Google Trends and Conditional Volatility:

Evidence from the Oil and Gold markets.

Christina Rouska

University of Macedonia

Interdepartmental Program of Postgraduate Studies in Economic Science

Thesis

Supervising Professor: Theodore Panagiotidis

February 2016, Thessaloniki

1 Contents

Abstract

This study explores the effect of Google search activity on the conditional volatility of the crude oil and gold price returns as a direct measure of investor sentiment. Two alternative approaches are adopted: a Google index and common factors, generated from a list of search queries. Within an EGARCH framework the empirical results support that there is evidence of asymmetry and persistence in the volatility and, in a certain degree, overlapped by our information demand variables.

Keywords: Google Trends, Volatility asymmetry, EGARCH, Principal components

1. Introduction

Within the last decades various approaches have been suggested in modeling the volatility in financial data; starting with the autoregressive conditional heteroskedasticity (ARCH) model (Engle 1982) and generalized ARCH (GARCH) model (Bollerslev 1986), many extensions of these models have been introduced in the literature. One of the commonly observed features of financial time series is the volatility clustering, that is periods of high volatility tend to be followed by periods of low volatility, and GARCH-type models have proved to be effective tools in capturing the volatility clustering in the data. Nelson (1991) introduced a modified model, called exponential GARCH (EGARCH), which allow us to capture the asymmetric effect of shocks on the conditional volatility. Furthermore, several studies investigate the relationship of volatility and information flow in the markets, making information demand on real-time economic activity a key factor for more accurate predictions, investment decisions and policy making in finance or economics. Various indicators have been constructed as a measure of economic activity and financial market "stress", but they are often released with a delay of a few weeks. These days, we have access to more contemporaneous data sources on, nearly, real-time economic activity. In our analysis we focus on Google Trends, a public web facility of Google Inc., which measures the search volume of particular queries individuals enter into Google search engine. With Google occupying the larger proportion of search engine popularity, globally and in the US, Google Trends is a useful tool to reveal information demand.

In this framework, we utilize the Google search volume information as a direct measure of individual attention (Da, Engelberg, and Gao 2010); using an EGARCH model we examine the effect of the internet information on the conditional volatility of crude oil and gold price

returns. Two proxies for the component of internet information are adopted; a Google Index constructed from the aggregated volume of a list of searching queries and common factors based on the method of principal components. The evidence indicate that excess internet search activity may reflect to an amplify volatility in the price returns, either as a result of noise trading (Da, Engelberg, and Gao 2011; Vlastakis and Markellos 2012), or increased hedging positions in times of turmoil due to the significant portfolio diversification properties of gold (Baur 2011). Additionally, Google search information variables seem to have a predictive ability beyond the alternative conventional sentiment indicators based on surveys, newspaper news or market-based variables measuring financial "stress" (Da, Engelberg, and Gao 2010; Kholodilin, Podstawski, and Siliverstovs 2010; Peri, Vandone, and Baldi 2014).

2. Literature Review

The vast amount of information that stems out of internet search volume, has already been exploited in wide fields of research. Many applications in financial markets analysis were carried out. Da *et al.* (2011) has employed a study that proposes the Google search frequency as a direct measure of investor attention. Also, Vlastakis & Markellos (2010) indicate the connection between internet search activity and noise trading and Da, Engelberg & Gao (2010) present the prospect of economy-related search queries constructing an index that captures investor sentiment. In epidemiology, Ginsberg *et al*. (2008), using millions of online-search data of US households, found that internet-information could help predict influenza outbreaks. Other applications of Google search query data include Choi & Varian (2012), predicting initial claims for unemployment, automobile demand and vacation destinations and Askitas & Zimmermann (2009) for finding intense correlation between keyword searches and unemployment rates in US, Germany and Israel. Finally, in the most recent work of Peri, Vandone & Baldi (2014), it is examined the relationship between internet-information flows and commodity future prices.

So, the overall literature is summarized in the following tables:

	Title	Author	Keywords	Data	Theme
1	In Search of Attention	(Da, Engelberg, and Gao 2011)		This paper used a data sample of 3000 stocks from 2004 to 2008 (Frank Russell and Company) and their corresponding weekly SVIs (search volume indexes) from Google Trends. Other variables were obtained from CRSP, Standard and Poor's COMPUTSTAT and I/B/E/S.	This paper proposes a new, direct measure of investor attention using search frequency in Google (SVI) instead of existing measures. In a sample of 3000 stocks it is found that SVI is correlated with investor attention, especially retail investors. Also the results provide evidence that an increase in SVI leads to temporary increase in stock prices, particularly in the case of IPOs.
2	Google Econometrics and Unemployment Forecasting	(Askitas and Zimmermann 2009)	Google, internet, keyword search, search engine, unemployment, predictions, time-series analysis	The data set consists of monthly unemployment rates of Germany from January 2004 to April 2009, which are obtained from the Federal Employment Agency, and search volume indexes for specific keywords collected from Google Insights.	This paper provides a new method of using data on internet activity in order to examine when, how and at what extent unemployment in Germany is affected after a long period of strong recovery. The results point out strong correlations between keyword searches and unemployment rates.
3	Google Internet search activity and volatility prediction in the market for foreign currency	(Smith 2012)	Google insights for Search ARCH (GARCH) Mixture of distributions hypothesis (MDH) Foreign currency Foreign exchange	Daily spot price data from Jan 2004 to Dec 2010 for the Australian dollar, Canadian dollar, Euro, Japanese yen, New Zealand dollar, Swiss franc and British pound, which were downloaded from Federal Reverse System. Also, the search volume data for the keywords economic crisis+financial crisis, inflation and recession were obtained from Google Insights.	This paper illustrates the predictive power of Google Internet searches for specific keywords over volatility in the market for foreign currency. Specifically, the data for the keywords economic crisis+financial crisis and recession has predictive power beyond the GARCH (1, 1).

	Title	Author	Keywords	Data	Theme
4	Information demand and stock market volatility	(Vlastakis and Markellos 2012)	Information demand Financial markets Volatility Risk aversion	The data used in this paper is a sample of 30 of the largest stocks traded on NYSE and NASDAQ (Dow Jones Industrial Average Index). Specifically, the sample consists of weekly closing stock prices and trading volumes with corresponding values of the S&P500 and VIX indexes. The search volume index (SVI) for these stocks was obtained from Google Insights for Search for the period Jan 2004 to Oct 2009.	This paper examines the information demand and supply at the firm and market level and its relation to stock market activity and risk aversion. The data included in the analysis are the 30 largest stocks traded in NYSE and NASDAQ, and the SVIs for their related keywords. The results indicate that information demand, approximated from the SVI, has a significant impact on individual stock and market level trading volume. Moreover, it is corroborated that there is a positive correlation between information demand and risk aversion.
5	Can Google date help predict French youth unemployment ?	(Fondeur and Karamé 2013)	Google econometrics, Forecasting, Nowcasting, Unemployment, Unobserved components, Diffuse initialization, Kalman filter, Univariate treatment of time series Smoothing, Multivariate models	The data set consists of monthly unemployment data, provided by the national employment agency of France for the period Jan 2004-Jul 2011, and a weekly Google series generated from Google search volume for the term “EMPLOI” considered to be related to the French labor market.	The purpose of this paper is to test the ability of Google search data to improve predictions of youth unemployment in France. The results infer that Google information enhance prediction models in terms of both level and accuracy.
6	Nowcasting with Google Trends in an Emerging Market	(Carrière-Swallow and Labbé 2013)	Nowcasting, Google Trends, forecast accuracy, emerging markets	This paper used monthly data on the volume of car sales provided from the national statistics agency in Chile, IMACEC, an index of economic activity released monthly by the Central Bank of Chile and a Google Trends-based index (GTAI) of automobile-related keywords.	The aim of this paper is to investigate whether Google activity correlates with consumer purchases in an emerging market. The results show that the models including GTAI outperform competing benchmark specifications in both in and out-of-sample nowcasting applications.

