Market information demand and volatility

Econometrics empirical project

Jaime Oliver

Universidad Carlos III de Madrid

Master in Economics
Universidad Carlos III
Madrid, Spain

Contents

1 Introduction

According to the 2017 SIFMA (Securities Industry and Financial Markets Association) Fact Book, all domestic stocks listed on U.S. exchanges were worth as much as $27.4 trillion. It is natural that there are interests to model correctly the returns and volatility of this assets.

According to Herbert Simon, actors begin their decision making processes by attempting to gather information [1]. And in today's world, information gathering often consists of searching online sources. In fact, around 3,000,000 Google searches are conducted worldwide each minute of everyday. So the question is, can we measure this demand for information? In that case, how can we include it in the model?

Google has begun to provide access to aggregated information on the volume of queries for different search terms and how these volumes change over time, via the publicly available service *Google Trends*. This information comes in the form of the Search Volume Index (SVI from now on), which is defined the absolute search volume for a term, relative to the number of searches received by Google in the moment of the query.

Past investigations have shown that the number of clicks on search results stemming from a given country correlate with the amount of investment in that country [2]. Further studies exploiting the temporal dimension of Google Trends data have demonstrated that changes in query volumes for selected search terms mirror changes in current numbers of influenza cases [3] and current volumes of stock market transactions [4]. Another study has shown that Internet users from countries with a higher per capita GDP are more likely to search for information about years in the future than years in the past [5]. Finally, a study has shown that there exists a relationship between Google querys and the stock market prices [6].

1.1 Aim of the study

In the first place, section 2 is dedicated towards the discussion of the data sources and description of the variables used.

We will base the project on previous work done by Vlastakis & Markellos [7]. They state that demand for information at the market level is significantly positively related to historical measures of asset return volatilities, even after controlling for market return. In section 3 we replicate this result.

In section 4 we proceed to model jointly returns and volatilities using a GARCH model, also including market information demand as an exogenous variable.

This study will be conducted for every firm in the Dow Jones Index. The data gathering process is explained in section 2.

2 Data and sample description

The firms conforming the Dow Jones Industrial Average index are shown in table 1. *DowDuPont Inc* was left out of our study, due to it's recent creation in December 2015 by the merger of *Dow Chemical Company* and it's competitor *DuPont Corporation*. This produces a final amount of 29 firms.

For the market data, we consider the S&P500 index. This is generally considered the large-cap stock index in the United States, given that the 500 companies that make up the index together make up 80% of all stock market value in the USA.

Symbol	Firm	Search Term	Symbol	Firm	Search Term
UNH	UnitedHealth Group	united health	BA	Boeing	boeing
PFE	Pfizer Inc.	pfizer	MCD	McDonald's Corporation	mcdonald's
MRK	Merck & Co. Inc.	merck	CSCO	Cisco Systems Inc.	cisco
GE	General Electric Co.	ge	SPY	S&P 500 Index	s&p 500
WMT	Wal-Mart Stores Inc.	walmart	TRV	Travelers Companies Inc	travelers companies
UTX	United Technologies Corp.	united technologies	MMM	3M Company	3m
INTC	Intel Corp.	intel	AAPL	Apple	apple
PG	Procter & Gamble Co.	procter gamble	JPM	JPMorgan Chase and Co.	jp morgan chase
VZ	Verizon Communications	verizon	CVX	Chevron Corp.	chevron
KO	Coca Cola Co.	coca cola	CAT	Caterpillar Inc.	caterpillar
AXP	American Express Co.	american express	JNJ	Johnson & Johnson	johnson and johnson
DIS	Walt Disney Co.	disney	V	Visa	visa
NKE	Nike Inc.	nike	MSFT	Microsoft Corp.	microsoft
GS	Goldman Sachs Group	goldman sachs	HD	Home Depot Inc.	home depot
XOM	Exxon Mobil Corp.	exxon mobil	IBM	International Business Machines Corp.	ibm

Table 1: List of series under study and their respective symbols and search terms

2.1 Information demand

As a proxy for information demand, we use monthly data for the period January 2011 to April 2018, which is publicly available through Google Trends. The Google Trends service provides SVI data for any keyword(s) the user inputs. For a comprehensive description and application of Google Trends data to forecasting a variety of economic variables, such as automobile and home sales, see Choi and Varian (2009) [8]. At the time of the analysis, the service provides data from 2004 onwards, at monthly frequency, or at weekly frequency for periods less than a quarter. The user has the ability to compare different queries and filter results according to category or geographical location.

