Can We Use Daily Internet Search Query Data to improve Predicting Power of EGARCH Models for Financial Time Series Volatility?

Dimche Risteski, and Danco Davcev

***Abstract*—**In this paper, we propose new model for daily volatility prediction of daily index returns of the French CAC40 index. We extend the conventional EGARCH model to daily EGARCH-SVI model by adding daily Search Volume Index (SVI) from Google Trend as an exogenous variable to EGARCH model. The main contribution of this paper is introducing daily SVI in EGARCH model to improve forecasting of daily volatility for French CAC40 index. The empirical results demonstrate that daily EGARCH-SVI model provides better daily index returns volatility prediction than basic EGARCH model. Also empirical results demonstrate that weekly EGARCH-SVI model have bigger improvement in predicting power than daily EGARCH-SVI model compared to conventional EGARCH model.

***Keywords*—**Volatility forecasting, financial time series, EGARCH, Google Trend

I. Introduction

INTERNET has completely changed the way of exchanging information. With using the Internet we also create additional information that can be used for other purpose. Today if we want to find some information we are using searching engines like Google. When we use Google to search some information we also leave additional information for our search of interest. From this information Google is creating Searching Volume Index (SVI) for the keywords that are used for search. SVI for a search term is the number of searches for that term scaled by its time-series average. Daily SVI is available only for ninety days. Weekly SVI is available for bigger period. Google product that makes SVI public is Google Trend [18]. Google's Chief Economist Hal Variant suggested that search data have the potential to capture the interest in different economic activities in real time and that offers an instantaneous picture of the economy [10]. Choi and Variant [11] provide evidence that search data can predict home sales, automotive sales, and tourism. Da and Engelberg [12] propose a new and straight measurement of investor's attention using search frequency SVI from Google Trend. They found out that increase in SVI predicts higher stock prices in the next two weeks and an eventual price reversal within a year. Dimpfl and Jang [13] found out that searching queries can help to improve predicting power of vector auto regressive models for forecasting realized volatility for Dow Jones index. Preis and Susannah [14], found out that by analyzing changes in Google query volumes for search terms related to finance, they find patterns that may be interpreted as "early warning signs" of stock market moves. Many other researches also are using Google Trends for different purpose [1, 2, 3, 4, 5, 15, 16, and 17].

Volatility analysis of financial time series is an important aspect in financial industry. The main purposes of forecasting volatility are option pricing, risk management and asset allocation and. In the last few decades, many volatility models have been developed, the most popular model is the Autoregressive Conditional Heteroskedasticity (ARCH) models by Engle [6] and it is extended to Generalize ARCH (GARCH) models by Bollerslev [7]. Nelson [8] suggested the Exponential GARCH (EGARCH) model to capture asymmetric variance effects and it performs well in modeling equity returns. In our previous paper [10] we are proposing a new model for improving predicting power of EGARCH model for volatility prediction of index returns of the German DAX index. We extend the conventional EGARCH model to EGARCH-SVI model by adding weekly SVI from Google Trend as an exogenous variable to the EGARCH model. The empirical results demonstrate that EGARCH-SVI model with weekly data provides better index returns volatility prediction than EGARCH model.

In this study, we analyze the daily and weekly returns of French CAC40 index that tracks the 40 most main companies on the French equities market. We propose new daily EGARCH-SVI model. We assess the influence of the daily and weekly Search Volume Index for the term CAC40 to the prediction power of the daily and weekly EGARCH model for volatility time series modeling. Our aim is the introduction of the using of daily SVI query as additional information in EGARCH to improve predicting power of conventional EGARCH model. The resulting EGARCH-SVI model show better volatility forecast compared to original EGARCH models for, in-sample as well as out-of-sample. The reminder of this paper is as follows. In section 2, we give detailed description of the data. In section 3 we formally define our daily EGARCH-SVI model. In section 4 we describe the empirical results of our volatility forecasting and finally, our conclusion is in the last section.

Dimche Risteski, Faculty of Computer Science and Engineering, Skopje, Macedonia (e-mail: rdimce@yahoo.com).

Danco Davcev, Faculty of Computer Science and Engineering, Skopje, Macedonia (e-mail: danco.davcev@finki.ukim.mk).

II. DATA

For our study, we consider daily and weekly returns on CAC40 index covering the sample period from January 2004 until March 2014. We take Bloomberg [19] as data provider for our data set. Figures 1 below show the daily return of the CAC40 index from January2004 until March 2014.