	Title	Author	Keywords	Data	Theme
7	Quantifying Trading Behavior in Financial Markets Using Google Trends	(Preis, Moat, and Stanley 2013)		The data are time series of closing prices of the Dow Jones Industrial Average from Jan 2004 until Feb 2011. The online data consists of the Google search volume of relative terms to the keyword "debt", restricted in the United States.	This study provides evidence of the relationship between search volume changes and stock market prices. The data point out that there are increases in search volume of financial market-related keywords before stock market falls. These early warning signs can be used in the construction of profitable trading strategies.
8	Forecasting Stock Returns: Do Commodity Prices Help?	(Black et al. 2014)	stock prices, commodity prices, forecasting, rolling	The data set is composed of quarterly index series for stocks (S&P500) and commodities (S&P GSCI) over the time period: from 1973:Q1 to 2012:Q2. Also included, results for the Dow Jones-UBS commodity index and several individual commodities, as Baltic Dry Index, Corn, Energy, Gold, LME and Precious Metals for the start dates 1991, 1985, 1979, 1983, 1978, 1984 and 1973, respectively, obtained from DataStream. Other variables used are the dividend yield and short-term interest rates.	This paper examines the relation between stock and commodity prices and whether this can be used to forecast stock returns. The evidence reveal the existence of a long-run cointegrating relationship. There is also evidence of change in the short-run relationship between returns, which has been affected following the dot.com crash. So, the predictive power is viewed over the full sample and subsamples identified by break tests. The comparison among forecast approaches, such as historical mean, recursive and rolling models suggests that when ignoring breaks the historical mean model outperforms all the others and when allowing parameter values to change within the forecast regression the forecast performance improves, while rolling approach performs better than the recursive.
9	Detecting influenza epidemics using search engine query data	(Ginsberg et al. 2009)		In this paper are used weekly time series, from Google database, for 50 million of the most common online web search queries in the United States from 2003 to 2008 and weekly data from the Centers for Disease Control and Prevention (CDC) including influenza-like illness (ILI) physician visits.	In this paper it is analyzed the relationship between millions of online web searches on non-seasonal flu outbreaks and the physician visits in the US. The study estimates precisely enough the level of weekly influenza activity in each region of the US and generates indicators, which are available formerly than more conventional statistical information, with reporting lag of one day. The early detection derived by this approach allows for better response to seasonal epidemics.

	Title	Author	Keywords	Data	Theme
10	Predicting the Present with Google Trends	(Choi and Varian 2012)		The data used are monthly 'Motor Vehicles and Parts Dealers' series from the US Census Bureau 'Advance Monthly Sales for Retail and Food Services' report, weekly seasonally adjusted initial claims data for unemployment benefits released from the US Department of Labor, monthly visitor data from US, Canada, Great Britain, Germany, France, Italy, Australia, Japan and India obtained from Hong Kong Tourism Board and Roy Morgan Consumer Confidence Index for Australia. Also included Google Trends data for the related queries to each index.	This paper exploits the timely data of online web search queries for short-term prediction of economic indicators. Examples comprise automobile sales, unemployment claims, travel destination planning and consumer confidence. The results indicate that simple seasonal AR models that include relevant Google Trends predictors tend to perform better than models which exclude them.
11	Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches	(Artola and Galan 2012)	Google, forecasting, nowcasting, tourism	The data are constituted of weekly time series of the Google Trends index for the searching term "Spain Holiday" (G-INDEX) restricted in the United Kingdom for the period Jan 2004-Sep 2011 and monthly data for the tourist arrivals from Britain in Spain, provided by the Tourism Studies Institute of Spain (IET).	This study focuses on exploring the prospect of forecasting foreign tourist inflows using Google Trends data. Specifically, the paper presents an application for the British tourist arrivals in Spain. Using the information contained in Google searches, an adjusted timely indicator for the tourist inflows is obtained (G-indicator). The results support that the forecast performance of models including the G-indicator depends on the model taken as a benchmark, for the estimation period 2006-2010.
12	Twitter mood predicts the stock market	(Bollen, Mao, and Zeng 2011)	Index Terms, stock market prediction, twitter, mood analysis	The data set was formed from a collection of public tweets, recorded from Feb 28 to Dec 19th, 2008. Using two assessment tools, OpinionFinder and GPOMS, 7 public mood time series were obtained. Additionally, a time series of daily DJIA closing values was extracted from Yahoo! Finance.	This paper demonstrates the predictive correlation between measurements of the public mood states from Twitter feeds and the DJIA values. It is observed that changes of the public mood, interpreted by mood dimensions, correspond to shifts in the DJIA values occurring 3 to 4 days later. Finally, a Self-Organizing Fuzzy Neural Network model (SOFNN) exhibits significant prediction accuracy.

	Title	Author	Keywords	Data	Theme
13	Using internet search data as economic indicators	(McLaren and Shanbhogue 2011)		For this study are used monthly unemployment data and other unemployment indicators: claimant count and the GfK consumer confidence question. Also included, monthly house price growth (HP) data and other indicators such as the house price growth balances from the Home Builders Federation and the Royal Institution of Chartered Surveyors. Finally, in this paper are downloaded weekly data for the related queries from the Google Insights for Search, for the time period Jun 2004 to Jan 2011, confined in the UK.	This article presents and evaluates the utility of internet search data as indicators of economic activity. More specifically, the study is focused on two specific markets: the labor and housing markets in the United Kingdom. The results suggest that internet search data can help predict changes in unemployment and house prices, at least as effectively as alternative indicators.
14	Predicting consumer behavior with Web search	(Goel et al. 2010)	culture, predictions	In this paper are used movie data on revenue, budget and number of opening screens in the US, obtained from IMDb for the period Oct 2008 to Sep 2009. Sales and critic ratings data for video games provided by VGChartz, from Sep 2008 to Sep 2009 and music data (the weekly Billboard rank), for the period Mar-Sep 2009. Search data for movies, video games and songs were collected from Yahoo!'s Web Search query logs for the US market and music.yahoo.com, respectively.	This paper investigates the ability of search activity data to predict consumer activity, by forecasting the opening weekend box-office revenue for feature films in the US, first-month sales of video games and the weekly Billboard rank. The outcome of the analysis is that search data have indeed predictive power, nevertheless, other existing indicators perform equally well or even better. Additional, models that include both search and baseline data, illustrate modest increase of the predictive accuracy, when publicly available data are used. Finally, in the absence of other data sources, search-based prediction models yield greater performance boost.
15	Learning Stock Volatility Using Keyword Search Volume	(Yang 2011)		The data used are daily stock prices of the sectors Technology, Energy and Financial, obtained from Yahoo! Finance. Additionally, weekly search volume data, from the US, for the corresponding stock tickers and the term "recession", from Google Insights for Search over the period Jan 2004 to Nov 2011.	This study examines how the volatility of a stock correlates with the search volume of the ticker and the word "recession". For this purpose, two approaches were used, the logistic regression and Gaussian discriminant analysis (GDA), with the later performing better in determining high volatility events.

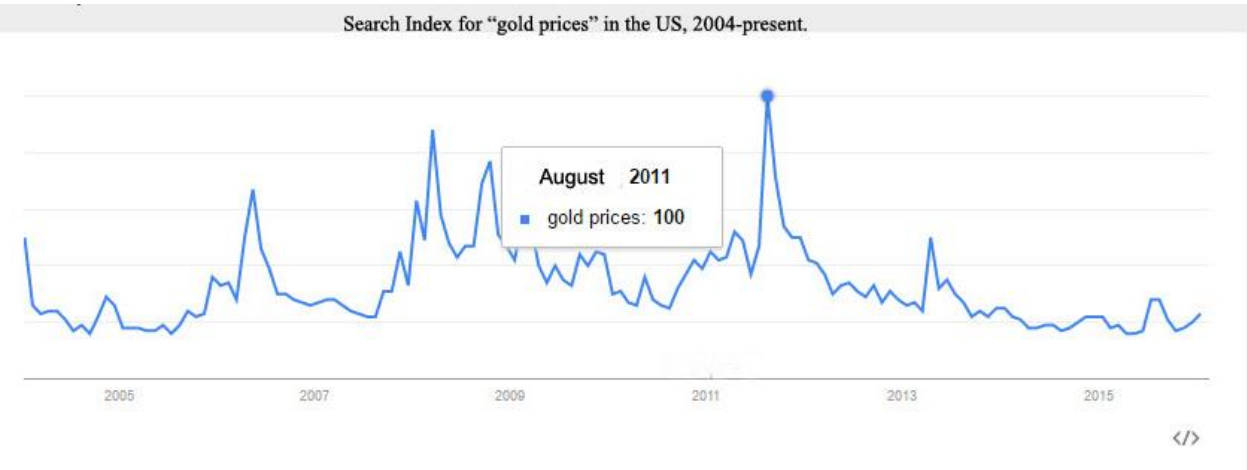
	Title	Author	Keywords	Data	Theme
16	The Sum of All FEARS: Investor Sentiment and Asset Prices	(Da, Engelberg, and Gao 2010)		This paper used data of S&P 500 index daily returns, treasury portfolio returns obtained from CRSP 10-year constant maturity Treasury File, exchange traded funds (ETFs) returns (NYSEARCA: SPY, NASDAQ: QQQQ, NYSEARCA: IWB, NYSEARCA: IWM), CRSP equally-weighted and value-weighted portfolio daily returns, realized market volatility computed using SPY intraday data from TAQ and mutual fund flows derived from Trim Tabs, for the period Jul 2004 to Oct 2009. Other variables used were the CBOE volatility index (VIX) of mutual fund flows, a news-based measure of economic policy uncertainty (EPU) and the Aruoba-Diebold-Scotti (ADS) business conditions index. Finally, the FEARS index was constructed using daily online search volume data from Google Trends result in the US, for the time period Jul-Dec 2009.	This paper explores the potential of internet search volume as a measure of investor sentiment. After a historical, regression-based approach, data for 30 search-terms were selected and with an econometric procedure, were modified and averaged to form a Financial and Economic Attitudes revealed by Search (FEARS) index. Afterwards, the relation of this index with asset prices, volatility and fund flows was examined. The results indicate that the FEARS index predicts short-term returns reversals, temporary increases in volatility and mutual fund flows, with investors switch from equity funds to bond funds after a spike in FEARS.
17	Internet, noise trading and commodity future prices	(Peri, Vandone, and Baldi 2014)	Noise trading, Corn price volatility, Information Mixture Distribution Hypothesis, EGARCH	The paper includes weekly data for internet search volume of the keyword “corn price(s)”, newspapers information for the same keyword and corn futures prices. The data sources are Google Insights, LexisNexis® Academic and CBOT, respectively, from Jan 2004 to Jul 2011.	This paper studies the relationship between internet, noise trading and commodity future prices. The facts reinforce the Mixture Distribution Hypothesis (MDH), which assumes a joint dependence of return volatility and information. Specifically, using an EGARCH model, the study captures the effect of information flows from internet and newspapers on the conditional volatility of corn futures prices. As the results denote, the internet search activity enhances the volatility produced by negative shocks, consistent with the notion that internet search mostly reflects the noise traders’ activity.

