University of Macedonia

INTERDEPARTMENTAL PROGRAMME OF POSTGRADUATE STUDIES IN ECONOMICS

MASTER THESIS

Investor Attention and the Predictability of Stock Returns

Author:

Konstantinos Vasilopoulos

Supervisor:

Theodore Panagiotidis

March, 2016

Investor Attention and the Predictability of Stock Returns

Konstantinos Vasilopoulos

Abstract

In this study we investigate whether web searches have any impact in trading activity and abnormal returns predictability. We use Goolge Search Volume as an direct proxy for investor's attention for all the stocks in S&P500 from 2008-2014. The search term is the stock's ticker. We assume that individuals use web searches when they want to buy the stock. We find that search volume increase trading activity after controlling for existing proxies of attention, such as absolute abnormal returns and return volatility. Search volume has also power into predicting returns one and two weeks ahead. It creates positive buying pressure at the second week and there is a total reversal within a year. We find that search volume apply negative pressure in one year returns, which can be interpreted as the reduction the information asymmetries and enhancement of market efficiency.

March, 2016

Contents

1 Introduction

The internet has already surpassed newspapers as a source Americans turn to for national and international news[1]. The number of internet users had reached the 3 billion at the end of 2014. Almost 59 percent of the United States population regularly use the Internet, this represents 39 percent of Internet usage worldwide. Search engines play a major role in the way we obtain information. According to a 2013 comScore[2] public release[3], as of December 2012, Google enjoyed a 65,2 percent share of web search volume worldwide. Google now process over 40.000 searches every second on average, which translates to over 3.5 billion searches per day and 1.2 trillion searches per year worldwide. At 2008 Google launched Google Insight for Search (now Google Trends), a tool that helps understanding searches behaviour. This task initiated a lot of search in how information is incorporated in stock prices, where Google searches proxy for information demand.

Some may assume that the stock market becomes more efficient when investor pay more attention, as stock prices incorporate information faster. However, other studies support that the greater the investor attention can exacerbate the effect of investor behavioural biases on the market prices.

We present an alternative to Efficient Market Hypothesis (EMH) by Shleifer and Summers (1990) investors are not fully rational and their demand for risky assets is affected by their beliefs or sentiments that are not fully justified by the fundamental news. Furthermore, arbitrage, i.e. trading by fully rational investors is not subject to such sentiment, therefore arbitrage is limited. Investors are divided into two distinct groups, arbitrageurs and noise traders. Arbitrageurs or rational speculators form fully rational expectations about security prices and they work to bring prices near fundamentals. Noise traders or irrational speculators may be subjected to systematic biases. These two assumptions conclude that changes in investor sentiment are not fully countered by arbitrageurs and so affect security returns.

There are two prevalent school of thoughts of how investor sentiment affects security prices. First, Merton (1987) with "information theory" introduces the notion that firms which attract less attention have to offer higher returns so that investors can be compensated for imperfect diversification due to incomplete information. On the other hand, Barber and Odean (2008) , with "attention theory" support that an increase in retail in-

[1] Rosentiel M. et al , "The Role of Internet."Pew Research Center , September 23, 2011

[2] www.comscore.com

[3] Meyer, David. "Microsoft down to fifth place in comScore's global search stats, thanks to Yandex." GigaOM. February 6, 2013.

vestor's attention affect stock prices. They test the propositions that individual investors are net buyers of attention grabbing stocks. This occurs because selling a stock require individuals to have owned the stock beforehand, whereas individuals who are willing to buy can choose from a large set of alternatives. Hence, assuming that retail investors are uniformed on average, an increase investor attention leads to net buying and thus to positive price pressure.

In this paper we employ aggregated Google search frequency for stock tickers as a direct proxy for investor attention. We construct a measure of Abnormal Search Volume (ASV), where we take the natural log of google search volume and we detract the median of the natural log of the previous 8 weeks (Da et al., 2011). We investigate whether Google Trends for stock tickers are related with trading activity and the cross section of stock returns.

First, we examine the relationship of ASV with some common indirect proxies of attention such as Abnormal Trading Volume, Absolute Abnormal Returns and two measures of Return Volatility. The first measure of volatility is derived from a GARCH(1,1) model and the second one is rolling standard deviation of the returns. These proxies make the assumption that if there is abnormal trading volume, extreme returns on excess volatility investor should have paid attention to it. In other words, in the absence of news, extreme returns or volatility became measures of investor attention themselves. However, that is an assertive argument. There can be increasing investor sentiment with no corresponding event news. Hence we examine the ASV concerning the existing proxies of attention, however if ASV can be explained by alternative measure of investor attention then it has no use as a direct proxy of attention.

Second, we examine whether ASV has any impact in trading activity. We define as trading activity the log of trading volume minus the log of the mean of the previous 2 months (Trad Act). With the construction of this measure we try to be consistent with ASV and capture primarily the outstanding trading volume in a time interval of two months. We find that ASV has a positive relationship with Trad Act, thus retail investors can increase trading activity.

Third, we try to to asses the power of ASV in predicting future abnormal returns for one to four weeks ahead. We also make a distinction between positive and negative ASV in order to check for asymmetric effects. The evidence indicate that ASV have power into predicting abnormal returns for the first two weeks. The first week the relationship is negative and low, however the second week we find a substantial buying pressure two time higher than the first week. Furthermore, we conclude that the pressure to future abnormal

returns comes mainly from negative ASV. Both results indicate that the investor delay into completing their buying positions after their web searches.

Finally, we use cumulative abnormal for the rest of the year. Our results, are partially in par with Da et al. (2011). We find a complete reversal which amount to 6 times the initial additive pressure of the first two weeks.

Generally, we find a short-run positive relationship and long-run negative relationship, which can interpreted as increasing searching volume has a buying pressure in the short-run and an increase of information flow in the long-run. The change of information flow reduces the information asymmetries which lead to the enhancement of pricing efficiency.

The rest of the paper is organized as follows. The next section provides the literature review. Section 3 details the data and methodology. Section 4 and section 5 present empirical findings and robustness analyses. The final section offers concluding remarks.

2 Literature Review

In empirical studies there have been several proxies of attention that investigate the relationship between investor attention and stock returns. Fang and Peress (2009) measure investor attention by media coverage and support that the more stocks are present in the news, the higher the attention should be. Stock with no media coverage and thus no attention earn higher abnormal returns. Barber and Odean (2008) proxy for investor attention by extreme returns, abnormal trading volume, new and headlines. The authors find that individual investors display attention driven buying behaviour. Individual investors are net buyers of high-volume days, following extremely negative and positive 1-day returns and days with news events.