The two obvious alternatives for the keyword used in the queries is the company name and the stock ticker. The company name is used in our analysis for two reasons. First, this allows us to derive a broad measure of information demand by investors which is related to the firm in general rather than only to the stock. Second, we avoid problems associated with the fact that several ticker names have generic meanings (for example, the tickers for Alcoa and Caterpillar are "AA" and "CAT", respectively).

Da et al. (2011) [9] argue that it is preferable to use the stock ticker instead of the company name on the basis of three reasons. First, people may search for the name of a company for reasons other than investment. Second, there are many different ways to spell the name of a company. Third, Google Trends does not allow alpharithmetic input, which would inhibit

the use of names for companies such as 3M (this is no longer the case).

We agree with Da et al. [9] that the SVI for the company name does include some irrelevant component, by people searching, for example, for products or support online. However, we assume that this component is either random noise or purely deterministic (ie., seasonality, time trend) and therefore with appropriate pre-processing of the data it should not influence the variable in a systematic manner.

In order to account for the possibility that the name of a company may be commonly spelled in a variety of ways, we adopt the following keyword selection procedure. We start by inserting the full company name and all the variations known to us to Google Insights for Search and check which keyword has the largest search volume. We also insert the keywords to an online service called Wordtracker that specializes in search engine keyword optimization. Wordtracker provides a tool that lists related keywords for every keyword inserted. We use this process to identify any name variations that we may have not included in our analysis. In order to avoid problems with noise arising from the fact that some of the search terms may have generic meanings, we examine the context in which these keywords are used in searches[1]. When a keyword is found to be massively implemented in search queries that are not relevant to the specific company, it is excluded from the analysis and the next most popular keyword is used. Table 1 presents a list of the companies in our sample along with the corresponding stock tickers and the search queries we finally adopted. In addition to a proxy for firm-specific information demand, our analysis employs a measure of demand for market-wide information on the basis of SVI for the keyword "S&P 500".

Once the data extracting process is finished, we apply a logarithmic transformation to the data. Some descriptive statistics of the SVI indexes obtained can be found in table 2

2.2 Stock market data

The data used to describe markets are the daily closing prices of all the indexes under study. Symbols used to retrieve the data from Yahoo finance can be found in table 1. The data obtained has a daily frequency excluding weekends, when the the New York Stock Exchange and NASDAQ are closed. This data is available for a much wider range of dates than the information demand variable, so we restrict the analysis to the range of the latter. In this way, we also omit the 2008 financial crisis from the study.

We will need to create some variables for the models that will be estimated in sections 3 and 4. The most immediate one in matter of financial models are the logarithmic returns, which are defined as follows:

$$r_t = \log(P_t) - \log(P_{t-1}) \tag{1}$$

where P_t are prices at time t.

Realized volatility is one of the most popular measures of historical volatility in the literature at present due to its accuracy and model-free nature [10][11] [12], therefore we first

[1]To this end, we examine both the "search terms" tool in Google Insights for Search and the "find keywords that include" tool in Wordtracker. These tools identify the most popular search queries that contain the inserted keywords

	max	min	mean	median	std	skew	kurtosis	J.B. pvalue
3M	4.62	3.91	4.19	4.15	0.14	1.1	0.58	0.0
American Express	4.62	3.76	4.22	4.26	0.19	-1.01	0.3	0.0
Apple	4.62	3.3	3.92	3.94	0.26	-0.02	0.02	0.99
Boeing	4.62	3.43	3.79	3.69	0.26	0.72	-0.54	0.0
Caterpillar	4.62	4.08	4.31	4.3	0.08	0.31	0.42	0.13
Chevron	4.62	3.66	4.05	3.98	0.2	0.84	-0.03	0.0
Cisco	4.62	3.3	3.89	3.87	0.33	0.32	-0.74	0.03
Coca-Cola	4.62	2.48	3.41	3.22	0.65	0.36	-1.36	0.0
Disney	4.62	3.91	4.3	4.3	0.16	-0.45	-0.19	0.05
Dow Jones	4.62	1.95	2.67	2.56	0.46	1.4	2.75	0.0
Exxon Mobil	4.62	3.43	3.99	4.04	0.25	-0.54	-0.46	0.01
General Electric	4.62	3.04	3.66	3.53	0.46	0.53	-1.05	0.0
Goldman Sachs	4.62	2.94	3.41	3.4	0.28	0.6	0.72	0.0
Home Depot	4.58	3.76	4.2	4.17	0.19	0.22	-0.93	0.02
IBM	4.62	2.3	3.26	3.11	0.69	0.39	-1.19	0.0
Intel	4.62	3.4	4.12	4.17	0.35	-0.33	-1.22	0.0
JPMorgan Chase	4.62	3.71	4.18	4.18	0.17	-0.1	0.15	0.8
Johnson & Johnson	4.62	2.83	3.63	3.64	0.52	0.3	-1.16	0.0
McDonald's	4.62	3.33	4.04	4.2	0.37	-0.33	-1.36	0.0
Merck	4.62	2.83	3.53	3.42	0.48	0.48	-1.07	0.0
Microsoft	4.62	3.09	3.84	3.85	0.41	-0.04	-1.07	0.02
Nike	4.62	3.74	4.15	4.04	0.24	0.48	-1.25	0.0
Pfizer	4.62	3.09	3.76	3.74	0.45	0.3	-1.12	0.0
Procter & Gamble	4.62	2.56	3.52	3.61	0.57	0.01	-1.02	0.02
Travelers Companies Inc	4.62	0.0	1.99	2.44	1.36	-0.46	-1.19	0.0
United Technologies	4.62	2.71	3.39	3.26	0.4	0.89	0.05	0.0
UnitedHealth	4.62	3.33	3.98	4.03	0.27	-0.48	-0.51	0.02
Verizon	4.62	3.43	4.07	4.12	0.24	-0.7	-0.14	0.0
Visa	4.62	4.32	4.47	4.47	0.06	0.15	-0.58	0.21
Wal-Mart	4.62	1.79	3.19	3.07	0.83	0.0	-1.51	0.0
s&p500	4.62	2.64	3.45	3.5	0.36	0.07	-0.59	0.27