	Title	Author	Keywords	Data	Theme
18	Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models	(Agnolucci 2009)	Oil price, GARCH, Implied Volatility (IV)	The data include daily returns from the generic light sweet crude oil future, based on the WTI traded at the NYMEX, obtained the Bloomberg Database. The sample goes from Dec 31, 1991 to May 2, 2005. Finally, the data for the estimation of the volatility implied by the observed price option have been provided by the Commodity Research Bureau and the US Treasury.	This paper compares the predictive efficiency of two approaches, which are used to forecast volatility in crude oil futures, the GARCH-type models (GARCH, EGARCH, APARCH, CGARCH, TGARCH) testing different error distributions and an implied volatility model. Also, the leverage effect is investigated. The parameters for the GARCH-type models have been estimated on a 5-year rolling sample. The article concludes that GARCH-type models outperform the IV model, but the results are improved when a composite estimator is used, combining GARCH and IV forecasts. Additionally to former studies, no leverage effect can be observed.
19	Do Google searches help in nowcasting private consumption? A real-time evidence for the US	(Kholodilin, Podstawski, and Siliverstovs 2010)	Google indicators, real-time nowcasting, principal components, US private consumption	The variables used are the monthly US real private consumption, from the ALFRED® database, the Consumer Sentiment Indicator produced by the University of Michigan, the Consumer Confidence Index by the US Conference Board, the 3-month US Treasury constant maturity rate, the 10-year US Treasury constant maturity rate, the year spread and the S&P500 index, downloaded from the Datastream. Google search data include 220 search time series related with private consumption, provided by Google Insights.	This article examines the predictive power of Google search data on the year-on-year growth rates of monthly US private consumption. The results are compared to those of an benchmark AR(1) model and models including conventional sentiment indicators. The Google indicators were constructed as common factors using principal components analysis. The study shows that the models including Google indicators perform better than the benchmark model, as well as the augmented models with the alternative sentiment indicators.

20	Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends	(Vosen and Schmidt 2011)	Google Trends, private consumption, forecasting, consumer sentiment indicators	The data obtained from Google Trends are the year-on-year growth rates of the aggregated indices of search queries from the selected 56 consumption-relevant categories. Also, the survey-based indicators are the University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index.	The article compares the predictive power of two common survey-based indicators to Google indicators, which consist of a number of extracted common factors. The results indicate that the performance of the Google-indicators model, including the first four factors, is better than the benchmark and the alternative models, in-sample and out-of-sample.

3. Google Trends

Google Trends provides a time series index, we can download as a CSV file, that represents the search volume of a keyword users enter into Google, for a given time period and geographic area. This query index, or SVI (Search Volume Index), is generated by the total query volume for a keyword, within a specific geographic region, divided by the total number of the queries in that region during the selected time period, for a random sample of individual users. So the data are scaled from 0 to 100, with the maximum query volume normalized to be 100 and 0 for insufficient data[1]. Queries with the same words in different sequence or different spelling produce separate SVI series. The trends data go back to January 1, 2004 at a weekly frequency[2]. The graph for a search term shows the term's popularity over time in (nearly) real time.



[1] The zero value does not indicate zero searches for the particular keyword, it suggests a very low number of searches.

[2] The data can also be available at daily frequency for a sample period less than or equal to a quarter, at http://www.google.com/trends/explore .

4. Data and Methodology

We try to capture the information from the Google search queries in two ways, constructing an index and through principal components analysis, using factors.

4.1 Construction of the Google Index

Our aim is to build a list of sentiment-reveal search terms towards gold and oil. The word "gold" is considered a *positive* word according to the Harvard IV-4 Dictionary that categorizes words as "positive", "negative", "strong", "weak" and so on. At first, we choose some primary keywords such as "gold", "gold price", "gold rate", "gold stock", "spot gold", "COMEX GOLD", "oil price", "oil stock", etc.

In order to understand how these terms might be searched in Google by individuals, we input each one in Google Trends which, among other results, returns ten "top searches" related to each term. For example, a search for "gold price" results in related searches "price of gold", "gold price today", "gold price India", "price for gold", "gold rate", etc. Also, we used the top 30 or 40 related searches from the Google Correlate[3]. This generates approximately 260 related words for each of the primary gold and oil search terms.

Next, we remove duplicates, terms with inadequate data (query series with zero search volume) and terms with no economic relation. For example, a search for "gold" results in related searches "how much gold", "grams of gold", "1 ounce", "youtube gold", "olympic bar", "green nike", etc. We keep the first three terms and remove the others. This leaves us with the final 61 gold-related and 49 oil-related search terms.

[3] Google Correlate is an application that finds search patterns, using a correlate algorithm returns the related searches to every query request users enter into Google search engine by geographic region. http://www.google.com/trends/correlate

We downloaded the weekly SVI for each of these terms over our sample period of 10/3/2004 – 10/26/2014 from Google Trends. We restrict the SVI results to the US.

An important feature of the Google Trends data is the sampling method that imports a measurement error into the outcome series for every query. This is because of the fact that requests for the same query on different days return slightly different results. So, to identify any irregularities in the data, we downloaded daily the SVI series for each keyword, for 60 days, following Carrière-Swallow and Labbé (2013). Figure 1 plots the cross-sectional mean for the keyword "gld" [4] across the 60 samples. The correlation of the samples for each term is above 99 percent. We assume that this sampling error has a small effect, and we use the cross-sectional mean of every keyword at time *t* for the Google Index construction. The resulting time series is used as the historical time series for each keyword. Figures 2&3 present the histograms of the cross-section mean series of "gld" and the mean-centered series "gld" for a random sample.