First, Da et al. (2011) introduce the search volume for stock ticker and company names as a direct measure of investor attention. They show that internet search volume measure attention more timely than other well-established attention variables, and primarly captures the individual investor attention. They report evidence that an increase in searching volume predicts higher abnormal returns in the short-run and an eventual price reversal within the year. The authors interpret this observation as support for the attention-induced price pressure hypothesis as proposed by Barber and Odean (2008). Finally, Da et al. (2011) examine IPO returns , where they found that searching volume contributes to the large first-day return and long-run performance for a sample of IPO stocks.

One of the first researches that employ online search queries is (Ginsberg et al., 2009) on detecting flu outbreaks in areas with large population of web searches. Later, in economics (Choi and Varian, 2009) provide evidence that google queries have the ability to predict home and automotive sales. However, (Choi and Varian, 2009) claim that google trends may help predict the present rather than the future. Predicting the present is a form of "contemporeaneous forecasting" or "nowcasting" as they describe in their paper.

Bank et al. (2011) analyze a sample of German stocks and find that variation in google search volume is significantly related to trading activity and thus to investor recognition. Tickers are not commonly used in Germany in order to identify stocks, hence the authors decide to use firm names as search term. They conclude that search volume primary captures the attention of retail investors, resulting in reduced information asymmetry, improved liquidity and short-term buying pressure. Joseph et al. (2011) found evidence that search volume predicts abnormal returns and trading volume. However, the sensitivity of returns to search intensity is positively related to the difficulty with which stocks can be arbitraged.

Vlastakis and Markellos (2012) study all the stocks in the Dow Jones Industrial Index and use google search volume to approximate information demand and public interest as the firm and market level. The authors found that internet search volume is positively related to volatility and returns volatility. Furthermore, they report that investors demand more information as their level of risk aversion increases. Arouri et al. (2013) use google search volume for firm names ans show that it is associated with increased liquidity and volatility in a sample of French 40 CAC stocks. . Dimpfl and Jank (2011) paper studies the dynamics of stock market volatility and retail investors attention measured by internet search querries. They find that investor's attention to the stock market rises in times of high market movements. Moreover, a rise in investor's attention is followed by higher volatility.

Fricke et al. (2014) use pre-earning announcement volume of Internet searches, as measure by google trends and show that an increase in google search volume improve the flow of information to uniformed investors and thus market efficiencyStorms et al. (2015b) develop a trading strategy exploiting limited investor attention. Trading signals for US S&P 500 are derived from google search volume data, taking a long position whenever investor attention is abnormally low in the previous week. The strategy can be augmented by combining a "buying a loser" approach. After controlling for systematic risk and transaaction costs their trading strategy seems to outperform S&P 500.

3 Data and Methodology

Our sample consists of the constituents of Standar's & Poor's 500 from January 2008 until December 2014. S&P 500 is an American stock market index based on the market capitalizations of 500 large companies having common stock listed on the NYSE or NASDAQ. It is one of the most commonly followed equity indices, and many consider it one of the best representations of the U.S. stock market. In order to make our sample free of "survivorship bias" (i.e. the addition & deletion of firms from the index) we take all the common stocks that were included during 2008-2014, thus we sum to 639 firms. The firm characteristics were obtained from Yahoo Finance[4] and the web searches from Google Trends[5]. The retrieving and aggregation of Google Trends search queries achieved through a data-mining program which were able to obtain data only in weekly frequency.

3.1 Google Trends

"Google Trends is a public web facility of Google Inc. base on Google Search, that shows how often a particular search-term is entered relative to the total search-volume across various regions of the world, and in various languages." Let $ABSV_{it}$ be the absolute search volume for keyword i and time t. The relative search volume then is computed by

$$RSV_{it} = \frac{ABSV_{it}}{\sum_t ABSV_{it}}$$

for each month or week. The month/week with the relative search volume receives a score of 100. The scores of the remained months/weeks are calculated relative to this.

$$GSV_{it} = \frac{RSV_{it}}{\max(RSV_{i1}, ..., RSV_{it})}$$

Where, $0 \leqslant GSV_{it} \leqslant 100$ with 100 be the highest search volume and 0 be no or little search volume. Also it is important to mention that daily searches are available ones, if the time interval is smaller than 93 days. However, there are not always data for all search queries, especially when you controlling for specific region, time or category.

[4]http://finance.yahoo.com/

[5]https://www.google.com/trends/

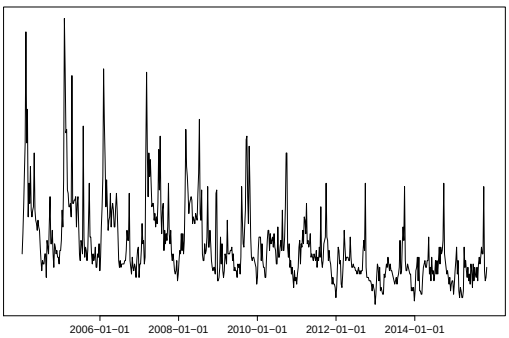
An empirical example

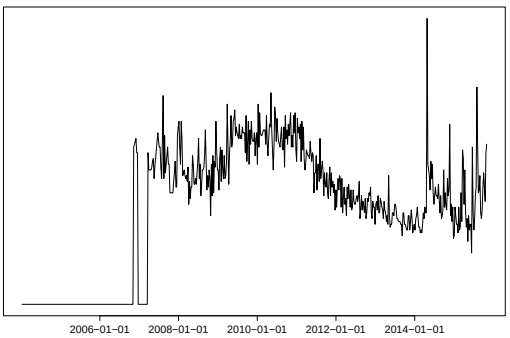
To familiarize the reader more with GSV_{it}, we create a chart with two google trends search queries , where the series have been independently drawn from Google Trends archive.

Figure 1

Google Search Volume for two search terms

Google Search Volume for the search term "afl" on the left and "agn" on the right. The geographical restriction is set to US. The time restriction has not been set, so the time spans from the beginning of Google Insights until the time of the research, i.e. 2004-2015.