Table 2: Descriptive statistics of the information demand obtained after a logarithmic transformation. In addition to some moments, the p-values for the Jarque-Berra normality test are included.

employ this measure of volatility in our analysis. In the calculation of realized volatilities for our series we follow the methodology of Andersen et al. [10].

In particular, we calculate daily logarithmic returns from inter-day stock price data. The returns are filtered by taking the residuals from a MA(1) model in order to remove the typically found negative serial correlation . Consistent with the findings in Andersen et al. [10], all estimated moving average coefficients are negative or not significant in our data (see appendix table 8). We then calculate weekly realized volatilities, by summing squared returns for each month. The realized volatility for month t is given by:

$$RV_t = \log\left[\sum_{i=1}^{N} r_{t,i}^2\right] \tag{2}$$

where $i = 1, \cdots, N$ stands for all the price observations during the month [2].

3 Realised Volatility and market information demand

Once we have the data we proceed to look for a relationship between the historical volatility and information demand. In order to do this, first we perform a stationarity anaysis and then we replicate the model from Vlastakis & Markellos [7].

3.1 Stationarity

The stationarity of the logarithmically transformed information demand variable (simply referred to as information demand hereafter) is assessed using two unit root tests: the Augmented Dickey–Fuller (ADF) test [13] and the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test [14]. Whereas for the ADF the null hypothesis is the existence of a unit root in the data, the null hypothesis for the KPSS test is stationarity, against the alternate hypothesis of a unit root. For this reason, the KPSS test complements the ADF test, especially in case the latter do not provide conclusive results. The results (shown in table 3) suggest that our information demand variable is an I(1) process, so we differentiate it in order to make it stationary. A visual example of this transformation can be observed in figures 1 and 2.

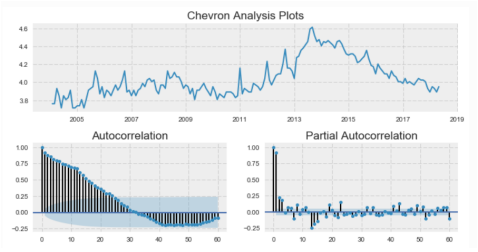

Figure 1: Chevron's $I(1)$ **SVI**

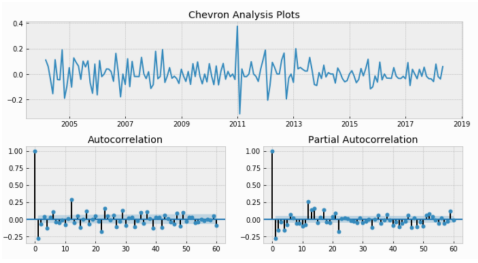

Figure 2: Chevron's $I(0)$ **SVI**.