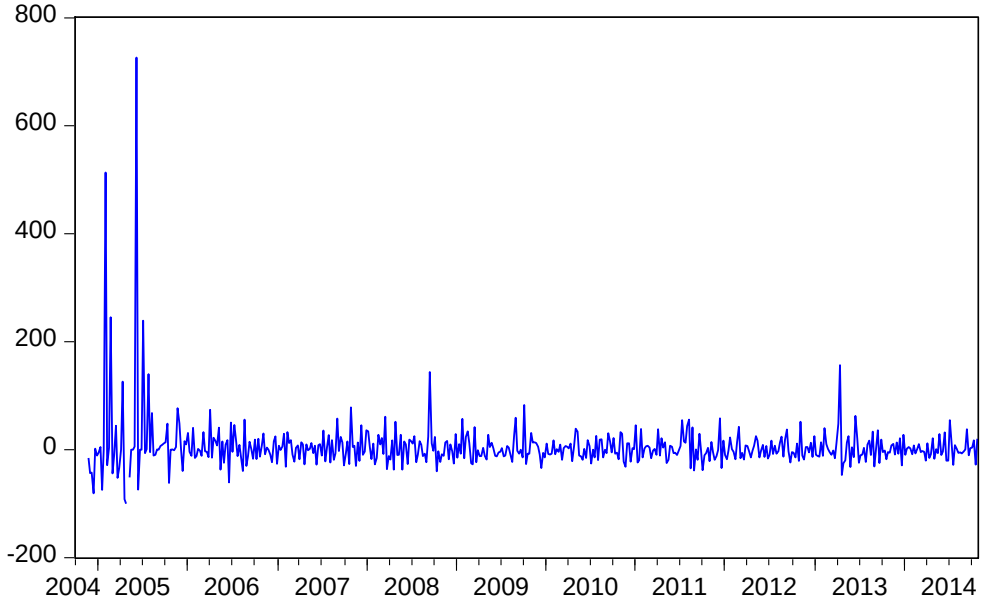
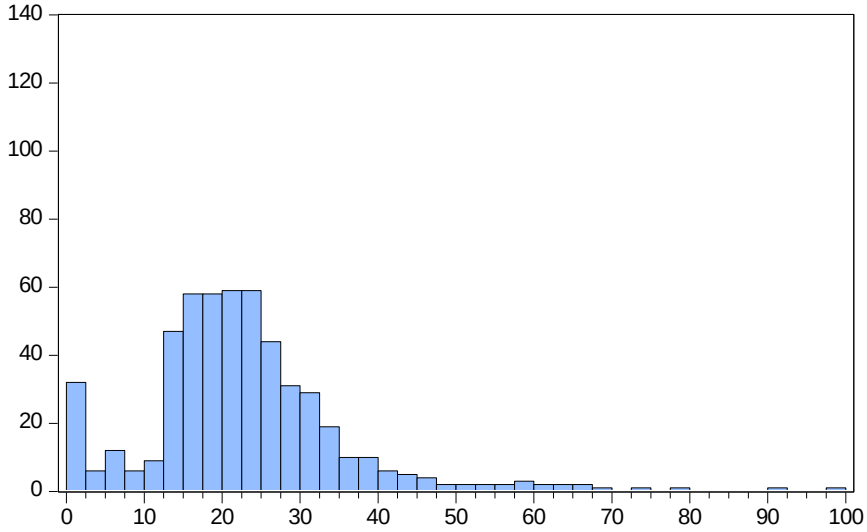


Figure 1. % Change cross-sectional mean for the keyword "gld"

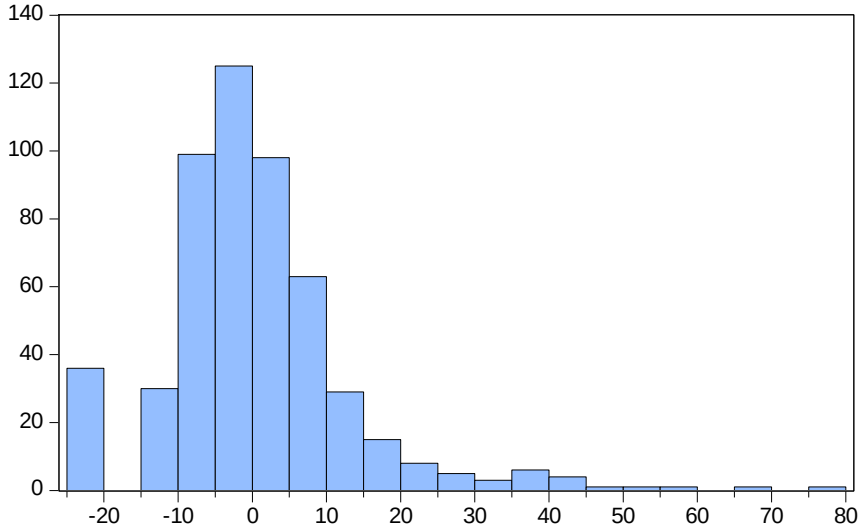
[4] For illustration proposes we show the results only for the keyword "gld". The conclusion is similar for the rest of the keywords.



Series: MEAN_GLD
Sample 10/03/2004 10/26/2014
Observations 526

Mean	22.79534
Median	21.41667
Maximum	99.98333
Minimum	0.000000
Std. Dev.	12.87689
Skewness	1.520515
Kurtosis	8.557901
Jarque-Bera	879.6939
Probability	0.000000

Figure 2. Histogram of the cross-section mean time series for the keyword "gld"



Series: S_GLD56
Sample 10/03/2004 10/26/2014
Observations 526

Mean	1.29e-14
Median	-1.891635
Maximum	77.10837
Minimum	-22.89163
Std. Dev.	13.01917
Skewness	1.522711
Kurtosis	8.549074
Jarque-Bera	878.1310
Probability	0.000000

Figure 3. Histogram of the mean-centered keyword "gld", sample#56

We keep only the SVIs with non-zero observations [5]and set the weekly change in search term ***i***

as: $\mathrm{DLSVI}_{i,t}$=$\ln(\mathrm{SVI}_{i,t})$-$\ln(\mathrm{SVI}_{i,t-1})$

At next, we plot two examples of weekly change for the keywords "gold" and "price of gold", during 2004-2006. Figures 4, 5 indicate some significant characteristics of the data. The difference in variance is obvious in Figure 4 and we can observe the presence of extreme values in Figure 5. Additionally, we run seasonality tests and identify seasonal patterns. So we adjust the data in the following way as in Da, Engelberg, and Gao (2010):

At first, we winsorize each of the DLSVIs at the 5% level. Then, we seasonally adjust the series by regressing DLSVIi,t on monthly dummies and keep the residual. Finally, to make each time-series comparable, we standardize them by subtracting the mean and divide the difference by the standard deviation. This results in an adjusted (winsorized, standardized and seasonally adjusted) weekly change of SVI, adj_DLSVI (adjusted DLSVI), for each search-term.

In order to specify which search terms are most significant for the market returns (gold market and oil market returns respectively), we run recursive rolling regressions of the adjusted DLSVI on market returns for every term separately, with an expanding window of 100 observations (2 years) and a step of 50 observations. We define the Google Index as:

$$\text{Google Index}_t = \sum_{\mathrm{j}} \mathrm{R}^j(\text{adj\_DLSVI}_t)$$

Where $\mathrm{R}^j(\text{adj\_DLSVI}_t)$ is the adj_DLSVI series for the search term which had t-statistic rank of j from the period October 2004 through the most recent 1-year period, with ranks run from smallest, j=1, to largest, j=20, for gold-terms and oil-terms.

[5] The list of the final search-terms that we use for the construction of the Index and the keyword samples for the Factors approach are available in the Appendix.

i.e. for a starting window of 100 observations, during the period October 3, 2004 - August 27, 2006, we run a recursive rolling regression of adj_DLSVI on contemporaneous market return for each of our terms.

Then we rank the t-statistics from every rolling outcome from smallest to largest. We select the most significant terms and use them to form our index for the next 50-observation period (1 year). For example, the regression during the period Oct 3, 2004 – Aug 27, 2006 gives us three significant search terms for gold: "gold prices", "gold price" and "price gold". We use their observations for the period September 3, 2006 – August 12, 2007. Google Index on week *t* during this period is simply the average adj_DLSVI of these three terms on week *t*. Finally, due to the need for an initial window, our Google Index begins in September 3, 2006[6].

[6] The full data sample for the index approach includes 426 observations and 526 observations for the factors approach for gold and crude oil data, respectively.

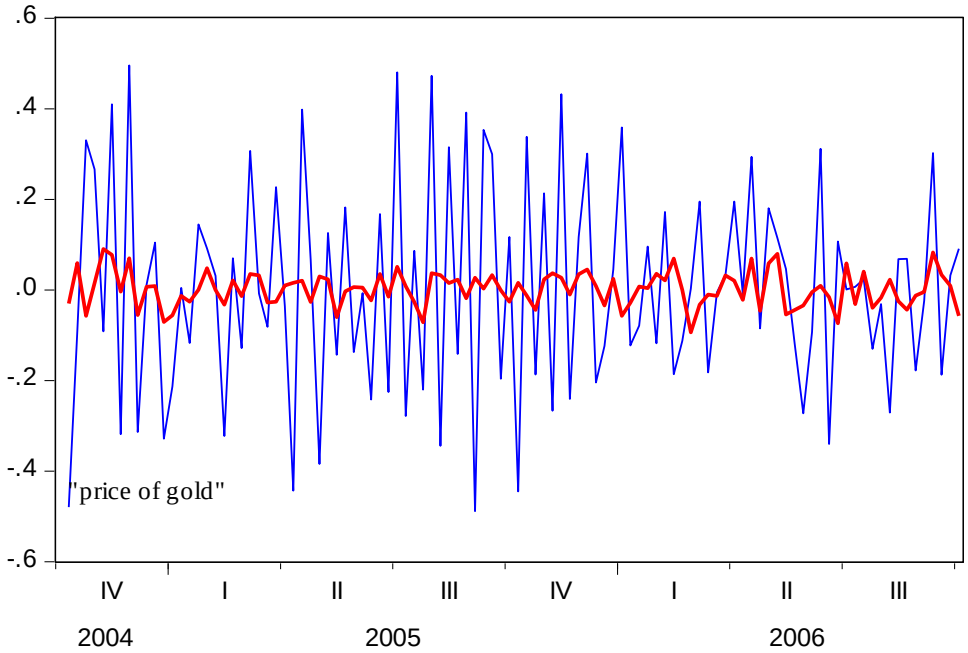


Figure 4. SVI change for the keywords "gold" and "price of gold"