The search-terms are put into google trends scrapper as "afl" & "agn", which are the corresponding stock tickers for "Aflac Inc." and "Allergan plc" in NYSE. The time restriction has not been set, so the time spans from the begining of Google Insights until the time of the research. i.e. 2004-2015. However, the geographical restriction is set only to US. In left figure we have the results of "afl" search term and in the right figure the result of "agn search term. In "afl" we observe that the series can be characterized by seasonality, probable January effect or some other calendar effect. In addition, as we see the search volume moves from 40 to 100 in only two or three weeks interval. So the series do not seem to be persistent. In "agn" we observe that there is no searching volume for two or three years. However, we can see that at 2007 there is a sudden jump in searching volume from 0 to 60. Generally, google trends series are suffering from seasonality (mostly intra-year) time trends and sudden jump. In our analysis we are trying to get rid this unwanted characteristics of google trend series by constructing abnormal search volume series, introduced by Da et al. (2011).

3.2 Google Search Volume as retail investor attention

Choosing S&P 500 has a dual purpose. First, since it constitutes an approximation for the whole market and second due to the fact that it is only comprised of large firms and not small and neglected stocks. Large firms are more widely followed by investors, media and analysts and therefore any change in fundamentals ought to be instantaneously incorporated into prices. As a search term it is more prominent by literature to use the stock ticker or the company's name, however in many papers alternative methodologies have been employed, where the search term was the ticker followed by the term "stock"(Storms et al., 2015a). For our research we choose as a web search term the stock's ticker because it specifies the goal of the search which is to inform the investor for the financials of the company. The main assumption is that an investor is using the search term only when they are interested to buy the stock and not to sell. The intuition behind that, is that when retail investors have already bought the stock they are already familiar with its financials, whereas they do not need to search again in order to sell their existing stocks. From the whole sample of 639 stocks we exclude tickers that are noisy, for example tickers that their search-term have a generic meaning (e.g. "A","DNA","EBAY").Furthermore, we exclude firms that their search queries are zero for more than fifty percent of the total series. We restrict the search only into the US because, stock ticker names may have different meaning in other languages. Finally, due to the unavailability of trend or financial data we conclude with a total number of 408 firms.

Construction of Abnormal Search Volume

We follow Da et al. (2011) methodology for the construction of Abnormal Searching Volume (ASV)a measure of direct investor attention. ASV is defined as,

$$ASV_{it} = log(GSV_{it}) - log[Med(GSV_{it-1}, ..., SVI_{it-8})]$$

where $log(GSV_{it})$ is the natural log of GSV_{it} (Google Search Volume) during the week t, and $log[Med(GSV_{it-1}, ..., SVI_{it-8})]$ is the natural log of the median value of GSV_{it} during the prior 8 weeks. This procedure allow Google Search Volume to be robust to recent jumps and remove low-frequency seasonalities and time trends, as we described before. In the rest of our research we will use $ASV^{+}(ASV > 0)$, $ASV^{-}(ASV < 0)$ and ASV^{2} in order to check the asymmetric effects and the non-linearities in trading activity and stock returns. More analytically, ASV^{+}/ASV^{-} explain that the current week's searching volume is higher/lower than at least five of the previous eight weeks.

3.3 Returns and Trading Volume

For the 408 firms we retrieve the corresponding closing prices (with the appropriate adjustments for capital changes such as split or stock dividend) and trading volume. Let $Price_{it}$ and Vol_{it} be the observed weekly adjusted closing prices and trading volume of stock i at time t. Trading volume constitutes an obvious measure of trading activity. However, in order to construct a measure of abnormal trading activity consistent with ASV we define trading activity as

$$TradAct_{it} = log(Vol_{it}) - \sum_{w=t-8}^{t-1} \frac{log(Vol_{wt})}{8}$$

where $log(Vol_{it})$ is the natural log of volume at time t and $\sum_{w=t-8}^{t-1} \frac{log(Vol_{wt})}{8}$ is the rolling mean of the trading volume with estimation window up to 8 weeks. Basically, we measure whether trading volume is different from the last two months. We use the same estimation window as in ASV, however instead of the median we take the mean because trading volume is an equilibrium outcome. Whereas we do not need to worry for exuberant variation or other abnormalities, as in ASV. We construct Trad Act because we do not know the ratio of arbitrageurs to noise trader, and the percentage of retail investors who are uniformed. So this variable try to capture the retail investors trading activity by measuring trading volume outstanding for the last two months. Furthermore we compute continuously compounded returns, by taking the first differences of the natural log of the adjusted closing prices.

$$Ret_{it} = log(Price_{it}) - log(Price_{it-1})$$

In addition we construct Carhart characteristic-based measure (Carhart, 1997) , which is a 4-factor Jensen measure, an extension of 3-factor model(Fama and French, 1993). It assumes that betas with respect to the returns of four factor-mimicking portfolios [6],

1. CRSP weighted index less T-Bbills (RMRF)
2. Small size minus big size (SMB)
3. High book-to-market minus low book-to-market(HML)
4. High prior-year return less prior-year return (PR1YR)

[6]Weekly factor data are available at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

are appropriate measure of multidimensional risk. According to this model, in the absence of stock selection or timing abilities the excess returns $Ret_{it} - Rf$, is the sum the product of betas with the factor risk premia, where Rf is the one month T-Bill.

$$AbnRet = \alpha_i = Ret_{it} - Rf - \hat{b}_i RMRF_t - \hat{s}_i SMB_t - \hat{h}_i HML_t - \hat{p}_i PR1YR_t$$

We estimate α_i as one-week abnormal return from the 4-factor model, where the 4-factor loadings are estimated over the whole sample period. With abnormal returns we are able to rule out the exposure to systematic risk factor and focus on the timing abilities of investor sentiment measures.

3.4 Indirect Proxies for Investor Attention

In order to make a comparable research we need to introduce some indirect proxies of attention. On the weeks when a stock experiences abnormally heavy volume, investors are paying more attention to it. We will use (Barber and Odean, 2008) measure for abnormal volume

$$AbnVol_{it} = \frac{Vol_{it}}{\overline{Vol_{it}}} \qquad \& \qquad \overline{Vol_{it}} = \sum_{w=t-52}^{t-1} \frac{Vol_{iw}}{52}$$

where Vol_{it} is the dollar volume for stock i traded on day t. Barber and Odean (2008), use daily volume in their paper however we converted it to weekly frequency in order to adjust it to our research. It is imperative to distinct between Trad Act and Abn Vol where in the first we try to find short horizon trading activity, while in the second we try to find abnormal one-year volume. We will see below more analytically how these two measure differ.

