stocks with significant *ASVI*-Volatility interdependence has also significant relationship between *ASVI* and Volume (but not vice versa). It has interesting implications to validity of Conjecture 2. In case of volume, I argued that *"insignificance may just suggest that the selected search query for given firm is not the one used by investors to search for company related information"*; however, this argument cannot be simply applied for *ASVI* coefficients in volatility equations, at least not to all of them. The logic behind this claim is that for eight stocks, the search volume for given query successfully predicts the stocks' trading volume; thus, the query volumes are able to capture the investor decision whether to trade or not, albeit they do not capture how the decision transforms into prices.

The weaker interdependence is confirmed by benchmarking with a restricted model that excludes *ASVI* from 5.3. The abolition of the restriction yields only 0.27% (0.17%) increase in average (median) $adjR^2$. Thus, *ASVI* brings far less information to the auto-regressive model of volatility, than it does to the auto-regressive model of volume.

$$\begin{aligned} VOLATILITY_t^{st} = \alpha + \sum_{i=1}^{5} \beta_i VOLATILITY_{t-i}^{st} + \beta_6 ASVI_t^{st} \\ + \beta_7 VOLUME_t^{st} + \beta_8 RETURN_t^{st} + \epsilon_t \end{aligned} \tag{5.4}$$

Furthermore, I control for the non-lagged trading volume and returns by fitting model 5.4. From the results available in Table C.8, it is apparent that *ASVI* most likely does not possess any significant amount of information about volatility than the one contained in those two variables (with the exception of three stocks, for which *ASVI* retains its significance); yet, all the coefficients - with one exception - are positive.

Time variation

Following the procedure for trading volume, I assess how the web search volume influences over volatility evolves over time, namely I compare the interdependence strength in pre-crisis, crisis and post-crisis periods as defined in Subsection 5.1.1. The results of regression 5.5 are presented in Table C.9.

$$\begin{aligned} VOLATILITY_t^{st} = {} & \alpha_1 PRE_t + \alpha_2 CRIS_t + \alpha_3 POST_t + \beta_1 ASVI_t^{PRE,st} \\ & + \beta_2 ASVI_t^{CRI,st} + \beta_3 ASVI_t^{POST,st} \\ & + \sum_{i=1}^{5} \beta_{3+i} VOLATILITY_{t-i}^{st} + \epsilon_t. \end{aligned} \tag{5.5}$$

It follows from the results that the interdependence between web-search-measured attention and stock price volatility was rather vague in the pre-crisis period. The emergence of the finical crisis boosted the strength of the interdependence, measured in terms of absolute size of the coefficients, circa twofold; yet, the effect of web search on price volatility was still mostly insignificant. Nevertheless, the crisis period was characterized by very high uncertainty. The high variance and overall complexity of the stock volatility in crisis (see for example Manda (2010)) made it really difficult for any measure to predict it with any significance. If I look at the results in conjunction with Andrei and Hasler (2011) model predictions, they seem to be rather contradicting. The authors argue that the low attention periods might exhibit negative interdependence between the attention and stock volatility, while the high attention periods should not. Yet, the results show that the number of stocks exhibiting negative effect of attention on volatility was 66% higher in crisis than in the other two periods; and one can hardly claim that the investor attention in the financial crisis was low. Finally, as the crisis perished, the significance of the relationship between the two variables increased. The rising number of significant *ASVI* coefficients in time supports the findings from Subsection 5.1.1; that is, the Google data quality improved due to increasing popularity of the service among Internet users as well as due to the overall enhancement of Internet popularity among people (including investors).

Granger-causality

The non-lagged interdependence between web search and volatility seems to be weaker than it is for web search and trading volume. In this subsection, I examine whether the fact holds also for lagged interdependence. Therefore, I run bi-variate VAR(5) model for the two variables and test for Granger-causality.

Results are presented in Table C.10.

The first look on the results indicates that for five stocks, the web search for company name Granger-causes the firm's stock price volatility; however, for several other stocks the no Granger-causality either cannot be rejected at 10% level or the rejection at 10% level is very tight. The opposite direction Granger-causality is, similarly to trading volume, weaker; it is significant for three stocks only. The feeble Granger-causality from price volatility to web search activity is, however, quite surprising; especially if one takes into account the rather decent Granger-casulity of trading volume to the web search. In my opinion, it seems more probable for traders to react to the price movements with their search activity, than to the previous trading volume. The predictive power web search possess over volatility is focused around the first lag, while the opposite way relation is rather spread between different lags. Similarly to the VAR model for trading volume, all lag-five coefficients of stock price volatility in the attention equation are negative, which might give modest support to the explanation offered in Section 5.1.1.

5.1.3 Daily returns

Correlation

Table C.11 shows correlation coefficients between each firm's returns and related web query. The correlation, contrary to volume and volatility, differs substantially in both sign and size, leading to average (median) correlation of -0.10% (0.27%). The correlation is significant for three stocks only (Boeing, Merck and Microsoft), for which it differs in sign as well. Therefore, the first look at the data casts doubts about the existence of any significant relation between the two variables.

Non-lagged interdependence

Contrary to *VOLUME* and *VOLATILITY*, *RETURN* time series exhibits short memory properties and is highly stationary. Testing for unit root by ADF test rejects the null for all 19 stocks; and Robinson's test show that returns are either close to I(0) or exactly I(0). Therefore, inclusion of few lags completely solves the residual autocorrelation and kernel-based autocorrelation-consistent error terms are not necessary in this case, albeit heteroskedasticity-consistent transformation is still needed. The number of lags was selected individually for each stock; I chose the minimum lag length that resolved the serial correlation of residuals in each equation. The model is specified as follows:

$$RETURN_t^{st} = \alpha + \sum_{i=1}^{k} \beta_i RETURN_{t-i} + \beta_4 ASVI_t^{st} + \epsilon_t;\ k \in \{1, 2, 3\}, \quad (5.6)$$

and results are presented in Table C.12.

The regression results confirm what the preliminary correlation analysis suggested; there is a scarce evidence of any significant relationship between the web search activity and daily returns. Out of the 19 regression, only 3 show significant ASVI coefficients, which, in addition, differ in sign. This finding is in line with Preis et al. (2010), who also used *normal* stock returns rather than abnormal ones, and found no interdependence for weekly S&P data. As can be seen from Table C.13, the three stocks' *ASVI* retain its significance even if one controls for trading volume and stock price volatility. All in all, the first look on the relationship between daily returns and web search volume does not bring much support to price pressure hypothesis of Barber and Odean (2008), however, it supports the Conjecture 3, which builds on the conclusions outlined by Preis et al. (2010).

Time variation

Since I claimed that the inability of Google data to *nowcast* returns stems from its incapability to distinguish between buying and selling intentions of investors, one might wonder whether the division of the sample might help with the *distinguishing* problem; that is whether in any of the three previously defined periods one of the intentions dominated. For example, Dorn and Weber (2013) argue that retail investors shifted their equity portfolios away from actively managed funds towards individual stocks, while Ma et al., (n.d.) show that individual investors' behavior change from the pre-crisis non-selling inclination in loss situation to cut-loss during the crisis. Thus, one might expect more significant and negative interdependence between web search activity and returns in crisis.

Turning now to the empirical evidence on the pre-crisis, crisis and post-crisis performance of *ASVI* in predicting returns, modeled by:

$$\begin{aligned} RETURN_t^{st} = {} & \alpha_1 PRE_t + \alpha_2 CRIS_t + \alpha_3 POST_t + \beta_1 ASVI_t^{PRE,st} \\ & + \beta_2 ASVI_t^{CRI,st} + \beta_3 ASVI_t^{POST,st} \\ & + \sum_{i=1}^{k} \beta_{3+i} RETURN_{t-i}^{st} + \epsilon_t, \end{aligned} \tag{5.7}$$

it can be seen that the *ASVI* coefficients in crisis are negative on average, and lower than in the other two periods. Yet, the higher significance is not present. In fact, none of the stocks show significant relation between the web search for their name and the daily returns in crisis. As Table C.14 displays, also for returns, *ASVI* performs best in the post-crisis period when it shows significance in four regression. Interestingly, all four are positive in sign.

Granger-causality

Due to the (almost) non-existing non-lagged interdependence between web search and daily returns, it would be surprising to find any interdependence in the lagged setting. To test whether any Granger-causality exists between the two time series, I estimate bi-variate VAR(p) model, with p between two and five for different stocks (selected based on Lütkepohl version of SBIC criterion (Lütkepohl, 2007)). The results are displayed in Table C.15, and match the expectations as there is not much evidence of Granger causality between the two time series, in any direction. More specifically, none of the stocks show any significant Granger-causality at 5% level, no matter what direction is considered. At 10% level, the web search Granger-causes daily returns of four stocks.

5.1.4 Conclusion

The preliminary analysis on the firm-by-firm level suggest that the web search activity positively influences current trading volume and the effect prevails even in lagged setting. For volatility, the relationship is significantly weaker. Yet, it is predominately positive - the higher is the web search, the more volatile stock prices are. For daily returns, there seem to be feeble or even non-existing relationship with the web search volume. The most probable reason is the difficulty of disentangling investors' buy/sell intentions. The validity of the previous claim is supported by the significant (positive) interdependence between the web search and absolute returns.

Interestingly, one might notice that it is usually the same set of firms that exhibits a link between the web search activity and different financial variables (for example Boeing and Merck). Presumably, these stocks are less infected by noise search - the search non-related to trading - as the stocks with most potential noise, such as Coca-Cola or Wal-Mart, show very poor performance of web search in all estimations.

All in all, I find support to all four conjectures that related to DJIA data set, i.e. Conjectures 1 to 4.

5.2 Panel setting

In the time series setting, I considered only few model specifications to conserve space (for example I focused mainly on the non-lagged interdependence). Now, I move to panel-data investigation of the sample. I will replicate the models examining the non-lagged relation between web search activity and financial variables to asses the overall significance. In addition, I will run similar models with lagged values of web search to examine its predictive power over future observations of financial variables (both in fixed effects and PVAR setting). Lastly, I ask whether the web search's inability to predict daily returns might be solved by introduction of interaction variables based on web search.

The section is divided in three parts, dealing with trading volume, stock price volatility and daily returns, respectively.

5.2.1 Trading volume

Non-lagged setting

Firstly, I replicate model 5.1 in panel-data setting to test for the overall non-lagged interdependence. Thus, I fit following fixed effects regression:

$$VOLUME_{i,t}^{st} = \alpha_i + \sum_{j=1}^{5} \beta_j VOLUME_{i,t-j}^{st} + \beta_6 ASVI_{i,t}^{st} + \epsilon_{i,t}. \tag{5.8}$$

Additionally, I control for stock price volatility and daily returns on the same day to check the robustness of the results; and introduce squared *ASVI* to the regression to see whether the relationship between web search and trading volume shows any traces of nonlinearity. See Table C.16 for results.

The web search, unsurprisingly, retains its statistical significance also in panel-data setting; even if controlled for current stock price volatility and daily returns. The coefficient sign is positive for *ASVI* in all equations, suggesting that high trading days correspond to days with high abnormal search activity. The overall effect of *ASVI* on trading volume is, however, quite weak in terms of substantive significance. If *ASVI* is introduced to the pure auto-regressive equation for trading volume, the goodness of fit increases only by 0.35%.

Lagged setting

I continue with examining the lagged relationship. First, I consider only a lag of one trading day, that is how yesterday's web search activity affects today's trading volume. The model corresponds to the fixed effect regression in the previous paragraph:

$$VOLUME_{i,t}^{st} = \alpha_i + \sum_{j=1}^{5} \beta_j VOLUME_{i,t-j}^{st} + \beta_6 ASVI_{i,t-1}^{st} + \epsilon_{i,t}; \tag{5.9}$$

see Table C.17 for the results.

Contrary to Fink and Johann (2013), who found only the non-lagged *ASVI* significant in predicting trading volume, the DJIA data set show significant (positive) reaction of trading volume to the one-day lagged web search as well. Similarly to model 5.8, the significance remains after controlling for one-day lagged stock price volatility and daily returns. Nevertheless, one must conclude that the substantive significance of the results is even lower than in the non-lagged case as the R^2 is changed by 0.12% only, if ASVI is included to the pure auto-regressive equation.

To assess how long does in take until the predictive power of *ASVI* perishes, I fit model the fixed effect regression also for web search lagged by more than one day:

$$VOLUME^{st}_{i,t} = \alpha_i + \sum_{j=1}^{5} \beta_j VOLUME^{st}_{i,t-j} + \beta_6 ASVI^{st}_{i,t-k} + \epsilon_{i,t}, \qquad (5.10)$$
$$k \in \{2, \ldots, 5\}$$

Table C.18 provides an overview of the results. Interestingly, the effect of *ASVI* over trading volume disappears very promptly, as already the two-day lagged values do not show any significance. Nevertheless, the effect comes back to the surface at lag five, but with an opposites sign (lag six is insignificant again). Following explanation matches the pattern in the results. When the investors get involved in a search activity for stock related information, they mostly trade on this information the same day or the next one. The positive two- and three-day lags suggest that some investor wait few days longer until trading on the information. Yet, after executing their trade decision based on the web search they undertook, they seem to relax their trading activity for a given stock for a while. Since, as written in the literature review chapter, *ASVI* measures mainly the attention of retail investors, the explanation make sense because retail investor hardly ever trade on everyday basis.

Time variation

Additionally, I examine the difference in the interdependence throughout the three previously specified periods. Since fixed effects do not perform well in the presence of slowly changing independent variables - see Beck (2001), who argues that the fixed effects will make it hard for such variables to appear either substantively or statistically significant - the inclusion of slope dummies to models 5.8 and 5.9 would yield poor estimates of *ASVI* coefficients. Therefore, I decided to fit the models 5.8 and 5.9 separately for each of the three periods, and compare the *ASVI* coefficients by F-test. The results are presented in Table C.19.

Surprisingly, the relative size of coefficient estimates between the periods differ in non-lagged and lagged settings. In the former setting, the strongest increase of trading volume, with a change in web search, can be observed in crisis period; while the weakest reaction corresponds to pre-crisis period. In addition, the F-test suggest that the difference between *ASVI* coefficients in the two periods is significant. Post-crisis, the relationship weakens, albeit it stays on higher level than pre-crisis; however, the difference is not significant this time.

In contrast, in the lagged setting, the *ASVI* coefficients are similar in size throughout the considered periods. The possible explanation for the results might be that investors payed more attention to trading in financial crisis, and thus reacted more quickly on any news they obtained from Google. This is certainly plausible as in periods with increased uncertainty, when a day-old information can be already outdated, traders are forced to incorporate new facts more rapidly. This explanation is further supported by the fact that in crisis the current web search activity had significantly higher impact on daily traded volume than the one-day lagged search; while neither in the pre-crisis period nor in the post-crisis period such relation held.

Panel vector-auto-regressive analysis

To examine the dynamics in the system more thoroughly, I employ bi-variate panel vector-auto-regressive model for *ASVI* and *VOLUME* in order to obtain the impulse responses (Figure C.1) and variance decompositions (Figure C.2) for the two variables. The lag length was set to five, in correspondence with time series models. Monte Carlo simulation with 1000 replications is used to estimate the 5% error bands for impulse response functions.

Figure C.1 shows that trading activity has positive response to shocks in *ASVI*, concentrated mainly around zero and one day horizon, and weakening afterward in linear way. The opposite way impulse response function is somewhat puzzling, as *ASVI* shows positive response to shocks in trading volume only one day after the shock, and then switches to negative response in two-to-three day horizon after the shock as well as in five day horizon. This encourages one to find possible explanation. It seems that high trading volume induces investors to search for the cause the next day; and the negative response in two-to-three day horizon follows as normally search is spread evenly between days, but increased activity in one day reduces the activity in following days as retail investors hardly search for firm related information on everyday basis. This explanation is supported by *ASVI* impulse response to its own shock, which is negative in the two-to-three day horizon.

The variance decomposition show that web search explains very low percentage of variance in trading volume; the explanatory power seems to raise in the first days, reaching the maximum of ca. 1.19% in four-day forecast horizon, and slightly decreases afterward. The opposite way explanatory power rises steeply until the seven-day forecast horizon, where the explanatory power of trading volume over the variance in web search reaches 0.12%; and it continues growing with the forecast horizon in a very slow linear pace.

5.2.2 Stock price volatility

Non-lagged setting

To test for the overall non-lagged interdependence between stock price volatility and web search volume, I employ a similar methodology as in 5.2.1. Thus, I fit regression:

$$VOLATILITY^{st}_{i,t} = \alpha_i + \sum_{j=1}^{5} \beta_j VOLATILITY^{st}_{i,t-j} + \beta_6 ASVI^{st}_{i,t} + \epsilon_{i,t}.In \quad (5.11)$$

In addition, I estimate models that include non-lagged trading volume, daily returns and squared *ASVI*.

Column (2) shows that, overall, stock price volatility is higher on days with high abnormal search volume. Column (3) provides more details on the nature of the relationship, suggesting that the non-lagged interdependence is quadratic. The results in Column (3) demonstrate that the stock price volatility increases both linearly and quadratically with *ASVI*, which is inconsistent with Andrei and Hasler (2011), who found the linear coefficient to be negative. Yet, the authors used what they call "*Focus on Economic News*" index as a proxy for attention; which measures Google search volumes on groups of words with financial or economic content - that is, their variable measures market-level attention whereas I use firm-specific attention. Columns (4) to (7) show that both *ASVI* variables remain significant even if controlled for trading volume, daily returns or both.

Lagged setting

Table C.21 lists the result of model 5.12 that corresponds to model 5.9 for trading volume. Interestingly, the exponential relationship between *ASVI* and stock price volatility is not significant for the lagged values of *ASVI*. Moreover, the coefficients of squared *ASVI* are of an opposite sign to the non-lagged setting, that is, negative. Thus, if the online attention increases, volatility is enhanced linearly-to-convexly on the same day, and linearly-to-concavely the day after.

$$VOLATILITY^{st}_{i,t} = \alpha_i + \sum_{j=1}^{5} \beta_j VOLATILITY^{st}_{i,t-j} + \beta_6 ASVI^{st}_{i,t-1} + \epsilon_{i,t} \quad (5.12)$$

I also fit the model with more lags of *ASVI*, to see whether the positive effect on volatility last longer than one day. Table C.22 provides an answer for the question; the influence, accordingly to the one on trading volume, diminishes after one day and reappears at lag-five with opposite sign. In addition, for stock price volatility, the negative relation at lag 5 is non-linear.

$$VOLATILITY^{st}_{i,t} = \alpha_i + \sum_{j=1}^{5} \beta_j VOLATILITY^{st}_{i,t-j} + \beta_6 ASVI^{st}_{i,t-k} + \epsilon_{i,t}, \quad k \in \{2, \dots, 5\} \quad (5.13)$$

Time variation

I proceed with fitting models 5.11 and 5.12 for pre-crisis, crisis and post-crisis periods separately, to assess how the mutual dependence between stock price volatility and web search evolved in time. The results, presented in Table C.23, bring several interesting discoveries. First, both for non-lagged and lagged setting, an increase in the web search activity in crisis is associated with a higher increase in stock price volatility than in the other two periods; although the differences are predominantly insignificant (with the exception of crisis to pre-crisis increase for lagged *ASVI*). Second, post-crisis, the more intense interdependence seems to partially prevail, that is, the coefficients are lower than they were in crisis, yet they exceed the pre-crisis levels. The sizes of the coefficients are in line[8] with the predictions of Andrei and Hasler (2011, p. 22) model which prognoses for high attention periods that *"the investor assigns a higher weight to news, hence the stock return volatility increases by accelerating revelation of news into prices"*. Third, the non-linear relationship between non-lagged *ASVI* and stock price volatility is present only in the pre-crisis period, which, on the other hand, contradicts the predictions of Andrei and Hasler (2011) model; the authors expect the quadratic relationship to be stronger in high attention periods. Fourth, the lagged relationship is linear in all three periods, however, the coefficients of squared *ASVI* retain the negative sign in all periods (but they are insignificant). Finally, in contrast to trading volume, for which the higher impact of current web search activity compared to the lagged one was driven by the crisis period, exactly the opposite holds for the stock price volatility - that is, the non-lagged coefficients are noticeably higher than the lagged ones for the pre-crisis and post-crisis periods, while during the crisis they are rather similar. It suggests that stock price volatility reacted more rapidly to changes in web search in the periods different form crisis than during its occurrence.

Panel vector-auto-regressive analysis

Lastly, I provide deeper examination of the system dynamics, by fitting bi-variate panel VAR model for *ASVI* and *VOLATILITY*. Lag length was set, accordingly to the trading volume model, to five. Monte Carlo simulation with 1000 replications is used to estimate the 5% error bands for impulse response functions. Results are displayed in Figures C.3 and C.4.

The impulse response function shows that stock price volatility reacts to shocks in web search very similarly as the trading volume, although it takes less time till the reaction diminishes; as the reaction approaches zero in five day horizon. Other minor difference between the reaction of price volatility and

[8]With the exception of a linear vs. a non-linear nature of the relationship.

trading volume is that the later responds most pronouncedly in one-day horizon, while the former during the day of the shock. Accordingly, *ASVI* seems to react to shocks in stock price volatility similarly as to shocks in trading volume, albeit the reaction is less severe - mostly on the edge of significance or even below.

The variance decomposition supports the findings as the predictive ability of *ASVI* over volatility peak at the three-day forecast horizon; when 0.25% of the variance in stock price volatility can be explained by *ASVI*. As for the amount of *ASVI* variance that can be explained by price volatility, it is almost two times lower than the amount awardable to the innovations in trading volume.

5.2.3 Daily returns

Non-lagged setting

Since the F-test suggest that the firm dummies are jointly equal to zero, and Breusch-Pagan Lagrange-multiplier test rejects the presence of random effects, I use pooled OLS instead of fixed or random effects to model daily returns in panel-setting. The Lütkepohl version of SBIC criterion (Lütkepohl, 2007) suggest that the sufficient number of lags of the dependent variable is two. Similarly to the previous subsections, I also control for volume, volatility, and squared *ASVI*.

$$RETURNS_{i,t}^{st} = \alpha + \sum_{j=1}^{2} \beta_j RETUNRS_{i,t-j}^{st} + \beta_3 ASVI_{i,t}^{st} + \epsilon_{i,t} \tag{5.14}$$

The results, presented in Table C.24, are in line with the time series analysis of daily returns' interdependence with web search volume. The *ASVI* coefficient is insignificant for all specifications and does not bring any improvement to the pure auto-regressive model. In addition, column (3) confirms that neither the non-linear relationship between the two variable exists.

Lagged setting

I proceed by fitting a model with one-day lagged web search, in order to control for the possibility that it takes some time until the increased attention transforms in a price change of specific direction. The results are available in Table C.25.

$$RETURNS_{i,t}^{st} = \alpha + \sum_{j=1}^{2} \beta_j RETUNRS_{i,t-j}^{st} + \beta_3 ASVI_{i,t-1}^{st} + \epsilon_{i,t} \tag{5.15}$$

Interestingly, in the lagged setting, *ASVI* coefficients are on the edge of 10% significance. Their positive value provide weak support to Barber and Odean (2008) model, which predicts a short-term price pressure if attention increases. Yet, the economic significance of the *ASVI* forecasting ability over returns is limited; the inclusion of a web search variable to a pure auto-regressive model enhances the goodness of fit by 2% only.

Interaction with sentiment and previous-day returns

As previously mentioned, *ASVI*'s failure in predicting returns stems from its inability to capture the buy/sell intention of the investors. Therefore, one might wonder whether the intention can be disentangled with a help form some other variable.

First, I consider investor sentiment and examine if its interaction with attention might shed some light on the problem. I take investor sentiment measure developed by Baker and Wurgler (2006) (monthly level of sentiment measure is used, see more details about the measure in Section 3.2.1 for more details) and construct an interaction variable with *ASVI* ($ASVI \times SENT$); that is $ASVI \times SENT > 0$ corresponds to an increase in attention in a positive sentiment period or a decrease in attention in a negative sentiment period. In addition, I divide the sentiment into quartiles and construct dummy variables for *positive, normal* and *negative* levels of sentiment;[9] and build interaction variables between the sentiment dummies and *ASVI*; $ASVI^{POSSENT}$, $ASVI^{NOSENT}$ and $ASVI^{NEGSENT}$. Second, I examine whether investor attention impact on daily returns changes with the size of previous day returns. To account for the possibility, I make interaction variable between one-day lagged returns and *ASVI* ($ASVI \times RET$). Moreover, I also introduce dummy variables for positive and

[9] In a way that *positive* sentiment dummy takes value of one if the level of sentiment exceeds the third quartile, *normal* sentiment dummy takes value of one if the level of sentiment is between the first and the third quartile, and finally *negative* sentiment dummy takes takes value of one if the level of sentiment is below the first quartile.

negative day-lagged returns to asses whether *ASVI* following negative returns has different predictions on current returns than *ASVI* following positive returns; $ASVI^{POSRET}$ and $ASVI^{NEGRET}$. Table C.26 lists the results.

We can see that the interaction with sentiment, in fact, enables the web search activity to successfully predict current daily returns; as the coefficient of $ASVI \times SENT$ (3) is positive, and highly significant. Column (4) clarifies the nature of the relationship between sentiment and attention. The results show that *ASVI* significantly interacts with sentiment only in the low sentiment periods and that the interaction is negative in sign. Thus, if investors increase their web search activity in a low sentiment period, it negatively impacts the same day returns. If we get back to the regression (3), the positive coefficient of $ASVI \times SENT$ can be predominantly assigned to a negative effect of jumps in the attention on daily returns whenever investor sentiment on the market is low ($\downarrow (ASVI^{+} \times SENT^{-}) \rightarrow \downarrow returns$); and only partially to the impact of increases in web search in high sentiment periods ($\uparrow (ASVI^{+} \times SENT^{+}) \rightarrow \uparrow returns$). Regressions (7) to (11) introduce sentiment dummy variables to the equations instead of a constant; $POSSENT$, $NOSENT$ and $NEGSENT$. It can be observed that (without an interaction with *ASVI*) returns tend to be lower in the high sentiment periods and higher in the low sentiment periods; consistent with the findings of Baker and Wurgler (2006).

If one combines the findings from regressions with *ASVI* and sentiment interactions, she gets very interesting picture which calls for an explanation. Yet, let me sum up few facts before proceeding to the explanation itself. First, the low sentiment periods are characterized by lower prices and higher returns than the periods without a distinctive sentiment; on the other hand, the high sentiment periods exhibit the opposite, that is, the returns are rather low and prices are elevated. Second, *ASVI* measures predominantly the attention of retail investors (Da et al., 2011). Third, retail investors tend to be more sentimental than institutional investors (Lee et al., 1991; Barber et al., 2009). Fourth, retail investors tend to be overly optimistic rather than overly pessimistic (Benartzi et al., 1999; Dimson et al., 2004). Now, I approach to the explanation. Let's consider the negative sentiment periods first. It follows from the first fact, that prices tend to move back to the *fundamental* values from the sentiment driven levels, i.e., the prices move up in the low sentiment times. Nevertheless, as column (4) shows, if the retail investors, which are likely to succumb to the negative market sentiment, increase their attention to the stocks, they push the prices back down. The truth is the opposite for the positive sentiment periods - the prices go down from their positive-sentiment induced values and the price adjustment is partially hindered when the optimistic retail investors increase their attention to the stocks - albeit the effect is insignificantly different form zero. From the fourth fact it follows that investors are more likely to be optimistic than pessimistic; thus, it requires more attention from the sentiment investor to succumb to the negative sentiment than to the positive one. Therefore, the more significant effect of investor attention on daily returns in negative sentiment periods than in the positive sentiment periods, seems valid.

Interestingly, as Table C.27 demonstrates, the lagged *ASVI* interaction with the investor sentiment has an opposite influence on returns than the non-lagged interaction. It suggests that the price change induced by the increased retail investor attention is only short-lived. Already a day after the change in returns, the effect is more or less reversed. Thus, it seems market restores the prices back towards the arguably more correct level.

Contrary to the sentiment, the interaction of *ASVI* with day-lagged returns does not help with the disentangling problem. The investor attention does not seem to affect the current day as well as the next day returns irrespectively to the sign and size of previous returns, or at least not significantly.

Time variation

The nature of the interaction between the investor sentiment and the online attention, and its influence on returns, is mostly confirmed when the sample is divided into the pre-crisis, crisis and post-crisis periods as in previous subsections. I run pooled OLS regression with ASVI slope dummies corresponding to each of the three periods. Results for the two versions of the model (first uses non-lagged *ASVI* (1) and the second one-day lagged *ASVI* (2)) are presented in Table C.28.

For the non-lagged setting, the web search activity seems to negatively impact the same-day returns in crisis period, which has $2.1\times$ higher incidence of low-sentiment months compared to the whole 2004 to 2010 sample. Conversely, post-crisis, an increase in the web search activity is associated with and increase in the same-day returns; while the post-crisis period shows the highest percentage of positive sentiment months out of the three periods. Lastly, pre-crisis show negative, yet insignificant coefficients of *ASVI*. The months in pre-crisis period mostly exhibit normal sentiment, nevertheless, the period also has higher than average occurrence of low sentiment months. The inequality of the three coefficients is confirmed by F-test.

Accordingly to the previous paragraph, the day-lagged web search activity has an opposite sign than the current web search activity. Nevertheless, only the web search variable from the pre-crisis period shows a significant coefficient. Yet, for lagged web search, the three coefficient's inequality is not confirmed by F-test.

Panel vector auto-regressive analysis

Finally, I asses how daily returns respond to shocks in online attention, and vice versa, by fitting a bi-variate PVAR(2) model for the two variables.

The IRF functions, displayed in Figure C.5, show that daily returns positively respond to shocks in search volume in one-day horizon, albeit the reaction is very weak. The reaction seems to disappear very quickly and we see that IRF bounds do not move from zero, in horizons higher than one. Contrary, the web search activity does not seem to react to shocks in returns significantly, within the five-day horizon.

Unsurprisingly, variance decomposition show that only a very minor portion of variance of any of the two variables might be explained by movements in the other one (Figure C.6). Moreover, the forecasting ability diminishes promptly.

Following findings in the previous paragraphs, I run the PVAR model for periods of high and low sentiment separately. The results change substantially. Figures C.7 and C.8 display the IRFs for positive and negative sentiment periods, respectively. The patterns mostly match model outcomes from the previous subsections; for negative sentiment periods, we see negative reaction at the day of the shock (on the edge of significance) and positive one the following day. Conversely, for the positive sentiment times, daily returns tend to react positively to shocks in *ASVI* on the day of the shock, and the effect perishes before the next day. On the other hand, *ASVI* does not react to shocks in returns in both negative and positive sentiment periods, accordingly.