"price of gold" "gold"

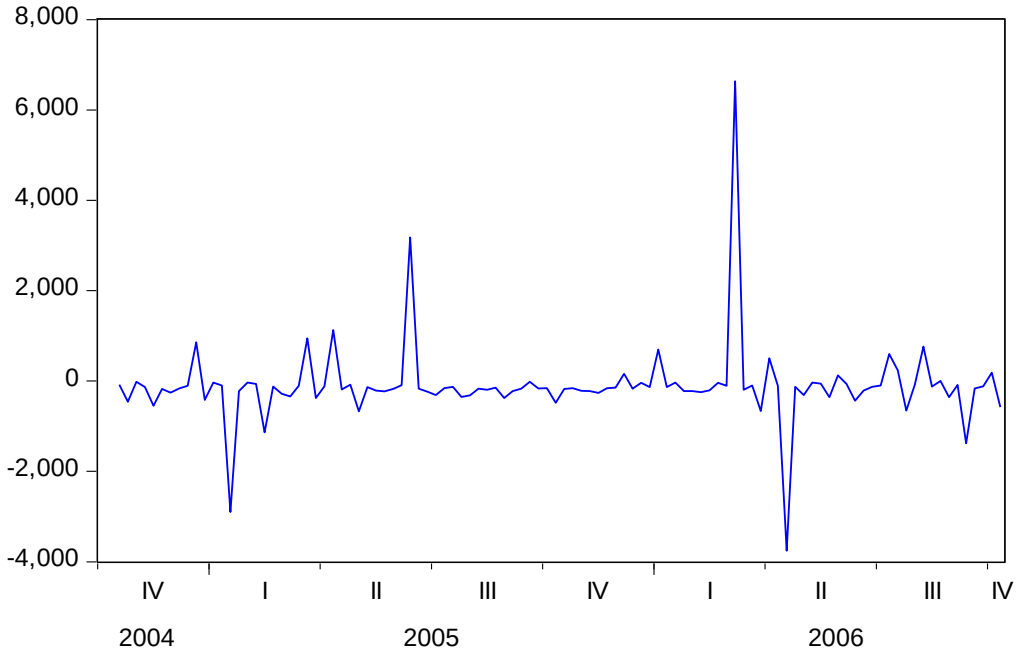


Figure 5. % SVI change for the keyword "price of gold"

4.2 Factors

Another approach is to use the large number of the search-volume series in principal components analysis and generate a reduced number of common factors from the selected keywords[7]. In order to choose the number of factors we use two criteria, the Kaiser criterion indicates that we can retain only factors with eigenvalues greater than 1. For example, Table 1 and Table 2 suggest that we would retain two factors from the 20 gold and oil related keywords. We can also conclude this graphically. Figures 6-9 plot the cumulative proportion of the eigenvalues in the total variance of the data sets. At next, we obtain the factors from every data set. The first factor, F1, accounts for the largest percentage of the common variance[8], while the other factors, account for the remaining proportion. Tables 3&4 present the factors for the 20 gold and oil terms, while Tables 7&8 for the 61 gold terms and 49 oil terms, respectively.

Eigenvalues: (Sum = 20, Average = 1)

Number	Value	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	12.27719	7.859023	0.6139	12.27719	**0.6139**
2	4.418171	3.708950	0.2209	16.69537	**0.8348**
3	0.709221	0.105239	0.0355	17.40459	0.8702
4	0.603982	0.192687	0.0302	18.00857	0.9004
5	0.411295	0.072042	0.0206	18.41986	0.9210
6	0.339253	0.060281	0.0170	18.75912	0.9380
7	0.278972	0.064586	0.0139	19.03809	0.9519
8	0.214386	0.024879	0.0107	19.25247	0.9626
9	0.189508	0.041899	0.0095	19.44198	0.9721
10	0.147609	0.046587	0.0074	19.58959	0.9795
11	0.101022	0.007587	0.0051	19.69061	0.9845
12	0.093435	0.026723	0.0047	19.78405	0.9892
13	0.066712	0.009324	0.0033	19.85076	0.9925
14	0.057388	0.011000	0.0029	19.90815	0.9954
15	0.046387	0.013662	0.0023	19.95454	0.9977
16	0.032726	0.019997	0.0016	19.98726	0.9994
17	0.012729	0.012722	0.0006	19.99999	1.0000
18	6.37E-06	2.48E-06	0.0000	20.00000	1.0000
19	3.89E-06	3.81E-06	0.0000	20.00000	1.0000
20	8.26E-08	---	0.0000	20.00000	1.0000

Table 1. Eigenvalues for the 20 gold-related keywords

[7] The data include the full keyword samples and the 20 non-zero volume keyword series for gold and oil.
[8] The proportion of variance that each term has in common with the other terms.

Eigenvalues: (Sum = 20, Average = 1)

Number	Value	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	15.47421	13.37391	0.7737	15.47421	**0.7737**
2	2.100305	1.237103	0.1050	17.57452	**0.8787**
3	0.863202	0.302882	0.0432	18.43772	0.9219
4	0.560320	0.258385	0.0280	18.99804	0.9499
5	0.301935	0.099377	0.0151	19.29997	0.9650
6	0.202558	0.050805	0.0101	19.50253	0.9751
7	0.151753	0.018441	0.0076	19.65428	0.9827
8	0.133312	0.062429	0.0067	19.78760	0.9894
9	0.070882	0.012224	0.0035	19.85848	0.9929
10	0.058658	0.028571	0.0029	19.91714	0.9959
11	0.030087	0.010959	0.0015	19.94722	0.9974
12	0.019129	0.006991	0.0010	19.96635	0.9983
13	0.012138	0.000320	0.0006	19.97849	0.9989
14	0.011818	0.005840	0.0006	19.99031	0.9995
15	0.005978	0.003942	0.0003	19.99629	0.9998
16	0.002036	0.001040	0.0001	19.99832	0.9999
17	0.000996	0.000316	0.0000	19.99932	1.0000
18	0.000679	0.000678	0.0000	20.00000	1.0000
19	1.47E-06	5.13E-07	0.0000	20.00000	1.0000
20	9.61E-07	---	0.0000	20.00000	1.0000

Table 2. Eigenvalues for the 20 oil-related keywords

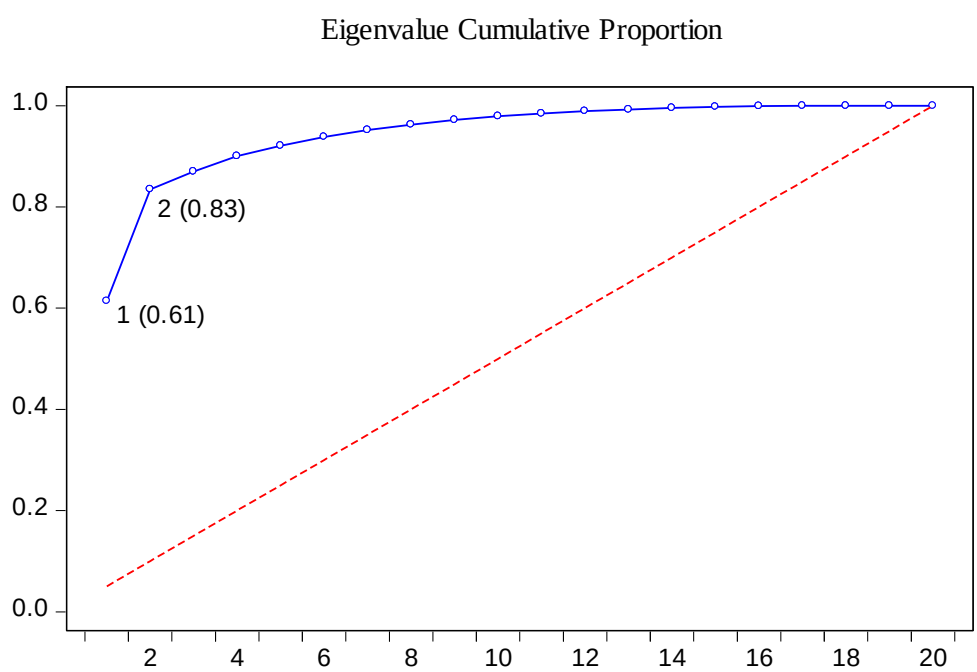


Figure 6.