[2] $N \approx 22$ because we use 5 out of 7 daily data and compute monthly realised volatility

	ADF c+t	ADF c	ADF	KPSS		ADF c+t	ADF c	ADF	KPSS
3M	0.79	0.04	0.45	0.01	JPMorgan Chase	0.00	0.00	0.75	0.06
American Express	0.59	0.94	0.16	0.02	Johnson & Johnson	0.95	0.26	0.00	0.01
Apple	0.82	0.33	0.98	0.01	McDonald's	0.92	0.81	1.00	0.02
Boeing	0.12	0.40	0.03	0.01	Merck	1.00	0.01	0.01	0.01
Caterpillar	0.15	0.32	0.63	0.05	Microsoft	0.05	0.88	0.00	0.10
Chevron	0.61	0.17	0.75	0.09	Nike	0.65	0.90	0.94	0.01
Cisco	0.09	0.48	0.01	0.01	Pfizer	0.92	0.07	0.00	0.01
Coca-Cola	0.98	0.54	0.00	0.01	S&P 500	0.00	0.82	0.89	0.10
Disney	0.19	0.97	0.28	0.02	Procter & Gamble	0.30	0.56	0.01	0.10
Exxon Mobil	0.05	0.98	0.15	0.01	Travelers Companies Inc	0.00	0.00	0.54	0.02
General Electric	1.00	0.10	0.16	0.01	United Technologies	0.81	0.17	0.03	0.01
Goldman Sachs	0.00	0.72	0.30	0.07	UnitedHealth	0.83	0.76	0.81	0.02
Home Depot	0.64	0.93	0.98	0.04	Verizon	0.91	0.99	0.31	0.01
IBM	0.99	0.02	0.00	0.01	Visa	0.30	0.84	0.85	0.02
Intel	0.63	1.00	0.00	0.01	Wal-Mart	0.09	0.81	0.11	0.10

Table 3: P-values of the Augmented Dickey–Fuller test with constant and trend (ADF c+t), with constant (ADF c) and wthout constant and trend (ADF) and the Kwiatkowski, Phillips, Schmidt, and Shin test (KPSS) for the information demand. We conclude that the variable is I(1).

The analysis for stationary for financial data is much more clear. Given that prices are an I(1) process, the series in difference will be I(0), and so we conclude that returns are an I(0) process. What is not so clear is that realized volatility is an I(0) too, even if it's a function of another I(0) (returns). We perform the same stationary tests as for the information demand, and we conclude that realized volatility is stationary (table 4).

	ADF c+t	ADF c	ADF	KPSS		ADF c+t	ADF c	ADF	KPSS
AAPL	0.09	0.72	0.84	0.10	MCD	0.01	0.06	0.69	0.09
AXP	0.00	0.00	0.32	0.08	MMM	0.11	0.09	0.75	0.05
BA	0.06	0.56	0.82	0.03	MRK	0.00	0.01	0.34	0.02
CAT	0.30	0.09	0.76	0.10	MSFT	0.39	0.48	0.68	0.04
CSCO	0.41	0.10	0.23	0.03	NKE	0.31	0.68	0.70	0.10
CVX	0.00	0.00	0.54	0.10	PFE	0.00	0.00	0.03	0.10
DIS	0.00	0.00	0.70	0.10	PG	0.00	0.01	0.45	0.10
GE	0.00	0.00	0.02	0.02	SPY	0.00	0.00	0.60	0.10
GS	0.40	0.11	0.54	0.01	TRV	0.02	0.03	0.48	0.05
HD	0.00	0.85	0.84	0.07	UNH	0.46	0.76	0.86	0.05
IBM	0.01	0.00	0.65	0.10	UTX	0.00	0.00	0.63	0.10
INTC	0.37	0.21	0.45	0.05	V	0.05	0.86	0.82	0.08
JNJ	0.00	0.07	0.58	0.08	VZ	0.00	0.00	0.33	0.10
JPM	0.11	0.02	0.48	0.02	WMT	0.00	0.05	0.54	0.10
KO	0.01	0.00	0.07	0.10	XOM	0.00	0.00	0.36	0.10

Table 4: P-values of the Augmented Dickey–Fuller test with constant and trend (ADF c+t), with constant (ADF c) and wthout constant and trend (ADF) and the Kwiatkowski, Phillips, Schmidt, and Shin test (KPSS) for the realized volatility. We conclude that the variable is I(0).

3.2 Model proposed by Vlastakis & Markellos

Vlastakis & Markellos [7] use the following regression in order to study the relationship between realized volatility and information demand, while controlling for the effect of market returns:

$$RV_t = \omega + \gamma\pi_t + \delta\phi_t + \lambda\nu_t + \theta\nu_{t-1} + \psi RV_{t-1} + \epsilon_t \tag{3}$$

where ω is the constant, π_t is idiosyncratic information demand at time t, ϕ_t is market-related information demand at time t, ν_t is the market return at time t and ϵ_t are the errors. The results from the estimation are shown in table 5.

From table 5 we can infer that at least one information demand variable enters the regression for all but seven of the stocks in the sample. Firm-specific information demand is a significant regresor in 9 cases, whereas market-related information demand is significant in 17 cases. The magnitude of estimated coefficients suggests that the effect of the two information demand variables is comparable.