Figure C.9 compares variance decompositions for *ASVI* and returns for the two different levels of sentiment. Interestingly, *ASVI* seems to explain more variance in daily returns in the low sentiment periods, while returns explain higher portion of variance in *ASVI* in the high sentiment periods. This results is very interesting. The higher portion of return variance explainable by *ASVI* in negative sentiment periods corresponds to the hypothesis that a higher attention is needed for the investors to succumb to the negative sentiment. The relative sizes of the opposite way relation, on the other hand, stems from the fact that returns are predominantly negative in positive sentiment periods; and (individual) investor pay more attention to negative returns than to positive ones Hacamo and Reyes (2012).

5.2.4 Conclusion

The panel data modeling brings results that are in line with finding from the one-by-one analysis presented in previous section. I confirm the positive relationship between trading volume and web attention, and show that it also holds for lagged values of the attention. In addition, I provide further support to the enhancement of the relationship with the outburst of crisis as well as the I demonstrate that the trading activity responded to web search more quickly in this period.

For the stock price volatility, I show that the positive relation with web search activity is significant overall and also sticks for day-old web search. Also the stock price volatility seemed to react more pronouncedly to changes in web search during the financial crisis, yet, the speed of the reaction decreased with the start of the crisis (to increase again as the crisis subsided).

Lastly, I find very little evidence for the existence of any significant relationship between online attention and daily returns. Nevertheless, if one takes into account the market sentiment, the relationship comes to the surface. I argue that the investors, whose attention is measure by Google search volume, succumb to market sentiment and so an increase in web search in negative (positive) sentiment periods reduces (increases) the daily returns; yet, the effect is only short-lived. This interaction of attention with behavioral biases is in line with findings of Ramos et al. (2013). The negative impact of attention on daily returns during the crisis, as well as the positive impact in the post crisis period, correspond with the results for market sentiment.

Thus, also the panel results seem to be in line with predictions stated in Conjectures 1 to 4.

5.3 Discussion

In this section, I briefly elaborate on the performance of different *ASVI* specifications in predicting various aspects of market behavior. In addition, I discuss web search data frequency and search term specification.

At first, I provide an example to show that the results are robust to *ASVI* specification. On purpose, I present a model equivalent to 5.8, for which $ASVI^{2,exc}$ that is used for the computational part of the thesis is not the best performing indicator. The results, available in Table C.29, are very similar for different specifications of *ASVI* with an exception of $ASVI^{k,inc}$ for $k = 1, 2, 3$. Thus, the exclusion of weekend values seems crucial for a good performance of *ASVI* in prediction of financial variables (similar behavior can be observed for other aspects of market behavior as well, but I do not present more robust-check results for the sake of brevity).

Second, the results for daily data are comparable to alternative frequency specification, as the comparison of results with other researchers suggests. To robust-check, I present an example for weekly trading volume and weekly *ASVI* (abnormal search volume compared to median of previous four weeks). Table C.30 lists the results. It is apparent that the results are very similar to those on daily frequency. Thus, it is up to researcher what she is interested in and whether the daily data, which are more difficult to obtain in reasonable time and extent, are necessary. It should be also noted that the interpretation is slightly different for daily and longer frequency data (i.e. weekly or monthly). While the daily data employment serves to examine financial variable reactions to the short-term changes in attention, the longer frequencies employment helps to investigate the reaction to trends in attention.

Third, I would like to discuss the specification of search term, that is, the utilization of search volume for stock ticker versus the utilization of search volume for the firm's name. This time, I do not provide any result comparison and rather leave it for future research. Nevertheless, there are certain questions that arise from the results in this chapter. I show in Subsection 5.2.3 that *ASVI* captures attention of sentiment investors and that the current sentiment on the market is able to disentangle the effect of *ASVI* on daily returns. Therefore, if search for ticker truly represents attention of more sophisticated investors, it might be interesting to estimate the model with ticker-based-*ASVI* and compare the results. Do more sophisticated investors also succumb to market sentiment? Do they react to previous day returns? Are the predictions on daily returns any different for such investors? In addition, one may try to construct a joint attention measure by summing up the web search for ticker, or other stock designations, with the web search for firm's name. Such specified index should be able to catch both the sophisticated and sentiment investors.

Chapter 6

Empirical results - IPO data set

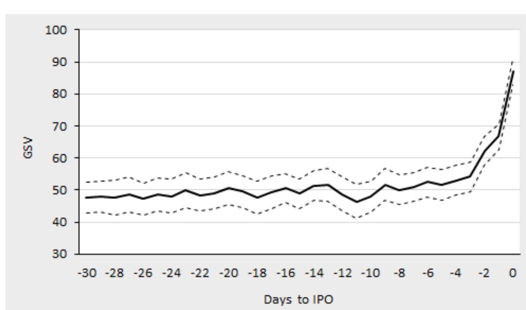
This chapter is devoted to web search usability in explaining IPO-related phenomena. I mainly follow Da et al. (2011) analysis on web search and IPOs with the difference that I employed data of a daily frequency.

6.1 Does investor attention react to IPOs?

Before approaching to the analysis per se, it might be interesting to examine, how the investor web search activity reacts to upcoming IPO. Thus, I graphed the average *GSV* for the names of the 75 companies in the IPO sample, for an interval starting 30 days prior IPO and ending on the emission date. Figure 6.1 clearly depicts the increase in *GSV* starting circa 5 days before the offering. Thus, investors are clearly interested in share IPOs.

Figure 6.1: Increase in investor attention prior IPO

The vertical axis show the average GSV for firms name for 75 firms in IPO sample. The horizontal axis show time to IPO. The dashed lines are 95% confidence intervals.



The next step is to identify the drivers of the enhanced attention prior IPO, so I regress web search activity on variables related to the offering; namely on offering size, the exchange on which the offering takes place[1] and on dummy variable that takes value of one for firms going public in the crisis period and zero otherwise.[2] Formally, the models are specified as follow:

All models in this chapter are estimated by OLS. Reported standard errors are either OLS standard errors, HC errors White (1980) or bootstrapped errors; the procedure of choosing the proper standard error estimator is described in Section 4.2, together with the procedure for dealing with influential observations.[3]

The results, presented in Table D.1, do not show investors to react any differently to the bigger emissions (1). Column (2) suggest that neither the exchange, on which the firm issues it shares, plays any role in in investor reaction to IPO. Therefore, it seems that it is only the year, in which the firm goes, that influences investor reaction to IPO; out of the considered variables. Results in column (3) clearly show that investor paid more attention to offerings in crisis years than to those in non-crisis ones.

6.2 Investor attention and IPO stylized facts

After assessing the drivers of ASVI, I proceed with the analysis of interest, i.e., whether the Google search data can shed some light on the two stylized facts about IPOs, that is, the high initial returns and long-term under-performance.

6.2.1 Initial returns

First, I analyze whether search volume may bring some information about the size of IPO first day returns. The investor sentiment theory on first day returns (Loughran and Ritter, 2002; Demers and Lewellen, 2003; Aggarwal et al., 2002) states that they tend to be higher in periods of positive sentiment. Da et al.

[1]AMEX was excluded as only 1 firm went public on this exchange. Thus it depicts difference in attention to NYSE and NASDAQ offerings.

[2]I also run the regression on year dummies, and obtained very similar results; i.e. there is only a difference in attention to IPOs in the crisis years and the non-crisis years.

[3]Due to omission of influential observations, the goodness of fit is incomparable between models (as the sample of observations differ). Thus, tables do not report R^2 and $adjR^2$ in this chapter.

(2011) argue that investor sentiment and investor attention is closely related for retail investors as those are prone to sentiment and attention is a necessary condition for sentiment. I already prove that the relationship is not that straightforward. Therefore, I measure both the effect of attention (firm specific) and sentiment (market level) on the first day returns.

Before proceeding to the regression analysis, I preliminary analyze how initial returns (IR) and investor attention (ASVI) are related. Thus, I divided the firms from my sample values into three groups based on their *ASVI* values prior IPO; namely to the high, medium and low attention groups.[4] The results show that the high attention group's average (median) initial return is 22.85% (21.29%), while low attention group's initial return only equals to 12.23% (6.65%). The difference is statistically significant at 5%. Thus, the first look at the data suggests that investor attention, very likely, drives the first day returns up.

I approach to regression analysis, and fit model 6.1 in order to estimate how an increase in attention prior IPO influences the size of the initial return in more detail. To check the robustness of the results, I control for the offering size and the investor sentiment (both in levels and change to previous month). Table D.2 provides the results.[5]

$$IR_i^{st} = \alpha + \beta_1 ASVI_i^{st,t} + \epsilon_i \tag{6.1}$$

Column (1) show that the steeper is the increase in attention prior IPO, the higher are corresponding initial returns. The effect is highly significant and has a notable size; a one standard deviation increment in *ASVI* leads to an increase in initial return by a magnitude of 41.4% of its standard deviation. Moreover, the goodness of fit is satisfactory, with $adjR^2$ equal to 12.6%.

Columns (2) to (9), which display the results of robust-check regressions, suggest that neither the offering size nor the investor sentiment (both in levels and changes from the previous month level) are able to predict initial returns. The insignificance of the offering size variable is in contradiction with results of Da et al. (2011), who used IPO data set with 185 firms that went public from 2004 to 2007. Thus it seems that the offering size effect over the initial return

[4] High attention group encompasses firms whose ASVI exceeds the third quartile, medium attention group encompasses firms whose ASVI lies between the first and the third quartile, and finally low attention group encompasses firms whose ASVI is below the first quartile.

[5] The different number of firms in each equation is caused by an exclusion of the overly influencing observations.

largely depends on selected sample of firms. The authors also found the change in investor sentiment modestly significant (at 10% level), which is not significant in my results either. The difference may be caused by very limited sample size in my analysis, as one seldom gets significant coefficients in small samples unless the effect of the regressors on the independent variable is very strong.

To test the sentiment hypothesis, I construct dummy variables for positive, normal and negative values of sentiment[6] and use them in the interaction with *ASVI* in regressions (10) to (13). The results show that attention significantly increases initial returns only in positive sentiment periods. For the negative and normal sentiment times, the attention boosts initial returns as well, albeit the effect is not significant. Nevertheless, the difference between the three coefficients in (13) is insignificant if tested by F-test. In addition, regressions (11) and (12) show that the results are robust if one controls for the original sentiment measures.

Lastly, I examine whether the predictive power of investor attention prevails if *lagged* values of *ASVI* are used.[7] Table D.3 shows that the predictive power of *ASVI* decreases with the forecasting horizon; with significant predictive power up to three days prior IPO. Yet, the results are highly dependent on the exact specification of *ASVI* and one may find the predictive power to disappear sooner or later for a different specification of ASVI.

6.2.2 Long-term returns

Second stylized fact about IPOs, that Google data might help to explain, is the long-term underpricing of IPO firms to their already traded peers. The sentiment-based hypothesis regarding high first day returns goes well with the subsequent long-term underperformance. The overoptimism of investors about the offering may lead to overly escalated initial returns, which should be followed by a price reversion towards the fundamental value afterward - that is, the long-term underperformance (Ljungqvist et al., 2006; Ritter and Welch, 2002).

[6]Th negative (positive) sentiment is defined as the sentiment below the 33% (above the 66%) centile, normal sentiment is the remainder.

[7]For the regressions (1) to (5) in Table D.3 I do not omit outliers to make the goodness of fit of the models comparable. I fit the models also with excluded outliers and the results changed only negligibly.

Since the comparison to the peers is not in scope of this thesis, I only attempt to model the negative long-term cumulative returns of the issuers following the emission.

I consider five different horizons for long-term performance for which I calculate cumulative log-returns: first day closing price to the (1) closing price one year, (2) half a year (3) and quarter of the year after IPO; and the closing price one month after IPO to (4) the closing price one year (5) and half a year after IPO . Figure D.1 provides an overview of the cumulative returns over the five specified horizons for the low and high attention IPOs. It seems that, with an exception of the shortest horizon, the high attention IPOs clearly under-perform the low attention ones in long-term. Thus, the first results are in line with (Da et al., 2011) findings and the attention/sentiment based theory on IPOs.

I proceed by regressing the long-term returns on the abnormal search volume on the IPO date. Table D.4 compares the predictive power of *ASVI* over the long-term cumulative returns (LR) for the five defined periods. The results provide only weak evidence for *ASVI* ability to forecast the negative LR returns. For half-year horizon (measure both from the opening day (2) and one month after IPO (5)), *ASVI* negatively correlates with the LR returns. Nevertheless, we see no significant effect on the one year (1, 4) or quarter of the year (3) cumulative returns; although all coefficients are negative in sign.

Da et al. (2011) wisely constructed an interaction variable between *ASVI* and initial return ($ASVI \times IR$); as the high initial return of IPOs that also experience increases in retail investor attention should be partly driven by the price pressure and hence revert in the long-term. I follow their procedure and regress the cumulative long-term returns on initial returns and the interaction variables. Table D.5 shows that there is, as expected, a higher price reversion for IPOs that experienced high initial returns (1,...,5); albeit the effect is significant only for cumulative returns measured from one month after IPO. The performance of the interaction variable (5,...,10) matches the findings of Da et al. (2011); it is obvious that high attention IPOs with high first day return experience severe price reversion in long-term. The effect is significant for all considered horizons with the exception of the quarter of the year horizon measured from the offering day. It seems, and the results from the other regression support this claim, that the quarter of the year horizon is too short for the prices to revert to the *fundamental* level.

Da et al. (2011) also constructed an interaction variable between initial returns and sentiment, and found no significant effect on the long-term returns. I, conversely, employ sentiment (dummy) interaction with *ASVI*, to account for the effect of attention on the long-term returns in positive, medium and negative sentiment periods. Thus, I regress the long-term returns on *ASVI* in different sentiment periods; results are provided in Table D.6. Interestingly, only IPOs that went public in high sentiment periods and get abnormal attention show the price reversion in long-term. Nevertheless, also the sentiment itself is able to predict the long-term reversal, albeit for fewer horizons and lower significance.

6.3 Investor attention in the setting of model by Ma and Tsai

Most researches use the terms initial return and underpricing interchangeably. Nevertheless, Ma and Tsai (2002) argue that under the sentiment hypothesis, the interchangeability is not correct. According to their definition, initial return has two components, true discount (*TD*) and market reaction (*MR*):

$$IR = \frac{(P_m - P_o)}{P_o} \tag{6.2}$$

$$IR = TD + MR \tag{6.3}$$

$$IR = \frac{P_e - P_o}{P_o} + \frac{P_m - P_e}{P_o} \tag{6.4}$$

here P_m is the first day closing price, P_o is the offer price and P_e is the equilibrium (fundamental) market price. Previous section showed that the price revision for high attention IPOs happens circa half a year after the offering. Moreover, if return variance is calculated for 30-day periods up to one year after IPO, the lowest variance corresponds to 150 to 180 day horizon. Therefore, I use the average price between $t + 150$ and $t + 180$, where t is the IPO date, as a estimate for P_e (Ma and Tsai (2002) used average price between one week and one moth after IPO, price two months after listing and one year after listing to be the equilibrium market price).

According to the authors, positive values of *MR* mean that investors over-react, while negative values suggest under-reaction of investors; true discount, on the other hand, corresponds to actual underpricing. Thus, I use this setting to confirm the results that *ASVI*, especially if combined with positive sentiment on the market, drives the investor overreaction. In contrast, I expect that *ASVI* should not possess any significant information about the underpricing term, *TD*. To see whether such expectations may be valid, I calculate mean *TD* and *MR* for high and low attention IPOs. Figure D.2 displays the comparison. As expected, the true discount does not seem to be influenced by attention. Conversely, market reaction and the attention devoted to an IPO show strong interdependence; in a way that market under-reacts to low attention IPOs and vice versa.

The relationship is mainly confirmed by the regression results. I regress *TD* and *MR* on the attention measured by *ASVI*, on the *ASVI* interaction with the initial return, and on the attention-sentiment interaction variables; results are presented in Table D.7. First, it can be observed that no attention based variable predicts the underpricing term. On the other hand, market seem to overreact on high attention IPOs, albeit the effect is significant only at 10%. The effect is more pronounced if we take into account the interaction with initial return, which is logical as *MR* is one of the two terms of which the initial return consists (thus, the stronger is the evidence against *ASVI* and *TD* interdependence, as the interaction term is insignificant in *TD* equation (3)). Surprisingly, we see only insignificant effect of the sentiment interaction variables and *MR*. While the coefficient is positive for the attention in positive sentiment periods, it is insignificant (albeit on the edge of 10% significance). Even more surprising is the positive coefficient for the attention in negative sentiment periods, as one would expect this term to be negative. It suggests that investor overreact to IPOs also in low sentiment period and that it is the attention that drives the overreaction and not sentiment. This is confirmed by regression (8), which shows that sentiment is not able to predict the market reaction on its own; and the insignificance is unquestionable. Nevertheless, it should be noted that the results in this section may differ if other P_e is specified, as its selection is rather arbitrary.

6.4 Conclusion

The results in this chapter provide support to the second bundle of my conjectures. First, Conjecture 5 seem to be valid as I find direct proportion between online attention and the size of initial returns. On the other hand, I do not find clear evidence for Conjecture 6, while I am not able to reject it either. Nevertheless, it seem that more than just attention is needed to forecast the long-term price reversion. Lastly, Conjectures 7 and 8 appear to valid as well, as *ASVI* is a reasonably significant predictor of the market reaction to IPO, while it does not bring significant amount of information about the true discount.

Chapter 7

Conclusion

This thesis contributes to a rapidly growing family of research on Google search volume influence on different aspects of financial markets behavior. Primarily, I investigate whether Google search volume for firm's name can be used to predict daily trading activity, stock price volatility and returns; for a sample of nineteen Dow Jones Industrial Average constituents. Secondly, I address the question whether Google data might be helpful in explaining two IPO stylized facts, the high initial returns and poor long-term performance; using sample of emerging growth firms that went public between year 2004 and 2010.

Google search volume and different aspects of financial markets behavior

I find that an increase in web search activity for a firm's name is associated with high trading activity the same day as well as the day after. Afterward, the effects of web search on trading activity diminish. I also demonstrate that the magnitude of the link between trading activity and online attention increased with the outbreak of recent financial crisis and remained on higher than pre-crisis level even when the crisis subsided. In addition, I show that during the crisis, investors reacted more quickly to the information obtained from web search than in the other periods.

Similarly to trading activity, the stock price volatility increases with both the same day and one-day lagged hikes in Google search volume. Yet, in contrast, I detect some evidence for non-linearity in the stock price's volatility interdependence with web search; it seems that an increase in web-based attention is followed by linear-to-convex increment in volatility the same day and linear-to-concave increment the day after. It can be observed that also volatility tended to react more pronouncedly to the changes in web search during the finical crisis, yet the difference to the pre-crisis and post-crisis periods is less significant. In contrast to trading volume, which reacted more rapidly to changes in web search during the crisis, the stock price volatility seem to react slower.

Daily returns, on the other hand, do not appear to react to innovations in current web search activity significantly; however, they show positive response to day-old surge in web search that is on the very fringes of significance. Thus, I find very little evidence for the short-term price pressure hypothesis proposed by Barber and Odean (2008). The ambiguity of the relationship between the two series seems to stem from Google data inability to distinguish between investor buying and selling intentions, as suggested by Preis et al. (2010). To overcome the vagueness, I look for phenomena that may help to disentangle the intentions in the interaction with attention and find that the market sentiment is able to shed some light on the interdependence while past returns are not.

The results suggest that in the negative sentiment periods, an increase in online attention leads to a significant negative reaction of current returns, while for the positive sentiment times, the adverse can be observed - that is, a boost in attention is associated with positive response from daily returns, yet on a very slight trend towards significance. Therefore, it gives the impression that the investors, whose attention is measured by Google search volume, tend to succumb to the market sentiment. Interestingly, the price effect is only short-lived and is partially reverted the next day, but the reversion is only significant for the negative sentiment.

The combined effect of sentiment and attention also comes to light if the nature of the interaction between daily returns and online attention is compared in different time periods. In crisis, when the sentiment was predominately negative, the effect of web search on daily returns was negative as well. Conversely, after the crisis, returns reacted positively to increases in Google search volume. Again, a subsequent reversion is present, albeit with varying (in)significance.

Google search volume and IPO stylized facts

Second part of the thesis, devoted to Google data and IPOs, mainly follows the findings of Da et al. (2011). I confirm that initial returns are higher for IPOs that receive above average attention, however, I argue that the effect is significantly present only for firms going public in positive sentiment periods. In addition, since I use daily data, I am able to demonstrate that Google search volume is capable of forecasting the initial returns within a few days horizon.

Contrary to Da et al. (2011), I observe a weak evidence for Google data ability to forecast (with negative sign) the long-term cumulative returns. Nevertheless, in line with the authors, I show that high attention IPOs leaving a lot of money on the table experience a price reversal in long-term. Correspondingly to the results for initial returns, the long-term cumulative returns seem to be inversely proportionate to investor attention to IPO only for firms that emitted shares in positive sentiment periods. The findings correspond to Derrien (2005) predictions claiming that it is the overoptimistic investors who leave the money on the table, rather than the issuing firms.

Finally, I test Google search volume in the setting of the model proposed by Ma and Tsai (2002), which questions the interchangeability of terms initial return and underpricing. The results suggest that the Google search volume is able to predict one part of initial returns - the market overreaction to the offering, while the other - the true IPO discount (i.e. the underpricing) - is unpredictable by Google data.

Possible extensions of the presented research

As the Google search volume for query *"Google econometrics"* indicates, the attention to similarly focused research is growing and I believe that the application of Internet-based data will soon become one of the hottest topics in the contemporary academic research. Moreover, I hope that this thesis might inspire some future research in this field. In relation, I can think of several extensions to the subject matter discussed in the thesis. Firstly, as mentioned in the discussion section in Chapter 5, the Google search volume interaction with market sentiment may be tested for different queries such as tickers, which are likely to capture the attention of more sophisticated investors. Secondly, one may augment the information source by Facebook posts, Tweets or Wikipedia edits to construct a combined measure of attention. Thirdly, the tests might be replicated on a larger sample, or a sample consisting of different types of firms (for example growth companies, if the search data are available). Nevertheless, the possibility of extension is undoubtedly far larger.

Bibliography

Achrekar, H., Gandhe, A., Lazarus, R., Yu, S.-H., Liu, B., 2011. Predicting flu trends using twitter data. In: Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on. IEEE, pp. 702–707.

Aggarwal, R. K., Krigman, L., Womack, K. L., 2002. Strategic ipo underpricing, information momentum, and lockup expiration selling. Journal of Financial Economics 66 (1), 105–137.

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

Anderson, T. W., Hsiao, C., 1982. Formulation and estimation of dynamic models using panel data. Journal of econometrics 18 (1), 47–82.

Andrei, D., Hasler, M., 2011. Investor's attention and stock market volatility. Available at SSRN 1761421.

Ap Gwilym, O., Hasan, I., Xie, R., Wang, Q., 2012. In search of concepts: The effects of speculative demand on returns and volume. Available at SSRN 2062813.

Arbel, A., Carvell, S., Strebel, P., 1983. Giraffes, institutions and neglected firms. Financial Analysts Journal, 57–63.

Arbel, A., Strebel, P., 1982. The neglected and small firm effects. Financial Review 17 (4), 201–218.

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. The review of economic studies 58 (2), 277–297.

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error-components models. Journal of econometrics 68 (1), 29–51.

Askitas, N., Zimmermann, K. F., Apr. 2009. Google econometrics and unemployment forecasting. Applied Economics Quarterly 55 (2), 107–120.

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

Baker, S., Fradkin, A., 2011. What drives job search? evidence from google search data. Evidence from Google Search Data (April 15, 2011).

Bank, M., Larch, M., Peter, G., Jul. 2011. Google search volume and its influence on liquidity and returns of german stocks. Financial Markets and Portfolio Management 25 (3), 239–264.

Barber, B. M., Odean, T., 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. Review of Financial Studies 21 (2), 785–818.

Barber, B. M., Odean, T., Zhu, N., 2009. Do retail trades move markets? Review of Financial Studies 22 (1), 151–186.

Beck, N., 2001. Time-series-cross-section data: what have we learned in the past few years? Annual review of political science 4 (1), 271–293.

Beck, N., Katz, J. N., 2011. Modeling dynamics in time-series-cross-section political economy data. Annual Review of Political Science 14, 331–352.

Benartzi, S., Kahneman, D., Thaler, R., 1999. Optimism and overconfidence in asset allocation decisions. News. morningstar. com.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., Weber, I., Jul. 2012. Web search queries can predict stock market volumes. PLoS ONE 7 (7), e40014.

Brav, A., Geczy, C., Gompers, P. A., 2000. Is the abnormal return following equity issuances anomalous? Journal of Financial Economics 56 (2), 209–249.

Breusch, T. S., Pagan, A. R., 1979. A simple test for heteroscedasticity and random coefficient variation. Econometrica: Journal of the Econometric Society, 1287–1294.

Cameron, A. C., Miller, D. L., 2013. A practitioner's guide to cluster-robust inference.

Caporale, G. M., Gil-Alana, L. A., 2010. Long memory and fractional integration in high frequency financial time series. Tech. rep., Discussion papers//German Institute for Economic Research.

Carhart, M. M., 1997. On persistence in mutual fund performance. The Journal of finance 52 (1), 57–82.

Carneiro, H., Mylonakis, E., Nov. 2009. Google trends: A web based tool for real time surveillance of disease outbreaks. Clinical Infectious Diseases 49 (10), 1557–1564.

Chamberlin, G., 2010. Googling the present. Economic & Labour Market Review 4 (12), 56.

Chen, S., 2011. Google Search Volume: Influence and Indication for the Dutch Stock Market. Erasmus University.

Choi, H., Varian, H., 2009a. Predicting initial claims for unemployment benefits. Google Inc.

Choi, H., Varian, H., 2009b. Predicting the present with google trends. Google Research Blog 2.

Choi, I., 2001. Unit root tests for panel data. Journal of international money and Finance 20 (2), 249–272.

Chordia, T., Huh, S.-W., Subrahmanyam, A., 2007. The cross-section of expected trading activity. Review of Financial Studies 20 (3), 709–740.

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

Chou, R. Y., Chou, H., Liu, N., 2010. Range volatility models and their applications in finance. In: Handbook of Quantitative Finance and Risk Management. Springer, pp. 1273–1281.

Colaco, H. M., De Cesari, A., Hegde, S. P., 2013. Retail investor sentiment and ipo valuation.

Cook, D. O., Kieschnick, R., Van Ness, R. A., Oct. 2006. On the marketing of IPOs. Journal of Financial Economics 82 (1), 35–61.

Cook, R. D., 2000. Detection of influential observation in linear regression. Technometrics 42 (1), 65–68.

Cooper, C. P., Mallon, K. P., Leadbetter, S., Pollack, L. A., Peipins, L. A., Jul. 2005. Cancer internet search activity on a major search engine, united states 2001-2003. Journal of Medical Internet Research 7 (3), e36.

Corley, C., Mikler, A. R., Singh, K. P., Cook, D. J., 2009. Monitoring influenza trends through mining social media. pp. 340–346.

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

D'Amuri, F., 2009. Predicting unemployment in short samples with internet job search query data.

D'Amuri, F., Marcucci, J., 2009. 'Google it!'Forecasting the US unemployment rate with a google job search index. Tech. rep., Nota di lavoro//Fondazione Eni Enrico Mattei: Global challenges.

Degiannakis, S., 2008. Arfimax and arfimax-tarch realized volatility modeling. Journal of Applied Statistics 35 (10), 1169–1180.

Della Penna, N., Huang, H., 2009. Constructing a consumer confidence index for the us using web search volume. Tech. rep., Working Paper.

Demers, E., Lewellen, K., Jun. 2003. The marketing role of IPOs: evidence from internet stocks. Journal of Financial Economics 68 (3), 413–437.

Derrien, F., 2005. Ipo pricing in hot market conditions: who leaves money on the table? The Journal of Finance 60 (1), 487–521.

Dickey, D. A., Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association 74 (366a), 427–431.

Dimpfl, T., Jank, S., 2011. Can internet search queries help to predict stock market volatility? Tech. rep., CFR working paper.

Dimson, E., Marsh, P., Staunton, M., 2004. Irrational optimism. Financial Analysts Journal, 15–25.

Dorn, D., Weber, M., 2013. Individual investors trading in times of crisis: Going it alone or giving up? Tech. rep., Citeseer.

Drakos, K., Konstantinou, P. T., 2011. Terrorism shocks and public spending: Panel var evidence from europe. Tech. rep., DIW Berlin, German Institute for Economic Research.

Dugas, A. F., Hsieh, Y.-H., Levin, S. R., Pines, J. M., Mareiniss, D. P., Mohareb, A., Gaydos, C. A., Perl, T. M., Rothman, R. E., Jan. 2012. Google flu trends: Correlation with emergency department influenza rates and crowding metrics. Clinical Infectious Diseases 54 (4), 463–469.

Eckbo, B. E., Norli, Ø., 2000. Leverage, liquidity and long-run ipo returns.

Ettredge, M., Gerdes, J., Karuga, G., Nov. 2005. Using web-based search data to predict macroeconomic statistics. Communications of the ACM 48 (11), 87–92.

Eysenbach, G., 2005. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. In: AMIA... Annual Symposium proceedings/AMIA Symposium. AMIA Symposium. pp. 244–248.

Fink, C., Johann, T., 2013. May i have your attention, please: The market microstructure of investor attention. Please: The Market Microstructure of Investor Attention (January 18, 2013).

Garman, M. B., Klass, M., 1980. On the estimation of security price volatilities from. The Journal of Business 53 (0).

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of time series analysis 4 (4), 221–238.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L., Nov. 2008. Detecting influenza epidemics using search engine query data. Nature 457 (7232), 1012–1014.

Giraitis, L., Kokoszka, P., Leipus, R., Teyssiere, G., 2003. Rescaled variance and related tests for long memory in volatility and levels. Journal of econometrics 112 (2), 265–294.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., Watts, D. J., Sep. 2010a. Predicting consumer behavior with web search. Proceedings of the National Academy of Sciences 107 (41), 17486–17490.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., Watts, D. J., 2010b. What can search predict. WWW 2010, April 26-30, 2010, Raleigh, North Carolina.

Granger, C. W., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424–438.

Guzman, G., 2011. Internet search behavior as an economic forecasting tool: The case of inflation expectations. Journal of Economic and Social Measurement 36 (3), 119–167.

Hacamo, I., Reyes, T., November 2012. Negativity bias in attention allocation: Retail investors reaction to stock returns.