In the absence of news or other information extreme returns or returns volatility can become news themselves. To analyse the behaviour of return volatility we construct two measures. First,

$$SD_{ROLL} = \sqrt{\sum_{w=t-24}^{t-1} (Ret_{it} - \overline{Ret_{it}})^2} \qquad \& \qquad \overline{Ret_{it}} = \sum_{w=t-24}^{t-1} \frac{Ret_{iw}}{24}$$

where SD_{ROLL} is the standard deviation of the returns with a 24-week rolling window estimation. The choice of the window is trying to capture the volatility of the last six months, however we did not want to make a bigger window because our series would be too

persistence but also not too small because it would be too volatile. Alternatively SD_{ROLL} may be considered a six-month trailing volatility. For the second volatility we construct the conditional variance which is derived from a GARCH(1,1) model (Bollerslev, 1986).

$$\begin{aligned} Ret_{it} &= \mu_i + e_{it} \\ e_{it} &= \sqrt{h_t}\epsilon_{it} \\ h_t &= \omega + \alpha h_{it-1} + \beta e_{it-1}^2 \\ h_t &= E_{t-1}(Ret_{it} - E_{t-1}Ret_{it})^2 = E_{t-1}(e_{it}^2|\Omega_{t-1}) \end{aligned}$$

ϵ_{it} is a sequence of $N(0,1)$ i.i.d random variables, h_{it} is the conditional volatility of returns Ret_{it} which equals $E_{t-1}(e_{it}^2|\Omega_{t-1})$, where Ω_{t-1} is information at time t-1. We impose the constraint $\omega > 0$, $\alpha > 0$, $\beta > 0$ to ensure positive variance and $\alpha + \beta < 1$ for stationarity. The construction of the series SD_{GARCH} is the standardized residuals of the GARCH(1,1) under the assumption that the residuals follow a t-distribution. The GARCH(1,1) assumes that the conditional variance is autoregressive - this give rise to volatility clustering. It is important to differentiate between these measures because they are depicting totally different information. Finally, we use $AbsAbnRet_{it}$ to define the absolute value of abnormal returns.

$$AbsAbnRet_{it} = |AbsAbnRet_{it}|$$

Abs Abn Ret represent the price impact factor. Alternatively instead of the absolute value of abnormal returns we can use the the absolute value of simple returns (Abs Ret). All these measure constitute indirect proxies for investor attention. Both Abn Vol and Abs Abn Ret has been established by the current literature as proxies for investor attention, however return volatility measures are newly introduced.

4 Results

4.1 Correlations

We calculate pairwise correlations for the variables of interest, the results are reported in Table 1. First, we examine how the constructed variables specified in section 3 are correlated with their corresponding initial variables. More analytically, ASV has

Table 1
Correlations

Correlations between different measure of attention at weekly frequency. The variables are defined at section 3. The sample period is from January 2008 to December 2014.

	ASV	$log(GSV)$	Vol	$TradAct$	$AbsAbnRet$	$AbnVol$	SD_{GARCH}	SD_{Roll}
ASV	1.0000							
GSV	0.2394	1.0000						
Vol	0.0156	-0.0379	1.0000					
$TradAct$	0.0201	0.0036	0.2504	1.0000				
$AbsAbnRet$	0.0517	0.0276	0.0019	-0.0150	1.0000			
$AbnVol$	0.0021	0.0006	0.1425	0.2224	0.0026	1.0000		
SD_{GARCH}	0.0028	-0.0038	-0.0019	0.0000	0.0029	-0.0017	1.0000	
SD_{Roll}	0.0007	0.0322	0.0307	-0.0070	0.4909	-0.0096	-0.0347	1.0000

been constructed with the help of GSV, nevertheless their correlation is about 24%. Furthermore, both trading activity (Trad Act) and Abnormal Trading Volume (Abn Vol) have 25% and 14% correlation with trading volume respectively.

These results indicate that with the current transformation, these variables are now able to capture different information than their initial corresponding variables. Furthermore, we examine the correlation of two measures of volatility. Although both these measures are returns volatility their correlation is small and negative. This result validate our assumption that measures contain completely different information, thus their correlation is close to zero.

As we proceed, we compare ASV to indirect proxies of attention such as Abs Abn Ret, Abn Vol, SD_{ROLL} and SD_{GARCH}. They all have positive correlation with ASV but the coefficients are extremely small. The bigger correlation is with Abs Abn Ret with a coefficient of only 5%. The low correlation of the indirect proxies and ASV could be attributed to the fact these variables may be a factor of many variables. For example, Abs Abn Ret and SD_{ROLL} are highly positively correlated with a coefficient of almost 50%. That means that both these variables may be driven by the same underlying factors.

4.2 Portfolio Formation

We next construct portfolios ,where every week, stocks are sorted by the level of GSV into four portfolios. The first portfolio is the one with the lowest GSV and the

Table 2

Firm characteristics and indirect measures of attention portfolios

Firm characteristics and indirect measures of attention have been sorted according to weekly levels of GSV into four portfolios. The first portolio has the lowest GSV and the fourth portfolio has the highest GSV. In addition the lowest and highest portfolio has been divided into two sub-portfolios. All variables have been standardized with mean zero and standard deviation one. Standard errors have been compute only for the portfolio spreads and are reported in parentheses.*,** and *** represent significance at the 10%,5% and 1% respectively .The sample period is from January 2008 to June 20014.

Portfolio	Vol	Ret	$AbnRet$	$AbsAbnRet$	$AbnVol$	SD_{Roll}	SD_{GARCH}
1A	0.0855	−0.0030	−0.0241	−0.0424	−0.0008	0.0624	−0.0046
1B	−0.0111	0.0005	−0.0088	−0.0309	−0.0110	−0.0006	−0.0049
1(Low)	0.0372	−0.0013	−0.0164	−0.0366	−0.0059	0.0309	−0.0047
2	−0.0250	−0.0034	0.0030	−0.0342	0.0132	−0.02454	−0.0020
3	−0.0354	0.0047	0.0021	−0.0181	−0.0015	−0.0223	0.0007
4(High)	0.0231	−0.0001	0.0116	0.0887	−0.0057	0.01604	0.0059
4A	−0.0348	0.0020	0.0079	0.0080	−0.0001	−0.0191	0.0049
4B	0.0809	−0.0021	0.0144	0.1693	−0.0115	0.0511	0.0070
4-1	−0.0141	0.0012	0.0276***	0.1253***	0.0002	−0.0149**	0.0107**
spread	(0.0179)	(0.0069)	(0.008)	(0.0081)	(0.0044)	(0.0079)	(0.0061)
4B-1A	−0.0046	0.0009	0.0385***	0.2117***	−0.0106	−0.0112	0.0116*
spread	(0.0113)	(0.0098)	(0.0118)	(0.0119)	(0.0074)	(0.0114)	(0.0085)
4A-1B	−0.0237**	0.0015	.01676	.0390***	.0111***	−.0185*	0.0098
spread	(.0102)	(0.0093)	(.0107)	(0.0109)	(0.0046)	(0.0100)	(0.0087)

fourth portfolio is the one with the biggest GSV. In addition, we split the portfolios with the highest and the lowest GSV into two sub-portfolio, in order to observe the behaviour in the extreme points.