As expected, realized volatility is highly persistent with almost all coefficients on the first lag being significantly positive. The adjusted R^2 coefficients, ranging between 9% for Coca Cola and 58 % for Home Depot suggest a reasonably good fit of the models. Overall, the results indicate that information demand at the market level has a significant positive association with realized volatility. Idiosyncratic information demand is also significant, but it has a secondary effect.

The results confirm what Vlastakis & Markellos [7] conclude in their research: the information demand at the market level is significantly positively related to realized volatility, even after controlling formarket return.

4 Returns and volatility jointly modeled with market information demand

Earlier contributions in the literature [15] [16] suggest the use of GARCH [17] models for approximating historical volatility. Kalev et al. [15] argue that modelling the relationship between information and volatility through a conditional heteroscedasticity process is a great improvement over previously used, unconditional volatility measures, such as absolute daily market returns. Although this approach can be expected to be less accurate than the realized volatility used previously, since the latter utilizes much more data, it has the advantage of being able to model the conditional mean and variance at the same time in a straightforward manner. Moreover, a wealth of empirical evidence exists on the application of GARCH models in finance.

But before estimating a model of this type, we must test for ARCH effects in our sample.

Symbol	ω	γ	δ		θ	ϕ	R^2	AIC
AAPL	2.86***	0.32	0.479	-6.81**	-2.58	0.24**	0.17	164.76
AXP	2.17***	-0.627	1.05**	-7.16***	-5.7**	0.206*	0.28	130.38
BA	1.7***	0.442	1.52***	-6.1**	-5.04*	0.583***	0.5	147.21
CAT	2.6***	-1.49	0.601	-3.04	-3.07	0.25**	0.15	131.59
CSCO	0.549***	1.37	0.677	-5.04	-3.12	0.212	0.13	159.21
CVX	1.61***	-0.742	0.701**	-5.35**	-0.696	0.533***	0.43	96.02
DIS	0.936***	1.16	1.61***	-5.67**	1.98	0.663***	0.53	138.27
GE	0.431***	3.32**	0.135	-4.46	-9.02***	0.11	0.26	148.45
GS	2.51***	0.504	1.4***	-5.97**	0.0452	0.469***	0.34	121.52
HD	1.15***	1.15	0.958**	-11.8***	-0.902	0.661***	0.58	133.97
IBM	4.31***	1.29	0.639	-0.948	-4.22	-0.0295	0.11	159.52
INTC	0.883***	2.84*	0.0603	-7.31***	-6.36**	0.312***	0.29	141.97
JNJ	0.978***	-1.42*	1.82***	-8.69***	-0.567	0.622***	0.53	141.5
JPM	0.926***	1.02*	1.99***	-7.45***	1.18	0.633***	0.46	112.16
KO	0.676***	0.687	0.852*	-1.58	-2.33	0.165	0.09	149.33
MCD	1.94***	2.6**	0.483	-5.95**	-5.57*	0.342***	0.25	152.93
MMM	1.97***	-2.86*	1.29**	-8.36***	-3.44	0.45***	0.4	154.22
MRK	1.25***	0.0661	1.26**	-6.96**	-1.11	0.345***	0.22	166.78
MSFT	1.16***	-2.6*	0.871	-6.14*	-5.23	0.412***	0.3	167.4
NKE	1.02***	1.55	-0.0678	-14.5***	-2.01	0.496***	0.39	178.61
PFE	0.224**	1.03*	1.12**	-2.05	0.138	0.56***	0.38	140.0
PG	1.83***	0.0928	0.931**	-3.46	-3.92	0.169	0.16	140.94
TRV	1.73***	0.0741	0.677	-9.34***	-1.94	0.393***	0.31	145.8
UNH	1.17***	0.101	1.27***	-3.42	-2.1	0.663***	0.51	150.86
UTX	2.11***	1.5***	0.488	-2.6	-4.01**	0.322***	0.41	86.41
V	1.09***	0.0982	1.48***	-5.46*	-3.1	0.546***	0.41	160.07
VZ	1.16***	0.0255	1.02**	-2.56	-2.77	0.218*	0.18	124.88
WMT	1.99***	0.74	0.0795	-4.02	-4.6	0.153	0.1	165.07
XOM	1.16***	0.71	0.805*	-4.47*	0.338	0.567***	0.39	129.34

Table 5: This table presents the results of OLS regressions between individual stock realized volatility, and information supply and demand variables. ω is the constant, γ is the coefficient for idiosyncratic information demand, δ is the coefficient for market-related information demand. λ is the coefficient for the market return, whereas θ is the coefficient of its first lag and ψ is the coefficient on the lagged realized volatility. The significance level of the estimated parameters is denoted as ***, ** and * if the null is rejected at 1%, 5% or 10% respectively.