Harris, D., Mccabe, B., Leybourne, S., 2008. Testing for long memory. Econometric Theory 24 (01), 143–175.

Hausman, J. A., 1978. Specification tests in econometrics. Econometrica: Journal of the Econometric Society, 1251–1271.

Hulth, A., Rydevik, G., Linde, A., 2009. Web queries as a source for syndromic surveillance. PloS one 4 (2), e4378.

Kenney, M., Patton, D., 2013. Firm database of emerging growth initial public offerings (ipos) from 1990 through 2010.

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

Kiviet, J. F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. Journal of econometrics 68 (1), 53–78.

Kristensen, I. P., Wawro, G., 2003. Lagging the dog? the robustness of panel corrected standard errors in the presence of serial correlation and observation specific effects. In: annual meeting of the Society for Political Methodology, University of Minnesota.

Kristoufek, L., 2013. Can google trends search queries contribute to risk diversification? Scientific reports 3.

Kulkarni, G., 2012. Patterns of consumer interest across developed and emerging markets. International Business & Economics Research Journal (IBER) 11 (6), 603–610.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., Shin, Y., 1992. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Journal of econometrics 54 (1), 159–178.

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

Li, X., Shang, W., Wang, S., 2013. Incorporation of social media data into macroeconomic forecast systems: A mixed frequency modelling approach.

Ljungqvist, A., Nanda, V., Singh, R., 2006. Hot markets, investor sentiment, and IPO pricing*. The Journal of Business 79 (4), 1667–1702.

Lo, A. W., 1989. Long-term memory in stock market prices. Tech. rep., National Bureau of Economic Research.

Loughran, T., Ritter, J. R., 1995. The new issues puzzle. The Journal of Finance 50 (1), 23–51.

Loughran, T., Ritter, J. R., 2002. Why don't issuers get upset about leaving money on the table in IPOs? Review of Financial Studies 15 (2), 413–444.

Love, I., Zicchino, L., 2006. Financial development and dynamic investment behavior: Evidence from panel var. The Quarterly Review of Economics and Finance 46 (2), 190–210.

Lütkepohl, H., 2007. New introduction to multiple time series analysis. Springer.

Ma, T., Chang, K.-W., Yin, C.-H., ???? The role of individual investors in liquidity risk and financial crisis.

Ma, T., Tsai, P., 2002. Are initial return and IPO discount the same thing? a comparison of direct public offerings and underwritten IPOs. A Comparison of Direct Public Offerings and Underwritten IPOs (November 2001). EFA.

Manda, K., 2010. Stock market volatility during the 2008 financial crisis. Ph.D. thesis, School of Business, New York University.

McLaren, N., Shanbhogue, R., 2011. Using internet search data as economic indicators. Bank of England Quarterly Bulletin (2011), Q2.

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

Mondria, J., Wu, T., 2011. Asymmetric attention and stock returns. In: AFA 2012 Chicago Meetings Paper.

Murphy, A., Izzeldin, M., 2009. Bootstrapping long memory tests: some monte carlo results. Computational Statistics & Data Analysis 53 (6), 2325–2334.

Newey, W. K., West, K. D., 1986. A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Newey, W. K., West, K. D., 1994. Automatic lag selection in covariance matrix estimation. The Review of Economic Studies 61 (4), 631–653.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica: Journal of the Econometric Society, 1417–1426.

Pelat, C., Turbelin, C., Bar-Hen, A., Flahault, A., Valleron, A.-J., 2009. More diseases tracked by using google trends. Emerging infectious diseases 15 (8), 1327.

Polgreen, P., Chen, Y., Pennock, D., Nelson, F., Dec. 2008. Using internet searches for influenza surveillance. Clinical Infectious Diseases 47 (11), 1443–1448.

Preis, T., Paul, W., Schneider, J. J., Jun. 2008. Fluctuation patterns in high-frequency financial asset returns. EPL (Europhysics Letters) 82 (6), 68005.

Preis, T., Reith, D., Stanley, H. E., Nov. 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.

Ramos, S. B., Veiga, H., Latoeiro, P., 2013. Predictability of stock market activity using google search queries. Tech. rep.

Ritter, J. R., 1991. The long-run performance of initial public offerings. The Journal of Finance 46 (1), 3–27.

Ritter, J. R., Welch, I., 2002. A review of IPO activity, pricing, and allocations. The Journal of Finance 57 (4), 1795–1828.

Robinson, P. M., 1995. Log-periodogram regression of time series with long range dependence. The annals of Statistics, 1048–1072.

Rogers, W., 1994. Regression standard errors in clustered samples. Stata technical bulletin 3 (13).

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

Shapiro, S. S., Wilk, M. B., 1964. An analysis of variance test for normality(complete samples). Ph.D. thesis, JSTOR.

Shi, R., Xu, Z., Chen, Z., Huang, J., 2012. Does attention affect individual investors investment return? China Finance Review International 2 (2), 143–162.

Shu, J., Zhang, J. E., 2006. Testing range estimators of historical volatility. Journal of Futures Markets 26 (3), 297–313.

Spiess, D. K., Affleck-Graves, J., 1995. Underperformance in long-run stock returns following seasoned equity offerings. Journal of Financial Economics 38 (3), 243–267.

StataCorp, 2013. Base Reference Manual. Stata Press, College Station, TX, stata 13 Edition.

Stern, R. L., Bornstein, P., 1985. Why new issues are lousy investments. Forbes 136, 152–190.

Suhoy, T., 2009. Query indices and a 2008 downturn: Israeli data. Research Department, Bank of Israel.

van Themaat, S. V., 2012. Daily Online Search Volume As A Timely Measure For Investor Attention: Predicting Daily Abnormal Returns. Erasmus University.

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

Vosen, S., Schmidt, T., 2011. Forecasting private consumption: survey-based indicators vs. google trends. Journal of Forecasting 30 (6), 565–578.

Watanabe, T., Ubukata, M., 2009. Option pricing using realized volatility and arch. Management Science 44, 1218–1233.

Webb, G. K., 2009. Internet search statistics as a source of business intelligence: Searches on foreclosure as an estimate of actual home foreclosures. Issues in Information Systems 10 (2), 82–87.

Welsch, R. E., Kuh, E., 1977. Linear regression diagnostics.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: Journal of the Econometric Society, 817–838.

Wooldridge, J., 2009. Introductory econometrics: A modern approach, 4th Edition. Cengage Learning.

Wu, L., Brynjolfsson, E., 2009. The future of prediction: How google searches foreshadow housing prices and sales. Available at SSRN 2022293.

Zhang, W., Shen, D., Zhang, Y., Xiong, X., Jul. 2013. Open source information, investor attention, and asset pricing. Economic Modelling 33, 613–619.

Zhou, X., Ye, J., Feng, Y., Mar. 2011. Tuberculosis surveillance by analyzing google trends. IEEE transactions on bio-medical engineering.

Appendix A

Complementary tables to Chapter 3

Table A.1: List of companies in DJIA sample and search queries

List of the companies in the DJIA sample (1), their corresponding tickers (2), employed search queries (3) and number of successfully obtained *GSV* observations (4).

Company (1)	Ticker (2)	Search query (3)	N (4)
3M Company	MMM	3M	2516
The Boeing Company	BA	Boeing	2516
Caterpillar Incorporated	CAT	Caterpillar	2516
Coca-Cola Company	KO	Coca Cola	2515
E. I. du Pont de Nemours and Company	DD	DuPont	2516
Exxon Mobil Company	XOM	Exxon	2428
General Electric Company	GE	GE	2516
Home Depot Incorporated	HD	Home Depot	2516
Intel Corporation	INTC	Intel	2516
International Business Machines	IBM	IBM	2516
Johnson & Johnson	JNJ	Johnson Johnson	2516
J.P. Morgan Chase	JPM	JP Morgan	2493
McDonald’s Corporation	MCD	McDonalds	2516
Merck & Co., Inc.	MRK	Merck	2516
Microsoft Corporation	MSFT	Microsoft	2516
Procter & Gamble Company	PG	P&G	2484
United Technologies Corporation	UTX	UTC	2516
Wal-Mart Stores Incorporated	WMT	WalMart	2516
Walt Disney Company	DIS	Disney	2516

Table A.2: Variable definition - DJIA

Definition of variable used in Chapter 5.

Variable	Definition
GSV	Original Google search volume for given keyword
$ASVI$	The log of *GSV* for given day minus the log of median *GSV* during the previous two days
$ASVI^{sq}$	Squared *ASVI*
$VOLUME$	Daily log trading volume in US dollars
$VOLATILITY$	Log of Garman-Klass daily volatility
$RETURN$	Daily log return
PRE	Dummy variable that takes value of one for days in interval $\langle 5,\ January\ 2004;\ 30,\ Novemver\ 2007\rangle$ and zero otherwise
CRI	Dummy variable that takes value of one for days in interval $\langle 3,\ December\ 2007;\ 30,\ June\ 2009\rangle$ and zero otherwise
$POST$	Dummy variable that takes value of one for days in interval $\langle 1,\ July\ 2009;\ 31,\ December\ 2013\rangle$ and zero otherwise
$ASVI^{PRE}$	Interaction variable that takes value of $ASVI$ for days in interval $\langle 5,\ January\ 2004;\ 30,\ Novemver\ 2007\rangle$ and zero otherwise
$ASVI^{CRI}$	Interaction variable that takes value of $ASVI$ for days in interval $\langle 3,\ December\ 2007;\ 30,\ June\ 2009\rangle$ and zero otherwise
$ASVI^{POST}$	Interaction variable that takes value of $ASVI$ for days in interval $\langle 1,\ July\ 2009;\ 31,\ December\ 2013\rangle$ and zero otherwise
$SENTIMENT$	Monthly time-varying aggregate market sentiment orthogonalized with respect to a set of macroeconomic conditions developed by Baker and Wurgler (2006)
$POSSENT$	Dummy variable that takes value of one if the level of $SENTIMENT$ exceeds the third quartile and zero otherwise
$NOSENT$	Dummy variable that takes value of one if the level of $SENTIMENT$ is between the first and the third quartile and zero otherwise
$NEGSENT$	Dummy variable that takes value of one if the level of $SENTIMENT$ is below the first quartile and zero otherwise
$ASVI \times SENT$	$ASVI$ and $SENTIMENT$ interaction variable
$ASVI^{POSSENT}$	Interaction variable that takes value of $ASVI$ if the level of $SENTIMENT$ exceeds the third quartile and zero otherwise
$ASVI^{NOSENT}$	Interaction variable that takes value of $ASVI$ if the level of $SENTIMENT$ is between the first and the third quartile and zero otherwise
$ASVI^{NEGSENT}$	Interaction variable that takes value of $ASVI$ if the level of $SENTIMENT$ is below the first quartile and zero otherwise
$ASVI \times RET$	$ASVI$ and $RETURN$ interaction variable
$ASVI^{POSRET}$	Interaction variable that takes value of $ASVI$ if previous day return was positive and zero otherwise
$ASVI^{NEGRET}$	Interaction variable that takes value of $ASVI$ if previous day return was negative and zero otherwise

Table A.3: IPO sample statistics

Statistics on the availability of daily search queries for IPOs. Column (2) shows a number of IPOs listed in Kenney and Patton (2013) for given year (1). Column (3) displays how many of these IPOs have the daily *GSV* for company name available on Google Trends. Column (4) presents similar information in percentage of IPOs in given year.

Year (1)	Number of IPOs (2)	Available query (3)	% of IPOs with available query (4)
2004	128	7	5%
2005	131	14	11%
2006	124	20	16%
2007	91	11	12%
2008	10	3	30%
2009	15	9	60%
2010	48	11	23%
Total	**547**	**75**	**14%**

Table A.4: Variable definition - IPOs

Definition of variable used in Chapter 6.

Variable	Definition
GSV	Original Google search volume for given keyword
$ASVI$	The log of *GSV* for given day minus the log of median *GSV* during previous 26 days
IR^{st}	Log initial return of IPO calculated from the offering price to the first day closing price
$LR^{(1)}$	Log cumulative return calculated from the first day closing price to the closing price one year after IPO
$LR^{(2)}$	Log cumulative return calculated from the first day closing price to the closing price half a year after IPO
$LR^{(3)}$	Log cumulative return calculated from the first day closing price to the closing price quarter the year after IPO
$LR^{(4)}$	Log cumulative return calculated from the closing price one month after IPO to the closing price one year after IPO
$LR^{(5)}$	Log cumulative return calculated from the closing price one month after IPO to the closing price half a year after IPO
TD_i^{st}	True discount of IPO defined as in Ma and Tsai (2002). $TD = \frac{P_e - P_o}{P_o}$, where P_o is the offering price and P_e is the so-called equilibrium price - in this case the average price between $t + 150$ and $t + 180$, where t is the IPO date.
MR_i^{st}	Market reaction to IPO defined as in Ma and Tsai (2002). $MR = \frac{P_m - P_e}{P_o}$ where P_o is the offering price, P_m is the first day closing price and P_e is the so-called equilibrium price - in this case the average price between $t + 150$ and $t + 180$, where t is the IPO date.
$POSSENT$	Dummy variable that takes value of one if the level of $SENTIMENT$ exceeds the third quartile and zero otherwise
$NOSENT$	Dummy variable that takes value of one if the level of $SENTIMENT$ is between the first and the third quartile and zero otherwise
$NEGSENT$	Dummy variable that takes value of one if the level of $SENTIMENT$ is below the first quartile and zero otherwise
$ASVI \times SENT$	$ASVI$ and $SENTIMENT$ interaction variable
$ASVI^{POSSENT}$	Interaction variable that takes value of $ASVI$ if the level of $SENTIMENT$ exceeds the third quartile and zero otherwise
$ASVI^{NOSENT}$	Interaction variable that takes value of $ASVI$ if the level of $SENTIMENT$ is between the first and the third quartile and zero otherwise
$ASVI^{NEGSENT}$	Interaction variable that takes value of $ASVI$ if the level of $SENTIMENT$ is below the first quartile and zero otherwise
$ASVI \times IR$	$ASVI$ and IR interaction variable
$Offering\ size$	Log size of the offering measured in US dollars
$NYSE$	Dummy variable that take one if the offering emits its shares at NYSE and zero if it emits its shares at NASDAQ
$Crisis$	Dummy variable that takes value of one for days in interval $\langle 3,\ December\ 2007;\ 30,\ June\ 2009 \rangle$ and zero otherwise
$Sentiment$	Monthly time-varying aggregate market sentiment orthogonalized with respect to a set of macroeconomic conditions developed by Baker and Wurgler (2006)
$\triangle Sentiment$	Month on month difference in time-varying aggregate market sentiment orthogonalized with respect to a set of macroeconomic conditions developed by Baker and Wurgler (2006)

Appendix B

Complementary tables to Chapter 4

Table B.1: Augmented Dickey Fuller test results

ADF test statistics. All p-values are <<1%, so they are not listed for the sake of brevity.

	ASVI	VOLUME	VOLATILITY	RETURNS
3M	-20.262	-9.605	-8.339	-21.303
Boeing	-23.971	-10.437	-8.401	-20.977
Caterpillar	-26.265	-8.597	-7.372	-20.451
Coca-Cola	-25.704	-9.271	-9.068	-21.900
DuPont	-25.677	-8.363	-7.276	-20.680
Exxon Mobil	-25.485	-6.247	-7.825	-21.802
General Electric	-26.200	-5.580	-6.386	-20.156
Home Depot	-23.442	-6.995	-7.264	-21.910
IBM	-23.174	-8.970	-7.963	-21.952
Intel	-26.210	-10.138	-8.787	-21.256
J.P. Morgan	-25.848	-4.757	-5.921	-24.311
Johnson & Johnson	-23.833	-8.875	-8.507	-20.821
McDonald's	-22.403	-9.462	-7.818	-21.588
Merck	-23.808	-9.575	-8.256	-20.040
Microsoft	-22.547	-11.342	-8.395	-22.074
Procter & Gamble	-26.955	-8.239	-8.726	-21.781
United Technologies	-27.187	-10.229	-8.273	-22.144
Wal-Mart	-20.31	-7.222	-8.371	-22.364
Walt Disney	-24.395	-9.490	-8.123	-21.402

Table B.2: Robinson's test results

Estimates of *d* for ASVI, VOLUME, VOLATILITY and RETURNS by firm. Standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively.

	ASVI	VOLUME	VOLATILITY	RETURNS
3M	0.00	0.36***	0.25***	-0.04**
	(0.17)	(17.41)	(12.39)	(-2.01)
Boeing	-0.07***	0.33***	0.26***	0.01
	(-3.1)	(16.79)	(13.58)	(0.3)
Caterpillar	-0.24***	0.35***	0.29***	0.00
	(-11.84)	(18.19)	(14.85)	(-0.24)
Coca-Cola	-0.33***	0.36***	0.27***	-0.04**
	(-15.57)	(18.06)	(13.47)	(-2.22)
DuPont	-0.26***	0.37***	0.27***	-0.02
	(-12.99)	(19.78)	(14.12)	(-0.87)
Exxon Mobil	-0.10***	0.37***	0.29***	-0.12***
	(-4.73)	(19.34)	(14.53)	(-6.62)
General Electric	-0.24***	0.43***	0.26***	0.02
	(-12.05)	(23.72)	(12.83)	(1.02)
Home Depot	-0.22***	0.37***	0.27***	-0.04**
	(-10.04)	(18.07)	(14.13)	(-2.22)
IBM	-0.13***	0.37***	0.28***	-0.01
	(-6.03)	(19.78)	(14.49)	(-0.46)
Intel	-0.20***	0.39***	0.26***	-0.06***
	(-9.71)	(20.3)	(13.13)	(-3.13)
J.P. Morgan	-0.17***	0.43***	0.34***	-0.12***
	(-8.4)	(22.91)	(17.09)	(-6.39)
Johnson & Johnson	-0.19***	0.39***	0.32***	-0.06***
	(-9.04)	(20.49)	(15.58)	(-3.24)
McDonald's	-0.11***	0.40***	0.27***	-0.10***
	(-5.61)	(19.8)	(14.03)	(-4.94)
Merck	-0.14***	0.45***	0.26***	-0.03*
	(-6.89)	(22.38)	(13.56)	(-1.77)
Microsoft	-0.01	0.38***	0.25***	-0.06***
	(-0.64)	(19.71)	(12.76)	(-2.97)
Procter & Gamble	-0.28***	0.35***	0.26***	-0.07***
	(-13.34)	(18.64)	(12.57)	(-3.71)
United Technologies	-0.20***	0.38***	0.25***	-0.06***
	(-9.64)	(19.14)	(12.57)	(-3.05)
Wal-Mart	0.08***	0.37***	0.24***	-0.05***
	(4.09)	(18.2)	(12.09)	(-2.67)
Walt Disney	-0.18***	0.40***	0.27***	-0.05**
	(-8.51)	(20.48)	(14.49)	(-2.46)

Appendix C

Complementary tables to Chapter 5

C.1 Time series setting

C.1.1 Trading volume

Table C.1: Pearson cross correlation coefficients for lag(0)

Cross correlation for $ASVI_{C_i}$ and $VOLUME_{C_j}$, where $C_i$and C_j denote specific firm. $VOLUME$ and $ASVI$ are defined in Table A.2. First column (1) show correlation of trading volume with firms own ASVI ($i = j$), second column (2) show median correlation on reshuffled data ($i \neq j$). The star denote 5% significance.

ρ	$i = j$ (1)	$i \neq j$(median) (2)
3M	4.60%*	2.30%
Boeing	10.15%*	2.59%
Caterpillar	6.87%*	2.11%
Coca-Cola	0.25%	0.73%
DuPont	3.54%	1.86%
Exxon Mobil	5.22%*	1.25%
General Electric	2.02%	0.33%
Home Depot	-8.01%*	1.87%
IBM	6.89%*	1.11%
Intel	4.23%*	2.39%
J.P. Morgan	4.72%*	1.32%
Johnson & Johnson	6.44%*	2.36%
McDonald's	6.60%*	2.88%
Merck	9.01%*	1.24%
Microsoft	-1.11%	0.20%
Procter & Gamble	1.30%	0.73%
United Technologies	4.30%*	1.31%
Wal-Mart	-1.14%	2.52%
Walt Disney	-2.54%	1.92%
Median	**4.30%**	**1.86%**

Table C.2: Trading volume and ASVI

The dependent variable in each regression is trading volume ($VOLUME_t^{st}$). $VOLUME_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations, and two bottom lines show % change in $R^2$and $adjR^2$against restricted model with $\beta_6 = 0$. T-statistics computed from Newey-West standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. Outlying values of ASVI were omitted. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{st}$	0.12***	0.15***	0.08***	0.01	0.06***	0.07***	0.05***	-0.07***	0.11***	0.06***	0.05***	0.10***	0.08***	0.10***	0.04**	0.03**	0.05**	-0.09***	-0.02
	(5.33)	(6.73)	(4.98)	(0.91)	(3.06)	(4.60)	(3.36)	(-4.86)	(6.69)	(2.86)	(4.20)	(5.31)	(5.07)	(4.40)	(2.04)	(2.03)	(2.50)	(-5.35)	(-1.12)
$VOLUME_{t-1}^{st}$	0.47***	0.46***	0.50***	0.46***	0.48***	0.47***	0.54***	0.51***	0.52***	0.52***	0.56***	0.47***	0.52***	0.54***	0.51***	0.46***	0.48***	0.50***	0.51***
	(19.86)	(23.69)	(27.12)	(23.33)	(23.15)	(19.12)	(24.27)	(25.45)	(24.01)	(22.70)	(28.17)	(21.47)	(22.24)	(28.02)	(24.89)	(18.16)	(23.05)	(23.99)	(23.18)
$VOLUME_{t-2}^{st}$	0.14***	0.12***	0.10***	0.14***	0.14***	0.16***	0.13***	0.12***	0.10***	0.09***	0.13***	0.12***	0.10***	0.10***	0.08***	0.13***	0.11***	0.13***	0.09***
	(6.11)	(5.43)	(5.66)	(7.12)	(6.06)	(7.70)	(5.98)	(5.82)	(4.88)	(4.73)	(6.24)	(5.16)	(4.31)	(5.27)	(3.83)	(5.10)	(5.34)	(6.64)	(4.64)
$VOLUME_{t-3}^{st}$	0.07**	0.11***	0.10***	0.09***	0.09***	0.10***	0.11***	0.11***	0.06**	0.04*	0.09***	0.13***	0.06**	0.06***	0.07***	0.10***	0.12***	0.12***	0.05**
	(2.38)	(5.42)	(5.07)	(3.80)	(3.98)	(4.78)	(4.42)	(4.63)	(2.51)	(1.65)	(4.42)	(5.76)	(2.00)	(3.15)	(3.52)	(4.04)	(4.96)	(4.89)	(2.30)
$VOLUME_{t-4}^{st}$	0.05***	0.06***	0.09***	0.00	0.06**	0.06***	0.03	0.08***	0.05**	0.07***	0.06**	0.05**	0.05**	0.05**	0.02	0.03	0.02	0.04*	0.05**
	(2.90)	(2.63)	(4.83)	(0.23)	(2.54)	(2.63)	(1.29)	(3.94)	(2.46)	(3.07)	(2.46)	(2.09)	(2.31)	(2.17)	(0.90)	(1.34)	(0.97)	(1.95)	(2.11)
$VOLUME_{t-5}^{st}$	0.09***	0.05***	0.06***	0.13***	0.08***	0.12***	0.14***	0.09***	0.12***	0.09***	0.12***	0.07***	0.10***	0.09***	0.08***	0.12***	0.08***	0.11***	0.11***
	(3.78)	(2.60)	(3.11)	(6.99)	(3.58)	(4.67)	(6.93)	(3.92)	(5.70)	(3.30)	(6.54)	(3.94)	(4.88)	(4.58)	(4.29)	(6.53)	(3.51)	(6.27)	(5.86)
Constant	-0.00	-0.00	-0.00	0.00	-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
	(-0.07)	(-0.03)	(-0.07)	(0.02)	(-0.01)	(0.57)	(0.04)	(-0.09)	(-0.08)	(-0.08)	(-0.23)	(0.03)	(-0.06)	(-0.30)	(-0.10)	(0.06)	(-0.03)	(-0.15)	(0.00)
N	2508	2502	2511	2510	2511	2402	2511	2511	2511	2511	2476	2509	2511	2506	2511	2471	2505	2508	2508
R^2	0.487	0.475	0.566	0.503	0.566	0.688	0.795	0.683	0.565	0.481	0.837	0.552	0.508	0.562	0.414	0.543	0.49	0.632	0.494
$adjR^2$	0.486	0.474	0.565	0.502	0.565	0.687	0.795	0.682	0.564	0.479	0.837	0.551	0.507	0.561	0.412	0.542	0.489	0.631	0.493
$\%\triangle R^2$ to restricted model without *ASVI*	1.88%	3.49%	1.25%	0.00%	0.53%	0.73%	0.25%	0.74%	1.99%	0.63%	0.24%	1.47%	1.20%	1.44%	0.24%	0.18%	0.41%	0.64%	0.00%
$\%\triangle adjR^2$ to restricted model without *ASVI*	1.89%	3.49%	1.25%	0.00%	0.53%	0.73%	0.25%	0.74%	1.99%	0.63%	0.24%	1.47%	1.20%	1.45%	0.24%	0.00%	0.41%	0.64%	0.00%

Table C.3: Trading volume, ASVI, volatility and returns

The dependent variable in each regression is trading volume ($VOLUME_t^{st}$). $VOLUME_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. T-statistics computed from Newey-West standard errors are in parentheses. Outlying values of ASVI were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{st}$	0.09***	0.11***	0.07***	0.01	0.05***	0.06***	0.04***	-0.07***	0.09***	0.04***	0.04***	0.08***	0.07***	0.08***	0.03*	0.03**	0.04**	-0.07***	-0.01
	(4.54)	(5.86)	(5.26)	(0.62)	(3.28)	(4.56)	(3.29)	(-5.30)	(6.73)	(2.60)	(4.06)	(5.15)	(4.46)	(4.27)	(1.89)	(2.11)	(2.15)	(-5.24)	(-0.85)
$VOLUME_{t-1}^{st}$	0.35***	0.35***	0.38***	0.38***	0.38***	0.38***	0.44***	0.43***	0.43***	0.44***	0.46***	0.40***	0.44***	0.48***	0.43***	0.40***	0.40***	0.40***	0.40***
	(17.19)	(19.91)	(20.26)	(20.69)	(19.06)	(17.28)	(21.86)	(21.70)	(21.04)	(19.84)	(25.52)	(19.55)	(16.90)	(23.79)	(23.11)	(15.73)	(20.51)	(22.73)	(19.19)
$VOLUME_{t-2}^{st}$	0.09***	0.08***	0.07***	0.11***	0.11***	0.13***	0.10***	0.09***	0.07***	0.08***	0.09***	0.10***	0.08***	0.08***	0.06***	0.12***	0.08***	0.09***	0.07***
	(4.49)	(4.37)	(4.45)	(6.45)	(5.55)	(6.16)	(4.76)	(4.76)	(3.99)	(4.52)	(5.62)	(4.76)	(3.72)	(4.52)	(3.34)	(5.03)	(4.22)	(5.45)	(4.18)
$VOLUME_{t-3}^{st}$	0.05**	0.07***	0.08***	0.09***	0.06***	0.08***	0.07***	0.09***	0.03	0.02	0.08***	0.11***	0.05*	0.05***	0.05***	0.09***	0.09***	0.10***	0.03*
	(2.24)	(3.80)	(4.09)	(4.17)	(3.30)	(4.26)	(3.72)	(4.24)	(1.54)	(0.98)	(4.55)	(4.88)	(1.74)	(2.66)	(3.22)	(3.72)	(4.23)	(4.73)	(1.71)
$VOLUME_{t-4}^{st}$	0.00	0.04**	0.06***	-0.01	0.03	0.03*	0.01	0.06***	0.02	0.05**	0.04**	0.04*	0.03*	0.05***	0.00	0.03	-0.01	0.03	0.03
	(0.22)	(2.06)	(3.58)	(-0.59)	(1.53)	(1.75)	(0.65)	(3.42)	(1.06)	(2.42)	(2.03)	(1.70)	(1.75)	(2.62)	(0.01)	(1.25)	(-0.29)	(1.35)	(1.21)
$VOLUME_{t-5}^{st}$	0.04*	0.03*	0.02	0.09***	0.05**	0.09***	0.10***	0.06***	0.08***	0.07***	0.09***	0.06***	0.07***	0.08***	0.06***	0.09***	0.03*	0.08***	0.08***
	(1.86)	(1.75)	(1.50)	(5.51)	(2.37)	(4.14)	(5.82)	(2.90)	(4.07)	(2.95)	(6.07)	(3.38)	(3.77)	(4.37)	(3.38)	(4.76)	(1.66)	(4.39)	(4.48)
$VOLATILITY_t^{st}$	0.43***	0.39***	0.38***	0.34***	0.35***	0.30***	0.30***	0.28***	0.35***	0.31***	0.26***	0.29***	0.30***	0.26***	0.34***	0.27***	0.37***	0.32***	0.35***
	(22.17)	(15.81)	(16.93)	(14.37)	(14.14)	(16.00)	(17.19)	(14.86)	(16.29)	(13.31)	(17.52)	(16.87)	(13.81)	(9.71)	(12.46)	(14.94)	(23.33)	(15.98)	(16.41)
$RETURN_t^{st}$	-0.06**	-0.01	-0.03	-0.02	-0.02	-0.03**	-0.02	0.02	-0.01	-0.04	-0.01	-0.01	0.01	-0.02	-0.01	-0.03	-0.04**	-0.03	0.02
	(-2.15)	(-0.51)	(-1.30)	(-0.95)	(-1.16)	(-2.40)	(-1.61)	(1.33)	(-0.52)	(-1.51)	(-0.76)	(-0.81)	(0.27)	(-0.89)	(-0.49)	(-1.49)	(-2.43)	(-1.39)	(0.67)
Constant	-0.00	-0.00	-0.00	0.00	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
	(-0.05)	(-0.02)	(-0.05)	(0.02)	(0.01)	(0.48)	(0.04)	(-0.03)	(-0.05)	(-0.05)	(0.00)	(0.01)	(-0.02)	(-0.13)	(-0.07)	(0.08)	(-0.05)	(-0.04)	(-0.00)
N	2508	2502	2511	2510	2511	2402	2511	2511	2511	2511	2476	2509	2511	2506	2511	2471	2505	2508	2508
R^2	0.620	0.596	0.671	0.600	0.654	0.749	0.845	0.737	0.653	0.565	0.871	0.620	0.586	0.624	0.517	0.607	0.596	0.710	0.590
$adjR^2$	0.619	0.595	0.670	0.599	0.653	0.748	0.844	0.736	0.652	0.563	0.871	0.619	0.584	0.623	0.515	0.606	0.595	0.709	0.589