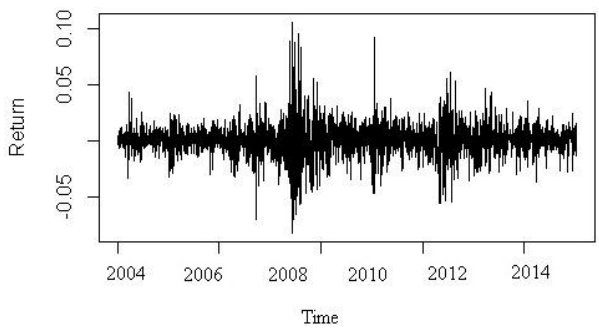

Fig. 1 Daily returns CAC40 January 2004- March 2014

The CAC index (CAC40) is the most widely-used indicator of the Paris market, reflects the performance of the 40 largest equities listed in France, measured by free-float market capitalization and liquidity [19].

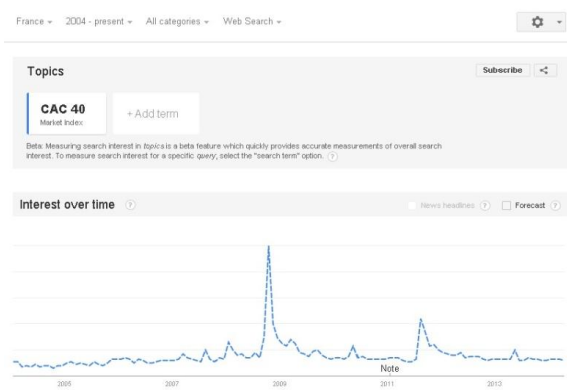

Fig. 2 Weekly SVI for "CAC40" January 2004- March 2014

From Google Trend, we use data Search Volume Index (SVI) for the keyword "CAC40". Fig.2 shows the weekly SVI data for the "CAC40" searching term from Google Trends from January 2004 until March 2014. The volume measure is based on the number of searches which were submitted within the France for CAC40 keyword. The data from Google Trend are relative in nature because they do not provide effective total number of searches, but only the search volume index. To create daily data for longer period from Google Trend first we download all quarterly daily data for period from January 2004 until March 2014. Google Trends have limitation of 90 days for daily data. To create daily data first we calculate the percentage change between the days. Then we use weekly data and percentage of daily change to create new daily index value [20]. From the daily SVI data we remove not trading dates. Fig.3 shows the daily SVI data for the "CAC40" searching term from Google Trends on the search volume from January 2014 until March 2014.

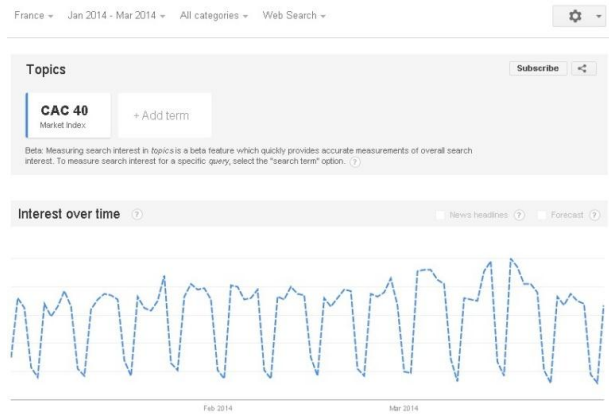

Fig. 3 Daily SVI for "CAC40" January 2014- March 2014

III. DAILY EGARCH-SVI MODEL

If p_t is the closing price of the index at the end of trading day then return of the index is defined as:

$$r_t = \log(p_t) - \log(p_{t-1}) \;\; t = 1,..,n \quad (1)$$

GARCH model was first developed by Bollerslev (1986) and extended to Exponential GARCH (EGARCH) by Nelson [8] to capture the "leverage effect" of equity returns. In this paper we consider straightforward EGARCH (1, 1) model, which is adequate for time series volatility modeling of asset returns. Equation (2) represents the conditional mean mode, each return r_t consist of a conditional mean, plus an uncorrelated, white noise innovations (ε_t):

$$\mathrm{r_t} = \mu + \varepsilon_t \quad \mathrm{t} = 1,..,\mathrm{n} \qquad (2)$$

$$\varepsilon_t = \mathrm{Z_t}\sigma_t \;\; \mathrm{t} = 1,..,\mathrm{n} \qquad (3)$$