4.1 Testing for ARCH effects

No ARMA effects were found for the returns, and only in 9 of the 29 stocks the intercept was found to be significant. The results are in concordance with the random walk hypothesis [18], but this is something out of the scope of this work.

Prior to estimating an ARCH model, we must perform a test for conditional heteroskedasticity or autocorrelation in the square residuals of the returns. Several tests exist for this matter, and in our case we implement Engle's Lagrange Multiplier test (ELM) [17], Ljung-Box test (LB) [19] and Durbin-Watson test (DW) [20].

The results of the test can be observed in table 6. Using as a reference the Ljung-Box test, we conclude that at a 10% significance level there are ARCH effects in 20 out of the 29 firms off the study. The other 9 firms will not be included in the ARCH estimation.

Symbol	**DW**	**LB**	**ELM**	**Symbol**	**DW**	**LB**	**ELM**
SPY	1.26	0.0	0.12	JNJ	1.56	0.03	1.0
AXP	1.82	0.0	0.16	CSCO	1.55	0.03	0.07
BA	1.56	0.0	0.68	PG	1.16	0.08	0.96
CAT	1.47	0.0	0.0	VZ	1.41	0.08	0.52
UTX	1.1	0.0	0.84	HD	1.54	0.09	0.34
UNH	1.73	0.0	0.0	NKE	1.4	0.09	1.0
PFE	0.91	0.0	0.0	CVX	1.4	0.12	0.32
GE	0.89	0.0	0.0	MSFT	1.35	0.16	0.64
GS	0.86	0.0	0.13	IBM	1.43	0.26	0.62
MRK	1.24	0.0	0.03	INTC	1.24	0.28	0.8
MMM	1.32	0.0	0.0	TRV	1.65	0.29	0.98
JPM	1.21	0.0	0.51	AAPL	1.63	0.39	0.47
KO	1.67	0.01	0.0	XOM	1.38	0.51	0.96
DIS	1.24	0.01	0.9	MCD	1.44	0.6	1.0
V	1.31	0.02	1.0	WMT	1.6	0.64	0.93

Table 6: Results for various conditional heteroskedasticity tests: DW is the Durbin-Watson statistic, LB is the pvalue for the Ljung-Box test for lag 24 (2 yeaers of sample), and ELM is the pvalue for the Engle's Lagrange Multiplier test. We take the Ljung-Box as the most robust, so we reject the null that the returns are independently distributed in 20 cases at a 10% significance rate.

4.2 Final model estimation

Financial theory suggests that an asset with a higher expected risk would pay a higher return on average. This excess return is referred as the risk premium. The relationship between investors' expected return and risk was presented in an ARCH framework by Engle et al. [21]. They introduced the ARCH in mean, or ARCH-M, model where the conditional mean is an explicit function of the conditional variance of the process.

On the other hand, to overcome some weaknesses of the GARCH model in handling financial time series, Nelson [22] proposes the exponential GARCH (EGARCH) model. In particular, to allow for asymmetric effects between positive and negative asset returns, he considers a weighted innovation function for the conditional variance.

Basing ourselves in the AIC criterion, Ljung-Box test for residual autocorrelations and significance of the parameters we estimate the following EGARCH-M(1,1) model:

$$r_t = \mu + \eta\sigma_t^2 + \epsilon_t, \epsilon_t|\Omega_{t-1} \approx N(0, \sigma_t^2) \tag{4}$$

$$\ln \sigma_t^2 = \omega + \alpha(\frac{\epsilon_{t-1}}{\sigma_{t-1}}) + \gamma\|\frac{\epsilon_{t-1}}{\sigma_{t-1}}\| + \beta \ln \sigma_{t-1}^2 + \delta\phi_t \tag{5}$$

where r_t is the stock return at interval t, μ and ω are constant terms, ϵ_t are the serially uncorrelated errors of stock returns with mean zero, σ_t^2 is the conditional variance of ϵ_t, whereas ϕ_t is market related information demand . In this framework we can interpret:

- δ represents the contemporaneous effect that market information demand has on conditional variance.
- β can be interpreted as the persistence of the model.
- α relates standardized shocks to volatility in an asymmetric style. For $\alpha < 0$ the future conditional variances will increase proportionally more as a result of a negative shock than for a positive shock of the same absolute magnitude.
- γ relates lagged standardized innovations to volatility in a symmetric way.
- η represents the risk premium, i.e., the increase in the expected rate of return due to an increase in the variance of the return.