Table C.4: Trading volume and ASVI: time variation

The dependent variable in each regression is trading volume ($VOLUME_t^{st}$). $VOLUME_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. Three bottom lines present F-test p-values for *ASVI* slope-dummy variable equality, *ASVI* slope-dummy variable joint significance and level-dummy variable equality; respectively. T-statistics computed from Newey-West standard errors are in parentheses. Outlying values of *ASVI* were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{CRI,st}$	0.08***	0.13***	0.05***	0.00	0.03	0.06***	0.01	-0.06***	0.11***	0.03	0.02*	0.08***	0.09***	0.09***	0.05**	0.03	0.03	-0.07**	-0.00
	(3.02)	(4.53)	(3.06)	(0.09)	(1.13)	(4.00)	(0.95)	(-3.22)	(5.04)	(1.21)	(1.66)	(3.26)	(3.64)	(2.85)	(2.18)	(1.11)	(1.03)	(-2.09)	(-0.15)
$ASVI_t^{CRI,st}$	0.18**	0.14*	0.05*	0.02	0.16***	0.05	0.14***	-0.03	0.13***	0.12***	0.12***	0.12**	0.06**	0.10**	0.06	0.07**	0.05	-0.09**	-0.07*
	(2.30)	(1.73)	(1.84)	(0.55)	(2.74)	(1.53)	(2.70)	(-0.85)	(4.25)	(2.76)	(3.95)	(2.09)	(2.52)	(2.56)	(1.64)	(2.32)	(1.20)	(-2.01)	(-1.87)
$ASVI_t^{POST,st}$	0.14***	0.15***	0.22***	0.03	0.07**	0.10***	0.06**	-0.12***	0.09***	0.10**	0.06***	0.11***	0.06***	0.08**	0.01	0.02	0.07***	-0.09***	-0.02
	(4.47)	(4.63)	(6.20)	(0.91)	(2.07)	(2.59)	(2.57)	(-4.73)	(3.74)	(2.51)	(2.95)	(4.41)	(2.76)	(2.03)	(0.53)	(0.67)	(2.85)	(-4.88)	(-0.79)
$VOLUME_{t-1}^{st}$	0.45***	0.43***	0.48***	0.44***	0.46***	0.46***	0.52***	0.50***	0.50***	0.51***	0.54***	0.46***	0.50***	0.51***	0.49***	0.44***	0.46***	0.48***	0.49***
	(20.35)	(22.35)	(26.70)	(22.14)	(21.16)	(19.27)	(23.88)	(24.50)	(23.43)	(22.59)	(26.34)	(20.80)	(20.96)	(24.50)	(24.21)	(17.95)	(22.33)	(24.82)	(22.33)
$VOLUME_{t-2}^{st}$	0.13***	0.10***	0.10***	0.13***	0.13***	0.16***	0.12***	0.12***	0.09***	0.09***	0.12***	0.11***	0.09***	0.08***	0.07***	0.12***	0.10***	0.11***	0.08***
	(5.68)	(4.47)	(5.29)	(6.59)	(5.34)	(7.41)	(5.53)	(5.65)	(4.46)	(4.49)	(5.76)	(4.76)	(3.92)	(4.36)	(3.58)	(4.66)	(4.92)	(5.81)	(4.28)
$VOLUME_{t-3}^{st}$	0.06**	0.09***	0.09***	0.07***	0.07***	0.09***	0.09***	0.10***	0.05**	0.03	0.08***	0.12***	0.05*	0.05***	0.06***	0.09***	0.11***	0.11***	0.04*
	(2.08)	(4.17)	(4.54)	(3.29)	(3.41)	(4.56)	(4.03)	(4.37)	(2.05)	(1.34)	(4.03)	(5.33)	(1.83)	(2.58)	(3.13)	(3.49)	(4.61)	(4.50)	(1.95)
$VOLUME_{t-4}^{st}$	0.04**	0.04*	0.07***	-0.01	0.04*	0.05**	0.02	0.07***	0.04*	0.06***	0.04*	0.04	0.04*	0.04	0.01	0.02	0.01	0.03	0.04*
	(2.32)	(1.80)	(3.87)	(-0.43)	(1.80)	(2.37)	(0.93)	(3.65)	(1.91)	(2.85)	(1.67)	(1.64)	(1.90)	(1.57)	(0.56)	(0.76)	(0.48)	(1.39)	(1.81)
$VOLUME_{t-5}^{st}$	0.08***	0.02	0.04**	0.11***	0.05**	0.11***	0.12***	0.08***	0.10***	0.08***	0.09***	0.06***	0.08***	0.07***	0.07***	0.10***	0.06***	0.09***	0.10***
	(3.23)	(0.79)	(2.18)	(5.95)	(2.54)	(4.35)	(5.88)	(3.42)	(4.94)	(2.96)	(4.68)	(2.89)	(4.12)	(3.70)	(3.60)	(5.57)	(2.79)	(5.16)	(5.07)
PRE	-0.06***	-0.16***	-0.12***	-0.12***	-0.15***	-0.00	-0.11***	-0.03	0.01	0.06***	-0.13***	-0.10***	-0.05	-0.15***	0.05**	-0.11***	-0.06***	0.00	-0.08***
	(-2.98)	(-5.55)	(-5.68)	(-3.56)	(-5.25)	(-0.31)	(-5.49)	(-1.30)	(0.45)	(3.23)	(-4.58)	(-3.09)	(-1.58)	(-5.00)	(1.99)	(-3.36)	(-2.91)	(0.19)	(-2.85)
CRI	0.22***	0.31***	0.17***	0.24***	0.21***	0.14***	0.13***	0.14***	0.22***	0.09***	0.15***	0.17***	0.20***	0.15***	0.17***	0.22***	0.26***	0.21***	0.21***
	(4.53)	(7.25)	(3.11)	(5.77)	(4.96)	(4.86)	(4.01)	(5.22)	(6.29)	(2.78)	(4.36)	(4.15)	(5.53)	(4.68)	(3.90)	(5.39)	(5.40)	(6.54)	(5.67)
POST	-0.03	0.03	0.04**	0.02	0.05**	-0.03	0.05***	-0.02	-0.09***	-0.09***	0.05***	0.03*	-0.03	0.07***	-0.10***	0.02	-0.04	-0.08***	-0.01
	(-0.94)	(1.56)	(2.09)	(0.85)	(2.04)	(-1.58)	(3.52)	(-1.33)	(-3.48)	(-2.79)	(3.31)	(1.69)	(-1.61)	(3.45)	(-3.82)	(1.16)	(-1.33)	(-3.30)	(-0.21)
N	2508	2502	2511	2510	2511	2402	2511	2511	2511	2511	2476	2509	2511	2506	2511	2471	2505	2508	2508
R^2	0.495	0.491	0.577	0.513	0.578	0.691	0.801	0.687	0.573	0.487	0.842	0.559	0.514	0.572	0.423	0.552	0.500	0.639	0.501
$adjR^2$	0.493	0.489	0.575	0.511	0.576	0.690	0.800	0.686	0.571	0.485	0.841	0.557	0.512	0.570	0.420	0.550	0.498	0.637	0.499
F-test for equality of ASVI slope dummies (p-value) (H_0 : *all coefficients are equal)*	0.27	0.97	0.00	0.76	0.10	0.57	0.02	0.04	0.60	0.07	0.01	0.47	0.62	0.94	0.43	0.40	0.61	0.79	0.29
F-test for joint significance of ASVI slope dummies (p-value) (H_0 : *all coefficients are equal to 0)*	0.00	0.00	0.00	0.77	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.07	0.01	0.00	0.24
F-test for equality of level dummies (p-value) (H_0 : *all coefficients are equal)*	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 C.5: Trading volume and ASVI: VAR(5)

The dependent variables are trading volume ($VOLUME_t^{st}$) in Panel A and web search activity ($ASVI_t^{st}$) in Panel B. $VOLUME_t^{st}$ and $ASVI_t^{st}$ are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. Two bottom lines in each panel present Granger-causality Wald tests with corresponding test statistics in parentheses. T-statistics are in parentheses. Outlying values of ASVI were omitted. *,**, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

Panel A - $VOLUME_t^{st}$																			
	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$VOLUME_{t-1}^{st}$	0.46***	0.46***	0.49***	0.45***	0.48***	0.47***	0.54***	0.52***	0.51***	0.52***	0.55***	0.46***	0.52***	0.52***	0.50***	0.45***	0.48***	0.50***	0.51***
	(23.14)	(22.75)	(24.64)	(22.93)	(24.03)	(22.74)	(26.92)	(26.00)	(25.46)	(26.17)	(27.58)	(22.97)	(25.98)	(26.39)	(24.97)	(22.26)	(24.12)	(24.88)	(25.59)
$VOLUME_{t-2}^{st}$	0.13***	0.10***	0.11***	0.14***	0.14***	0.15***	0.12***	0.12***	0.10***	0.09***	0.12***	0.12***	0.09***	0.09***	0.09***	0.13***	0.11***	0.12***	0.09***
	(5.84)	(4.40)	(4.86)	(6.55)	(6.22)	(6.75)	(5.51)	(5.18)	(4.54)	(4.02)	(5.39)	(5.55)	(4.00)	(3.93)	(3.96)	(5.79)	(4.79)	(5.27)	(4.15)
$VOLUME_{t-3}^{st}$	0.07***	0.12***	0.09***	0.09***	0.09***	0.10***	0.11***	0.10***	0.06***	0.03	0.10***	0.12***	0.06***	0.07***	0.07***	0.10***	0.12***	0.12***	0.05**
	(3.15)	(5.25)	(4.17)	(4.19)	(4.08)	(4.42)	(4.77)	(4.62)	(2.66)	(1.41)	(4.40)	(5.65)	(2.66)	(3.03)	(3.15)	(4.63)	(5.65)	(5.21)	(2.30)
$VOLUME_{t-4}^{st}$	0.07***	0.06***	0.10***	0.01	0.06***	0.05**	0.04*	0.07***	0.07***	0.07***	0.06**	0.07***	0.05**	0.06***	0.03	0.04*	0.02	0.05**	0.06**
	(3.08)	(2.90)	(4.55)	(0.26)	(2.68)	(2.38)	(1.71)	(3.25)	(2.96)	(3.16)	(2.43)	(3.07)	(2.08)	(2.91)	(1.24)	(1.83)	(1.12)	(2.20)	(2.51)
$VOLUME_{t-5}^{st}$	0.08***	0.05***	0.05***	0.13***	0.08***	0.12***	0.14***	0.09***	0.10***	0.09***	0.12***	0.07***	0.10***	0.10***	0.07***	0.13***	0.07***	0.11***	0.11***
	(4.08)	(2.60)	(2.62)	(6.48)	(4.15)	(5.65)	(6.84)	(4.37)	(5.21)	(4.51)	(5.78)	(3.49)	(5.02)	(4.95)	(3.68)	(6.25)	(3.75)	(5.45)	(5.44)
$ASVI_{t-1}^{st}$	0.06***	0.02	0.03**	0.07***	0.01	0.01	0.02*	0.02	0.05***	-0.03**	0.02***	0.07***	0.02	0.05***	0.06***	0.06***	0.04**	-0.01	0.03*
	(3.60)	(1.11)	(2.18)	(4.81)	(0.61)	(0.75)	(1.85)	(1.57)	(3.11)	(-2.17)	(2.66)	(4.43)	(1.01)	(3.39)	(3.51)	(4.19)	(2.51)	(-0.65)	(1.75)
$ASVI_{t-2}^{st}$	-0.02	0.02	-0.03*	0.01	0.02	-0.01	0.02	0.01	-0.01	-0.01	0.00	-0.02	0.00	0.01	-0.01	0.03*	-0.01	0.01	0.01
	(-1.39)	(1.00)	(-1.92)	(0.88)	(1.28)	(-0.64)	(1.64)	(1.11)	(-0.36)	(-0.76)	(0.23)	(-1.16)	(0.29)	(1.04)	(-0.59)	(1.79)	(-0.72)	(0.38)	(0.91)
$ASVI_{t-3}^{st}$	0.04***	0.01	0.00	0.04**	0.01	-0.00	0.02	0.02	0.02	0.01	0.00	0.03	-0.01	0.03**	0.04**	0.05***	-0.01	-0.03*	-0.01
	(2.78)	(0.56)	(0.01)	(2.47)	(0.74)	(-0.17)	(1.54)	(1.23)	(1.07)	(0.88)	(0.46)	(1.63)	(-0.32)	(2.23)	(2.04)	(2.80)	(-0.93)	(-1.72)	(-0.62)
$ASVI_{t-4}^{st}$	-0.02	0.00	-0.00	-0.01	0.02	0.01	0.01	0.01	-0.02	-0.00	0.00	-0.03**	0.00	0.01	-0.03*	-0.02	-0.02	0.01	0.02
	(-1.54)	(0.05)	(-0.14)	(-0.36)	(1.06)	(0.64)	(1.13)	(0.79)	(-1.26)	(-0.06)	(0.31)	(-1.99)	(0.31)	(0.42)	(-1.68)	(-1.01)	(-1.13)	(0.78)	(1.11)
$ASVI_{t-5}^{st}$	0.01	-0.00	-0.03**	0.01	-0.02	0.01	-0.02**	-0.01	0.01	0.05***	-0.00	-0.02	-0.01	-0.04**	0.01	0.01	-0.03*	-0.03**	-0.03*
	(0.63)	(-0.11)	(-2.01)	(0.64)	(-1.54)	(0.42)	(-2.17)	(-0.48)	(0.35)	(3.34)	(-0.17)	(-1.56)	(-0.67)	(-2.57)	(0.42)	(0.88)	(-1.87)	(-2.00)	(-1.73)
Constant	-0.00	-0.00	-0.00	0.00	-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
	(-0.03)	(-0.09)	(-0.04)	(0.07)	(-0.00)	(0.80)	(0.03)	(-0.10)	(-0.10)	(-0.11)	(-0.22)	(-0.04)	(-0.07)	(-0.37)	(-0.11)	(0.02)	(-0.04)	(-0.14)	(-0.00)
Granger-causality ($ASVI \rightarrow VOLUME$)	0.00***	0.63	0.02**	0.00***	0.26	0.82	0.01***	0.29	0.02**	0.00***	0.19	0.00***	0.86	0.00***	0.00***	0.00***	0.02**	0.30	0.10
	(18.33)	(3.46)	(13.69)	(25.05)	(6.54)	(2.21)	(15.79)	(6.18)	(13.15)	(20.61)	(7.50)	(30.73)	(1.90)	(27.37)	(18.39)	(26.16)	(13.74)	(6.10)	(9.15)
N	2506	2500	2509	2503	2509	2313	2509	2509	2509	2509	2435	2507	2509	2504	2509	2398	2503	2506	2506

	Panel B - $ASVI_t^{st}$																		
	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$VOLUME_{t-1}^{st}$	0.02	-0.02	0.04	0.04	0.00	0.06*	0.07*	-0.08**	0.00	-0.02	0.09**	0.04	0.03	0.05**	-0.08***	0.06**	0.04	0.02	0.03
	(0.96)	(-0.88)	(1.29)	(1.38)	(0.07)	(1.69)	(1.76)	(-2.48)	(0.03)	(-0.73)	(2.02)	(1.47)	(1.13)	(2.01)	(-3.77)	(2.28)	(1.44)	(0.85)	(1.39)
$VOLUME_{t-2}^{st}$	-0.07***	-0.03	0.01	-0.02	0.00	-0.06	-0.06	-0.02	0.02	-0.05*	-0.08	-0.00	-0.02	-0.05	0.06**	-0.04	-0.06**	0.07**	-0.06**
	(-2.95)	(-1.10)	(0.47)	(-0.68)	(0.02)	(-1.58)	(-1.19)	(-0.51)	(0.59)	(-1.74)	(-1.52)	(-0.10)	(-0.62)	(-1.64)	(2.34)	(-1.30)	(-2.07)	(2.54)	(-2.29)
$VOLUME_{t-3}^{st}$	-0.01	0.01	-0.06*	0.01	0.00	-0.04	0.00	0.04	-0.01	-0.05*	0.04	-0.07**	-0.03	-0.01	-0.02	-0.00	0.01	-0.03	0.03
	(-0.23)	(0.42)	(-1.95)	(0.33)	(0.12)	(-1.12)	(0.08)	(1.06)	(-0.36)	(-1.78)	(0.79)	(-2.40)	(-1.15)	(-0.31)	(-1.00)	(-0.00)	(0.26)	(-0.98)	(1.10)
$VOLUME_{t-4}^{st}$	0.05**	0.00	0.07**	0.05*	-0.01	0.01	0.08*	0.06*	0.05	0.06**	-0.03	0.03	-0.03	0.02	0.05*	0.02	0.01	-0.04	0.02
	(2.25)	(0.00)	(2.14)	(1.77)	(-0.41)	(0.29)	(1.83)	(1.82)	(1.56)	(2.19)	(-0.50)	(0.98)	(-1.13)	(0.65)	(1.84)	(0.77)	(0.36)	(-1.41)	(0.66)
$VOLUME_{t-5}^{st}$	-0.05**	-0.00	-0.06**	-0.08***	-0.03	0.01	-0.14***	-0.01	-0.11***	0.03	-0.04	-0.02	0.03	-0.10***	-0.06**	-0.06**	0.01	0.02	-0.03
	(-2.39)	(-0.08)	(-2.27)	(-2.86)	(-1.10)	(0.21)	(-3.44)	(-0.25)	(-4.04)	(0.99)	(-0.91)	(-0.79)	(1.20)	(-3.66)	(-2.56)	(-2.05)	(0.23)	(0.76)	(-1.38)
$ASVI_{t-1}^{st}$	0.06***	0.24***	-0.12***	-0.08***	-0.07***	-0.04*	-0.07***	0.11***	0.10***	0.01	-0.06***	-0.01	0.19***	0.08***	0.31***	-0.21***	-0.04**	0.32***	0.02
	(3.79)	(13.86)	(-5.80)	(-3.81)	(-3.25)	(-1.84)	(-3.64)	(5.38)	(5.04)	(0.62)	(-3.29)	(-0.30)	(9.44)	(4.03)	(15.49)	(-10.35)	(-2.14)	(18.33)	(0.99)
$ASVI_{t-2}^{st}$	-0.22***	-0.32***	-0.35***	-0.37***	-0.37***	-0.35***	-0.36***	-0.41***	-0.39***	-0.34***	-0.34***	-0.36***	-0.30***	-0.34***	-0.41***	-0.44***	-0.30***	-0.22***	-0.29***
	(-13.55)	(-17.71)	(-17.23)	(-18.70)	(-18.37)	(-17.34)	(-18.03)	(-20.35)	(-19.15)	(-17.00)	(-17.82)	(-18.39)	(-14.88)	(-18.00)	(-19.94)	(-21.34)	(-15.96)	(-12.87)	(-16.16)
$ASVI_{t-3}^{st}$	-0.03*	0.03*	-0.13***	-0.07***	-0.11***	-0.09***	-0.12***	-0.05**	-0.09***	-0.12***	-0.06***	-0.08***	0.04**	-0.02	-0.03	-0.21***	-0.06***	0.06***	-0.04**
	(-1.81)	(1.82)	(-6.38)	(-3.36)	(-4.99)	(-4.16)	(-5.67)	(-2.18)	(-4.16)	(-5.99)	(-3.16)	(-3.78)	(2.05)	(-0.77)	(-1.19)	(-9.66)	(-2.97)	(3.28)	(-1.97)
$ASVI_{t-4}^{st}$	0.01	-0.09***	-0.11***	-0.13***	-0.07***	-0.11***	-0.10***	-0.08***	-0.05**	-0.12***	-0.12***	-0.11***	-0.05***	-0.09***	0.03	-0.17***	-0.08***	-0.06***	-0.09***
	(0.75)	(-4.95)	(-5.41)	(-6.47)	(-3.60)	(-5.37)	(-5.18)	(-4.09)	(-2.48)	(-6.18)	(-6.59)	(-5.58)	(-2.63)	(-4.96)	(1.31)	(-8.43)	(-4.11)	(-3.32)	(-4.92)
$ASVI_{t-5}^{st}$	0.09***	0.01	0.02	0.01	-0.03*	-0.06***	0.04*	0.15***	0.12***	0.04*	-0.01	0.08***	0.06***	0.07***	0.12***	-0.03*	-0.02	0.04**	0.01
	(5.98)	(0.49)	(1.20)	(0.62)	(-1.65)	(-2.96)	(1.90)	(7.81)	(6.15)	(1.82)	(-0.41)	(4.09)	(3.00)	(3.71)	(6.29)	(-1.70)	(-1.13)	(2.39)	(0.49)
Constant	0.00	-0.01	0.00	-0.00	-0.00	0.00	-0.00	0.00	0.00	0.00	-0.00	-0.01	-0.00	-0.01	-0.00	-0.01	-0.00	0.00	-0.01
	(0.26)	(-0.55)	(0.16)	(-0.18)	(-0.00)	(0.06)	(-0.06)	(0.08)	(0.04)	(0.06)	(-0.19)	(-0.30)	(-0.06)	(-0.33)	(-0.02)	(-0.27)	(-0.05)	(0.23)	(-0.32)
Granger-causality ($VOLUME \rightarrow ASVI$)	0.00***	0.39	0.06*	0.06*	0.56	0.23	0.01***	0.03***	0.00***	0.00***	0.28	0.10*	0.23	0.00***	0.00***	0.10	0.44	0.01***	0.18
	(21.38)	(5.26)	(10.80)	(10.53)	(3.94)	(6.91)	(16.96)	(12.03)	(19.69)	(17.95)	(6.34)	(9.27)	(6.84)	(29.38)	(25.02)	(9.20)	(4.81)	(16.33)	(7.61)
N	2506	2500	2509	2503	2509	2313	2509	2509	2509	2509	2435	2507	2509	2504	2509	2398	2503	2506	2506

C.1.2 Stock price volatility

Table C.6: Pearson cross correlation coefficients for lag(0)

Cross correlation for $ASVI_{C_i}$ and $VOLATILITY_{C_j}$, where $C_i$and C_j denote specific firm. $VOLATILITY$ and $ASVI$ are defined in Table A.2. First column (1) show correlation of stock price volatility with firms own ASVI ($i = j$), second column (2) show median correlation on reshuffled data ($i \neq j$). The star denote 5% significance.

ρ	$i = j$ (1)	$i \neq j$(median) (2)
3M	2.84%	0.62%
Boeing	6.08%*	2.16%
Caterpillar	3.65%	0.97%
Coca-Cola	0.86%	0.05%
DuPont	1.19%	1.65%
Exxon Mobil	1.97%	1.18%
General Electric	1.80%	0.75%
Home Depot	-2.27%	1.08%
IBM	1.61%	0.11%
Intel	3.16%	0.33%
J.P. Morgan	3.48%	1.16%
Johnson & Johnson	3.79%	0.79%
McDonald's	2.68%	0.18%
Merck	6.64%*	0.08%
Microsoft	0.20%	0.68%
Procter & Gamble	0.39%	0.49%
United Technologies	1.88%	1.63%
Wal-Mart	-1.30%	0.53%
Walt Disney	-2.67%	0.86%
Median	**1.88%**	**0.75%**

Table C.7: Stock price volatility and ASVI

The dependent variable in each regression is stock price volatility ($VOLATILITY_t^{st}$). $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations, and two bottom lines show % change in $R^2$and $adjR^2$against restricted model with $\beta_6 = 0$. T-statistics computed from Newey-West standard errors are in parentheses. *,**, and *** represent significance at the 10%, 5%, and 1% level, respectively. Outlying values of ASVI were omitted. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{st}$	0.06***	0.09***	0.05**	0.02	0.02	0.03*	0.04**	-0.01	0.03**	0.03**	0.04***	0.05***	0.03*	0.08***	0.02	0.01	0.02	-0.02	-0.02
	(3.31)	(5.25)	(2.57)	(0.91)	(1.20)	(1.85)	(2.39)	(-1.02)	(2.56)	(2.03)	(2.90)	(3.52)	(1.94)	(4.07)	(1.19)	(0.67)	(1.25)	(-0.90)	(-1.03)
$VOLATILITY_{t-1}^{st}$	0.29***	0.29***	0.32***	0.30***	0.30***	0.33***	0.29***	0.28***	0.28***	0.28***	0.37***	0.31***	0.30***	0.30***	0.26***	0.28***	0.25***	0.25***	0.34***
	(11.38)	(12.65)	(15.51)	(13.22)	(13.64)	(12.66)	(11.75)	(12.13)	(11.40)	(11.82)	(14.70)	(12.99)	(14.68)	(14.24)	(11.03)	(11.75)	(11.17)	(9.94)	(12.84)
$VOLATILITY_{t-2}^{st}$	0.18***	0.18***	0.18***	0.19***	0.16***	0.17***	0.17***	0.19***	0.18***	0.16***	0.13***	0.17***	0.16***	0.17***	0.14***	0.14***	0.16***	0.18***	0.13***
	(8.70)	(7.69)	(9.67)	(7.84)	(7.29)	(7.33)	(8.12)	(10.49)	(7.85)	(7.22)	(6.01)	(8.35)	(7.54)	(8.52)	(7.22)	(6.46)	(7.51)	(9.67)	(6.09)
$VOLATILITY_{t-3}^{st}$	0.12***	0.15***	0.10***	0.07***	0.14***	0.10***	0.18***	0.14***	0.16***	0.15***	0.15***	0.14***	0.11***	0.11***	0.16***	0.13***	0.14***	0.09***	0.07***
	(5.48)	(6.44)	(4.27)	(3.05)	(8.02)	(4.98)	(8.80)	(7.51)	(8.67)	(7.15)	(6.98)	(5.41)	(5.20)	(5.24)	(7.62)	(5.37)	(7.18)	(5.16)	(2.86)
$VOLATILITY_{t-4}^{st}$	0.11***	0.09***	0.11***	0.12***	0.12***	0.08***	0.11***	0.12***	0.10***	0.11***	0.14***	0.09***	0.13***	0.11***	0.14***	0.12***	0.12***	0.14***	0.15***
	(5.31)	(4.46)	(6.71)	(4.64)	(6.28)	(4.51)	(5.99)	(5.48)	(5.94)	(5.71)	(7.22)	(4.45)	(7.01)	(5.09)	(7.09)	(5.45)	(5.64)	(6.31)	(6.30)
$VOLATILITY_{t-5}^{st}$	0.11***	0.11***	0.14***	0.12***	0.13***	0.16***	0.14***	0.14***	0.12***	0.10***	0.12***	0.11***	0.13***	0.13***	0.12***	0.13***	0.15***	0.15***	0.15***
	(4.90)	(5.02)	(6.75)	(5.41)	(6.76)	(8.56)	(6.11)	(6.96)	(6.74)	(4.36)	(6.25)	(5.32)	(6.83)	(5.84)	(7.46)	(5.99)	(7.57)	(7.67)	(7.90)
Constant	-0.00	-0.00	-0.00	-0.00	-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
	(-0.07)	(-0.06)	(-0.05)	(-0.02)	(-0.07)	(0.52)	(-0.08)	(-0.11)	(-0.03)	(-0.06)	(-0.15)	(0.07)	(-0.14)	(-0.16)	(-0.00)	(0.17)	(0.02)	(-0.14)	(-0.05)
N	2508	2502	2511	2510	2511	2402	2511	2511	2511	2511	2476	2509	2511	2506	2511	2471	2505	2508	2508
R^2	0.429	0.433	0.5	0.391	0.496	0.498	0.596	0.522	0.485	0.393	0.651	0.439	0.444	0.451	0.407	0.386	0.411	0.393	0.454
$adjR^2$	0.427	0.431	0.499	0.389	0.495	0.497	0.595	0.521	0.484	0.392	0.65	0.438	0.443	0.45	0.406	0.384	0.41	0.391	0.452
$\%\triangle R^2$ to restricted model without *ASVI*	0.47%	1.64%	0.40%	0.26%	0.00%	0.20%	0.34%	0.00%	0.21%	0.26%	0.31%	0.46%	0.23%	1.35%	0.00%	0.26%	0.00%	0.00%	0.22%
$\%\triangle adjR^2$ to restricted model without *ASVI*	0.47%	1.41%	0.40%	0.00%	0.00%	0.20%	0.17%	0.00%	0.21%	0.26%	0.15%	0.46%	0.23%	1.35%	0.00%	0.00%	0.00%	-0.26%	0.00%