All portfolio contain equal number of stock i.e. 102, and the two sub-portfolios contain 51 each. All the variables have been standardized, with mean zero and standard deviation one. Afterwards, we compute the mean of all the variables , which in our case are the firm's characteristics Vol,Ret and $AbnRet$, and all the indirect proxies of attention $AbsAbnRet$, $AbnVol$,SD_{Roll} and SD_{GARCH}.

This method allows us to see at every portfolio how many standard deviations each variable is from the mean. At the end we compute the spread between the 4-1, 4B-1A portfolio in order to find whether the difference of high-low sorted portfolios are statistically significant. In addition, in some cases excluding the two sub-portfolio in the extremes may provide us with more intuitive results. Therefore we also compute 4A-1B spread where the two sub-portfolios 4B & 1A are not taken into account. The 4B-1A spread help us examine the variables that exhibit unorthodox behaviour in the extreme searching volume portfolios.

Table 2 show the results from this procedure. We see that in almost all portfolios in returns, returns are near to zero and the spreads are not statistically significant. However, if we examine Abnormal Returns(Abn Ret) we see substantial differences, where Abn Ret is increasing as it goes from low ASV to high ASV portfolios. Examining the spreads, we see that the 4-1 differential is almost 0.03 sd and the 4B-1A differential 0.04. Although all variables are standarized and we do not have the real values, it is considerably profound that ASV generates higher abnormal returns as a timely measure of attention. Next we examine the portfolios of trading volume, where a U-shape pattern arises, where there is high trading volume concentrated in the extreme portfolios. As we test for equality of means between portfolios we see that only 4A-1B differential is statistical significant.

Moving further to indirect proxies of attention (Abn Ret , Abn Vol, SD_{Roll},SD_{GARCH}). We see that all measure besides Abn Vol have are relatively small in portfolio 1 and relatively big in portfolio 4, with statistically significant 4-1 spreads. Abn Vol has the same behaviour with Vol, where it forms a U-shape from portfolio 1 to 4 and only the 4A-1B spread is statistical significant. This indicate that Abn Vol is sensitive to extreme observations. In the next section, we will discuss more thoroughly the relation between ASV and the alternative measures of attention.

4.3 Abnormal Searching Volume (ASV) and Indirect proxies of attention

In this section we use panel regression in order to compare the Abnormal Searching Volume (ASV) with the Indirect proxies for attention. In Table 3 we report the panel regression results, where the dependent variable is always ASV. All regression reported in this table contain weekly fixed effects, and the robust standard errors are clustered by firm.

In regression 1 we use as independent variables the most common measure of attention which are Abnormal Volume (Abn Vol) and Absolute Abnormal returns (Abs Abn Ret). We see that ASV is positively correlated with Abs Abn Ret and Abn Vol is not statistcally significant. Suggesting that absolute abnormal returns, have a major impact in Abnormal Search Volume, and the relation of Abnormal Trading Volume with ASV is close to zero.

In regression 2 and 3 we add the measures of returns volatility interchangeably. In regression 2, we add the return volatility which derive from the GARCH(1,1) and in equation 3, the trailing volatility SD_{ROLL}. The relation of ASV with SD_{GARCH} is statistical insignificant while with with SD_{ROLL} negative and significant at 1% . In equation 3, we observe that with the addition of SD_{ROLL} ,Abs Abn Ret has been increased a lot. As we mentioned before Abs Abn Ret has a correlation of almost 50% with SD_{ROLL}. We believe that these variables contain the same information, hence in the rest of our research we choose SD_{ROLL} as a more appropriate control variable.

In the final regression we include all the variables and we see that an increase in 1% in Abs Abn Ret will have an increase of 0.46% in ASV. On the other hand, an increase of SD_{ROLL} by 1% will have a decrease of 0.29% in ASV. The R^2 in all regression are no more than 2,2%. The existing proxies have very little power into explaining ASV variation. If the R^2 was substantially bigger the information that derived from ASV would be interpreted through indirect proxies of attention, hence there would be no need to use it in our research.

In sum, we the positive relationship of Abs Abn Ret with ASV indicate a major price impact on the searching volume, while the negative relationship of SD_{ROLL} with ASV indicate that the trailing volatility reduces the web searches. Interesting enough the abnormal volatility and the volatility clustering which have been proxied by SD_{GARCH} have no statistical significance in ASV.

Table 3

Abnormal Searching Volume and Indirect measures of attention

The dependent variable in each regression is the abnormal searching volume (ASV). The independent variable are abnormal volume(Abn Vol), rolling return volatility(SD_{ROLL}), return volatility from GARCH(1,1) model (SD_{GARCH}) and Absolute Abnormal Returns (Abs Abn Ret). We contain weekly fixed effects in each regression. Robust standard errors clustered by firm are reported in parentheses.*,** and *** represent significance at the 10%,5% and 1% respectively. The sample period is from January 2008 to June 20014.

	(1)	(2)	(3)	(4)
$AbsAbnRet$	0.297*	0.302*	0.458**	0.472***
	(0.160)	(0.161)	(0.180)	(0.177)
$AbnVol$	-0.00123	-0.00118	-0.00173	-0.00172
	(0.00212)	(0.00218)	(0.00200)	(0.00206)
SD_{GARCH}		2.05e-06		-8.03e-06
		(2.46e-05)		(2.47e-05)
SD_{ROLL}			−0.290***	−0.301***
			(0.0749)	(0.0713)
Constant	0.00253	0.00233	0.0128***	0.0130***
	(0.00463)	(0.00462)	(0.00268)	(0.00274)
Time Fixed Effects	YES	YES	YES	YES
Observations	148,466	148,466	148,354	148,354
Firms	366	366	366	366
R^2	0.021	0.021	0.021	0.022

4.4 Trading Activity and Investor Sentiment

In this section, we will examine the relation of trading activity with ASV and Indirect measures of attention. We use Trad Act,to define retail investors trading activity. Trading activity however can be related to alternative proxies for investor attention and various firm characteristics measures.