The results of the estimation are shown in table 7.

4.3 Interpretation of results

The contemporaneous effect that market information demand has on conditional variance (δ) is significant at a 5% level in 17 out of the 20 cases. It's effect is positive in most of them, which is consistent with the work of Vlastakis & Markellos [7]. The conditional variance of the returns is very persistent, as in the majority of the cases is very close to unity. This could be an indication for us to estimate an IGARCH model, but this would imply for the model to become technically complex to estimate, so it is left for future studies.

The coefficient for the asymmetric term is negative in most of the cases, which is an indication that the future conditional variances will increase proportionally more as a result of a negative shock than for a positive shock of the same absolute magnitude. Finally, we notice that the risk premium term is positive and significant in all but 5 cases. In principle, this is consistent with the argument that a higher risk asset should yield higher expected returns. We may interpret the 5 negative risk premium assets as having lower returns in times of high volatility in the markets.

Symbol	μ	η	ω	α	β	γ	δ
AXP	-0.00692**	0.135**	0.00209**	-0.168**	0.999**	-0.137**	0.0638**
BA	-0.0328**	0.658**	-0.156**	-0.127**	0.971**	-0.138**	0.247**
CAT	0.0295**	-0.311**	-0.403**	-0.214*	0.918**	0.421**	0.667
CSCO	-0.077**	1.11**	-0.882**	-0.222**	0.834**	-0.206**	-0.502**
DIS	0.0317**	-0.38**	-1.54	-0.241*	0.727**	0.271	1.48**
GE	-0.0141**	0.147**	-0.184**	-0.257**	0.965**	-0.194**	0.335**
GS	-0.0614**	0.873**	-0.481**	-0.233**	0.909**	-0.1**	0.8**
HD	-0.0292**	0.704**	-0.459**	-0.17**	0.919**	-0.175**	-0.295**
JNJ	0.0539**	-1.07**	-0.35**	0.0685**	0.946**	-0.174**	0.333**
JPM	-0.0594**	0.866**	-0.503**	-0.229**	0.903**	-0.108**	0.884**
KO	-0.0307**	0.787**	-0.207**	-0.115**	0.968**	-0.0728**	0.591**
MMM	0.000427	0.15	-1.1	-0.182*	0.814**	0.183*	-0.471
MRK	-0.0304**	0.571**	-0.794**	-0.364**	0.863**	-0.292**	-1.12**
NKE	-0.0369**	0.666**	-0.381**	-0.175**	0.929**	-0.134**	0.297**
PFE	-0.00725	0.249	-1.17**	-0.276**	0.804**	-0.0433	-0.35
PG	-0.0376**	0.946**	-1.23**	-0.302**	0.81**	-0.302**	-0.135**
UNH	-0.0291**	0.673**	-0.464**	-0.223**	0.916**	-0.122**	0.616**
UTX	-0.034**	0.671**	-0.627**	-0.22**	0.892**	-0.177**	-0.727**
V	0.0784**	-0.951**	-0.423**	0.227**	0.925**	-0.0225**	1.2**
VZ	0.0452**	-0.843**	-6.9**	0.228	-0.147**	-0.473	1.94**

Table 7: Parameters resulting from the EGARCH-M(1,1) estimation. The mean equation is $r_t = \mu + \eta\sigma_t^2 + \epsilon_t, \epsilon_t$, while the variance equation is $\ln \sigma_t^2 = \omega + \alpha(\frac{\epsilon_{t-1}}{\sigma_{t-1}}) + \gamma\|\frac{\epsilon_{t-1}}{\sigma_{t-1}}\| + \beta \ln \sigma_{t-1}^2 + \delta\phi_t$. The significance level are denoted as as ***, ** and * if the null is rejected at 1%, 5% or 10% respectively.

5 Future work

In this work we have estimated a basic model for returns and volatilities of the firms conforming the Dow Jones Index. We have also derived a significant measure of information demand that positively relates to assets volatility, and it has been successfully included in the model.

But so many paths have been left opened for future work:

- It has been shown that the effect of market information demand on volatility is positive in the majority of the cases, but not for all and not with the same magnitude. The most negative effects of the information demand on volatility are for United Technologies Co., Merck & Co., Cisco Systems, 3M and Pfizer. This are very big firms belonging to the software/technology and pharmaceutical sectors. This could be an indication that this effect may vary with the firm's sector, structure, size, etc...

- We have included in the model the effect of information demand, but not the effect of information supply. This effect could be measured by means of analyzing news sources like twitter, reddit, or even newspapers like the financial times [23][24][25].

References

[1] Herbert A. Simon. A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69, 1995.