Table C.8: Stock price volatility, ASVI, volume and returns

The dependent variable in each regression is stock price volatility ($VOLATILITY_t^{st}$). $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. T-statistics computed from Newey-West standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. Outlying values of ASVI were omitted. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{st}$	0.01	0.04***	0.02	0.01	0.01	0.01	0.02	0.01	0.00	0.02	0.02*	0.03**	0.01	0.06***	0.02	0.01	0.01	0.00	-0.01
	(0.82)	(3.19)	(1.13)	(0.84)	(0.36)	(0.34)	(1.55)	(0.90)	(0.22)	(1.35)	(1.88)	(2.12)	(0.42)	(3.92)	(1.35)	(0.39)	(0.45)	(0.15)	(-0.63)
$VOLATILITY_{t-1}^{st}$	0.15***	0.18***	0.20***	0.20***	0.19***	0.22***	0.15***	0.19***	0.18***	0.21***	0.24***	0.23***	0.22***	0.24***	0.18***	0.21***	0.16***	0.14***	0.24***
	(6.40)	(9.30)	(10.38)	(10.51)	(9.62)	(8.85)	(7.50)	(8.99)	(7.86)	(8.59)	(10.55)	(12.24)	(12.32)	(10.96)	(8.87)	(9.49)	(8.02)	(5.87)	(9.73)
$VOLATILITY_{t-2}^{st}$	0.10***	0.12***	0.12***	0.14***	0.11***	0.11***	0.09***	0.14***	0.13***	0.12***	0.07***	0.13***	0.12***	0.14***	0.11***	0.10***	0.10***	0.12***	0.09***
	(5.66)	(6.26)	(6.88)	(6.75)	(5.47)	(4.94)	(4.97)	(7.60)	(6.53)	(5.57)	(3.58)	(6.89)	(6.41)	(7.94)	(6.75)	(5.45)	(5.49)	(7.12)	(4.71)
$VOLATILITY_{t-3}^{st}$	0.07***	0.11***	0.06***	0.04**	0.11***	0.06***	0.11***	0.10***	0.13***	0.12***	0.09***	0.12***	0.09***	0.10***	0.14***	0.10***	0.09***	0.05***	0.05**
	(4.09)	(5.61)	(3.45)	(2.26)	(6.36)	(2.86)	(5.83)	(5.34)	(7.07)	(6.21)	(5.27)	(5.39)	(4.91)	(4.62)	(7.50)	(4.87)	(4.96)	(2.83)	(2.45)
$VOLATILITY_{t-4}^{st}$	0.10***	0.06***	0.08***	0.10***	0.10***	0.05***	0.07***	0.09***	0.07***	0.10***	0.10***	0.08***	0.11***	0.11***	0.13***	0.10***	0.09***	0.09***	0.12***
	(5.14)	(3.28)	(4.70)	(4.39)	(5.48)	(3.10)	(3.94)	(4.25)	(4.32)	(5.47)	(5.51)	(4.16)	(6.54)	(5.38)	(7.60)	(5.13)	(4.69)	(4.36)	(5.71)
$VOLATILITY_{t-5}^{st}$	0.07***	0.09***	0.11***	0.10***	0.10***	0.11***	0.08***	0.11***	0.09***	0.09***	0.07***	0.09***	0.13***	0.13***	0.11***	0.11***	0.11***	0.11***	0.12***
	(3.85)	(4.70)	(5.75)	(5.05)	(5.53)	(6.13)	(4.19)	(6.18)	(4.92)	(4.10)	(4.34)	(4.83)	(7.53)	(5.97)	(7.13)	(5.72)	(6.61)	(5.65)	(6.70)
$VOLUME_t^{st}$	0.45***	0.41***	0.40***	0.37***	0.37***	0.39***	0.44***	0.34***	0.37***	0.33***	0.39***	0.32***	0.32***	0.28***	0.35***	0.32***	0.38***	0.42***	0.36***
	(19.08)	(16.88)	(16.08)	(12.85)	(15.07)	(10.76)	(13.46)	(13.22)	(16.94)	(14.32)	(15.94)	(12.26)	(8.46)	(10.31)	(14.65)	(10.35)	(18.09)	(14.20)	(16.03)
$RETURN_t^{st}$	-0.04***	-0.03**	-0.05***	-0.03*	-0.06***	-0.07***	-0.02	-0.01	-0.06***	-0.03*	-0.01	0.01	-0.04***	-0.05***	0.01	-0.03	-0.00	0.02	-0.02
	(-2.93)	(-2.06)	(-2.70)	(-1.83)	(-3.51)	(-3.81)	(-1.39)	(-0.65)	(-3.78)	(-1.87)	(-0.70)	(0.71)	(-2.99)	(-3.32)	(0.77)	(-1.58)	(-0.18)	(1.38)	(-1.29)
Constant	-0.00	-0.00	-0.00	-0.00	-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
	(-0.02)	(-0.04)	(-0.00)	(-0.04)	(-0.05)	(-0.25)	(-0.05)	(-0.05)	(-0.00)	(-0.03)	(-0.16)	(0.02)	(-0.09)	(-0.05)	(0.01)	(-0.07)	(0.02)	(-0.08)	(-0.03)
N	2508	2502	2511	2510	2511	2402	2511	2511	2511	2511	2476	2509	2511	2506	2511	2471	2505	2508	2508
R^2	0.566	0.559	0.611	0.503	0.592	0.591	0.683	0.597	0.584	0.487	0.715	0.521	0.527	0.522	0.516	0.470	0.514	0.510	0.553
$adjR^2$	0.565	0.558	0.609	0.502	0.591	0.590	0.682	0.595	0.582	0.485	0.714	0.520	0.526	0.520	0.514	0.468	0.512	0.509	0.552

Table C.9: Stock price volatility and ASVI: time variation

The dependent variable in each regression is stock price volatility ($VOLATILITY_t^{st}$). $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. Three bottom lines present F-test p-values for *ASVI* slope-dummy variable equality, *ASVI* slope-dummy variable joint significance and level-dummy variable equality; respectively. T-statistics computed from Newey-West standard errors are in parentheses. Outlying values of ASVI were omitted. *,**, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{CRI,st}$	0.05*	0.09***	0.01	0.01	-0.01	0.04	0.03	-0.00	0.03	0.00	0.01	0.04*	0.02	0.09***	0.00	0.01	0.03	0.05	-0.01
	(1.75)	(3.00)	(0.32)	(0.45)	(-0.47)	(1.64)	(1.43)	(-0.24)	(1.26)	(0.25)	(0.94)	(1.87)	(0.92)	(2.77)	(0.00)	(0.62)	(1.33)	(1.56)	(-0.24)
$ASVI_t^{CRI,st}$	0.10***	0.13***	0.03	0.00	0.06	-0.01	0.06*	0.03	0.04	0.03	0.12***	0.05	0.06	0.05**	0.04	-0.00	-0.02	-0.04	-0.13***
	(2.91)	(2.69)	(1.02)	(0.00)	(1.20)	(-0.40)	(1.75)	(0.87)	(1.45)	(0.68)	(4.29)	(0.89)	(1.37)	(2.10)	(0.73)	(-0.11)	(-0.37)	(-0.91)	(-3.04)
$ASVI_t^{POST,st}$	0.05*	0.08***	0.19***	0.02	0.06**	0.04	0.03	-0.06***	0.03*	0.10***	0.04	0.07***	0.03	0.08**	0.02	0.01	0.02	-0.04**	-0.01
	(1.92)	(3.32)	(5.62)	(1.00)	(2.28)	(1.31)	(1.00)	(-2.83)	(1.69)	(3.29)	(1.52)	(3.22)	(1.24)	(2.41)	(1.08)	(0.37)	(0.64)	(-2.03)	(-0.22)
$VOLATILITY_{t-1}^{st}$	0.27***	0.26***	0.30***	0.27***	0.28***	0.31***	0.27***	0.25***	0.25***	0.25***	0.34***	0.29***	0.26***	0.27***	0.23***	0.26***	0.23***	0.21***	0.32***
	(11.16)	(12.15)	(14.50)	(12.23)	(13.38)	(11.54)	(11.05)	(10.78)	(10.21)	(11.39)	(13.53)	(13.07)	(12.06)	(13.27)	(9.96)	(11.20)	(10.46)	(8.19)	(12.67)
$VOLATILITY_{t-2}^{st}$	0.16***	0.16***	0.18***	0.17***	0.14***	0.15***	0.15***	0.16***	0.16***	0.13***	0.11***	0.16***	0.13***	0.15***	0.11***	0.12***	0.14***	0.14***	0.11***
	(7.91)	(7.06)	(9.32)	(7.06)	(6.63)	(6.57)	(7.27)	(8.54)	(7.22)	(6.22)	(5.10)	(7.73)	(6.14)	(7.07)	(5.56)	(5.51)	(7.05)	(7.62)	(5.55)
$VOLATILITY_{t-3}^{st}$	0.10***	0.13***	0.09***	0.05**	0.13***	0.09***	0.17***	0.11***	0.14***	0.12***	0.13***	0.13***	0.09***	0.09***	0.13***	0.11***	0.12***	0.06***	0.05**
	(4.69)	(5.96)	(3.95)	(2.38)	(7.09)	(4.25)	(8.42)	(6.37)	(7.53)	(6.10)	(5.89)	(5.20)	(3.85)	(4.52)	(6.57)	(4.89)	(6.35)	(3.47)	(2.14)
$VOLATILITY_{t-4}^{st}$	0.10***	0.07***	0.10***	0.09***	0.10***	0.07***	0.10***	0.09***	0.08***	0.09***	0.12***	0.08***	0.10***	0.09***	0.11***	0.10***	0.10***	0.10***	0.13***
	(4.55)	(3.52)	(5.84)	(3.85)	(5.44)	(3.61)	(5.17)	(4.18)	(4.70)	(4.48)	(6.14)	(3.97)	(5.30)	(4.08)	(5.64)	(4.74)	(4.71)	(4.71)	(5.80)
$VOLATILITY_{t-5}^{st}$	0.09***	0.09***	0.12***	0.09***	0.11***	0.14***	0.12***	0.10***	0.10***	0.07***	0.09***	0.09***	0.10***	0.10***	0.08***	0.11***	0.13***	0.12***	0.13***
	(3.97)	(4.07)	(6.00)	(4.36)	(5.64)	(7.75)	(5.08)	(5.00)	(5.35)	(2.96)	(4.13)	(4.64)	(4.74)	(4.78)	(5.21)	(5.12)	(6.49)	(5.56)	(6.79)
PRE	-0.05***	-0.07***	-0.05***	-0.06**	-0.08***	0.04**	-0.10***	-0.05**	-0.04*	-0.04	-0.12***	-0.04*	0.06***	-0.03	-0.12***	-0.03	-0.06***	0.01	-0.06***
	(-2.80)	(-3.37)	(-3.14)	(-2.30)	(-3.99)	(2.03)	(-5.16)	(-2.42)	(-1.75)	(-1.64)	(-3.83)	(-1.82)	(2.59)	(-1.58)	(-4.06)	(-1.32)	(-2.79)	(0.31)	(-2.97)
CRI	0.28***	0.30***	0.23***	0.33***	0.28***	0.25***	0.23***	0.40***	0.31***	0.39***	0.32***	0.23***	0.35***	0.34***	0.38***	="0.31***	"=0.29*""	0.41***	0.28***
	(4.58)	(4.58)	(3.74)	(5.50)	(4.46)	(5.27)	(3.95)	(6.89)	(5.02)	(5.72)	(5.21)	(3.93)	(5.96)	(6.82)	(6.08)	(4.19)	(4.46)	(5.78)	(3.95)
POST	-0.06*	-0.04	-0.04	-0.07**	-0.03	-0.10***	0.00	-0.10***	-0.08***	-0.11***	-0.01	-0.04*	-0.17***	-0.10***	-0.03	-0.08***	-0.05*	-0.15***	-0.05*
	(-1.86)	(-1.52)	(-1.36)	(-2.45)	(-1.22)	(-3.22)	(0.21)	(-3.40)	(-2.91)	(-3.38)	(-0.53)	(-1.83)	(-4.72)	(-3.02)	(-1.04)	(-3.07)	(-1.85)	(-4.88)	(-1.79)
N	2508	2502	2511	2510	2511	2402	2511	2511	2511	2511	2476	2509	2511	2506	2511	2471	2505	2508	2508
R^2	0.439	0.444	0.511	0.404	0.507	0.508	0.603	0.538	0.496	0.412	0.661	0.447	0.462	0.465	0.424	0.398	0.422	0.413	0.464
$adjR^2$	0.437	0.442	0.509	0.401	0.505	0.506	0.601	0.536	0.494	0.409	0.659	0.445	0.459	0.462	0.421	0.395	0.420	0.411	0.462
F-test for equality of ASVI slope dummies (p-value) *(H_0 : all coefficients are equal)*	0.43	0.67	0.00	0.88	0.14	0.38	0.65	0.05	0.92	0.03	0.00	0.51	0.69	0.60	0.70	0.88	0.71	0.05	0.03
F-test for joint significance of ASVI slope dummies (p-value) *(H_0 : all coefficients are equal to 0)*	0.00	0.00	0.00	0.75	0.08	0.21	0.12	0.03	0.10	0.01	0.00	0.00	0.24	0.00	0.64	0.91	0.51	0.06	0.02
F-test for equality of level dummies (p-value) *(H_0 : all coefficients are equal)*	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 C.10: Stock price volatility and ASVI: VAR(5)

The dependent variables are stock price volatility ($VOLATILITY_t^{st}$) in Panel A and web search activity ($ASVI_t^{st}$) in Panel B. $VOLATILITY_t^{st}$ and $ASVI_t^{st}$ are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. Two bottom lines in each panel present Granger-causality Wald tests with corresponding test statistics in parentheses. T-statistics are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. Outlying values of ASVI were omitted. The sample period is from January 2004 to December 2013.

	Panel A - $VOLATILITY_t^{st}$																		
	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$VOLATILTY_{t-1}^{st}$	0.28***	0.29***	0.31***	0.30***	0.30***	0.34***	0.29***	0.28***	0.27***	0.28***	0.37***	0.30***	0.29***	0.29***	0.26***	0.28***	0.25***	0.25***	0.34***
	(14.17)	(14.55)	(15.88)	(14.96)	(15.18)	(16.77)	(14.61)	(14.18)	(13.85)	(14.34)	(18.49)	(15.24)	(14.88)	(14.80)	(13.07)	(14.02)	(12.76)	(12.57)	(17.16)
$VOLATILTY_{t-2}^{st}$	0.18***	0.17***	0.18***	0.19***	0.16***	0.15***	0.16***	0.19***	0.18***	0.15***	0.13***	0.17***	0.16***	0.17***	0.14***	0.13***	0.16***	0.18***	0.13***
	(8.72)	(8.18)	(8.82)	(9.33)	(7.76)	(6.96)	(7.91)	(9.40)	(8.92)	(7.53)	(5.93)	(8.32)	(7.83)	(8.14)	(7.00)	(6.39)	(7.83)	(8.81)	(6.18)
$VOLATILTY_{t-3}^{st}$	0.12***	0.16***	0.10***	0.07***	0.15***	0.12***	0.19***	0.14***	0.16***	0.15***	0.15***	0.14***	0.11***	0.12***	0.16***	0.13***	0.14***	0.09***	0.07***
	(5.97)	(7.45)	(4.83)	(3.50)	(7.11)	(5.37)	(9.24)	(6.74)	(7.91)	(7.16)	(7.15)	(6.92)	(5.53)	(5.55)	(7.86)	(6.33)	(6.88)	(4.64)	(3.24)
$VOLATILTY_{t-4}^{st}$	0.12***	0.09***	0.12***	0.12***	0.12***	0.08***	0.12***	0.12***	0.11***	0.11***	0.14***	0.09***	0.13***	0.12***	0.14***	0.13***	0.12***	0.14***	0.15***
	(5.86)	(4.36)	(5.67)	(5.62)	(5.99)	(3.81)	(5.64)	(5.95)	(5.22)	(5.35)	(6.55)	(4.55)	(6.36)	(5.89)	(6.63)	(6.24)	(6.05)	(6.85)	(7.12)
$VOLATILTY_{t-5}^{st}$	0.11***	0.11***	0.14***	0.12***	0.12***	0.16***	0.14***	0.13***	0.12***	0.10***	0.12***	0.11***	0.13***	0.13***	0.12***	0.12***	0.15***	0.15***	0.15***
	(5.71)	(5.48)	(7.06)	(5.87)	(6.29)	(7.91)	(7.18)	(6.82)	(6.17)	(5.08)	(5.90)	(5.38)	(6.61)	(6.45)	(5.94)	(6.16)	(7.47)	(7.68)	(7.74)
$ASVI_{t-1}^{st}$	0.05***	-0.00	0.02	0.05***	0.03*	-0.00	0.02	-0.00	0.03*	-0.01	0.02*	0.03*	0.02	0.04**	0.03	0.00	0.01	0.02	0.03
	(2.72)	(-0.04)	(1.32)	(2.96)	(1.81)	(-0.05)	(1.22)	(-0.07)	(1.86)	(-0.43)	(1.70)	(1.80)	(1.11)	(2.34)	(1.43)	(0.25)	(0.53)	(0.87)	(1.64)
$ASVI_{t-2}^{st}$	0.01	0.00	-0.01	-0.01	-0.00	-0.00	0.02	-0.01	0.04**	-0.02	-0.01	0.00	-0.00	-0.01	0.01	0.02	-0.02	-0.02	0.01
	(0.55)	(0.16)	(-0.78)	(-0.42)	(-0.04)	(-0.07)	(1.57)	(-0.48)	(2.26)	(-1.22)	(-0.96)	(0.10)	(-0.17)	(-0.41)	(0.73)	(0.95)	(-1.14)	(-0.92)	(0.91)
$ASVI_{t-3}^{st}$	0.03	-0.01	-0.01	0.01	-0.00	-0.03*	0.02	0.05***	0.00	-0.01	-0.00	-0.02	-0.01	0.02	-0.01	-0.00	-0.02	0.00	-0.01
	(1.62)	(-0.44)	(-0.75)	(0.47)	(-0.24)	(-1.95)	(1.10)	(2.81)	(0.28)	(-0.82)	(-0.08)	(-1.06)	(-0.54)	(1.17)	(-0.48)	(-0.21)	(-1.04)	(0.25)	(-0.71)
$ASVI_{t-4}^{st}$	-0.01	0.02	0.01	-0.02	-0.01	0.02	0.03**	0.00	0.01	0.01	-0.01	-0.03	0.01	-0.02	0.03	-0.03*	-0.01	-0.01	0.02
	(-0.43)	(1.39)	(0.37)	(-1.31)	(-0.52)	(1.04)	(1.97)	(0.29)	(0.76)	(0.45)	(-0.60)	(-1.62)	(0.65)	(-1.25)	(1.36)	(-1.93)	(-0.54)	(-0.43)	(1.36)
$ASVI_{t-5}^{st}$	-0.00	-0.01	-0.02	-0.01	-0.00	-0.01	-0.02	-0.01	0.03**	-0.01	-0.01	-0.03	0.00	-0.03*	-0.03*	-0.02	-0.03	-0.02	-0.02
	(-0.02)	(-0.39)	(-1.60)	(-0.34)	(-0.31)	(-0.44)	(-1.26)	(-0.52)	(2.15)	(-0.56)	(-0.80)	(-1.59)	(0.27)	(-1.69)	(-1.72)	(-1.31)	(-1.54)	(-1.44)	(-1.51)
Constant	-0.00	-0.00	-0.00	-0.00	-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
	(-0.07)	(-0.11)	(-0.09)	(-0.04)	(-0.06)	(0.39)	(-0.05)	(-0.13)	(-0.08)	(-0.08)	(-0.12)	(0.00)	(-0.17)	(-0.22)	(-0.01)	(0.02)	(-0.02)	(-0.15)	(-0.06)
Granger-causality ($ASVI \rightarrow VOLATILITY$)	0.07*	0.85	0.29	0.04**	0.47	0.35	0.08*	0.02**	0.02**	0.7	0.39	0.03**	0.75	0.03**	0.26	0.11	0.47	0.52	0.13
	(10.32)	(2.00)	(6.21)	(11.68)	(4.55)	(5.53)	(9.70)	(13.86)	(13.77)	(3.0274)	(5.2104)	(12.232	(2.65)	(12.46)	(6.48)	(9.04)	(4.58)	(4.21)	(8.51)
N	2506	2500	2509	2503	2509	2313	2509	2509	2509	2509	2435	2507	2509	2504	2509	2398	2503	2506	2506

Panel B - $ASVI_t^{st}$																			
	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$VOLATILTY_{t-1}^{st}$	-0.02	-0.02	-0.01	0.01	-0.00	0.03	-0.02	-0.05**	-0.02	-0.01	0.05	0.02	0.02	0.03	-0.06***	0.03	0.01	-0.01	-0.01
	(-1.19)	(-0.90)	(-0.34)	(0.28)	(-0.16)	(0.97)	(-0.77)	(-2.10)	(-0.77)	(-0.38)	(1.51)	(0.78)	(0.65)	(1.41)	(-2.79)	(1.35)	(0.37)	(-0.60)	(-0.46)
$VOLATILTY_{t-2}^{st}$	-0.03	-0.01	0.03	0.01	-0.03	-0.05*	-0.03	-0.01	-0.01	-0.07***	-0.05	0.01	-0.01	-0.03	0.06***	-0.03	0.01	0.01	0.01
	(-1.46)	(-0.56)	(1.22)	(0.24)	(-1.12)	(-1.67)	(-0.96)	(-0.23)	(-0.33)	(-2.76)	(-1.47)	(0.39)	(-0.47)	(-1.43)	(2.72)	(-1.15)	(0.46)	(0.66)	(0.49)
$VOLATILTY_{t-3}^{st}$	-0.01	0.04*	-0.04	0.01	0.02	0.00	0.04	0.02	0.00	0.02	-0.01	-0.00	-0.02	0.01	-0.03	-0.03	0.02	-0.00	0.02
	(-0.44)	(1.89)	(-1.29)	(0.51)	(0.61)	(0.07)	(1.51)	(0.86)	(0.06)	(0.75)	(-0.22)	(-0.10)	(-0.70)	(0.42)	(-1.16)	(-1.43)	(0.79)	(-0.18)	(0.72)
$VOLATILTY_{t-4}^{st}$	0.02	0.03	0.01	0.02	0.00	0.01	0.04	0.03	0.00	0.03	0.03	-0.03	0.02	0.02	0.01	0.03	-0.00	-0.01	0.02
	(1.13)	(1.32)	(0.50)	(0.91)	(0.06)	(0.32)	(1.27)	(1.27)	(0.02)	(1.28)	(1.03)	(-1.33)	(0.68)	(0.85)	(0.36)	(1.07)	(-0.21)	(-0.73)	(0.71)
$VOLATILTY_{t-5}^{st}$	0.01	-0.06***	-0.01	-0.04*	-0.00	-0.01	-0.06**	-0.02	-0.00	0.03	-0.05	0.00	-0.01	-0.09***	-0.00	-0.02	-0.02	0.03*	-0.04
	(0.68)	(-2.89)	(-0.49)	(-1.82)	(-0.00)	(-0.53)	(-2.20)	(-0.64)	(-0.07)	(1.08)	(-1.57)	(0.13)	(-0.35)	(-3.84)	(-0.20)	(-0.88)	(-1.01)	(1.80)	(-1.64)
$ASVI_{t-1}^{st}$	0.07***	0.24***	-0.11***	-0.08***	-0.06***	-0.03*	-0.07***	0.11***	0.10***	0.01	-0.06***	-0.00	0.19***	0.09***	0.31***	-0.21***	-0.04**	0.32***	0.01
	(3.93)	(14.05)	(-5.69)	(-3.81)	(-3.19)	(-1.66)	(-3.38)	(5.79)	(5.19)	(0.70)	(-3.09)	(-0.16)	(9.65)	(4.52)	(15.46)	(-10.24)	(-2.12)	(18.43)	(0.83)
$ASVI_{t-2}^{st}$	-0.22***	-0.32***	-0.34***	-0.37***	-0.37***	-0.35***	-0.36***	-0.40***	-0.38***	-0.34***	-0.34***	-0.36***	-0.30***	-0.34***	-0.42***	-0.44***	-0.30***	-0.23***	-0.28***
	(-13.61)	(-18.02)	(-17.15)	(-18.64)	(-18.33)	(-17.50)	(-17.86)	(-20.42)	(-19.07)	(-17.08)	(-17.92)	(-18.47)	(-14.97)	(-18.05)	(-20.04)	(-21.28)	(-16.14)	(-13.10)	(-16.00)
$ASVI_{t-3}^{st}$	-0.03*	0.03	-0.13***	-0.07***	-0.10***	-0.09***	-0.11***	-0.05**	-0.09***	-0.13***	-0.06***	-0.08***	0.04**	-0.01	-0.03	-0.21***	-0.06***	0.06***	-0.04**
	(-1.78)	(1.58)	(-6.43)	(-3.31)	(-4.90)	(-4.33)	(-5.45)	(-2.22)	(-4.06)	(-6.13)	(-2.99)	(-4.01)	(1.98)	(-0.66)	(-1.21)	(-9.59)	(-3.11)	(3.29)	(-1.98)
$ASVI_{t-4}^{st}$	0.02	-0.09***	-0.10***	-0.13***	-0.07***	-0.11***	-0.10***	-0.08***	-0.04*	-0.12***	-0.12***	-0.11***	-0.06***	-0.09***	0.03	-0.17***	-0.08***	-0.05***	-0.09***
	(1.04)	(-5.24)	(-5.22)	(-6.42)	(-3.56)	(-5.41)	(-4.89)	(-4.14)	(-1.94)	(-6.03)	(-6.63)	(-5.58)	(-2.80)	(-4.83)	(1.38)	(-8.26)	(-4.16)	(-3.21)	(-5.00)
$ASVI_{t-5}^{st}$	0.09***	0.01	0.02	0.02	-0.03*	-0.06***	0.04**	0.16***	0.12***	0.04*	-0.01	0.08***	0.06***	0.07***	0.13***	-0.03	-0.02	0.04**	0.01
	(5.82)	(0.61)	(1.26)	(0.75)	(-1.69)	(-3.00)	(1.99)	(7.87)	(5.97)	(1.84)	(-0.38)	(4.07)	(2.98)	(3.60)	(6.50)	(-1.60)	(-1.09)	(2.30)	(0.55)
Constant	0.00	-0.01	0.00	-0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00	-0.00	-0.01	-0.00	-0.01	-0.00	-0.01	-0.00	0.00	-0.01
	(0.25)	(-0.55)	(0.16)	(-0.18)	(-0.01)	(-0.02)	(-0.06)	(0.09)	(0.02)	(0.06)	(-0.20)	(-0.31)	(-0.06)	(-0.35)	(-0.03)	(-0.32)	(-0.04)	(0.23)	(-0.32)
Granger-causality ($VOLATILITY \rightarrow ASVI$)	0.23	0.03**	0.69	0.58	0.87	0.53	0.07*	0.24	0.86	0.07*	0.21	0.80	0.93	0.00***	0.02**	0.24	0.85	0.52	0.60
	(6.92)	(12.73)	(3.06)	(3.78)	(1.83)	(4.14)	(10.16)	(6.77)	(1.93)	(10.23)	(7.18)	(2.32)	(1.39)	(21.92)	(13.54)	(6.71)	(2.00)	(4.24)	(3.66)
N	2506	2500	2509	2503	2509	2313	2509	2509	2509	2509	2435	2507	2509	2504	2509	2398	2503	2506	2506

C.1.3 Daily returns

Table C.11: Pearson cross correlation coefficients for lag(0)

Cross correlation for $ASVI_{C_i}$ and $RETURN_{C_j}$, where $C_i$and C_j denote specific firm. $RETURN$ and $ASVI$ are defined in Table A.2. First column (1) show correlation of daily returns with firms own ASVI ($i = j$), second column (2) show median correlation on reshuffled data ($i \neq j$). The star denote 5% significance.

ρ	$i = j$ (1)	$i \neq j$(median) (2)
3M	2.41%	0.12%
Boeing	5.90%*	0.58%
Caterpillar	1.21%	0.97%
Coca-Cola	-0.81%	1.20%
DuPont	-1.07%	0.97%
Exxon Mobil	-2.99%	1.11%
General Electric	1.16%	0.61%
Home Depot	-1.78%	0.35%
IBM	1.24%	1.38%
Intel	-1.63%	0.95%
J.P. Morgan	1.90%	-0.21%
Johnson & Johnson	3.65%	2.31%
McDonald's	-3.77%	0.90%
Merck	-11.25%*	0.23%
Microsoft	4.63%*	1.37%
Procter & Gamble	0.27%	0.88%
United Technologies	-0.66%	0.42%
Wal-Mart	-0.75%	1.19%
Walt Disney	0.50%	0.18%
Median	**0.27%**	**0.90%**

Table C.12: Daily returns and ASVI

The dependent variable in each regression are daily returns ($RETRUN_t^{st}$). $RETURN_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations, and two bottom lines show % change in $R^2$and $adjR^2$against restricted model with $\beta_4 = 0$. HAC standard errors are in parentheses. Outlying values of ASVI were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{st}$	0.03	0.07***	0.01	-0.01	-0.01	-0.03	0.01	-0.02	0.01	-0.01	-0.00	0.04*	-0.04**	-0.02	0.05***	0.00	-0.01	0.03	0.00
	(1.25)	(2.79)	(0.42)	(-0.39)	(-0.55)	(-1.31)	(0.54)	(-0.90)	(0.62)	(-0.75)	(-0.15)	(1.78)	(-1.99)	(-0.88)	(2.59)	(0.16)	(-0.53)	(1.09)	(0.21)
$RETURN_{t-1}^{st}$	-0.06**	-0.02	0.01	-0.07	-0.04	-0.17***	-0.02	0.00	-0.01	-0.07**	-0.11**	-0.06	-0.08***	-0.03	-0.08**	-0.10***	-0.08**	-0.06**	-0.07**
	(-2.08)	(-0.55)	(0.46)	(-1.57)	(-1.12)	(-3.54)	(-0.36)	(0.11)	(-0.45)	(-2.34)	(-2.17)	(-1.47)	(-2.87)	(-0.93)	(-2.35)	(-3.04)	(-2.49)	(-1.97)	(-2.08)
$RETURN_{t-2}^{st}$	-0.05			-0.04		-0.13**						-0.10**	-0.08***		-0.05*	-0.09**	-0.05	-0.06*	-0.07*
	(-1.49)			(-0.86)		(-2.27)						(-2.03)	(-2.62)		(-1.68)	(-2.51)	(-1.47)	(-1.82)	(-1.86)
$RETURN_{t-3}^{st}$												0.05							
												(1.46)							
Constant	0.00	-0.00	0.00	0.00	0.00	-0.01	0.00	-0.00	0.00	0.00	-0.00	0.00	-0.00	0.01	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.02)	(-0.07)	(0.02)	(0.02)	(0.01)	(-0.26)	(0.01)	(-0.01)	(0.00)	(0.00)	(-0.17)	(0.03)	(-0.03)	(0.57)	(-0.00)	(-0.02)	(-0.08)	(-0.05)	(-0.02)
N	2510	2505	2514	2512	2514	2402	2514	2514	2514	2514	2478	2510	2513	2509	2513	2473	2507	2510	2510
R^2	0.007	0.004	0.000	0.006	0.002	0.040	0.000	0.000	0.000	0.005	0.013	0.018	0.014	0.001	0.011	0.016	0.009	0.007	0.009
$adjR^2$	0.005	0.003	-0.000	0.005	0.001	0.039	-0.000	-0.000	-0.000	0.004	0.012	0.017	0.013	0.000	0.009	0.015	0.007	0.006	0.008
$\%\triangle R^2$ to restricted model without *ASVI*	17%			0%	0%	3%				25%	0%	6%	17%	0%	38%	0%	0%	0%	0%
$\%\triangle adjR^2$ to restricted model without *ASVI*	0%			-17%	0%	3%				0%	0%	13%	8%		29%	-6%	-13%	0%	0%