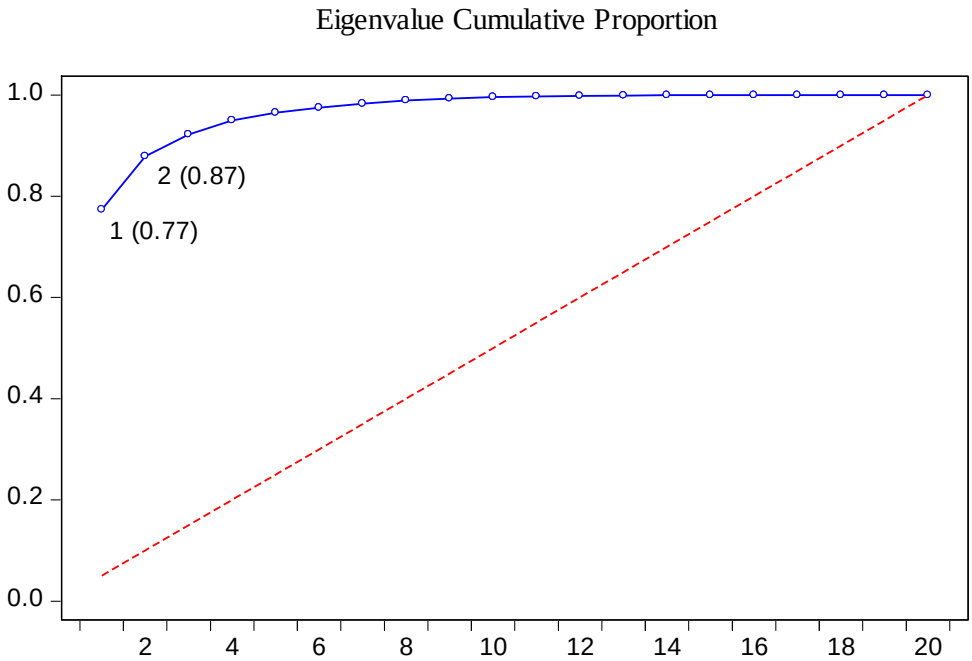


Figure 7. 20 oil-related keywords

Factor	Variance	Cumulative	Difference	Proportion	Cumulative
F1	12.12194	12.12194	7.891349	0.741288	0.741288
F2	4.230591	16.35253	---	0.258712	1.000000
Total	16.35253	16.35253		1.000000	

Table 3. Factors from the 20 gold-related keywords. Factor Method: Unweighted Least Squares, Number of Factors: Kaiser-Guttman

Factor	Variance	Cumulative	Difference	Proportion	Cumulative
F1	15.38262	15.38262	13.55481	0.893796	0.893796
F2	1.827811	17.21043	---	0.106204	1.000000
Total	17.21043	17.21043		1.000000	

Table 4. Factors from the 20 oil-related keywords

Eigenvalues: (Sum = 49, Average = 1)

Number	Value	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	37.31033	33.56703	0.7614	37.31033	**0.7614**
2	3.743298	1.492610	0.0764	41.05363	**0.8378**
3	2.250688	1.281021	0.0459	43.30432	**0.8838**
4	0.969667	0.198156	0.0198	44.27398	0.9036
5	0.771511	0.209246	0.0157	45.04549	0.9193
6	0.562264	0.132260	0.0115	45.60776	0.9308
7	0.430005	0.075731	0.0088	46.03776	0.9395
8	0.354273	0.073815	0.0072	46.39204	0.9468
9	0.280458	0.014656	0.0057	46.67249	0.9525
10	0.265802	0.015206	0.0054	46.93830	0.9579
11	0.250597	0.043531	0.0051	47.18889	0.9630
12	0.207066	0.023768	0.0042	47.39596	0.9673
13	0.183298	0.039484	0.0037	47.57926	0.9710
14	0.143813	0.005595	0.0029	47.72307	0.9739
15	0.138218	0.014806	0.0028	47.86129	0.9768
16	0.123412	0.015742	0.0025	47.98470	0.9793
17	0.107670	0.010195	0.0022	48.09237	0.9815
18	0.097475	0.007324	0.0020	48.18985	0.9835
19	0.090151	0.005886	0.0018	48.28000	0.9853
20	0.084265	0.005294	0.0017	48.36426	0.9870
21	0.078972	0.008147	0.0016	48.44323	0.9886
22	0.070825	0.007669	0.0014	48.51406	0.9901
23	0.063155	0.008961	0.0013	48.57721	0.9914
24	0.054195	0.005705	0.0011	48.63141	0.9925
25	0.048490	0.002986	0.0010	48.67990	0.9935
26	0.045504	0.000503	0.0009	48.72540	0.9944
27	0.045000	0.005611	0.0009	48.77040	0.9953
28	0.039389	0.003476	0.0008	48.80979	0.9961
29	0.035914	0.003471	0.0007	48.84571	0.9969
30	0.032443	0.002043	0.0007	48.87815	0.9975
31	0.030400	0.005942	0.0006	48.90855	0.9981
32	0.024458	0.008281	0.0005	48.93301	0.9986
33	0.016177	0.005116	0.0003	48.94918	0.9990
34	0.011061	0.001430	0.0002	48.96024	0.9992
35	0.009631	0.002054	0.0002	48.96987	0.9994
36	0.007577	0.000279	0.0002	48.97745	0.9995
37	0.007298	0.001685	0.0001	48.98475	0.9997
38	0.005613	0.002091	0.0001	48.99036	0.9998
39	0.003521	0.000633	0.0001	48.99388	0.9999
40	0.002888	0.001066	0.0001	48.99677	0.9999
41	0.001821	0.001009	0.0000	48.99859	1.0000
42	0.000813	0.000262	0.0000	48.99940	1.0000
43	0.000551	0.000516	0.0000	48.99996	1.0000
44	3.52E-05	2.90E-05	0.0000	48.99999	1.0000
45	6.20E-06	4.93E-06	0.0000	49.00000	1.0000
46	1.27E-06	3.09E-07	0.0000	49.00000	1.0000
47	9.62E-07	1.28E-07	0.0000	49.00000	1.0000
48	8.34E-07	8.34E-07	0.0000	49.00000	1.0000
49	1.64E-16	---	0.0000	49.00000	1.0000