In equation (3) Zt is a sequence of independent and identically distributed (i.i.d.) random variables with zero mean and unit variance:

$$\ln \sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_i \left[\varphi Z_{t-j} + \psi\left(|z_{t-j}| - E|z_{t-j}|\right)\right] + \sum_{i=1}^{p} \beta_i \ln \sigma_{t-i}^2 \quad t = 1,.,n \qquad (4)$$

Equation (4) is case of asymmetric model and represents the conditional variance model where conditional variance depends on both size and the sign of lagged residuals. EGARCH (p, q) model may be defined as a combination of Equation (3) and Equation (4). When p=1 and q=1 than we have simple EGARCH (1, 1):

$$ln\, \sigma_t^2 = \alpha_0 + \varphi Z_{t-1} + \psi(|z_{t-1}| - E|z_{t-1}|) + \beta_1\, ln\, \sigma_{t-1}^2 \quad t = 1,..,n \qquad (5)$$

We add one exogenous variable daily SVI from Google Trend as additional information to the conditional variance.

The new conditional variance of model EGARCH-SVI is defined as:

$$ln\,\sigma_t^2 = \alpha_0 + \varphi Z_{t-1} + \psi(|z_{t-1}| - E|z_{t-1}|) + \beta_1\, ln\,\sigma_{t-1}^2 + \gamma_1 SVI_{t-2}\; t = 1..n \quad (6)$$

SVI – represents daily changes in Google query volumes for search terms related to the asset that we want to model but with lag 2. Google Trend has two days delay in providing the data. In our case of CAC40 index, we are using SVI for the term "CAC40". In the weekly EGARCH-SVI model when we add one exogenous variable weekly SVI from Google Trend as additional information to the conditional variance we are using lag 1[10]. Maximum likelihood estimation is applied for estimation of the parameters of the model.

IV. Empirical Results

We test daily and weekly EGARCH-SVI model with the data set for CAC40 index from France. The model was implemented in R version 3.1.0. For the purpose of comparison we are testing the both models daily and weekly EGARCH and daily and weekly EGARCH -SVI.

We are interested in analyzing the daily volatility of the index so we are calculating the daily return of the CAC40 index from January 2004 until March 2014. From Google Trends we are taking the daily SVI for "CAC40" for every quarter for the same period. Then with using the weekly data [20] we generate daily index.

We are also looking for any evidence of conditional heteroskedasticity in the time series by using ACF and PACF plot. The ACF (Autocorrelation) plot is a bar chart of the coefficients of correlation between a time series and lags of itself and the PACF (Partial Autocorrelation) plot looks at the amount of partial correlation coefficients between the series and lags of itself.

Figure 4 shows the ACF and PACF plots for daily close price as the ACF decrease slowly and the PACF shows only pick for the 1th lag.

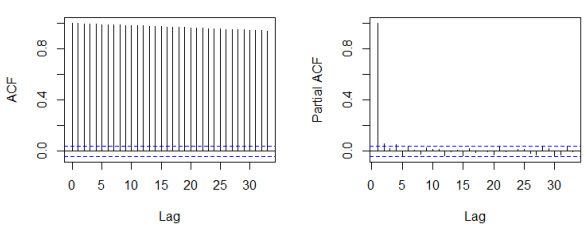

Fig. 4 ACF and PCF close price CAC40

Figure 5 ACF and PACF plots for daily CAC40 index returns show only significant autocorrelation in the 1th lag.

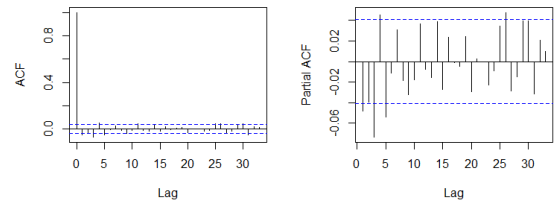

Fig. 5 ACF and PCF for CAC40 returns

Figure 6 confirms the similar pattern of the ACF and PACF for Google Trend and close price as the ACF decrease and the PACF shows only pick for the 1st lag.

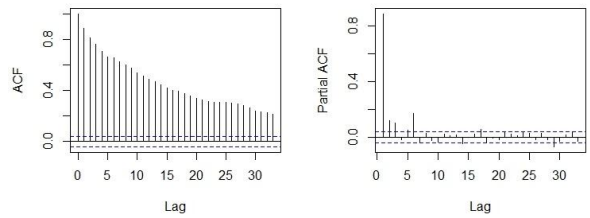

Fig. 6 ACF and PCF for Google Search Index for CAC40

Table I shows optimal parameters for EGARCH (1, 1) model.