Table C.13: Daily returns, ASVI, volume and volatility

TThe dependent variable in each regression are daily returns ($RETURN_t^{st}$). $RETURN_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. T-statistics computed from White standard errors are in parentheses. Outlying values of ASVI were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{st}$	0.04	0.07***	0.02	-0.01	-0.01	-0.02	0.01	-0.02	0.01	-0.01	-0.00	0.04*	-0.04**	-0.02	0.05***	0.00	-0.01	0.02	0.00
	(1.64)	(2.94)	(0.74)	(-0.38)	(-0.59)	(-1.17)	(0.58)	(-0.78)	(0.66)	(-0.69)	(-0.18)	(1.79)	(-1.96)	(-0.89)	(2.93)	(0.16)	(-0.44)	(1.21)	(0.19)
$RETURN_{t-1}^{st}$	-0.07***	-0.02	0.01	-0.08**	-0.05*	-0.19***	-0.02	0.00	-0.02	-0.07*	-0.11**	-0.06	-0.09***	-0.03	-0.08**	-0.10**	-0.08***	-0.06	-0.07***
	(-3.04)	(-0.84)	(0.25)	(-2.04)	(-1.84)	(-3.03)	(-0.46)	(0.17)	(-0.91)	(-1.71)	(-2.19)	(-1.18)	(-3.66)	(-1.03)	(-2.33)	(-2.54)	(-2.97)	(-1.63)	(-2.93)
$RETURN_{t-2}^{st}$	-0.06*			-0.04		-0.14*						-0.10*	-0.08***		-0.06	-0.10**	-0.06	-0.06*	-0.08**
	(-1.67)			(-1.20)		(-1.67)						(-1.91)	(-3.04)		(-1.21)	(-2.52)	(-1.50)	(-1.89)	(-2.14)
$RETURN_{t-3}^{st}$												0.05							
												(1.27)							
$VOLUME_t^{st}$	-0.07	0.00	-0.01	0.00	0.02	-0.02	-0.01	0.02	0.04	-0.04	0.01	-0.02	0.03	-0.00	-0.04	-0.03	-0.05*	-0.06*	0.04
	(-1.52)	(0.01)	(-0.32)	(0.09)	(0.70)	(-0.55)	(-0.12)	(0.65)	(1.04)	(-0.99)	(0.27)	(-0.72)	(0.99)	(-0.03)	(-0.99)	(-1.05)	(-1.83)	(-1.68)	(1.05)
$VOLATILITY_t^{st}$	-0.04	-0.06**	-0.07*	-0.05	-0.10***	-0.11***	-0.03	-0.02	-0.09***	-0.03	-0.02	0.00	-0.07**	-0.06***	-0.00	-0.04	-0.01	0.04	-0.07*
	(-1.31)	(-2.41)	(-1.71)	(-1.63)	(-2.82)	(-3.98)	(-0.94)	(-0.70)	(-2.73)	(-1.23)	(-0.51)	(0.08)	(-2.51)	(-2.60)	(-0.14)	(-1.39)	(-0.38)	(1.32)	(-1.78)
Constant	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	-0.00
	(0.02)	(-0.07)	(0.02)	(0.03)	(0.01)	(-0.08)	(0.01)	(-0.01)	(0.00)	(0.00)	(-0.22)	(0.04)	(-0.03)	(0.65)	(-0.00)	(0.02)	(-0.09)	(-0.06)	(-0.02)
N	2510	2505	2514	2512	2514	2402	2514	2514	2514	2514	2478	2510	2513	2509	2513	2473	2507	2510	2510
R^2	0.017	0.008	0.006	0.008	0.009	0.054	0.002	0.001	0.006	0.008	0.013	0.019	0.017	0.005	0.012	0.020	0.012	0.009	0.012
$adjR^2$	0.015	0.007	0.005	0.006	0.008	0.052	0.000	-0.001	0.004	0.007	0.011	0.016	0.015	0.003	0.010	0.018	0.010	0.007	0.010

Table C.14: Daily returns and ASVI: time variation

The dependent variable in each regression are daily returns ($RETURN_t^{st}$). $RETURN_t^{st}$ and independent variables are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. Three bottom lines present F-test p-values for *ASVI* slope-dummy variable equality, *ASVI* slope-dummy variable joint significance and level-dummy variable equality; respectively. T-statistics computed from White standard errors are in parentheses. Outlying values of ASVI were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$ASVI_t^{CRI,st}$	0.01	0.07**	0.03	-0.01	0.01	-0.02	0.02	-0.01	0.03	-0.02	-0.00	0.03	-0.04	-0.04	0.01	-0.02	-0.04*	0.05	0.03
	(0.47)	(2.39)	(1.32)	(-0.31)	(0.66)	(-0.75)	(1.52)	(-0.47)	(1.01)	(-0.93)	(-0.13)	(1.29)	(-1.29)	(-1.14)	(0.61)	(-0.96)	(-1.79)	(1.02)	(1.15)
$ASVI_t^{CRI,st}$	0.06	-0.06	-0.04	-0.02	-0.10	-0.08	-0.13	-0.08	0.04	0.00	-0.08	0.11	-0.06	-0.09	0.10	0.02	0.01	0.03	-0.07
	(0.64)	(-0.46)	(-0.43)	(-0.19)	(-0.96)	(-1.06)	(-1.31)	(-0.89)	(0.49)	(0.04)	(-0.69)	(0.99)	(-1.02)	(-1.36)	(1.26)	(0.22)	(0.10)	(0.42)	(-0.61)
$ASVI_t^{POST,st}$	0.04	0.10***	0.04	-0.01	-0.02	-0.01	0.10***	0.00	-0.00	-0.01	0.03	0.02	-0.03	0.06**	0.06**	0.04	0.02	0.01	-0.01
	(1.16)	(2.69)	(0.91)	(-0.17)	(-0.54)	(-0.37)	(2.63)	(0.03)	(-0.16)	(-0.21)	(1.12)	(0.74)	(-1.17)	(2.13)	(2.37)	(1.22)	(0.77)	(0.45)	(-0.24)
$RETURN_{t-1}^{st}$	-0.06**	-0.02	0.01	-0.07	-0.04	-0.17***	-0.02	0.00	-0.01	-0.07**	-0.12**	-0.06	-0.08***	-0.03	-0.08**	-0.10***	-0.08**	-0.06**	-0.07**
	(-2.15)	(-0.64)	(0.35)	(-1.62)	(-1.19)	(-3.57)	(-0.56)	(0.05)	(-0.47)	(-2.39)	(-2.27)	(-1.50)	(-2.89)	(-1.08)	(-2.37)	(-3.14)	(-2.53)	(-1.98)	(-2.17)
$RETURN_{t-2}^{st}$	-0.05			-0.04		-0.13**						-0.10**	-0.08***		-0.06*	-0.09**	-0.06	-0.06*	-0.07*
	(-1.56)			(-0.90)		(-2.28)						(-2.08)	(-2.65)		(-1.70)	(-2.56)	(-1.52)	(-1.84)	(-1.90)
$RETURN_{t-3}^{st}$												0.05							
												(1.44)							
PRE	-0.02	0.02	0.01	0.00	-0.01	0.02	0.01	-0.03	-0.01	-0.01	0.00	0.00	0.03	0.03	0.01	0.02	0.01	-0.03	-0.01
	(-0.65)	(0.73)	(0.46)	(0.08)	(-0.40)	(0.86)	(0.73)	(-1.40)	(-0.51)	(-0.37)	(0.33)	(0.10)	(0.86)	(1.34)	(0.30)	(0.71)	(0.31)	(-0.99)	(-0.45)
CRI	-0.08	-0.13*	-0.10	-0.08	-0.09	-0.08	-0.15	-0.04	-0.02	-0.06	-0.08	-0.07	-0.05	-0.10	-0.07	-0.12	-0.09	-0.01	-0.09
	(-1.01)	(-1.66)	(-1.30)	(-0.94)	(-1.05)	(-0.94)	(-1.54)	(-0.49)	(-0.27)	(-0.71)	(-0.75)	(-0.95)	(-0.68)	(-1.22)	(-0.80)	(-1.45)	(-1.13)	(-0.15)	(-1.01)
POST	0.04	0.03	0.03	0.03	0.04	-0.00	0.04*	0.04*	0.02	0.03	0.01	0.03	-0.01	0.03	0.02	0.02	0.02	0.03	0.04
	(1.61)	(1.13)	(1.02)	(1.10)	(1.53)	(-0.12)	(1.72)	(1.88)	(0.71)	(1.24)	(0.67)	(1.01)	(-0.27)	(1.37)	(0.71)	(0.97)	(0.85)	(1.15)	(1.51)
N	2510	2505	2514	2512	2514	2402	2514	2514	2514	2514	2478	2510	2513	2509	2513	2473	2507	2510	2510
R^2	0.009	0.009	0.003	0.008	0.005	0.041	0.010	0.003	0.001	0.006	0.015	0.020	0.015	0.007	0.012	0.020	0.011	0.008	0.012
$adjR^2$	0.005	0.006	0.001	0.005	0.002	0.038	0.007	-0.000	-0.002	0.003	0.012	0.016	0.012	0.004	0.009	0.017	0.008	0.005	0.009
F-test for equality of ASVI slope dummies (p-value)	0.79	0.45	0.73	0.99	0.44	0.71	0.04	0.68	0.66	0.91	0.43	0.74	0.85	0.02	0.24	0.29	0.24	0.81	0.46
(H_0 : all coefficients are equal)																			
F-test for joint significance of ASVI slope dummies (p-value)	0.58	0.00	0.43	0.98	0.65	0.61	0.01	0.80	0.74	0.82	0.63	0.36	0.25	0.05	0.06	0.48	0.28	0.70	0.63
(H_0 : all coefficients are equal to 0)																			
F-test for equality of level dummies (p-value)	0.15	0.15	0.30	0.42	0.18	0.45	0.13	0.06	0.66	0.39	0.67	0.44	0.53	0.29	0.62	0.24	0.40	0.32	0.19
(H_0 : all coefficients are equal)																			

Table C.15: Daily returns and ASVI: VAR(p)

The dependent variables are daily returns ($RETURN_t^{st}$) in Panel A and web search activity ($ASVI_t^{st}$) in Panel B. $RETURN_t^{st}$ and $ASVI_t^{st}$ are defined in Table A.2. Each column show results for one stock; stock tickers are presented to save space. N is number of observations. Two bottom lines in each panel present Granger-causality Wald tests with corresponding test statistics in parentheses. T-statistics are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013.

	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
								Panel A - $RETRUN_t^{st}$											
$RETURN_{t-1}^{st}$	-0.06***	-0.02	0.01	-0.07***	-0.04**	-0.17***	-0.02	0.00	-0.01	-0.06***	-0.11***	-0.06***	-0.08***	-0.03*	-0.08***	-0.10***	-0.08***	-0.06***	-0.07***
	(-3.10)	(-0.98)	(0.65)	(-3.60)	(-2.01)	(-8.33)	(-0.77)	(0.22)	(-0.72)	(-3.15)	(-5.73)	(-3.00)	(-4.14)	(-1.67)	(-3.89)	(-4.80)	(-3.94)	(-3.18)	(-3.45)
$RETURN_{t-2}^{st}$	-0.05**	-0.01	-0.02	-0.04**	-0.02	-0.13***	0.01	-0.05**	0.01	-0.02	-0.03	-0.10***	-0.08***	-0.04**	-0.05***	-0.09***	-0.05***	-0.06***	-0.07***
	(-2.53)	(-0.51)	(-0.97)	(-2.04)	(-1.18)	(-6.32)	(0.66)	(-2.33)	(0.44)	(-0.95)	(-1.45)	(-5.05)	(-4.01)	(-2.34)	(-2.66)	(-4.56)	(-2.71)	(-2.89)	(-3.58)
$RETURN_{t-3}^{st}$			0.05**				-0.01	0.02	0.03	0.06***		0.05**				0.02			
			(2.50)				(-0.69)	(0.97)	(1.53)	(3.03)		(2.51)				(1.04)			
$RETURN_{t-4}^{st}$								-0.04**		-0.07***		-0.01				-0.05**			
								(-2.06)		(-3.57)		(-0.50)				(-2.26)			
$RETURN_{t-5}^{st}$								-0.03											
								(-1.28)											
$ASVI_{t-1}^{st}$	0.00	0.00	0.01**	-0.00	0.00	-0.00	0.01	0.00	-0.00	-0.01	0.00	-0.00	-0.00	-0.00	0.00	0.00	0.00**	0.00	0.00
	(0.84)	(1.22)	(2.55)	(-0.65)	(0.71)	(-0.04)	(0.77)	(0.00)	(-0.69)	(-0.86)	(0.32)	(-0.46)	(-1.05)	(-0.73)	(0.63)	(0.28)	(2.18)	(0.87)	(0.53)
$ASVI_{t-2}^{st}$	-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**	-0.01**	0.00	0.00	-0.00	0.00
	(-0.94)	(-0.55)	(-0.25)	(-0.03)	(0.86)	(-0.30)	(-0.13)	(1.34)	(-1.64)	(0.08)	(1.17)	(0.37)	(2.05)	(-2.23)	(-2.04)	(0.51)	(0.62)	(-0.10)	(0.07)
$ASVI_{t-3}^{st}$			0.00				-0.00	0.00	0.00	0.01		-0.00				-0.00			
			(0.23)				(-0.61)	(0.79)	(0.06)	(0.89)		(-0.41)				(-0.10)			
$ASVI_{t-4}^{st}$								0.00		0.01		0.00				0.00			
								(0.25)		(1.54)		(1.29)				(0.84)			
$ASVI_{t-5}^{st}$								0.01											
								(1.50)											
Constant	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
	(1.18)	(1.45)	(0.99)	(1.49)	(0.91)	(1.53)	(0.22)	(1.38)	(1.18)	(0.01)	(0.46)	(1.92)	(2.94)	(1.08)	(0.75)	(1.58)	(1.59)	(1.08)	(1.63)
Granger-causality ($ASVI \rightarrow RETURN$)	0.50	0.45	0.07**	0.81	0.55	0.96	0.70	0.56	0.30	0.24	0.48	0.72	0.09*	0.05*	0.12	0.93	0.08*	0.67	0.86
	(1.39)	(1.6)	(7.00)	(0.43)	(1.21)	(0.09)	(1.44)	(3.91)	(3.68)	(5.55)	(1.45)	(2.09)	(4.79)	(5.82)	(4.21)	(0.87)	(5.15)	(0.80)	(0.29)
N	2509	2503	2511	2509	2512	2360	2511	2509	2511	2510	2458	2508	2512	2507	2512	2413	2506	2509	2509

	Panel B - $ASVI_t^{st}$																		
	MMM	BA	CAT	KO	DD	XOM	GE	HD	IBM	INTC	JPM	JNJ	MCD	MRK	MSFT	PG	UTX	WMT	DIS
$RETURN_{t-1}^{st}$	-0.22*	-0.20	-0.00	0.02	0.14	0.09	-0.01	-0.07	-0.02	0.06	-0.24	-0.42*	0.07	-0.18	0.07	0.04	-0.10	0.06	-0.10
	(-1.66)	(-1.26)	(-0.03)	(0.09)	(1.28)	(0.35)	(-0.13)	(-0.91)	(-0.21)	(1.30)	(-1.58)	(-1.77)	(0.43)	(-1.14)	(1.04)	(0.11)	(-0.48)	(0.35)	(-0.93)
$RETURN_{t-2}^{st}$	0.01	0.12	-0.04	-0.20	-0.07	0.41	0.01	-0.05	-0.04	0.03	-0.18	-0.02	-0.20	0.07	0.06	0.17	0.27	0.22	-0.09
	(0.11)	(0.77)	(-0.39)	(-0.89)	(-0.64)	(1.61)	(0.23)	(-0.60)	(-0.45)	(0.58)	(-1.19)	(-0.10)	(-1.21)	(0.43)	(0.89)	(0.44)	(1.29)	(1.36)	(-0.83)
$RETURN_{t-3}^{st}$			0.03				-0.11**	0.07	-0.08	0.04		-0.18				0.04			
			(0.31)				(-2.05)	(0.91)	(-0.81)	(0.76)		(-0.76)				(0.11)			
$RETURN_{t-4}^{st}$								0.03		-0.03		0.24				-0.07			
								(0.34)		(-0.59)		(1.03)				(-0.19)			
$RETURN_{t-5}^{st}$								-0.01											
								(-0.15)											
$ASVI_{t-1}^{st}$	0.09***	0.24***	-0.10***	-0.05***	-0.03	0.01	-0.06***	0.12***	0.10***	0.01	-0.03*	-0.01	0.18***	0.10***	0.33***	-0.20***	-0.02	0.31***	0.03*
	(5.18)	(14.64)	(-5.12)	(-2.58)	(-1.36)	(0.31)	(-3.02)	(5.83)	(5.20)	(0.47)	(-1.71)	(-0.49)	(9.66)	(5.35)	(19.04)	(-10.13)	(-1.09)	(18.24)	(1.72)
$ASVI_{t-2}^{st}$	-0.24***	-0.30***	-0.31***	-0.33***	-0.34***	-0.31***	-0.33***	-0.40***	-0.39***	-0.35***	-0.30***	-0.37***	-0.29***	-0.32***	-0.48***	-0.43***	-0.28***	-0.20***	-0.26***
	(-15.75)	(-17.83)	(-16.52)	(-17.32)	(-18.00)	(-16.30)	(-17.75)	(-20.41)	(-21.13)	(-17.61)	(-16.70)	(-18.99)	(-14.94)	(-17.89)	(-27.10)	(-21.61)	(-15.51)	(-12.86)	(-15.35)
$ASVI_{t-3}^{st}$			-0.13***				-0.12***	-0.05**	-0.14***	-0.14***		-0.11***				-0.20***			
			(-6.72)				(-6.12)	(-2.20)	(-6.93)	(-7.07)		(-5.68)				(-9.76)			
$ASVI_{t-4}^{st}$								-0.08***		-0.12***		-0.11***				-0.17***			
								(-4.25)		(-5.96)		(-5.81)				(-8.26)			
$ASVI_{t-5}^{st}$								0.15***											
								(7.79)											
Constant	-0.00	-0.01**	-0.00	-0.00*	-0.00	-0.01***	-0.00	-0.00	-0.00**	-0.00	-0.01***	-0.00**	-0.00	-0.01**	-0.00	-0.02***	-0.01**	-0.00	-0.00
	(-1.15)	(-2.06)	(-1.35)	(-1.64)	(-1.49)	(-2.61)	(-1.06)	(-0.19)	(-2.30)	(-1.25)	(-3.11)	(-2.03)	(-0.59)	(-2.27)	(-1.26)	(-4.21)	(-2.06)	(-0.22)	(-1.24)
Granger-causality $(RETURN \rightarrow ASVI)$	0.25	0.33	0.97	0.66	0.35	0.27	0.23	0.81	0.83	0.60	0.17	0.35	0.42	0.47	0.42	0.99	0.37	0.38	0.48
	(2.78)	(2.22)	(0.25)	(0.82)	(2.11)	(2.60)	(4.28)	(2.25)	(0.89)	(2.75)	(3.58)	(4.45)	(1.74)	(1.53)	(1.74)	(0.26)	(1.98)	(1.91)	(1.47)
N	2509	2503	2511	2509	2512	2360	2511	2509	2511	2510	2458	2508	2512	2507	2512	2413	2506	2509	2509

C.2 Panel setting

C.2.1 Trading volume

Table C.16: Trading volume and ASVI: non-lagged models

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLUME_t^{st}$. $VOLUME_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_t^{st}$		0.024*** (4.330)	0.024*** (4.276)	0.019*** (4.306)	0.024*** (4.310)	0.019*** (4.292)
$ASVI_t^{sq,st}$			0.001 (0.184)			
$VOLUME_{t-1}^{st}$	0.507*** (60.058)	0.507*** (60.234)	0.507*** (60.373)	0.418*** (43.833)	0.507*** (60.788)	0.418*** (44.128)
$VOLUME_{t-2}^{st}$	0.115*** (22.855)	0.118*** (23.498)	0.118*** (23.543)	0.091*** (22.531)	0.118*** (22.883)	0.091*** (22.130)
$VOLUME_{t-3}^{st}$	0.090*** (14.380)	0.090*** (14.480)	0.090*** (14.462)	0.070*** (12.088)	0.090*** (14.520)	0.070*** (12.129)
$VOLUME_{t-4}^{st}$	0.056*** (11.160)	0.054*** (11.035)	0.054*** (11.030)	0.034*** (7.064)	0.054*** (11.005)	0.034*** (7.035)
$VOLUME_{t-5}^{st}$	0.102*** (15.076)	0.103*** (15.134)	0.103*** (15.274)	0.073*** (11.218)	0.103*** (15.230)	0.073*** (11.316)
$VOLATILITY_t^{st}$				0.170*** (32.366)		0.169*** (32.577)
$RETURN_t^{st}$					-0.015*** (-4.729)	-0.008*** (-3.848)
Constant	-0.000*** (-13.331)	-0.000*** (-11.279)	-0.000*** (-3.293)	-0.001*** (-13.653)	-0.000*** (-11.275)	-0.001*** (-13.638)
N	47493	47493	47493	47493	47493	47493
R^2	0.6020	0.6041	0.6041	0.6790	0.6049	0.6792

Table C.17: Trading volume and ASVI: lagged models

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLUME_t^{st}$. $VOLUME_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_{t-1}^{st}$		0.013***	0.013***	0.013***	0.014***	0.014***
		(6.185)	(6.193)	(6.181)	(6.184)	(6.182)
$ASVI_{t-1}^{sq,st}$			-0.006			
			(-1.111)			
$VOLUME_{t-1}^{st}$	0.506***	0.503***	0.503***	0.489***	0.500***	0.487***
	(59.687)	(59.282)	(59.496)	(47.882)	(57.431)	(46.806)
$VOLUME_{t-2}^{st}$	0.116***	0.117***	0.117***	0.118***	0.119***	0.120***
	(23.402)	(23.492)	(23.535)	(23.530)	(23.736)	(23.756)
$VOLUME_{t-3}^{st}$	0.090***	0.092***	0.092***	0.092***	0.093***	0.092***
	(14.965)	(15.203)	(15.209)	(15.155)	(14.849)	(14.815)
$VOLUME_{t-4}^{st}$	0.056***	0.057***	0.056***	0.056***	0.057***	0.057***
	(11.388)	(11.343)	(11.320)	(11.386)	(11.439)	(11.479)
$VOLUME_{t-5}^{st}$	0.102***	0.101***	0.101***	0.100***	0.102***	0.101***
	(14.624)	(14.533)	(14.479)	(14.155)	(14.549)	(14.201)
$VOLATILITY_{t-1}^{st}$				0.013***		0.012***
				(4.432)		(4.330)
$RETURN_{t-1}^{st}$					-0.020***	-0.020***
					(-11.072)	(-11.344)
Constant	-0.000***	-0.000***	-0.001***	-0.000***	-0.000***	-0.000***
	(-11.299)	(-10.655)	(-4.512)	(-11.101)	(-10.818)	(-11.212)
N	47491	47491	47491	47491	47491	47491
R^2	0.6007	0.6014	0.6014	0.6017	0.6029	0.6032

Table C.18: Trading volume and ASVI: lagged ASVI

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLUME_t^{st}$. $VOLUME_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)
$ASVI_{t-1}^{st}$	0.013***				
	(6.185)				
$ASVI_{t-2}^{st}$		0.002			
		(1.175)			
$ASVI_{t-3}^{st}$			0.002		
			(1.340)		
$ASVI_{t-4}^{st}$				-0.003	
				(-1.631)	
$ASVI_{t-5}^{st}$					-0.007**
					(-2.740)
$VOLUME_{t-1}^{st}$	0.503***	0.507***	0.507***	0.507***	0.507***
	(59.282)	(58.763)	(58.509)	(59.232)	(58.301)
$VOLUME_{t-2}^{st}$	0.117***	0.114***	0.114***	0.115***	0.115***
	(23.492)	(22.200)	(22.899)	(22.775)	(23.139)
$VOLUME_{t-3}^{st}$	0.092***	0.091***	0.090***	0.091***	0.090***
	(15.203)	(14.948)	(14.463)	(14.656)	(14.879)
$VOLUME_{t-4}^{st}$	0.057***	0.057***	0.057***	0.056***	0.057***
	(11.343)	(11.811)	(11.686)	(11.350)	(11.326)
$VOLUME_{t-5}^{st}$	0.101***	0.102***	0.102***	0.102***	0.102***
	(14.533)	(14.528)	(14.719)	(14.324)	(14.915)
Constant	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(-10.655)	(-8.537)	(-9.180)	(-12.672)	(-7.090)
N	47491	47490	47489	47470	47451
R^2	0.6014	0.6007	0.6005	0.6009	0.6005

Table C.19: Trading volume and ASVI: time variation

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLUME_t^{st}$. $VOLUME_t^{st}$ and independent variables are defined in Table A.2. Slope dummy equality tests whether the $ASVI_t^{st}$ and $ASVI_{t-1}^{st}$ coefficients from (2) and (3), respectively (5) and (6); are higher than coefficients from (2), respectively (4). T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample periods are from January 2004 to November 2007 (1,4); December 2007 to June 2009 (2,5); July 2009 to December 2014 (3,6) . N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_t^{st}$	0.018***	0.036***	0.026***			
	(4.028)	(4.613)	(3.448)			
$ASVI_{t-1}^{st}$				0.011***	0.014***	0.016***
				(4.638)	(3.819)	(4.840)
$VOLUME_{t-1}^{st}$	0.468***	0.527***	0.487***	0.463***	0.522***	0.482***
	(52.990)	(32.604)	(59.948)	(57.048)	(32.455)	(54.844)
$VOLUME_{t-2}^{st}$	0.097***	0.135***	0.102***	0.096***	0.131***	0.103***
	(11.592)	(12.902)	(13.004)	(11.492)	(12.999)	(13.598)
$VOLUME_{t-3}^{st}$	0.105***	0.055***	0.059***	0.108***	0.057***	0.060***
	(10.744)	(5.179)	(7.875)	(11.233)	(5.158)	(8.176)
$VOLUME_{t-4}^{st}$	0.043***	0.026*	0.044***	0.045***	0.031**	0.046***
	(5.218)	(2.079)	(7.279)	(5.635)	(2.646)	(7.342)
$VOLUME_{t-5}^{st}$	0.066***	0.072***	0.096***	0.065***	0.071***	0.094***
	(6.084)	(7.289)	(10.792)	(6.086)	(7.289)	(10.131)
Constant	-0.042***	0.083***	0.000	-0.042***	0.085***	0.000
	(-11.682)	(8.022)	(1.237)	(-11.581)	(8.396)	(1.329)
N	18434	7535	21524	18432	7535	21524
R^2	0.6007	0.6014	0.6014	0.6017	0.6029	0.6032
F-test for inequality of ASVI slope dummies		0.049	0.390		0.529	0.288
(H_0 : coefficient is higher than in pre-crisis period)		(3.893)	(0.740)		(0.397)	(1.129)

Figure C.1: Trading volume and ASVI: IFRs

The figures show impulse response functions of the vertical axis variable to a shock of one standard deviation in the title variable. Vertical axis displays the magnitude of the response (in terms of % of standard deviation), horizontal axis show the time horizon in days.

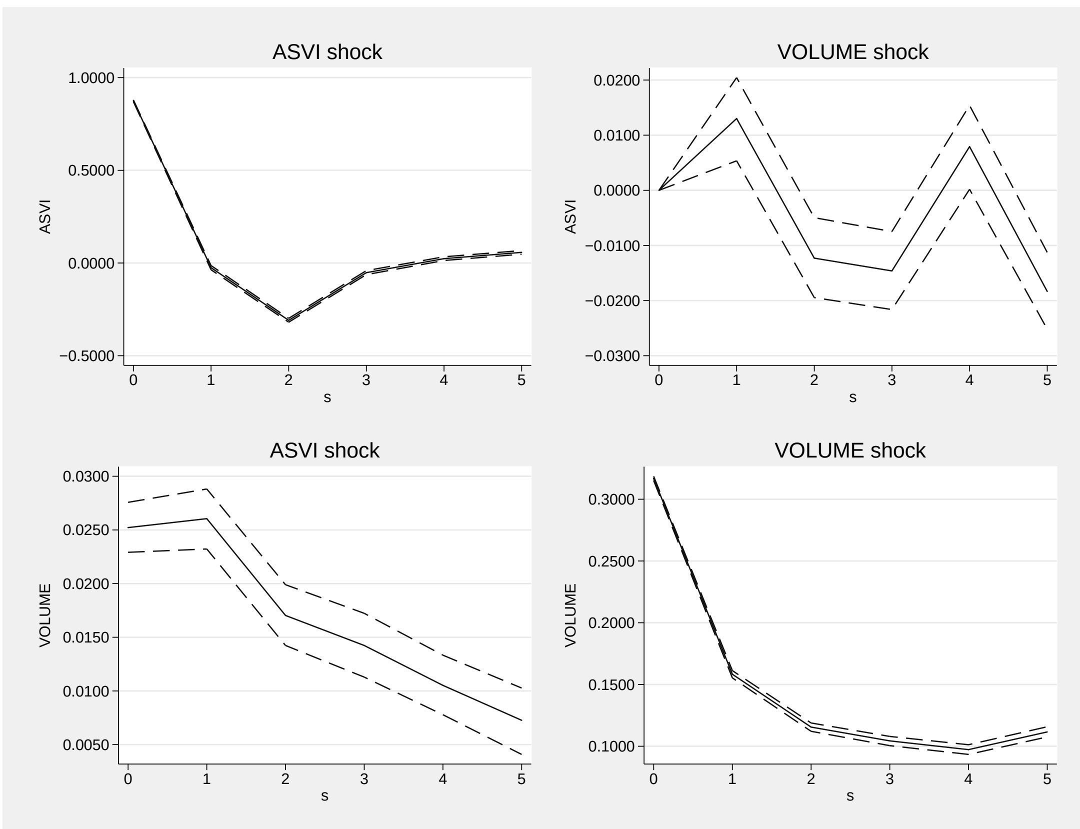
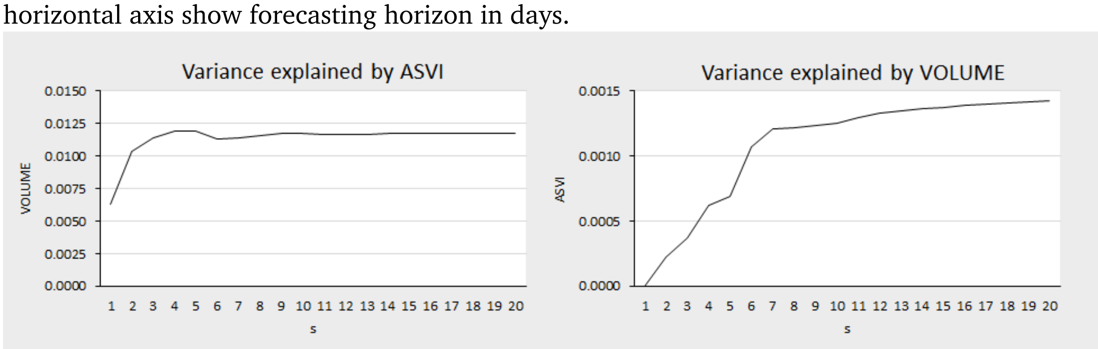


Figure C.2: Trading volume and ASVI: variance decomposition

The first figure show fraction of variation in *VOLUME* explained by *ASVI*. The second figure show fraction of variation in *ASVI* explained by *VOLUME*. Vertical axis show the fraction of variation, horizontal axis show forecasting horizon in days.