In our research, we will use Absolute Abnormal Returns (Abs Abn Ret) instead of Absolute Returns (Abs Ret) as a factor of price impact. Furthermore, we will use returns

Table 4

Trading activity and it's relation with measures of attention.

The dependent variable in each regressions is the Trad Act defined in section 3. The independent variables are the abnormal searching volume (ASV), absolute returns (Abs Ret) as price impact factor, and the returns volatility measures SD_{ROLL} and SD_{GARCH}. We also include ASV^+, ASV^- and ASV^2 in order to check for asymmetric effects and non-linearities. We contain weekly fixed effects in each regression. Robust standard errors clustered by firm are reported in parentheses.*,** and *** represent significance at the 10%,5% and 1% respectively . The sample period is from January 2008 to June 20014.

	(1)	(2)	(3)	(4)	(5)
ASV	0.0125***	0.0125***	0.0118***		
	(0.00400)	(0.00400)	(0.00397)		
ASV^+				0.0149***	
				(0.00487)	
ASV^-				0.00745	
				(0.00517)	
ASV^2					0.00148**
					(0.000826)
$AbsRet$	-0.0441	-0.0514	-0.0366	-0.0440	-0.0440
	(0.0367)	(0.0418)	(0.0346)	(0.0397)	(0.0397)
SD_{GARCH}		-4.99e-05		-4.86e-05	-4.87e-05
		(5.63e-05)		(5.86e-05)	(5.86e-05)
SD_{ROLL}			−0.0723**	−0.0718**	−0.0712**
			(0.0309)	(0.0309)	(0.0310)
Constant	−0.00623***	−0.00601***	-0.00306	-0.00334	-0.00296
	(0.00142)	(0.00157)	(0.00197)	(0.00213)	(0.00209)
Assymetric test				1.45	
Time Fixed Effects	YES	YES	YES	YES	YES
Firms	366	366	366	366	366
Observations	145,407	145,407	145,299	145,299	145,299
R^2	0.254	0.259	0.259	0.259	0.259

volatility measures SD_{GARCH} and SD_{Roll} as alternative proxies of investor attention. We do not include Abnormal Trading Volume (Abn Vol) as a control variable because any possible relationship would be meaningless. Finally, we include ASV^+, ASV^- and ASV^2 in order to check for asymmetric effects and non-linearities. Results are reported in table 4, where all dependent variables are Trad Act. All regression reported in this table contain weekly fixed effects, and the robust standard errors are clustered by firm.

In regression 1 we use ASV and Abs Ret as an independent variable. We observe that there positive relationship with statistical significance at 1% level. An increase of 1% in abnormal searching volume(ASV) will increase the trading activity by 0.013%. In regression 2,3 we add the returns volatility controls interchangeably. Abs Ret is statistically insignificance in all regressions thus is does not appear to be any price impact in Trad Act. Of the two volatility measures only rolling standard deviation SD_{ROLL} has statistically significant coefficient. The relationship of SD_{ROLL} with Trad Act seems to be negative , an increase of 1% in SD_{ROLL} will have a decrease of 0.072% in Trad Act. Volatility seems to have greater impact to Trad Act (in absolute value) than ASV. This results is not very surprising because volatility can be mechanically related to trading activity.

In regression 4 and 5 we check for asymmetric effects and non-linearity respectively. Although ASV^+ seems statistically significant and and bigger than ASV^- , the asymmetric test with value 1.45 does not indicate any asymmetric effects. On the other hand, in regression 5 we observe that the coefficient of ASV^2 indicate that there in non-linear relationship between searching volume and trading activity.

An increase in investor attention approximated with ASV leads to an increase of trading activity, while an increase of trailing volatility leads to a reduction of trading activity. R^2 amount to 25% in almost all regressions. So large fraction of trading activity's variation can be explained mainly through ASV. The non-linearities suggest that there might be an intermediate channel which lead to the relationship of trading activity with searching volume.

4.5 The Predictability of Stock Returns

In this section, we will examine the behaviour of Abnormal Returns regarding the investor sentiment. The methodology is similar to Da et al. (2011). We are considering both Abnormal Searching Volume (ASV) and alternative measures of attention as predictors. However we make a differentiation in abnormal searching volume (ASV) between

positive ASV^+ and negative ASV^- in order to check for asymmetric effects. The time horizon is 1, 2, 3, 4 weeks ahead and the remaining year - week 5 to 52. We use Fama and MacBeth (1973) cross sectional regression to account for time-specific errors shocks with errors-in-variables. Each week, we regress future Abnormal Returns (Abn Ret) at different horizons on ASV and other measures of attention. The regression coefficients are the averages of the cross-sectional regression estimates. Standards errors are computed using the Newey and West (1987) formula with 5 lags. The regression results are reported in Table 5.

In regression 1 and 2, the dependent variables are first week's abnormal returns. Although statistically significant the coefficients of ASV are low and negative, also ASV^+ and ASV^- seems to be zero. However, we find the coefficient of SD_{ROLL} negative and relatively high. It seems that trailing volatility have negative impact in future abnormal returns, which is completely justifiable. Increasing volatility of the last months, scare the investors and reduce abnormal returns. This confirms that investors have mean-variance preferences and they are risk averse.

In regression 3 and 4, we have the second week's abnormal returns. All measures of attention besides ASV seems to become statistically insignificant. ASV now is positive and larger than week 1. An increase in ASV will lead to an increase of abnormal returns after 2 weeks. A possible explanation, is that retail investors may be hesitant to carry their buying positions. So after searching for financial information they need extra time in order to complete their trading strategies. We find that ASV has also asymmetric effects, hence the positive pressure in returns is attribute solely to ASV^-. The nature of ASV^- indicates that in the previous 2 month we had at least 5 weeks with higher searching volume than today. This results confirms our theory that search volume need more time in order to be into incorporated into prices.

The positive coefficient of ASV in week 1 is almost 2 times larger than the negative coefficient in week 1. Although our result are not conclusive we can make the deduction that in the short-term the buying pressure is larger Barber and Odean (2008).

As we proceed, we find that R^2 is 8.8% for the first week and 1% on average for the next weeks, something completely normal for models which predict future returns. All other control variables seems to statistically insignificant for time horizon more than 1 week . The interesting fact is that Abnormal Trading Volume (Abn Vol) has no power into predicting future returns at every time horizon.