TABLE I
Optimal parameters for EGARCH model

Parametars	Estimate	Std.error	t-value	Pr(>|t|)
mu	0.000016	0.000198	0.081723	0.93487
omega	-0.255066	0.041966	-6.07798	0
alpha1	-0.178954	0.016581	-10.7924	0
beta1	0.970712	0.004879	198.9773	0
gamma1	0.10283	0.01616	6.363092	0

Table II shows optimal parameters for EGARCH-SVI model.

TABLE II
Optimal parameters for EGARCH-SVI model

Parametars	Estimate	Std.error	t-value	Pr(>|t|)
mu	0.000018	0.000236	0.074927	0.940272
omega	-0.63777	0.103646	-6.15332	0
alpha1	-0.20888	0.01867	-11.1877	0
beta1	0.944824	0.008405	112.4103	0
gamma1	0.095307	0.015484	6.155305	0
svi	0.06191	0.014261	4.341207	0.000014

In our new model EGARCH-SVI, new variable svi is statistically significant with t-value = 4.341207.

To assess the forecasting performance, we compute following point estimation performance: Mean Square Error (MSE), Mean Absolute Error (MAE), Information criteria (Akaike, Bayes, Shibata and Hannan-Quinn) and Diebold-Mariano.

Table III and Table IV shows the evaluation of the in-sample and out-sample forecasting of the model with using

Akaike, Bayes, Shibata and Hannan-Quinn information criteria.

TABLE III
DAILY IN SAMPLE TEST RESULTS

Measure	EGARCH	EGARCH-SVI	Improvement
Akaike	-5.958598	-5.970183	0.2 %
Bayes	-5.945927	-5.954978	0.15 %
Shibata	-5.958607	-5.970197	0.2%
H.-Quinn	-5.953974	-5.964634	0.18%

From the results in the Table III, we can see that our new model generates better in sample forecasting compared with the simple EGARCH. Information criteria Akaike and Shibata information criteria are 0.2% better for the new daily EGARCH-SVI model. Also Bayes and Hannan-Quinn information criteria is better (0.15%, 0.18%) for the new model.

Table IV demonstrates that for out of sample 5 EGARCH-SVI generates better 0.02% (0.1%) better MSE (MAE) than EGARCH and for out of sample 10 EGARCH generates 0.01% (0.01%) better MSE (MAE) than basic EGARCH. MSE is improved 0.01% with forecasting EGARCH-SVI for the out of sample 15. To compare both models, we also calculate the Diebold Mariano (DM) value with using random walk method as a benchmark. EGARCH-SVI model has DM value of 20.4114 that is bigger than DM value of 20.4112 for EGARCH model and both are statistically significant i.e. daily EGARCH-SVI predicts volatility better than EGARCH.

TABLE IV
DAILY OUT SAMPLE TEST RESULTS

Measure	EGARCH	EGARCH-SVI	Improvement
MSE(5)	0.000115659	0.000115641	0.02%
MAE(5)	0.0095041	0.0095036	0.01 %
MSE(10)	0.0000868	0.0000867	0.01 %
MAE(10)	0.007774	0.007772	0.02 %
MSE(15)	0.0000819	0.0000818	0.01%
MAE(15)	0.007649	0.007648	0.01%

We are also interested in analyzing the weekly volatility of the CAC40 index so we are calculating the weekly return of the CAC40 index from January 2004 until March 2014. From Google Trends we are taking the weekly SVI for "CAC40" from the same period.

First we are checking conditional heteroskedasticity in the time series by using ACF and PACF plot.

Figure 4 shows the ACF and PACF plots for daily close price as the ACF decrease slowly and the PACF shows only pick for the 1th lag.

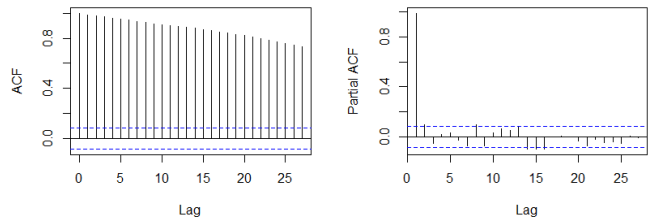

Fig. 4 ACF and PCF weekly close price CAC40

Figure 5 ACF and PACF plots for daily CAC40 index returns show only significant autocorrelation in the 1th lag.