C.2.2 Stock price volatility

Table C.20: Stock price volatility and ASVI: non-lagged models

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLATILITY_t^{st}$. $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_t^{st}$		0.028***	0.029***	0.016***	0.029***	0.017***
		(4.329)	(4.413)	(4.091)	(4.374)	(4.074)
$ASVI_t^{sq,st}$			0.019**	0.045***	0.019**	0.045***
			(2.654)	(4.268)	(2.683)	(4.318)
$VOLATILITY_{t-1}^{st}$	0.296***	0.297***	0.297***	0.198***	0.296***	0.198***
	(42.151)	(42.436)	(42.406)	(28.223)	(41.816)	(28.030)
$VOLATILITY_{t-2}^{st}$	0.165***	0.166***	0.166***	0.115***	0.166***	0.115***
	(38.965)	(38.762)	(39.014)	(24.493)	(38.834)	(24.627)
$VOLATILITY_{t-3}^{st}$	0.129***	0.129***	0.129***	0.093***	0.129***	0.093***
	(17.403)	(17.444)	(17.476)	(14.495)	(17.145)	(14.439)
$VOLATILITY_{t-4}^{st}$	0.120***	0.120***	0.120***	0.092***	0.119***	0.092***
	(30.866)	(29.921)	(29.857)	(21.036)	(30.225)	(20.833)
$VOLATILITY_{t-5}^{st}$	0.130***	0.131***	0.131***	0.101***	0.131***	0.101***
	(34.313)	(34.895)	(35.184)	(23.798)	(34.743)	(23.523)
$VOLUME_t^{st}$				0.684***		0.683***
				(29.079)		(29.395)
$RETURN_t^{st}$					-0.033***	-0.025***
					(-4.777)	(-4.513)
Constant	-0.000***	-0.000***	0.000	0.001**	0.000	0.001**
	(-11.164)	(-9.134)	(0.524)	(2.844)	(0.533)	(2.876)
N	47493	47493	47493	47493	47493	47493
R^2	0.4699	0.4707	0.4708	0.5636	0.4720	0.5643

Table C.21: Stock price volatility and ASVI: lagged models

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLATILITY_t^{st}$. $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_{t-1}^{st}$		0.016*** (3.938)	0.016*** (3.856)	0.014*** (3.522)	0.016*** (3.897)	0.014*** (3.592)
$ASVI_{t-1}^{sq,st}$			-0.011 (-1.248)	-0.004 (-0.531)	-0.011 (-1.249)	-0.004 (-0.538)
$VOLATILITY^{st}_{t-1}$	0.296*** (40.990)	0.296*** (41.490)	0.296*** (41.322)	0.256*** (35.573)	0.291*** (41.493)	0.252*** (36.708)
$VOLATILITY^{st}_{t-2}$	0.165*** (38.854)	0.165*** (38.539)	0.165*** (38.613)	0.154*** (32.947)	0.166*** (39.592)	0.156*** (33.852)
$VOLATILITY^{st}_{t-3}$	0.130*** (17.583)	0.130*** (17.620)	0.130*** (17.626)	0.125*** (17.733)	0.132*** (17.762)	0.127*** (17.899)
$VOLATILITY^{st}_{t-4}$	0.121*** (32.838)	0.121*** (32.835)	0.121*** (32.877)	0.117*** (33.082)	0.121*** (31.556)	0.117*** (32.463)
$VOLATILITY^{st}_{t-5}$	0.130*** (34.623)	0.130*** (34.306)	0.130*** (34.153)	0.128*** (36.618)	0.130*** (35.227)	0.128*** (36.907)
$VOLUME_{t-1}^{st}$				0.155*** (10.526)		0.154*** (10.470)
$RETURN_{t-1}^{st}$					-0.063*** (-12.982)	-0.063*** (-12.887)
Constant	-0.001*** (-19.829)	-0.001*** (-18.637)	-0.001*** (-4.190)	-0.001*** (-4.108)	-0.001*** (-4.301)	-0.001*** (-4.276)
N	47491	47491	47491	47491	47491	47491
R^2	0.4702	0.4704	0.4705	0.4744	0.4749	0.4788

Table C.22: Stock price volatility and ASVI: lagged ASVI

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLATILITY_t^{st}$. $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)
$ASVI_{t-1}^{st}$	0.016*** (3.856)				
$ASVI_{t-1}^{sq,st}$	-0.011 (-1.248)				
$ASVI_{t-2}^{st}$		0.003 (0.702)			
$ASVI_{t-2}^{sq,st}$		0.008 (1.338)			
$ASVI_{t-3}^{st}$			-0.004 (-1.404)		
$ASVI_{t-3}^{sq,st}$			-0.012 (-1.565)		
$ASVI_{t-4}^{st}$				-0.006 (-1.190)	
$ASVI_{t-4}^{sq,st}$				-0.002 (-0.230)	
$ASVI_{t-5}^{st}$					-0.012*** (-3.528)
$ASVI_{t-5}^{sq,st}$					-0.016** (-2.695)
$VOLATILITY_{t-1}^{st}$	0.296*** (41.322)	0.297*** (40.607)	0.297*** (41.530)	0.297*** (41.664)	0.297*** (40.565)
$VOLATILITY_{t-2}^{st}$	0.165*** (38.613)	0.164*** (37.566)	0.164*** (38.264)	0.165*** (38.086)	0.163*** (37.203)
$VOLATILITY_{t-3}^{st}$	0.130*** (17.626)	0.130*** (17.633)	0.130*** (17.683)	0.130*** (17.810)	0.130*** (17.599)
$VOLATILITY_{t-4}^{st}$	0.121*** (32.877)	0.121*** (32.291)	0.120*** (31.037)	0.120*** (30.162)	0.120*** (31.102)
$VOLATILITY_{t-5}^{st}$	0.130*** (34.153)	0.130*** (34.392)	0.130*** (33.879)	0.130*** (31.596)	0.131*** (33.099)
Constant	-0.001*** (-4.190)	-0.000* (-1.889)	-0.001*** (-4.434)	-0.001*** (-6.310)	-0.001*** (-8.678)
N	47491	47490	47489	47470	47451
R^2	0.4705	0.4701	0.4699	0.4702	0.4703

Table C.23: Stock price volatility and ASVI: time variation

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLATILITY_t^{st}$. $VOLATILITY_t^{st}$ and independent variables are defined in Table A.2. Slope dummy equality tests whether the $ASVI_t^{(sq),st}$ and $ASVI_{t-1}^{(sq),st}$ coefficients from (2) and (3), respectively (5) and (6); are higher than coefficients from (2), respectively (4). T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample periods are from January 2004 to November 2007 (1,4); December 2007 to June 2009 (2,5); July 2009 to December 2014 (3,6) . N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_t^{st}$	0.023***	0.035**	0.032***			
	(3.920)	(2.498)	(3.653)			
$ASVI_t^{sq,st}$	0.027**	0.002	0.016			
	(2.487)	(0.131)	(1.709)			
$ASVI_{t-1}^{st}$				0.009*	0.028***	0.019**
				(2.021)	(4.847)	(2.823)
$ASVI_{t-1}^{sq,st}$				-0.011	-0.015	-0.008
				(-0.789)	(-0.756)	(-1.109)
$VOLATILITY_{t-1}^{st}$	0.219***	0.374***	0.264***	0.218***	0.372***	0.263***
	(20.483)	(27.021)	(37.100)	(19.980)	(27.232)	(37.533)
$VOLATILITY_{t-2}^{st}$	0.110***	0.166***	0.149***	0.109***	0.165***	0.149***
	(15.449)	(16.334)	(22.919)	(15.073)	(15.920)	(23.493)
$VOLATILITY_{t-3}^{st}$	0.094***	0.110***	0.103***	0.096***	0.110***	0.103***
	(8.443)	(9.867)	(10.521)	(8.524)	(9.934)	(10.597)
$VOLATILITY_{t-4}^{st}$	0.081***	0.101***	0.095***	0.082***	0.103***	0.097***
	(11.260)	(8.242)	(17.461)	(11.645)	(8.449)	(17.120)
$VOLATILITY_{t-5}^{st}$	0.068***	0.109***	0.119***	0.067***	0.108***	0.118***
	(9.695)	(12.766)	(14.917)	(9.695)	(12.579)	(14.489)
Constant	-0.082***	0.150***	-0.057***	-0.082***	0.150***	-0.058***
	(-17.637)	(18.715)	(-19.852)	(-16.685)	(18.883)	(-20.722)
N	18434	7535	21524	18432	7535	21524
R^2	0.6007	0.6014	0.6014	0.6017	0.6029	0.6032
F-test for inequality of ASVI slope dummies		0.436	0.374		0.010	0.222
(H_0 : coefficient is higher than in pre-crisis period)		(0.608)	(0.791)		(6.684)	(1.492)
F-test for inequality of $ASVI^{sq}$ slope dummies		0.133	0.469		0.885	0.853
(H_0 : coefficient is higher than in pre-crisis period)		(2.259)	(0.524)		(0.021)	(0.034)

Figure C.3: Stock price volatility and ASVI: IRFs

The figures show impulse response functions of the vertical axis variable to a shock of one standard deviation in the title variable. Vertical axis displays the magnitude of the response (in terms of % of standard deviation), horizontal axis show the time horizon in days.

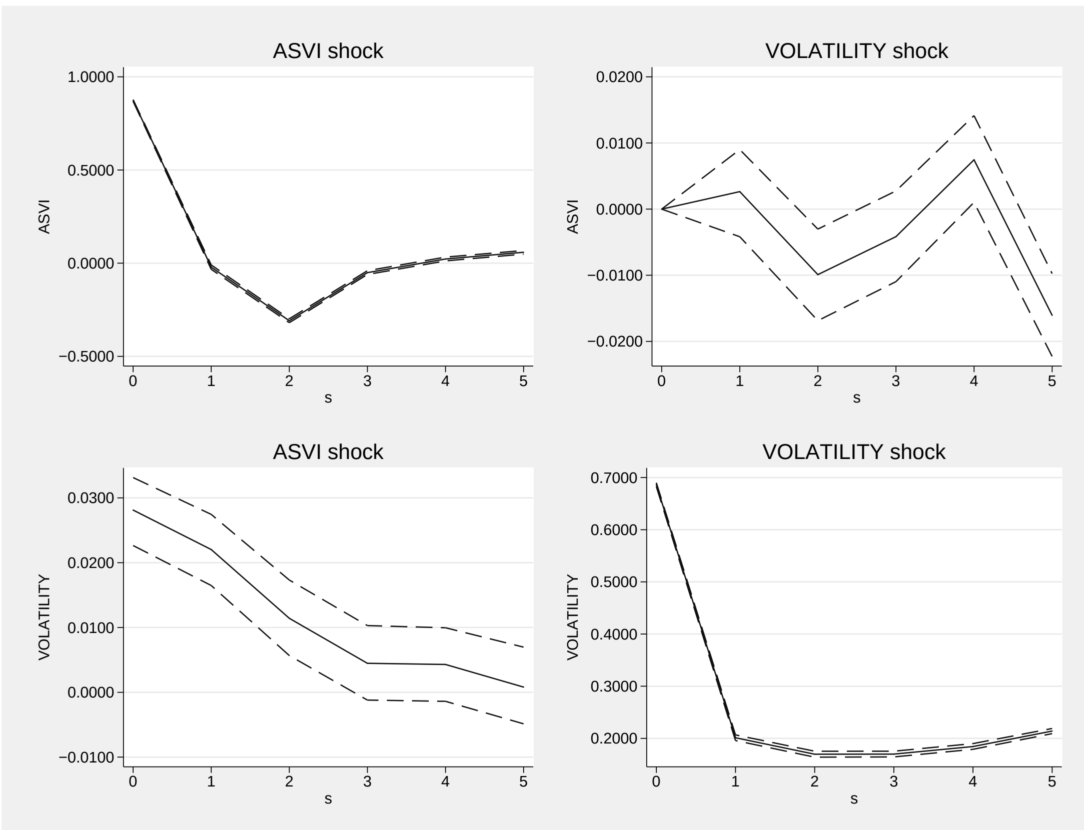
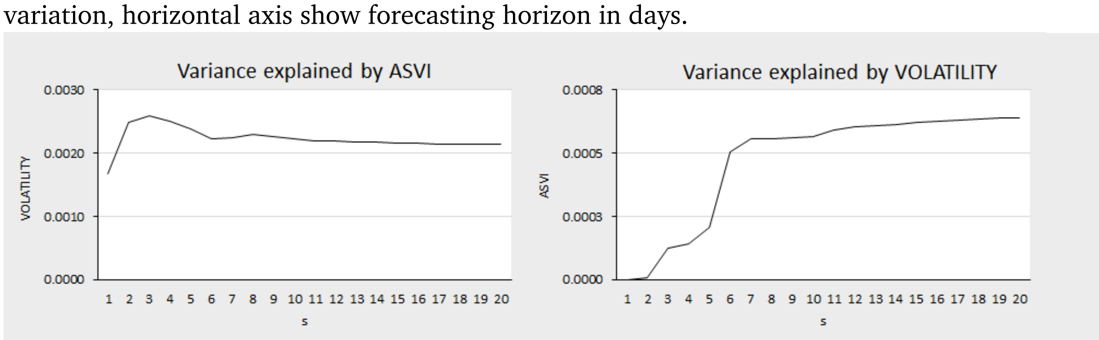


Figure C.4: Stock price volatility and ASVI: variance decomposition

The first figure show fraction of variation in *VOLATILITY* explained by *ASVI*. The second figure show fraction of variation in *ASVI* explained by *VOLATILITY*. Vertical axis show the fraction of variation, horizontal axis show forecasting horizon in days.



C.2.3 Daily returns

Table C.24: Daily returns and ASVI: non-lagged models

Pooled OLS estimation is used to obtain the results. The dependent variable in each regression is $RETURN_t^{st}$. $RETURN_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_t^{st}$		0.002	0.002	0.003	0.003	0.003
		(0.369)	(0.362)	(0.425)	(0.498)	(0.507)
$ASVI_t^{sq,st}$			-0.000			
			(-0.029)			
$RETURN_{t-1}^{st}$	-0.056***	-0.056***	-0.056***	-0.057***	-0.060***	-0.060***
	(-4.342)	(-4.334)	(-4.335)	(-4.406)	(-4.648)	(-4.662)
$RETURN_{t-2}^{st}$	-0.038***	-0.038***	-0.038***	-0.038***	-0.041***	-0.041***
	(-5.103)	(-5.101)	(-5.102)	(-5.181)	(-5.446)	(-5.450)
$VOLUME_t^{st}$				-0.022***		-0.007**
				(-7.000)		(-2.442)
$VOLATILITY_t^{st}$					-0.046***	-0.043***
					(-6.340)	(-5.891)
Constant	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(-0.090)	(-0.088)	(-0.091)	(-0.051)	(-0.022)	(-0.023)
N	47528	47528	47528	47528	47528	47528
R^2	0.0044	0.0044	0.0044	0.0049	0.0065	0.0066

Table C.25: Daily returns and ASVI: lagged models

Pooled OLS estimation is used to obtain the results. The dependent variable in each regression is $RETURN_t^{st}$. $RETURN_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)
$ASVI_{t-1}^{st}$		0.009	0.010*	0.009*	0.009*	0.009*
		(1.732)	(1.761)	(1.747)	(1.763)	(1.767)
$ASVI_{t-1}^{sq,st}$			0.015			
			(1.339)			
$RETURN_{t-1}^{st}$	-0.055***	-0.056***	-0.056***	-0.056***	-0.056***	-0.056***
	(-4.576)	(-4.585)	(-4.582)	(-4.603)	(-4.685)	(-4.686)
$RETURN_{t-2}^{st}$	-0.038***	-0.038***	-0.037***	-0.038***	-0.038***	-0.038***
	(-5.042)	(-5.016)	(-5.011)	(-5.038)	(-5.097)	(-5.098)
$VOLUME_{t-1}^{st}$				-0.005*		-0.002
				(-1.970)		(-0.779)
$VOLATILITY_{t-1}^{st}$					-0.010*	-0.009
					(-1.831)	(-1.708)
Constant	-0.000	-0.000	0.000	-0.000	-0.000	-0.000
	(-0.018)	(-0.012)	(0.136)	(-0.012)	(-0.001)	(-0.000)
N	47527	47527	47527	47527	47527	47527
R^2	0.0042	0.0043	0.0044	0.0043	0.0044	0.0044

Table C.26: Daily returns and ASVI: non-lagged disentangling

Pooled OLS estimation is used to obtain the results. The dependent variable in each regression is $RETURN_t^{st}$. $RETURN_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$ASVI_t^{st}$		-0.007						-0.007			
		(-0.581)						(-0.574)			
$ASVI \times SENT_t^{st}$			0.016***						0.015***		
			(3.349)						(3.247)		
$ASVI_t^{POSSENT,st}$				0.014						0.013	
				(1.360)						(1.311)	
$ASVI_t^{NOSENT,st}$				-0.006						-0.006	
				(-0.550)						(-0.533)	
$ASVI_t^{NEGSENT,st}$				-0.013***						-0.012***	
				(-3.630)						(-3.515)	
$ASVI \times RET_t^{st}$					0.009						0.009
					(0.903)						(0.900)
$ASVI_t^{POSRET,st}$						-0.006					
						(-0.499)					
$ASVI_t^{NEGRET,st}$						-0.003					
						(-0.429)					
$RETURN_{t-1}^{st}$	-0.063***	-0.063***	-0.063***	-0.063***	-0.063***	-0.063***	-0.063***	-0.064***	-0.063***	-0.064***	-0.064***
	(-4.629)	(-4.638)	(-4.616)	(-4.632)	(-4.649)	(-4.662)	(-4.697)	(-4.706)	(-4.684)	(-4.699)	(-4.718)
$RETURN_{t-2}^{st}$	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***	-0.054***
	(-5.773)	(-5.779)	(-5.734)	(-5.707)	(-5.772)	(-5.789)	(-5.858)	(-5.864)	(-5.819)	(-5.793)	(-5.857)
$POSSENT_t$							-0.051***	-0.051***	-0.050***	-0.050***	-0.051***
							(-6.181)	(-6.181)	(-6.114)	(-6.112)	(-6.156)
$NOSENT_t$							0.004	0.004	0.004	0.004	0.004
							(0.681)	(0.675)	(0.696)	(0.644)	(0.681)
$NEGSENT_t$							0.011*	0.011*	0.011*	0.011*	0.012**
							(2.092)	(2.068)	(1.978)	(1.916)	(2.107)
Constant	-0.008**	-0.008**	-0.008**	-0.008**	-0.008**	-0.008**					
	(-2.390)	(-2.395)	(-2.389)	(-2.426)	(-2.362)	(-2.393)					
N	33241	33241	33241	33241	33241	33241	33241	33241	33241	33241	33241
R^2	0.0065	0.0065	0.0067	0.0068	0.0065	0.0065	0.0070	0.0071	0.0072	0.0073	0.0071

Table C.27: Daily returns and ASVI: lagged disentangling

Pooled OLS estimation is used to obtain the results. The dependent variable in each regression is $RETURN_t^{st}$. $RETURN_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$ASVI_{t-1}^{st}$		0.015*						0.015*			
		(1.957)						(1.969)			
$ASVI \times SENT_{t-1}^{st}$			-0.017**						-0.018**		
			(-2.219)						(-2.265)		
$ASVI_{t-1}^{POSSENT,st}$				-0.001						-0.001	
				(-0.107)						(-0.159)	
$ASVI_{t-1}^{NOSENT,st}$				0.006						0.006*	
				(1.734)						(1.748)	
$ASVI_{t-1}^{NEGSENT,st}$				0.018*						0.018*	
				(2.058)						(2.084)	
$ASVI \times RET_{t-1}^{st}$					-0.003						-0.003
					(-0.366)						(-0.366)
$ASVI_{t-1}^{POSRET,st}$						0.002					
						(0.341)					
$ASVI_{t-1}^{NEGRET,st}$						0.014					
						(1.626)					
$RETURN_{t-1}^{st}$	-0.061***	-0.061***	-0.060***	-0.060***	-0.061***	-0.063***	-0.061***	-0.061***	-0.061***	-0.061***	-0.061***
	(-4.853)	(-4.845)	(-4.829)	(-4.821)	(-4.868)	(-4.631)	(-4.928)	(-4.920)	(-4.903)	(-4.895)	(-4.943)
$RETURN_{t-2}^{st}$	-0.053***	-0.052***	-0.053***	-0.052***	-0.053***	-0.054***	-0.053***	-0.053***	-0.053***	-0.053***	-0.053***
	(-5.606)	(-5.571)	(-5.606)	(-5.549)	(-5.605)	(-5.787)	(-5.682)	(-5.647)	(-5.682)	(-5.624)	(-5.680)
$POSSENT_t$							-0.049***	-0.049***	-0.049***	-0.049***	-0.049***
							(-5.678)	(-5.665)	(-5.755)	(-5.689)	(-5.687)
$NOSENT_t$							0.004	0.004	0.004	0.004	0.004
							(0.689)	(0.702)	(0.678)	(0.693)	(0.687)
$NEGSENT_t$							0.012**	0.012**	0.012**	0.012**	0.012*
							(2.113)	(2.125)	(2.183)	(2.218)	(2.084)
Constant	-0.007**	-0.007**	-0.007**	-0.007**	-0.007**	-0.008**					
	(-2.211)	(-2.182)	(-2.218)	(-2.161)	(-2.203)	(-2.386)					
N	33240	33240	33240	33240	33240	33194	33240	33240	33240	33240	33240
R^2	0.0061	0.0063	0.0063	0.0065	0.0061	0.0066	0.0066	0.0068	0.0069	0.0070	0.0066

Table C.28: Daily returns and ASVI: time variation

Pooled OLS estimation is used to obtain the results. The dependent variable in each regression is $RETURN_t^{st}$. $RETURN_t^{st}$ and independent variables are defined in Table A.2. The bottom line present F-test for equality of *ASVI* slope-dummy variables. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1)	(2)
$ASVI_t^{PRE,st}$	-0.000	
	(-0.028)	
$ASVI_t^{CRI,st}$	-0.014**	
	(-2.477)	
$ASVI_t^{POST,st}$	0.013**	
	(2.561)	
$ASVI_{t-1}^{PRE,st}$		0.008*
		(2.009)
$ASVI_{t-1}^{CRI,st}$		0.013
		(0.819)
$ASVI_{t-1}^{POST,st}$		-0.002
		(-0.373)
$RETURN_{t-1}^{st}$	-0.057***	-0.055***
	(-4.339)	(-4.604)
$RETURN_{t-2}^{st}$	-0.038***	-0.037***
	(-5.144)	(-4.863)
Constant	-0.000	-0.000
	(-0.088)	(-0.011)
N	47528	47527
R^2	0.005	0.004
F-test for equality of ASVI slope dummies (p-value)	0.006	0.283
(H_0 : all coefficients are equal)	(6.764)	(1.353)

Figure C.5: Daily returns and ASVI: IRFs

The figures show impulse response functions of the vertical axis variable to a shock of one standard deviation in the title variable. Vertical axis displays the magnitude of the response (in terms of % of standard deviation), horizontal axis show the time horizon in days.

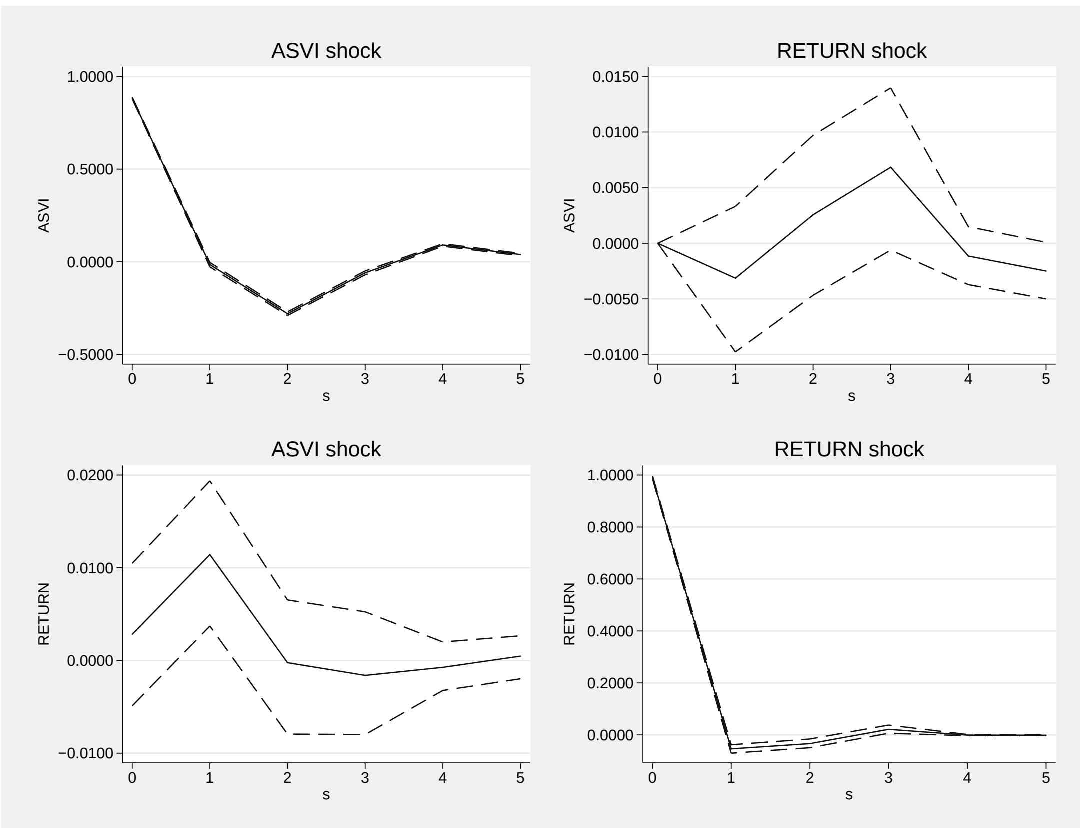


Figure C.6: Daily returns and ASVI: variance decomposition

The first figure show fraction of variation in *RETURN* explained by *ASVI*. The second figure show fraction of variation in *ASVI* explained by *RETURN*. Vertical axis show the fraction of variation, horizontal axis show forecasting horizon in days.

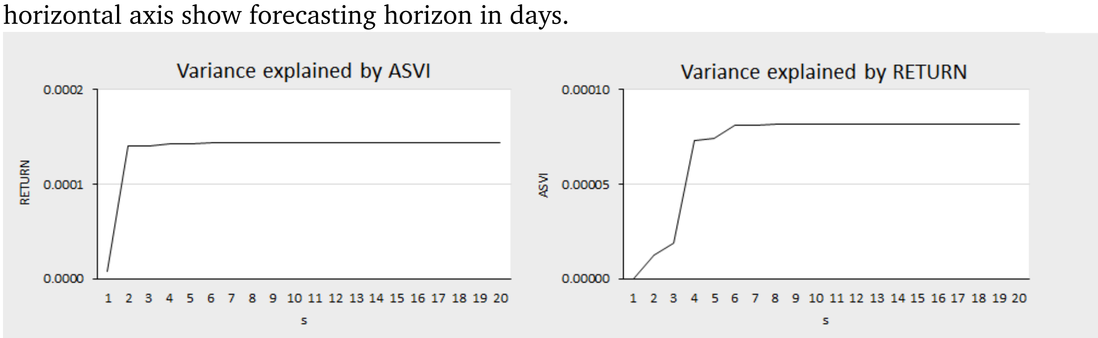


Figure C.7: Daily returns and ASVI: IRFs (positive sentiment)

The figures show impulse response functions of the vertical axis variable to a shock of one standard deviation in the title variable. Vertical axis displays the magnitude of the response (in terms of % of standard deviation), horizontal axis show the time horizon in days.

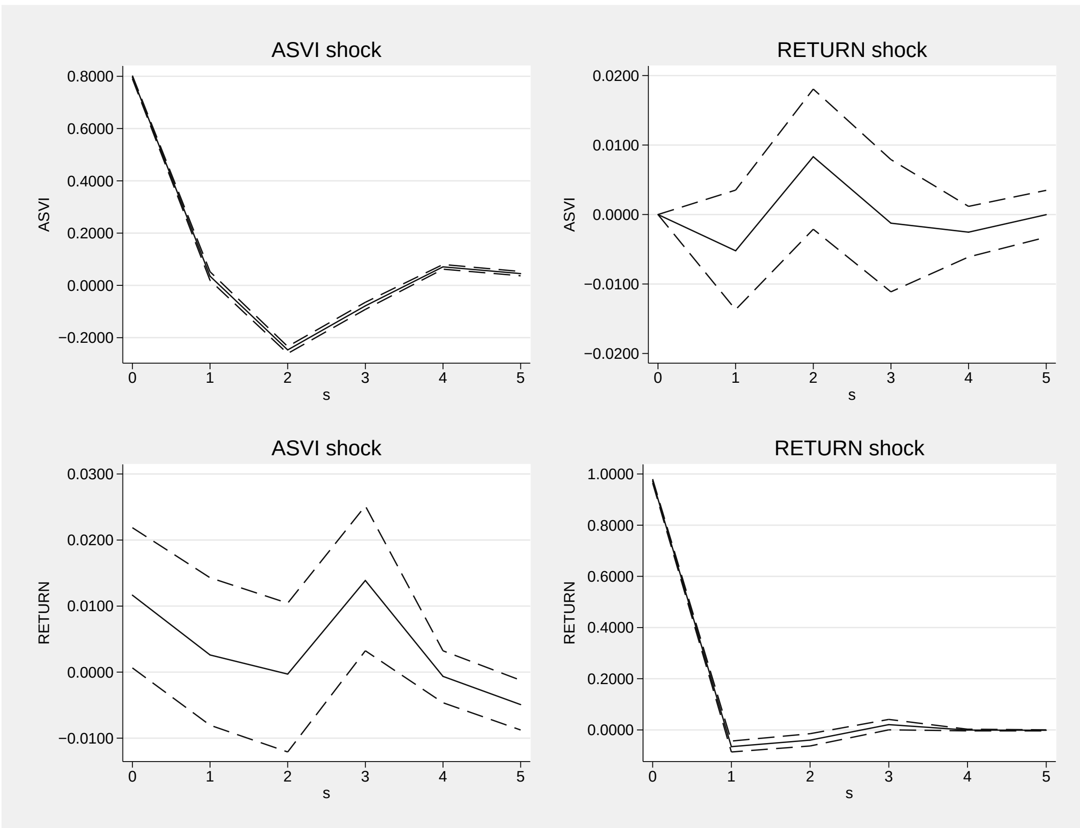


Figure C.8: Daily returns and ASVI: IRFs (negative sentiment)

The figures show impulse response functions of the vertical axis variable to a shock of one standard deviation in the title variable. Vertical axis displays the magnitude of the response (in terms of % of standard deviation), horizontal axis show the time horizon in days.