In regression 9 and 10 we use cumulative abnormal returns for the rest of the year - week 5 to 52. The relationship is negative and almost 6 times bigger than the additive

Table 5

Abnormal Returns Predictability

The dependant variable is Abnormal Returns (Abn Ret) at week 1,2,3 and 4. The independent variables are Abnormal Searching Volume (ASV). The other variables are Abnormal Volume (Abn Vol) and return volatility measures SD_{ROLL} and SD_{GARCH}. We also include ASV^{+}, ASV^{-} in order to check for asymmetric effects. Standard errors are computed using the (Newey and West, 1987) formula with 5 lags.*,** and *** represent significance at the 10%,5% and 1% respectively. The sample period is from January 2008 to June 20014.

Week	Abn Ret(t+1)		Abn Ret(t+2)		Abn Ret(t+3)		Abn Ret(t+4)		Abn Ret(t+5,t+52)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ASV	-0.000992^{*}		0.00198^{**}		-0.000754		0.000765		-0.00633^{**}	
	(0.000511)		(0.000840)		(0.000580)		(0.000556)	(0.00313)		
ASV^{+}		-0.000901		0.000418		0.000315		0.000556		-0.00474
		(0.000723)		(0.000660)		(0.000993)		(0.000717)		(0.00537)
ASV^{-}		0.000168		0.00477^{**}		-0.00258^{**}		0.000304		-0.0151^{*}
		(0.00134)		(0.00206)		(0.00115)		(0.00147)		(0.00772)
$AbnVol$	0.00102	0.00101	0.000137	0.000104	0.00403	0.00412	-0.00128	-0.00121	-0.0114	-0.0119
	(0.00377)	(0.00378)	(0.00152)	(0.00151)	(0.00530)	(0.00532)	(0.00191)	(0.00190)	(0.0151)	(0.0153)
SD_{ROLL}	-0.0739^{***}	-0.0733^{***}	0.00107	0.00248	0.00493	0.00385	0.00918	0.00961	-0.0222	-0.0211
	(0.0239)	(0.0240)	(0.00589)	(0.00589)	(0.00511)	(0.00512)	(0.00575)	(0.00587)	(0.0393)	(0.0393)
SD_{GARCH}	-1.00e-05	-1.01e-05	-6.18e-06	-6.07e-06	-1.01e-05	-9.61e-06	1.20e-05*	$1.23e-05^{**}$	$4.28e-05$	$4.35e-05$
	(1.12e-05)	(1.12e-05)	(9.30e-06)	(9.24e-06)	(7.33e-06)	(7.36e-06)	(6.09e-06)	(6.16e-06)	(4.86e-05)	(4.90e-05)
$Constant$	0.00182	0.00185	-0.000876	-0.000710	-0.00496	-0.00512	0.000490	0.000410	-0.0255	-0.0251
	(0.00441)	(0.00442)	(0.00150)	(0.00148)	(0.00518)	(0.00519)	(0.00200)	(0.00197)	(0.0215)	(0.0217)
Assymetric Effects		0.24		3.97^{**}		3.60^{*}		0.01		1.22
Observations	148,349	148,349	148,342	148,342	148,341	148,341	148,340	148,340	146,516	146,516
Firms	365	365	365	365	365	365	365	365	365	365
R^2	0.079	0.081	0.009	0.012	0.011	0.014	0.008	0.011	0.010	0.012

positive pressure of the first two weeks. Suggesting that the initial pressure had been completely reversed after one year.

Although in the short-run the results are not very persuasive we can say that we find a small positive pressure Barber and Odean (2008). However, in the long-run our results perfectly align with "Information Theory" Merton (1987).

An increase in searching volume, increases the information flow. As investor are searching for specific stock for a long time they became acquainted with the firm's fundamentals. Then are able to make more rational decisions and reduce the abnormal returns. This result to the reduction of information asymmetries and to pricing efficiency. In the long-run we observe no asymmetric effects because investors have enough time to complete their trading strategies. Furthermore, we test in every regression whether non-linearities appear. The results are non-statistical so we choose to omit them.

In sum, an increase of retail investor attention lead to short-term buying pressure and a complete reversal within a year. The individual investors are late into completing their position hence the results of buying pressure is more obvious in the second week. Due to the increase of information flow information asymmetries are reduced. Thus investors in the long run become more rational and lead the price near to fundamentals.

5 Robustness

In this section we will examine whether our results are robust to different specifications. We are trying different estimation windows, alternative specification methodologies and we divide our sample into two sub-periods.

We think that both Abnormal Returns (Abn Ret) and Abnormal Trading Volume (Abn Vol) have been firmly established as indirect proxies of attention by the current literature and there is no need for further investigation. Furthermore, they do not have any serious impact in our research. However, the other measure are prone to critic.

First, we would like to check the estimation window of Abnormal Searching Volume (ASV). Our results remain robust only for bigger window(11 weeks) and not for smaller window(5 weeks). Small window may not be able to diminish the effects of the sudden jumps and dives of searching volume, thus they are very noisy. Furthermore, ASV with 11 weeks estimation window seems to adjust better than 8 weeks to our model, something that confirm our rationale. They same conclusion apply to trading activity (Trad Act).

As we proceed we check both our return volatility measures specifications. For the first measure, SD_{ROLL} we try again different estimation windows. Our results hold

for both smaller (12 weeks) and bigger (36 weeks) windows. For our second measure, SD_{GARCH} we use an GJR-GARCH(1,1) methodology to estimate the conditional variance (Glosten et al., 1993). Both our return volatility measures are robust to different specifications

Finally, we divide our sample into two sub-periods. Our whole sample comprised of 7 years, so we make the decision to split it into the first four years 2008-2011 and the last 3 year 2012-1014. This division is also due to the fact that the first period contain the aftermath of 2007 sub-prime crisis, while the second period is moderate and uneventful. Our results hold for both sub-periods.

6 Conclusion

Recent research in financial markets concurs that web search constitute a direct measure of investor's attention. In this research we use google search volume as direct measure of attention while controlling with alternative indirect proxies such absolute abnormal returns, trading volume and two return volatility measures. In a sample of S&P 500 from 2008-2014, we first show that ASV is correlated with alternative measures of attention, however they are able to explain only a fraction of ASV. Furthermore, we find that ASV related with trading activity, thus an increase of retail investors attention proxied by ASV increases the trading activity.

We analyse the contribution of ASV in abnormal returns predictability. We find that ASV creates a positive pressure in abnormal return in the second week, and there is a complete reversal within the first year, as it is described in Da et al. (2011). Individual investors are late into completing their transaction, hence we find the pressure being bigger on the second week and not the first. This assumption is also validated through the fact that there is an asymmetric effect in ASV,i.e. only the ASV^{-} is statistically important. Finally, the negative pressure of ASV into long-run abnormal returns indicate that ASV reduces the asymmetries in information and enhances market efficiency.