[2] Yi Zhang Jordi Mondria, Thomas Wu. The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, 82, 2010.

[3] Rajan S. Patel Lynnette Brammer Mark S. Smolinski Larry Brilliant Jeremy Ginsberg, Matthew H. Mohebbi. Detecting influenza epidemics using search engine query data. *Nature*, 457, 2009.

[4] H. Eugene Stanley Tobias Preis, Daniel Reith. Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical transactions of the royal society A*, 2010.

[5] H. Eugene Stanley Steven R. Bishop Tobias Preis, Helen Susannah Moat. Quantifying the advantage of looking forward. *Scientific Reports*, 2, 2012.

[6] H. Eugene Stanley Tobias Preis, Helen Susannah Moat. Quantifying trading behavior in financial markets using google trends. *Scientific Reports*, 3, 2013.

[7] Raphael N. Markellos Nikolaos Vlastakis. Information demand and stock market volatility. *Journal of Banking & Finance*, 36, 2012.

[8] Hal Varian Hyunyoung Choi. Predicting the present with google trends. *Google Inc*, 2009.

[9] Engelberg J. Gao P. Da, Z. In search of attention. *The Journal of Finance*, 66, 2011.

[10] Bollerslev T. Diebold F.X. Ebens H. Andersen, T.G. The distribution of realized stock return volatility. *Journal of Financial Economics*, 2001.

[11] Bollerslev T. Diebold F.X. Labys P. Andersen, T.G. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 2001.

[12] Shephard Barndorff Nielsen, O. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society B*, 2002.

[13] Fuller W.A. Dickey, D.A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 1979.

[14] Phillips P.C.B. Schmidt P. Shin Y. Kwiatkowski, D. Testing the null hypothesis of stationary against the alternative of a unit root. *Journal of Econometrics*, 1992.

[15] Liu W. Pham P.K. Jarnecic E. Kalev, P. Public information arrival and volatility of intraday stock returns. *Journal of Banking and Finance*, 2004.

[16] A.N. Bomfim. Pre-announcement effects, news effects, and volatility: monetary policy and the stock market. *Journal of Banking and Finance*, 2001.

[17] R.F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 1982.

[18] A. Kendall, M. G.; Bradford Hill. The analysis of economic time-series-part i: Prices. *Journal of the Royal Statistical Society*, 1953.

[19] G. M. Ljung; G. E. P. Box. On a measure of a lack of fit in time series models. *Biometrika*, 1978.

[20] G. S. Durbin, J.; Watson. Testing for serial correlation in least squares regression.iii. *Biometrika*, 1971.

[21] Lilien D.M. Robins R.P. Engle, R.F. Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica*, 1987.

[22] D.B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 1987.

[23] V. Vance Roley Grant McQueen. Stock prices, news, and business conditions. *The review of financial studies*, 2015.

[24] Murray Z. Frank Werner Antweiler. Is all that talk just noise? the information content of internet stock message boards. *The journal of finance*, 2005.

[25] Xiaojun Zeng Johan Bollen, Huina Mao. Twitter mood predicts the stock market. *Journal of Computational Science*, 2011.

A Realised Volatility

The coefficients found for the MA(1) filter applied to the returns are shown in table 8.

Symbol	**MA(1) coefficient**	**p-value**
AAPL	0.022557	2.757361e-01
AXP	-0.032954	1.114356e-01
BA	0.005247	8.161807e-01
CAT	0.005337	7.921444e-01
CSCO	-0.030267	1.474866e-01
CVX	-0.059200	4.281851e-03
DIS	-0.011861	5.646949e-01
GE	-0.026319	2.079888e-01
GS	-0.036154	8.080090e-02
HD	0.001103	9.598348e-01
IBM	0.006155	7.603312e-01
INTC	-0.060036	3.410168e-03
JNJ	-0.070125	7.111677e-04
JPM	-0.100084	1.840416e-06
KO	-0.048472	1.912861e-02
MCD	-0.052263	1.356720e-02
MMM	-0.028250	1.611836e-01
MRK	-0.018213	3.828227e-01
MSFT	-0.083645	8.187378e-05
NKE	-0.032102	1.240827e-01
PFE	-0.019121	3.696674e-01
PG	-0.075078	4.267197e-04
SPY	-0.056944	7.605430e-03
TRV	-0.120285	1.786011e-09
UNH	-0.013861	5.001650e-01
UTX	-0.031659	1.271020e-01
V	-0.094429	1.650034e-05
VZ	0.013166	5.342568e-01
WMT	-0.066415	1.795043e-03
XOM	-0.142565	3.557150e-11

Table 8: Coefficients found for the MA(1) filter applied to the returns