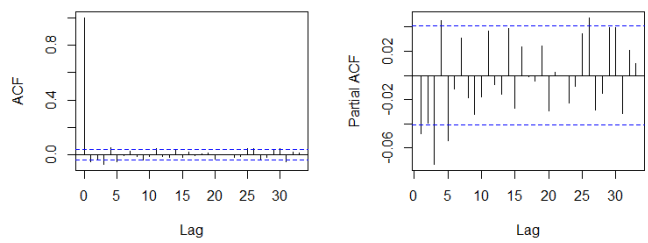

Fig. 5 ACF and PCF for weekly CAC40 returns

Figure 6 confirms the similar pattern of the ACF and PACF for Google Trend and close price as the ACF decrease and the PACF shows only pick for the 1st lag.

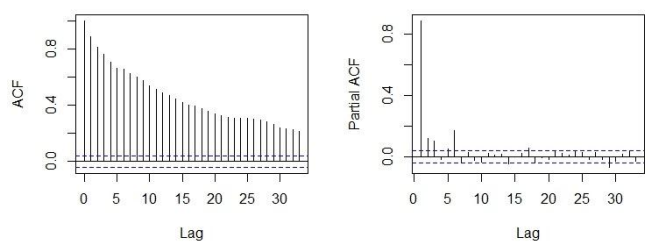

Fig. 6 ACF and PCF for Google Search Index for CAC40

Table V shows optimal parameters for weekly EGARCH model.

TABLE V
OPTIMAL PARAMETERS FOR WEEKLY EGARCH MODEL

Parametars	Estimate	Std.error	t-value	Pr(>|t|)
mu	0.000243	0.00095	0.25632	0.797703
omega	-0.454702	0.1089	-4.17543	0.00003
alpha1	-0.255134	0.037772	-6.75466	0
beta1	0.938177	0.01518	61.80304	0
gamma1	0.118428	0.038745	3.0566	0.002239

Table VI shows optimal parameters for EGARCH-SVI model. In our new model, the new variable SVI is statistically significant with t-value =4.07651.

TABLE VI
OPTIMAL PARAMETERS FOR EGARCH-SVI MODEL

Parametars	Estimate	Std.error	t-value	Pr(>|t|)
mu	0.000428	0.000885	0.48311	0.629021
omega	-3.27659	0.750481	-4.36599	0.000013
alpha1	-0.37122	0.049185	-7.5475	0
beta1	0.716276	0.064413	11.12007	0
gamma1	-0.03633	0.055678	-0.65254	0.514051
svi	0.483922	0.11871	4.07651	0.000046

Table VII and Table VIII show the evaluation of the in-sample and out-sample forecasting of the model with using Akaike, Bayes, Shibata and Hannan-Quinn information criteria.

TABLE VII
WEEKLY IN SAMPLE TEST RESULTS

Measure	EGARCH	EGARCH-SVI	Improvement
Akaike	4.506216	-4.580298	1.6%
Bayes	-4.466080	-4.532134	1.5%
Shibata	-4.506390	-4.580548	1.6%
H.-Quinn	-4.490510	-4.561451	1.6%

From the results in the Table VII, we can see that our new model generates better in sample forecasting compared with the simple EGARCH.

TABLE VIII
WEEKLY OUT SAMPLE TEST RESULTS

Measure	EGARCH	EGARCH-SVI	Improvement
MSE(5)	0.000530328	0.000529244	0.2%
MAE(5)	0.02117608	0.02114117	0.2%
MSE(10)	0.0003670	0.0003642	0.8%
MAE(10)	0.0162871	0.0161688	0.7%
MSE(15)	0.0003975	0.0003957	0.5%
MAE(15)	0.0164730	0.163756	0.6%

Table VIII demonstrates that for out of Sample 5 EGARCH -SVI generates better 0.2% (0.2%), better MSE (MAE) than EGARCH and for out of sample 10 EGARCH generates 0.8% (0.7%) better MSE (MAE) than basic EGARCH. MSE (MAE) is improved 0.5% (0.6%) with forecasting EGARCH-SVI for the out of sample 15. Weekly EGARCH-SVI model has DM value of 9.2038 that is bigger than DM value of 9.1947 for EGARCH model and both are statistically significant i.e. weekly EGARCH-SVI predicts volatility better than EGARCH.