This research illustrates the usefulness of search data. Overall, we conclude that google search volume primary capture the attention of retail investors, who are comprised in average of uniformed individuals . Resulting in increased trading activity, positive short-run buying pressure and long-run reduction of information asymmetries.

Bibliography

Arouri, M., Aouadi, A., Foulquier, P., and Teulon, F. (2013). Can information demand help to predict stock market liquidity? google it! Technical report, Department of Research, Ipag Business School.

Bank, M., Larch, M., and Peter, G. (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. and 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.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

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

Choi, H. and Varian, H. (2009). Predicting initial claims for unemployment benefits.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

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

Dimpfl, T. and Jank, S. (2011). Can internet search queries help to predict stock market volatility? *European Financial Management*.

Fama, E. F. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics*, 33(1):3–56.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, pages 607–636.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, 64(5):2023–2052.

Fricke, E., Fung, S., and Goktan, M. S. (2014). Google search, information uncertainty, and post-earnings announcement drift. *Journal of Accounting and Finance*, 14(2):11.

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

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, 48(5):1779–1801.

Joseph, K., Wintoki, M. B., and Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27(4):1116–1127.

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

Newey, W. K. and West, K. D. (1987). Hypothesis testing with efficient method of moments estimation. *International Economic Review*, pages 777–787.

Shleifer, A. and Summers, L. H. (1990). The noise trader approach to finance. *The Journal of Economic Perspectives*, pages 19–33.

Storms, K., Kapraun, J., and Rudolf, M. (2015a). Can retail investor attention enhance market efficiency? insights from search engine data. *Insights from Search Engine Data (July 28, 2015)*.

Storms, K., Kapraun, J., and Rudolf, M. (2015b). In search of alpha-trading on limited investor attention.

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

7 Appendix

An Insight into Fama-Macbeth Procedure (Cochrane, 2009)

Instead of estimating a single cross-sectional regression with the sample averages, we now run a cross sectional regression at each time period,i.e.

$$R_t^i = \beta_i' x_t + e_{it}, \qquad i = 1, ..., n \quad for \ \ each \ \ t$$

We write the case of a single regressor for simplicity, but is easy to extend the model to multiple regressors. Then we estimate x and e_i as the averages of the cross-sectional regression estimates,

$$\hat{x} = \frac{1}{T}\sum_{t=1}^{T} \hat{x}_t \qquad \hat{e_i} = \frac{1}{T}\sum_{t=1}^{T} \hat{e_t}$$

They suggest that we use the standard deviation of the cross-sectional regression estimates to generate te sampling errors for these estimates,

$$\sigma^2(\hat{x}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{x}_t - \hat{x})^2 \qquad \sigma^2(\hat{e_i}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{e_{it}} - \hat{e_i})^2$$

Is is $1/T^2$ because we are finding standard errors of sample means σ^2/T. Fama-Macbeth calculate the standard errors, corrected for cross-sectional correlation. The formula assumes that the time series is not autocorrelated, but one could easily extend the idea to estimate $\hat{x}_t$ that are correlated over time by using a long-run variance matrix,i.e.,

$$\sigma^2(\hat{x}) = \frac{1}{T}\sum_{j=-\infty}^{\infty} cov(\hat{x}_t, \hat{x_{t-j}})^2$$

One should use some sort of weighting matrix or a parametric description of the autocorrelations of $\hat{x}$. It is natural to use the sampling theory to test whether all the pricing errors are jointly zero. The covariance matrix of the sample pricing errors is.

$$cov(\hat{e}) = \frac{1}{T^2}\sum_{t=1}^{T}(\hat{e_t} - \hat{e})(e_t - \hat{e})'$$

and then use the test

$$\hat{e}' cov(\hat{e})^{-1}\hat{e} \sim x_{N-1}^2$$

The Fama-Macbeth procedure was also important because it allowed changing betas, which a single cross-sectional or a time-series regression cannot test easily.

Table 7

Variable Description

(All variables are in weekly frequency)

Symbol	Description	Calculation
GSV	Google Search Volume	Aggregate search frequency
ASV	Abnormal Search Volume	Log of GSV for stock ticker minus to the log of the median of last 8 weeks
Vol	Trading Volume	Current trading volume minus to the log of the mean of last 8 weeks
Trad Act	Trading Activity	Current trading volume
Ret	Raw Returns	Continuously compounded returns
Abn Ret	Abnormal Returns	Absolute value of continuously compounded returns
Abs Ret	Absolute Returns	Continuously compounded returns filtered through 4-factor Carhart model
Abs Abn Ret	Absolute Abnormal Returns	Absolute continuously compounded returns filtered through 4-factor Carhart model
Abn Vol	Abnormal Trading Volume	Current trading volume divided by the rolling mean of the last year
SD_{Roll}	Standard Deviation ROLL	Rolling standard deviation with estimation window 24 weeks
SD_{GARCH}	Standard Deviation GARCH	Volatility measure derived from a GARCH(1,1) model

Table 8

Summary Statistics

(All variables are in weekly frequency)

Statistics	log(GSV)	ASV	Ret	Abn Ret	log(Vol)	Abs Ret	Abs Abn Ret	Abn Vol	SD_{ROLL}	SD_{GARCH}
Mean	3.4698	0.0096	0.0012	-0.0008	14.8423	0.0377	0.0280	0.9982	0.0493	-1.1507
Std Deviation	0.7376	0.3520	0.0678	0.0562	1.4006	0.0564	0.0487	0.2052	0.0411	32.5894
Skewness	-1.6463	1.58512	-16.4197	-10.3904	-2.8345	37.9844	28.9821	164.6417	8.7555	-58.69954
Kurtosis	7.5788	81.4043	2348.371	1093.681	32.3681	4796.387	1866.872	35478.86	179.2883	12346.2
Min	0	-0.4659	-0.1765	-0.1272	11.8336	0.0004	0.0003	0.9069	0.01504	-82.05293
25th percentile	3.0910	-0.0687	-0.0220	-0.0179	14.1202	0.0109	0.0079	0.9769	0.0277	-13.06622
Median	3.6109	0	0.0028	-0.0003	14.8679	0.02446	0.0176	0.9969	0.0394	-0.3184886
75th percentile	3.9890	0.0668	0.0265	0.0173	15.6014	0.0474	0.0343	1.0185	0.0580	12.16843
Max	4.4998	0.7721	0.1570	0.1224	17.9010	0.2172	0.1632	1.0956	0.1910	67.87394