Table 5. Eigenvalues for the 49 oil-related keywords

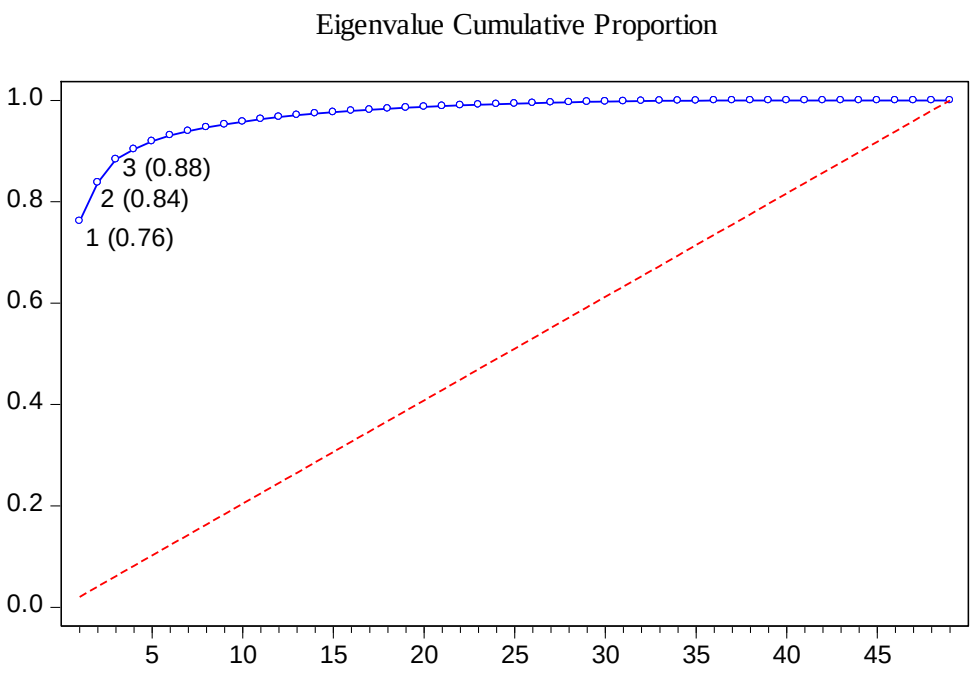


Figure 8. 49 oil-related keywords

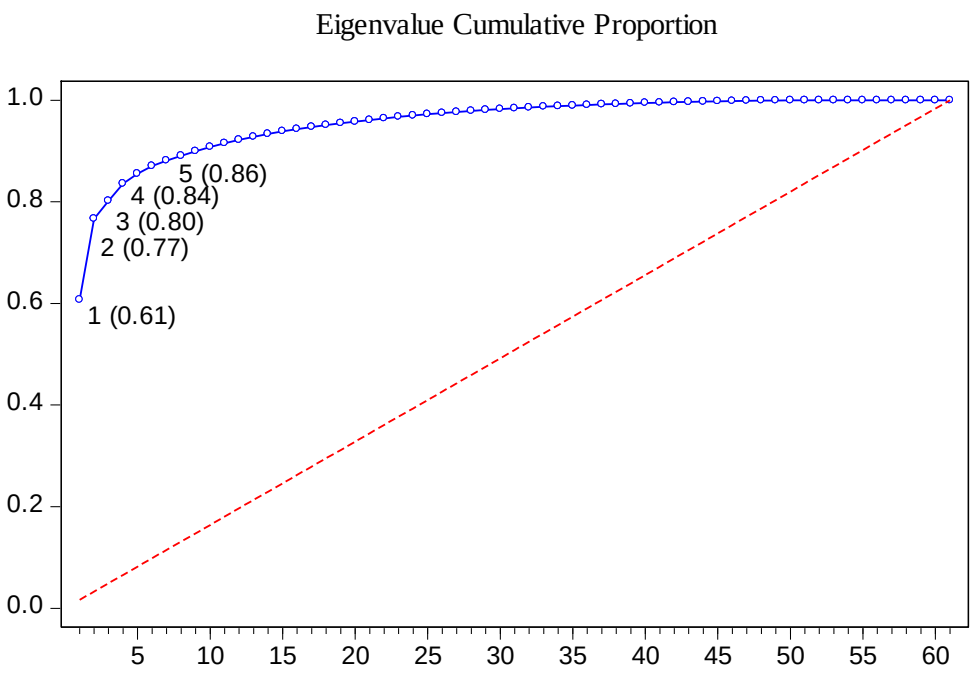


Figure 9. 61 gold-related keywords

Eigenvalues: (Sum = 61, Average = 1)

Number	Value	Difference	Proportion	Cumulative Value	Cumulative Proportion
1	37.04873	27.35548	0.6074	37.04873	**0.6074**
2	9.693249	7.535231	0.1589	46.74198	**0.7663**
3	2.158018	0.075850	0.0354	48.90000	**0.8016**
4	2.082169	0.890012	0.0341	50.98217	**0.8358**
5	1.192156	0.269702	0.0195	52.17432	**0.8553**
6	0.922454	0.284556	0.0151	53.09678	0.8704
7	0.637899	0.050398	0.0105	53.73468	0.8809
8	0.587501	0.045658	0.0096	54.32218	0.8905
9	0.541843	0.030427	0.0089	54.86402	0.8994
10	0.511416	0.052727	0.0084	55.37544	0.9078
11	0.458689	0.066380	0.0075	55.83413	0.9153
12	0.392309	0.020632	0.0064	56.22644	0.9217
13	0.371678	0.021973	0.0061	56.59811	0.9278
14	0.349705	0.038122	0.0057	56.94782	0.9336
15	0.311583	0.034786	0.0051	57.25940	0.9387
16	0.276798	0.030601	0.0045	57.53620	0.9432
17	0.246197	0.013736	0.0040	57.78240	0.9473
18	0.232461	0.017105	0.0038	58.01486	0.9511
19	0.215356	0.016457	0.0035	58.23021	0.9546
20	0.198898	0.002131	0.0033	58.42911	0.9579
21	0.196767	0.007331	0.0032	58.62588	0.9611
22	0.189437	0.013116	0.0031	58.81531	0.9642
23	0.176320	0.004252	0.0029	58.99164	0.9671
24	0.172068	0.025577	0.0028	59.16370	0.9699
25	0.146491	0.011838	0.0024	59.31019	0.9723
26	0.134653	0.004292	0.0022	59.44485	0.9745
27	0.130360	0.008416	0.0021	59.57521	0.9766
28	0.121944	0.004439	0.0020	59.69715	0.9786
29	0.117505	0.017683	0.0019	59.81466	0.9806
30	0.099822	0.001245	0.0016	59.91448	0.9822
31	0.098577	0.003223	0.0016	60.01306	0.9838
32	0.095354	0.004378	0.0016	60.10841	0.9854
33	0.090976	0.013037	0.0015	60.19939	0.9869
34	0.077940	0.002456	0.0013	60.27733	0.9882
35	0.075484	0.006762	0.0012	60.35281	0.9894
36	0.068721	0.001662	0.0011	60.42153	0.9905
37	0.067059	0.006410	0.0011	60.48859	0.9916
38	0.060649	0.001595	0.0010	60.54924	0.9926
39	0.059054	0.005262	0.0010	60.60829	0.9936
40	0.053792	0.006456	0.0009	60.66209	0.9945
41	0.047336	0.001373	0.0008	60.70942	0.9952
42	0.045964	0.005302	0.0008	60.75539	0.9960
43	0.040662	0.003406	0.0007	60.79605	0.9967
44	0.037256	0.006965	0.0006	60.83330	0.9973
45	0.030291	0.002004	0.0005	60.86360	0.9978
46	0.028287	0.002138	0.0005	60.89188	0.9982
47	0.026149	0.002980	0.0004	60.91803	0.9987
48	0.023170	0.005590	0.0004	60.94120	0.9990
49	0.017580	0.002460	0.0003	60.95878	0.9993
50	0.015120	0.003841	0.0002	60.97390	0.9996
51	0.011279	0.002879	0.0002	60.98518	0.9998
52	0.008400	0.002234	0.0001	60.99358	0.9999
53	0.006167	0.006020	0.0001	60.99975	1.0000
54	0.000147	9.17E-05	0.0000	60.99990	1.0000

Table 6. Eigenvalues for the 61 gold-related keywords

55	5.50E-05	3.26E-05	0.0000	60.99995	1.0000
56	2.23E-05	1.08E-05	0.0000	60.99997	1.0000
57	1.15E-05	4.20E-06	0.0000	60.99998	1.0000
58	7.28E-06	2.06E-06	0.0000	60.99999	1.0000
59	5.22E-06	1.71E-06	0.0000	61.00000	1.0000
60	3.51E-06	3.43E-06	0.0000	61.00000	1.0000
61	7.71E-08	---	0.0000	61.00000	1.0000

Table 6. Eigenvalues for the 61 gold-related keywords (*cont.*)

Factor	Variance	Cumulative	Difference	Proportion	Cumulative
F1	36.91387	36.91387	27.37941	0.719614	0.719614
F2	9.534461	46.44834	7.517108	0.185869	0.905483
F3	2.017353	48.46569	0.156604	0.039327	0.944810
F4	1.860749	50.32644	0.890434	0.036274	0.981084
F5	0.970314	51.29675	---	0.018916	1.000000
Total	51.29675	51.29675		1.000000	

Table 7. Factors for the 61 gold-related keywords

Factor	Variance	Cumulative	Difference	Proportion	Cumulative
F1	37.20835	37.20835	33.65929	0.869319	0.869319
F2	3.549060	40.75741	1.504768	0.082919	0.952238
F3	2.044292	42.80171	---	0.047762	1.000000
Total	42.80171	42.80171		1.000000	

Table 8. Factors for the 49 oil-related keywords

4.3 Other Data

For the weeks ending on Sunday between October 3, 2004 and October 26, 2014, we downloaded from the website of the Federal Reserve Economic Data-FRED-St. Louis Fed [9]the aggregated weekly spot price data for gold: Gold Fixing Price 10:30 A.M., London time, in London Bullion Market, based in US dollars per troy ounce. Also, the weekly price data for oil are collected from the US Energy Information Administration [10]and include the crude oil prices: West Texas Intermediate (WTI), based in US dollars per barrel.