V. CONCLUSION

In this paper, we present new daily EGARCH-SVI model for improving predicting power for daily volatility prediction of index return. We add daily data from Google Trends as exogenous variable to the conditional variance of the EGARCH model. We assess the extension of our model for prediction of CAC40 index daily volatility; we use the Searching Volume Index for a search term "CAC40" from Google Trend as additional information to the conditional variance on the CAC40 index. We conclude from our empirical results that EGARCH with additional daily SVI information have better forecasting power for time series daily volatility than basic EGARCH. The new model daily EGARCH-SVI has better results in both volatility forecasting in sample and out of sample compared to EGARCH. Empirical results demonstrate that weekly EGARCH-SVI model have bigger improvement in predicting power that daily EGARCH-SVI model compared to conventional EGARCH model. For future research, we plan to investigate the influence of mixed daily and weekly data on improvement of volatility predicting power of EGARCH-SVI model.

REFERENCES

[1] Ladislav Kristoufek ."BitCoin meets Google Trends and Wikipedia: Quantifying the relation-ship between phenomena of Internet era" Scientific Reports 3, no: 3415 Dec. 2013.

[2] Ladislav Kristoufek "Can Google Trends search queries contribute to risk diversification?" Scientific Reports 3, no:2713 Sep. 2013

[3] J.Mondria,T.Wu,Y.Zhang."The determinants of international investment and attention allocation: Using internet search query data" Journal of International Economics, Vol.82, pp 85-95,Sep. 2010. http://dx.doi.org/10.1016/j.jinteco.2010.04.007

[4] T.Preis,D.Reith, H.Stanley." Complex dynamics of our economic life on different scales: in-sights from search engine query data" Phil. Trans. R pp. 5707–5719, 2010.

[5] M.Drake,D.Roulstone,J.Thornock."Investor Information Demand: Evidence from Google Searches Around Earnings Announcements" Journal of Accounting Research , Jan. 2012 http://dx.doi.org/10.1111/j.1475-679X.2012.00443.x

[6] Engle, Robert F. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," Econometrica, 50:4, pp. 987-1007, 1982. http://dx.doi.org/10.2307/1912773

[7] Bollerslev,T."Generalized Autoreg. Conditional Heteroskedasticity" J.Econometrics, 31,pp 307-327, 1986 http://dx.doi.org/10.1016/0304-4076(86)90063-1

[8] D.Nelson. "Conditional Heteroskedasticity in Asset Returns: A new Approach," Econometrica, Vol 59, No 2, pp.347-370. 1991. http://dx.doi.org/10.2307/2938260

[9] "Micro starts, macro effects", The Economist, November 24th 2012;

[10] D.Risteski,A.Sadoghi,D.Davcev "Improving predicting power of EGARCH models for financial time series volatility by Google Trends" International Conference on FEEMCE Shanghai pp.312-317, Nov. 2013.

[11] H. Choi, H. Varian "Predicting the Present with Google Trends," Economic Record, vol.88, pp 2-9, 2012. http://dx.doi.org/10.1111/j.1475-4932.2012.00809.x

[12] Zhin Da, Joseph Engelberg , and Pengjie Ca. "In Search of Attention," J. Finance, LXVI, 5, pp 1461-1499, 2011. http://dx.doi.org/10.1111/j.1540-6261.2011.01679.x

[13] Thomas Dimpfl and Stephan Jank . "Can internet search queries help to predict stock market volatility" unpublished paper, 2013.

[14] Tobias Preis, Helen Susannah Moat & H.Eugene Stanley. "Quantifying Trading Behavior in Financial Markets Using Google Trends," Scientific Reports, 2013.

[15] M. Bank, M. Larch, and G. Peter, "Google Search Volume and its Influence on Liquidity and Returns of German Stocks," Financial Markets and Portfolio Management, vol.25, pp. 239-264, 2011 http://dx.doi.org/10.1007/s11408-011-0165-y

[16] Bordino I, Battiston S,Galdarelli G."Web Search Queries Can Predict Stock Market Volumes" PLoS ONE 7(7): e40014. doi:10.1371/journal.pone.0040014, 2012.

[17] 17. Andrei Daniel and Hasler Michael. "Investor Attention and Stock Market Volatility"-Rotman School of Management, unpublished, 2013.

[18] http://www.google.com/trends

[19] http://www.bloomberg.com

[20] http://erikjohansson.blogspot.com.