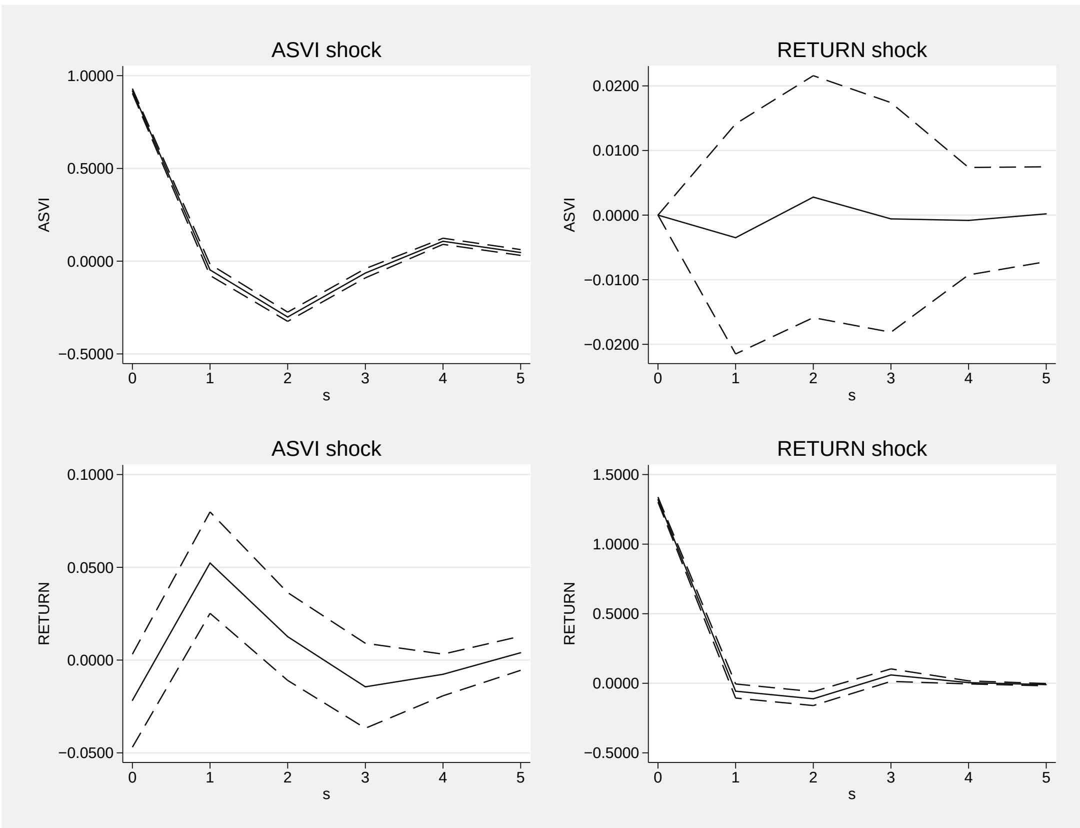
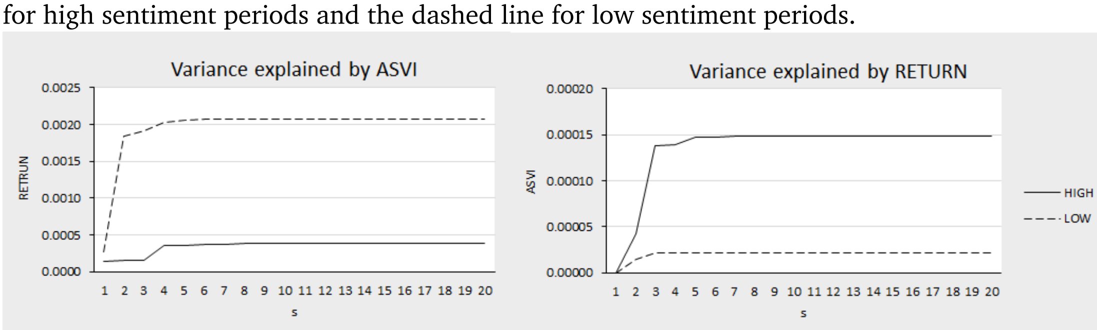


Figure C.9: Daily returns and ASVI: variance decomposition for positive and negative sentient

The first figure show fraction of variation in *RETURN* explained by *ASVI*. The second figure show fraction of variation in *ASVI* explained by *RETURN*. Vertical axis show the fraction of variation, horizontal axis show forecasting horizon in days. The solid line show variances decompositions for high sentiment periods and the dashed line for low sentiment periods.



C.3 Discussion

Table C.29: Different ASVI performance comparison

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLUME_t^{st}$. $VOLUME_t^{st}$ and independent variables are defined in Table A.2. Second row depicts specific *ASVI* used as independent variable in the regression. AIC and BIC are Akaike and Schwarz information criteria. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	(1) $ASVI_{i,t}^{1,week}$	(2) $ASVI_{i,t}^{2,week}$	(3) $ASVI_{i,t}^{3,week}$	(4) $ASVI_{i,t}^{4,week}$	(5) $ASVI^{1,inc}$	(6) $ASVI^{2,inc}$	(7) $ASVI^{3,inc}$	(8) $ASVI^{4,inc}$	(9) $ASVI^{5,inc}$	(10) $ASVI^{6,inc}$	(11) $ASVI^{7,inc}$	(12) $ASVI^{1,exc}$	**(13)** $ASVI^{2,exc}$
ASVI	0.026***	0.024***	0.022***	0.022***	-0.009**	-0.002	0.006*	0.018***	0.024***	0.028***	0.028***	0.016***	**0.024*** **
	(5.233)	(5.196)	(5.543)	(4.866)	(-2.545)	(-0.594)	(1.766)	(3.895)	(4.723)	(5.193)	(5.236)	(3.027)	**(4.330)**
$VOLUME_{t-1}^{st}$	0.502***	0.503***	0.503***	0.504***	0.507***	0.507***	0.507***	0.506***	0.505***	0.504***	0.504***	0.509***	**0.507*** **
	(59.701)	(58.054)	(58.247)	(58.862)	(59.719)	(59.108)	(60.570)	(60.319)	(60.199)	(60.053)	(60.363)	(58.828)	**(60.234)**
$VOLUME_{t-2}^{st}$	0.114***	0.114***	0.115***	0.115***	0.115***	0.115***	0.116***	0.118***	0.118***	0.117***	0.117***	0.116***	**0.118*** **
	(23.131)	(22.819)	(22.554)	(22.347)	(23.074)	(22.644)	(23.483)	(23.077)	(23.331)	(23.672)	(23.332)	(22.734)	**(23.498)**
$VOLUME_{t-3}^{st}$	0.090***	0.091***	0.090***	0.091***	0.090***	0.090***	0.090***	0.089***	0.092***	0.092***	0.092***	0.089***	**0.090*** **
	(14.255)	(14.552)	(14.733)	(14.772)	(14.641)	(14.573)	(14.884)	(14.659)	(14.914)	(14.944)	(14.826)	(14.371)	**(14.480)**
$VOLUME_{t-4}^{st}$	0.058***	0.057***	0.056***	0.056***	0.057***	0.056***	0.056***	0.054***	0.054***	0.055***	0.056***	0.055***	**0.054*** **
	(11.480)	(11.258)	(11.016)	(10.850)	(11.607)	(11.751)	(11.544)	(11.302)	(10.877)	(10.905)	(10.927)	(11.342)	**(11.035)**
$VOLUME_{t-5}^{st}$	0.107***	0.105***	0.105***	0.105***	0.100***	0.101***	0.102***	0.103***	0.102***	0.102***	0.103***	0.102***	**0.103*** **
	(16.593)	(15.993)	(15.714)	(15.622)	(14.588)	(15.152)	(15.026)	(15.199)	(15.080)	(15.185)	(15.725)	(15.023)	**(15.134)**
Constant	-0.000***	-0.000***	-0.000	-0.000**	-0.001***	-0.001***	-0.001***	-0.000***	-0.000***	-0.000***	-0.000***	-0.001***	**-0.000*** **
	(-7.378)	(-5.237)	(-0.672)	(-2.137)	(-17.636)	(-16.391)	(-19.766)	(-12.323)	(-8.776)	(-8.954)	(-8.337)	(-13.320)	**(-11.279)**
N	47279	47295	47183	47086	47423	47465	47454	47484	47475	47485	47486	47473	**47493**
R^2	0.6029	0.6035	0.6034	0.6035	0.602	0.6018	0.6019	0.6029	0.6038	0.6044	0.6047	0.6033	**0.6041**
AIC	25675	25751	25759	25754	25945	26037	25979	25916	25790	25742	25729	25894	**25810**
BIC	25727	25804	25812	25807	25998	26090	26032	25968	25842	25795	25781	25947	**25863**

	(14) $ASVI^{3,exc}$	(15) $ASVI^{4,exc}$	(16) $ASVI^{5,exc}$	(17) $ASVI^{6,exc}$	(18) $ASVI^{7,exc}$	(19) $ASVI_{i,t}^{1,weighted}$	(20) $ASVI_{i,t}^{2,weighted}$	(21) $ASVI_{i,t}^{3,weighted}$	(22) $ASVI_{i,t}^{4,weighted}$	(23) $ASVI_{i,t}^{5,weighted}$	(24) $ASVI_{i,t}^{16,weighted}$	(25) $ASVI_{i,t}^{7,weighted}$
ASVI	0.027***	0.029***	0.029***	0.029***	0.029***	0.028***	0.031***	0.028***	0.029***	0.028***	0.028***	0.025***
	(4.885)	(5.200)	(5.008)	(5.010)	(5.296)	(4.407)	(4.776)	(4.911)	(4.848)	(4.803)	(4.748)	(4.536)
$VOLUME_{t-1}^{st}$	0.504***	0.503***	0.503***	0.502***	0.502***	0.506***	0.503***	0.502***	0.502***	0.502***	0.502***	0.502***
	(59.658)	(59.550)	(59.336)	(58.945)	(58.988)	(59.800)	(59.327)	(58.759)	(58.392)	(57.688)	(58.222)	(59.369)
$VOLUME_{t-2}^{st}$	0.119***	0.118***	0.117***	0.116***	0.116***	0.118***	0.118***	0.118***	0.116***	0.115***	0.115***	0.114***
	(24.068)	(23.659)	(23.364)	(23.174)	(22.911)	(23.744)	(23.594)	(23.574)	(23.425)	(22.792)	(22.863)	(22.469)
$VOLUME_{t-3}^{st}$	0.091***	0.092***	0.092***	0.092***	0.091***	0.091***	0.093***	0.093***	0.093***	0.092***	0.092***	0.092***
	(14.843)	(14.953)	(14.953)	(14.820)	(14.782)	(14.850)	(14.907)	(14.921)	(14.922)	(15.015)	(14.954)	(14.989)
$VOLUME_{t-4}^{st}$	0.055***	0.056***	0.056***	0.056***	0.056***	0.055***	0.056***	0.057***	0.058***	0.058***	0.058***	0.058***
	(11.069)	(10.923)	(10.963)	(10.834)	(10.977)	(10.978)	(10.918)	(11.157)	(11.260)	(11.301)	(11.148)	(11.095)
$VOLUME_{t-5}^{st}$	0.103***	0.103***	0.104***	0.105***	0.105***	0.102***	0.102***	0.102***	0.103***	0.104***	0.105***	0.105***
	(15.309)	(15.355)	(15.768)	(16.019)	(15.907)	(14.842)	(14.840)	(14.951)	(15.319)	(15.651)	(15.766)	(15.635)
Constant	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.001***	-0.000***	-0.001***	-0.000***	-0.001***	-0.000***	-0.000***
	(-9.485)	(-10.802)	(-9.444)	(-8.475)	(-6.355)	(-18.090)	(-14.996)	(-21.512)	(-8.064)	(-14.929)	(-6.236)	(-10.239)
N	47478	47487	47478	47460	47425	47588	47565	47547	47510	47479	47463	47437
R^2	0.6041	0.6047	0.6048	0.6049	0.6048	0.6047	0.6048	0.6043	0.6042	0.6039	0.6037	0.6034
AIC	25780	25741	25715	25696	25676	25879	25839	25862	25859	25869	25884	25903
BIC	25833	25794	25768	25748	25728	25932	25891	25915	25912	25922	25937	25955

Table C.30: ASVI and trading volume: weekly data

Fixed-effects estimation is used to obtain the results. The dependent variable in each regression is $VOLUME_t^{st}$. $VOLUME_t^{st}$ and independent variables are defined in Table A.2. T-statistics computed from robust standard errors clustered by firm are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from January 2004 to December 2013. N is number of observations.

	$VOLUME_t^{st}$
ASVI	0.052***
	(3.791)
$VOLUME_{t-1}^{st}$	0.490***
	(42.396)
$VOLUME_{t-2}^{st}$	0.125***
	(9.582)
$VOLUME_{t-3}^{st}$	0.061***
	(5.146)
$VOLUME_{t-4}^{st}$	0.057***
	(5.505)
$VOLUME_{t-5}^{st}$	0.151***
	(18.553)
$Constant$	0.000***
	(3.865)
N	9880
R^2	0.6266

Appendix D

Complementary tables to Chapter 6

D.1 Does investor attention react to IPOs?

Table D.1: ASVI drivers prior IPO

OLS estimation is used to obtain the results. The dependent variable in each regression is $ASVI_i^{st}$. $ASVI_i^{st}$ and independent variables are defined in Table A.4. T-statistics computed from bootstrapped standard errors are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1)	(2)	(3)
$Offering\ size_i^{st}$	-0.0259 (-0.31)		
$NYSE_i$		0.0875 (0.54)	
$Crisis_i$			0.330** (2.16)
$Constant$	-0.156* (-1.68)	-0.229** (-2.08)	-0.0485 (-0.41)
N	71	69	70

D.2 Investor attention and IPO stylized facts

D.2.1 Initial returns

Table D.2: IPO first-day return and ASVI

OLS estimation is used to obtain the results. The dependent variable in each regression is the IPO first day return IR_i^{st}. IR_i^{st} and independent variables are defined in Table A.4. T-statistics computed from OLS standard errors are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
$ASVI_i^{st}$	0.414***				0.400***	0.437***	0.397***	0.404***	0.357***				
	(3.311)				(3.138)	(3.768)	(3.169)	(3.217)	(3.084)				
$Offering\ size_i^{st}$		-0.119			-0.094		-0.015	-0.130	-0.068				
		(-1.090)			(-0.929)		(-0.153)	(-1.269)	(-0.685)				
$Sentiment_i^{st}$			0.112			0.003		-0.020			-0.039		
			(0.876)			(0.025)		(-0.177)			(-0.242)		
$\triangle Sentiment_i^{st}$				0.044					0.082			-0.037	
				(0.345)					(0.711)			(-0.291)	
$ASVI_i^{POSSENT,st}$										0.297**	0.275**	0.268**	0.344***
										(2.600)	(2.034)	(2.276)	(2.815)
$ASVI_i^{NEGSENT,st}$										0.163	0.152	0.136	0.280
										(1.231)	(1.253)	(1.260)	(1.539)
$ASVI_i^{NOSENT,st}$													0.268
													(1.365)
$Constant$	0.003	-0.045	-0.044	0.023	-0.065	-0.034	-0.019	-0.073	-0.089	-0.011	-0.048	-0.028	0.064
	(0.026)	(-0.443)	(-0.398)	(0.203)	(-0.681)	(-0.369)	(-0.190)	(-0.763)	(-0.936)	(-0.109)	(-0.484)	(-0.275)	(0.574)
N	70	72	67	70	67	65	67	63	66	69	68	67	66

Table D.3: IPO first-day return and lagged ASVI

OLS estimation is used to obtain the results. The dependent variable in each regression is the IPO first day return IR_i^{st}. IR_i^{st} and independent variables are defined in Table A.4. T-statistics computed from OLS standard errors are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1)	(2)	(3)	(4)	(5)
$ASVI_{i,t-1}^{st}$	0.411***				
	(3.85)				
$ASVI_{i,t-2}^{st}$		0.349***			
		(3.18)			
$ASVI_{i,t-3}^{st}$			0.246**		
			(2.17)		
$ASVI_{i,t-4}^{st}$				0.137	
				(1.18)	
$ASVI_{i,t-5}^{st}$					0.0318
					(0.27)
$Constant$	-8.63e-10	-1.36e-09	1.96e-10	6.36e-10	3.24e-10
	(-0.00)	(-0.00)	(0.00)	(0.00)	(0.00)
N	75	75	75	75	75

D.2.2 Long-term returns

Figure D.1: Long-term cumulative returns for low and high attention IPOs

Average and median cumulative log-returns: first day closing price to the (1) closing price one year, (2) half a year (3) and 91 days after IPO; and the closing price one month after IPO to (4) the closing price one year (5) and half a year after IPO. The vertical axis show the return magnitude and the horizontal axis show the period over which the return is calculated.

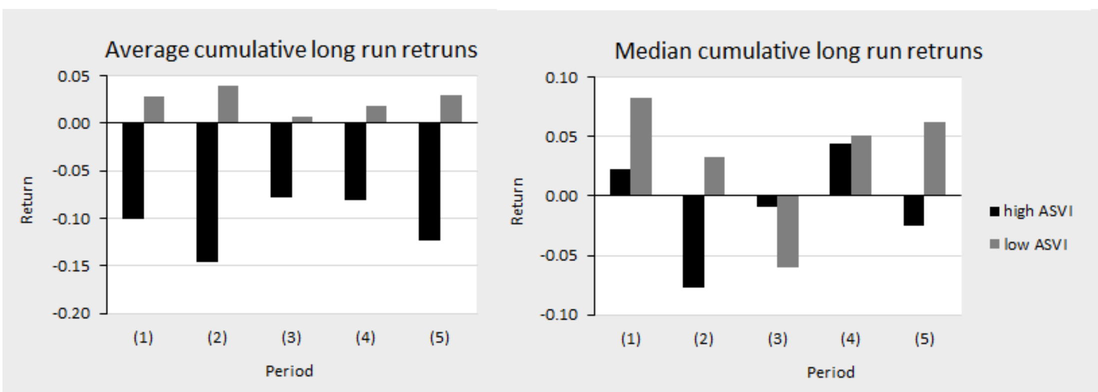


Table D.4: IPO long-term performance and ASVI

OLS estimation is used to obtain the results. The dependent variable in each regression is cumulative long-term return LR_i^{st}. LR_i^{st} and independent variables are defined in Table A.4. The columns show over which period the cumulative return is calculated: first day closing price to the (1) closing price one year, (2) half a year (3) and 91 days after IPO; and the closing price one month after IPO to (4) the closing price one year (5) and half a year after IPO. T-statistics computed from bootstrapped standard errors (1,3,4,5) and OLS standard errors (2) are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1)	(2)	(3)	(4)	(5)
$ASVI_i^{st}$	-0.171	-0.204*	-0.0662	-0.190	-0.187**
	(-1.19)	(-1.97)	(-0.63)	(-1.29)	(-2.15)
$Constant$	0.0292	0.0711	0.102	0.0265	0.0775
	(0.22)	(0.84)	(1.30)	(0.19)	(1.00)
N	59	60	59	59	60

Table D.5: IPO long-term performance, ASVI and initial returns

OLS estimation is used to obtain the results. The dependent variable in each regression is cumulative long-term return LR_i^{st}. LR_i^{st} and independent variables are defined in Table A.4. The columns show over which period the cumulative return is calculated: first day closing price to the (1) closing price one year, (2) half a year (3) and 91 days after IPO; and the closing price one month after IPO to (4) the closing price one year (5) and half a year after IPO. T-statistics computed from bootstrapped standard errors (1,3,4,5) and OLS standard errors (2) are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
IR_i^{st}	-0.143	-0.053	-0.020	-0.221**	-0.162**					
	(-1.263)	(-0.566)	(-0.237)	(-2.083)	(-2.052)					
$ASVI \times IR_i^{st}$						-0.387*	-0.317**	0.112	-0.411**	-0.293**
						(-1.94)	(-2.19)	(1.07)	(-2.47)	(-2.16)
$Constant$	0.221**	0.218***	0.197***	0.195**	0.200***	0.0185	0.0768	0.176***	-0.0229	0.104
	(2.477)	(3.094)	(3.048)	(2.245)	(3.185)	(0.14)	(0.93)	(2.76)	(-0.18)	(1.35)
N	56	56	57	57	57	58	59	58	60	60

Table D.6: IPO long-term performance, ASVI and sentiment

OLS estimation is used to obtain the results. The dependent variable in each regression is cumulative long-term return LR_i^{st}. LR_i^{st} and independent variables are defined in Table A.4. The columns show over which period the cumulative return is calculated: first day closing price to the (1) closing price one year, (2) half a year (3) and 91 days after IPO; and the closing price one month after IPO to (4) the closing price one year (5) and half a year after IPO. T-statistics computed from OLS standard errors are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
$ASVI_i^{POSSENT,st}$	-0.388**	-0.206**	0.007	-0.326***	-0.094					
	(-2.340)	(-2.300)	(0.067)	(-2.932)	(-0.813)					
$ASVI_i^{NOSENT,st}$	-0.033	-0.019	0.247	-0.008	-0.019					
	(-0.189)	(-0.099)	(1.127)	(-0.048)	(-0.101)					
$ASVI_i^{NEGSENT,st}$	0.056	-0.052	0.034	0.085	0.013					
	(0.449)	(-0.573)	(0.410)	(0.729)	(0.134)					
$POSSENT_i$						-0.414**	-0.058	-0.003	-0.423**	-0.053
						(-2.066)	(-0.273)	(-0.014)	(-2.116)	(-0.252)
$NOSENT_i$						0.189	-0.073	-0.044	0.163	-0.118
						(0.900)	(-0.329)	(-0.200)	(0.780)	(-0.538)
$NEGSENT_i$						0.309	0.158	0.055	0.350	0.206
						(1.363)	(0.665)	(0.232)	(1.550)	(0.866)
$Constant$	0.031	0.111	0.208**	0.056	0.160*					
	(0.281)	(1.232)	(2.302)	(0.552)	(1.823)					
N	54	57	56	55	55	62	62	62	62	62

D.3 Investor attention in the setting of model by Ma and Tsai

Figure D.2: True discount and market overreaction for low and high attention IPOs

Average and median trued discount and market reaction for low and high attention IPOs.

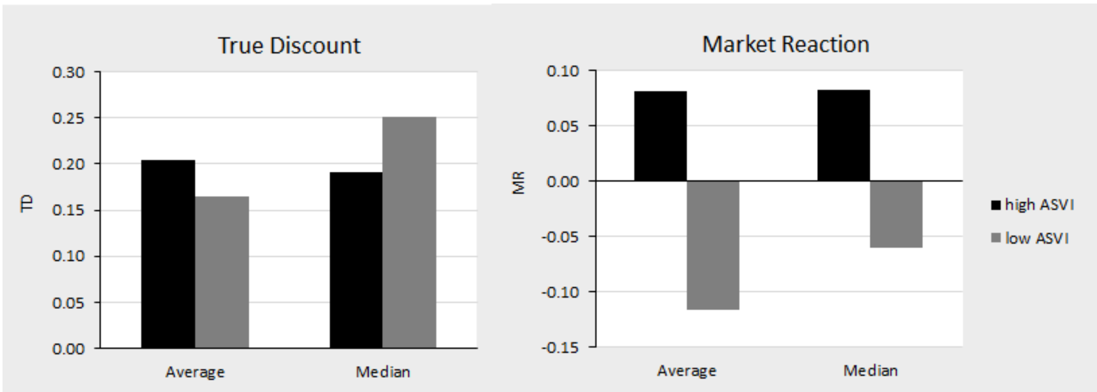


Table D.7: Ma-Tsai model and ASVI

OLS estimation is used to obtain the results. The dependent variables are true discount TD_i^{st} and market reactionMR_i^{st} as defined by Ma and Tsai (2002). TD_i^{st}, MR_i^{st} and independent variables are defined in Table A.4. T-statistics computed from bootstrapped standard errors are in parentheses. Outlying values were omitted. *, **, and *** represent significance at the 10%, 5%, and 1% level, respectively. N is number of observations.

	(1) TD_i^{st}	(2) MR_i^{st}	(3) TD_i^{st}	(4) MR_i^{st}	(5) TD_i^{st}	(6) MR_i^{st}	(7) TD_i^{st}	(8) MR_i^{st}
$ASVI_i^{st}$	0.00754	0.221*						
	(0.06)	(1.88)						
$ASVI \times IR_i^{st}$			0.109	0.428*				
			(0.62)	(1.86)				
$ASVI_i^{POSSENT,st}$					-0.130	0.252		
					(-0.717)	(1.641)		
$ASVI_i^{NOSENT,st}$					-0.081	0.083		
					(-0.494)	(0.482)		
$ASVI_i^{NEGSENT,st}$					0.025	0.153		
					(0.214)	(1.136)		
$POSSENT_i$							0.098	-0.013
							(0.464)	(-0.063)
$NOSENT_i$							-0.130	-0.042
							(-0.590)	(-0.190)
$NEGSENT_i$							0.027	0.066
							(0.114)	(0.276)
$Constant$	-0.0451	0.0946	-0.0406	0.0389	-0.067	0.051		
	(-0.39)	(0.83)	(-0.39)	(0.31)	(-0.590)	(0.457)		
I	58	56	56	57	56	55	62	62

Appendix E

Thesis proposal

Master Thesis Proposal

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Author:	**Bc. Tomáš Vakrman**	Supervisor:	PhDr. Ladislav Krištoufek
E-mail:	Tomas.vakrman@gmail.com	E-mail:	kristoufek@gmail.com
Phone:	724039080	Phone:	607684417
Specialization:	*Economics and Finance*	Defense Planned:	February 2014

Proposed Topic:

Google searches and financial markets: IPOs and uncertainty

Topic Characteristics:

The thesis will focus on possible utilization of Google Trends dataset in describing phenomena on financial markets. Among others, I will examine whether trading volume, volatility and stock prices can be predicted by correspondent search data.

The existing literature suggests that the first two variables should be predictable by such data. On the other hand, price should not. The logic behind these results is simple. Higher search volume should indicate higher trading volume and volatility, but should not say anything about the price movement, since the surge in search volume can signify both market's propensity to sell and buy.

Substantial part of the thesis should be devoted to IPOs. In this part, I would like to build on the work of Ma & Tsai (2001), who impugn the current theory concerning under-pricing and initial return. The theory usually uses initial return as a proxy to measure the under-pricing of an issue. However, Ma & Tsai argue that initial return has two parts, trued discount and market over-/under- reaction. Thus, suggesting that initial return and IPO discount are the same only if markets are efficient.

Concerning this, I would like to examine whether Google Trends data can be used as a measure of investors' interest in the issue and resultant market over-/under- reaction. If this proves right, Google Trends can be used as predictor of price movements after IPOs.

Furthermore, I would like to examine whether the effect of investor interest on post-IPO price movements differs for various industries.

Hypotheses:

1. Google search data predicts trading volume of stocks.
2. Google search data predicts volatility of stocks.
3. Google search data does not predict price of stocks.
4. The change in web search volume for company's name prior the IPO leads to market overreaction on the stock emission and consequently to higher first day return on company's stocks.

Methodology:

I will employ econometric methods to assess Google search data's usability for describing variables at financial markets. Firstly, cross correlation between variables will be assessed to examine some trivial relationship. Secondly, I will use VAR model to control for bilateral relationship between variables. Lastly, in- and out- sample forecasting ability of Google data will be investigated.

Outline:

1. Introduction
2. Google Trends data & trading volume
3. Google Trends data & volatility
4. Google Trends data & price movements
5. Case study – Google Trends & IPOs
6. Conclusion

Core Bibliography:

Allen, F. & Faulhaber, G.R., 1989. Signalling by Underpricing in the IPO Market. Journal of financial Economics, 23(2), s.303–323.

Askitas, N. & Zimmermann, K., 2009. Google econometrics and unemployment forecasting.

Bank, M., Larch, M. & Peter, G., 2011. Google search volume and its influence on liquidityand returns of German stocks. Financial Markets and Portfolio Management, s.1–26.

Bordino, I. et al., 2011. Web search queries can predict stock market volumes. Arxiv preprint arXiv:1110.4784.

D'Amuri, F. & Marcucci, J., 2010. "Google it!" Forecasting the US unemployment rate with a Google job search index. Fondazione Eni Enrico Mattei Working Papers, s.421.

Dimpfl, T. & Jank, S., 2011. Can internet search queries help to predict stock market volatility?

Ginsberg, J. et al., 2008. Detecting influenza epidemics using search engine query data. Nature, 457(7232), s.1012–1014.

Chamberlin,G, 2010. Googling the present. Economic and Labour Market Review, 4(12),s.59–95.

Choi, H. & Varian, H., 2009a. Predicting initial claims for unemployment benefits. Google Inc.

Choi, H. & Varian, H., 2009b. Predicting the present with Google trends. Google inc, s.1–23.

Ma, T. & Tsai, P.R., 2001. Are initial return and IPO discount the same thing? A comparison of direct public offerings and underwritten IPOs, Working Paper. National Sun Yat-sen University.

McLaren, N. & Shanbhogue, R., 2011. Using internet search data as economic indicators. Bank of England Quarterly Bulletin No. 2011 Q2.

Preis, T., Reith, D. & 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),s.5707–5719.

Schmidt, T. & Vosen, S., 2009. Forecasting Private Consumption: Survey Based Indicators Vs. Google Trends, Rheinisch-Westfälisches Institut für Wirtschaftsforschung.

William A. Reese, 1998. IPO Underpricing, Trading Volume, and Investor Interest. SSRN eLibrary. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=115348.

Vah

Author

Ladislav Kristoufek
cn=Ladislav Kristoufek, o=Academy of Sciences of the Czech Republic, ou=Institute of Information Theory and Automation, email=kristouf@utia.cas.cz, c=CZ
2013.09.04 13:00:20 +02'00'

Supervisor