Weekly rates of return are calculated by taking the weekly nominal percentage return series, i.e. $\mathrm{r}_t = 100 \ln\left(\frac{\mathrm{P}_t}{\mathrm{P}_{t-1}}\right)$, where P_t is the gold and crude oil price at time *t*. Table 9 reports the descriptive statistics for the price return data; the excess kurtosis and the Jarque-Bera probability indicate that the returns exhibit a significant departure from normality, suggesting a GARCH approach. Additionally, the ARCH test on the lagged squared residuals rejects the null hypothesis of 'no ARCH effects', implying heteroskedasticity in the returns.

[9] The FRED data are available at https://research.stlouisfed.org/fred2/
[10] Independent Statistics and Analysis, US Energy Information Administration http://www.eia.gov

Descriptive Statistics: gold and oil price data		
	gold price returns	crude oil price returns
Mean	0.210111	0.095614
Median	0.413061	0.247219
Maximum	8.745335	25.12470
Minimum	-11.49791	-19.09960
Std. Dev.	2.255860	4.142654
Skewness	-0.637040	-0.189216
Kurtosis	6.384873	7.540376
Jarque-Bera	286.1391	454.0862
Probability	0.000000	0.000000
Sum	110.3082	50.19755
Sum Sq. Dev.	2666.587	8992.668
Observations	525	525

Table 9. Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as $\chi^2(2)$.The 5% critical value is 5.99

Figures 10&11 show the gold and oil price returns and the first factors, F1, from the 20-keywords sample, F1 (20), and the full keyword sample for gold and oil, respectively (F1 (61) for gold and F1 (49) for oil). Figures 12&13 present the gold and oil price returns with the Google Indexes. The graphs reveal volatility clustering in the returns.

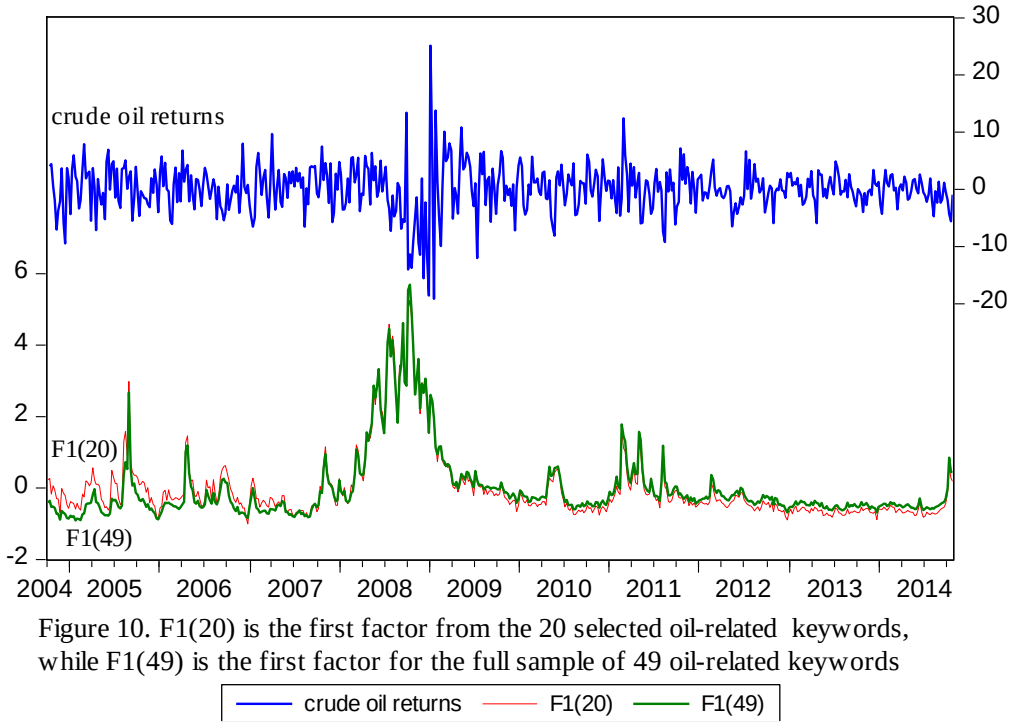


Figure 10. F1(20) is the first factor from the 20 selected oil-related keywords, while F1(49) is the first factor for the full sample of 49 oil-related keywords

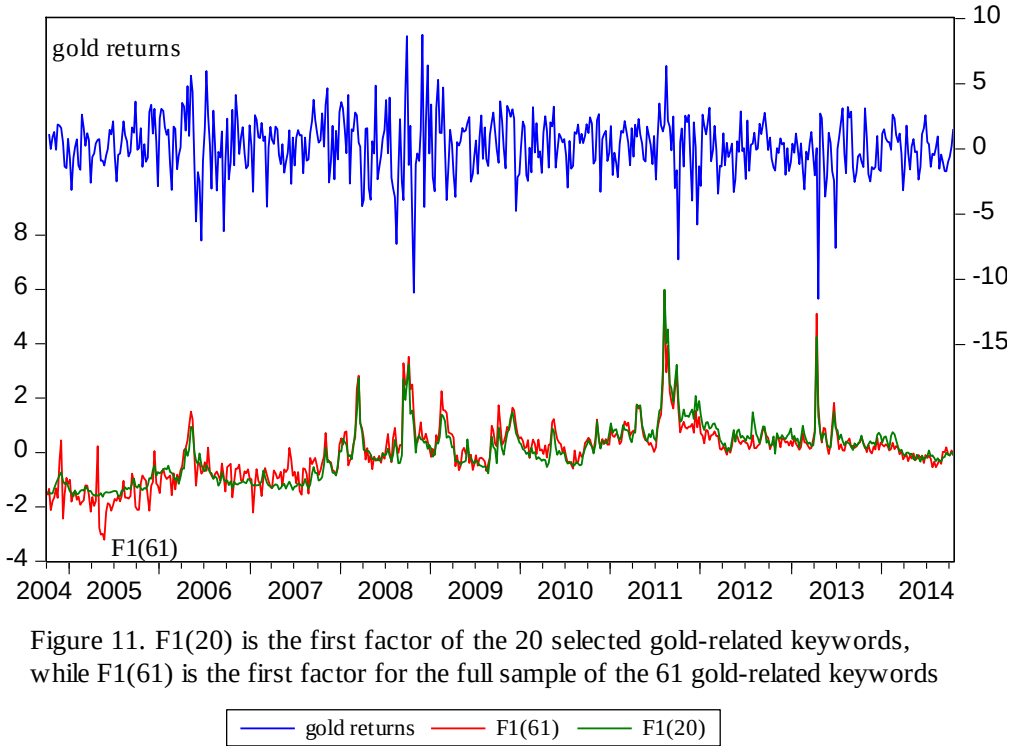


Figure 11. F1(20) is the first factor of the 20 selected gold-related keywords, while F1(61) is the first factor for the full sample of the 61 gold-related keywords

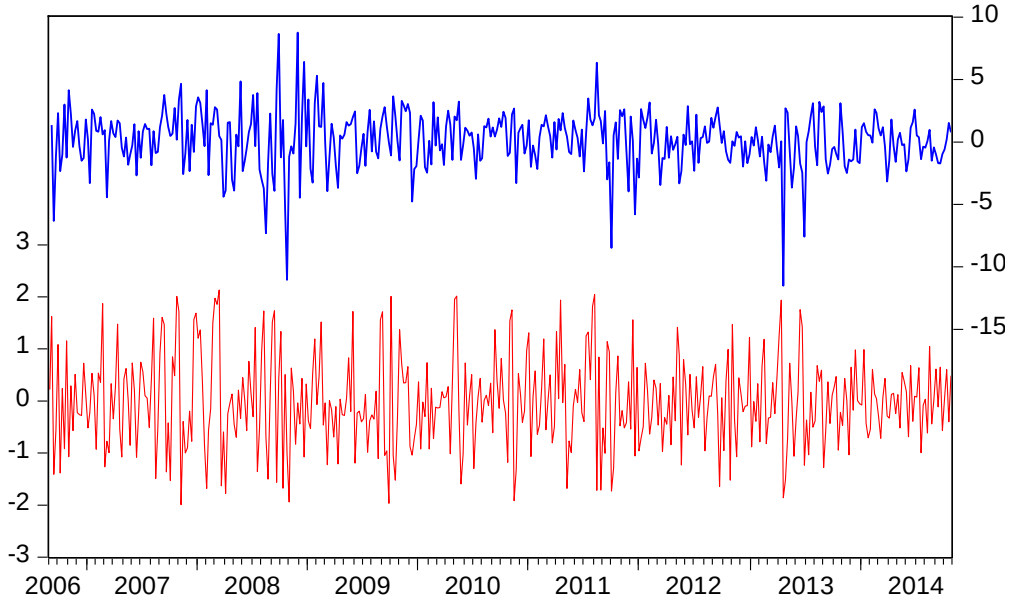


Figure 12.

gold returns — Google Index-gold

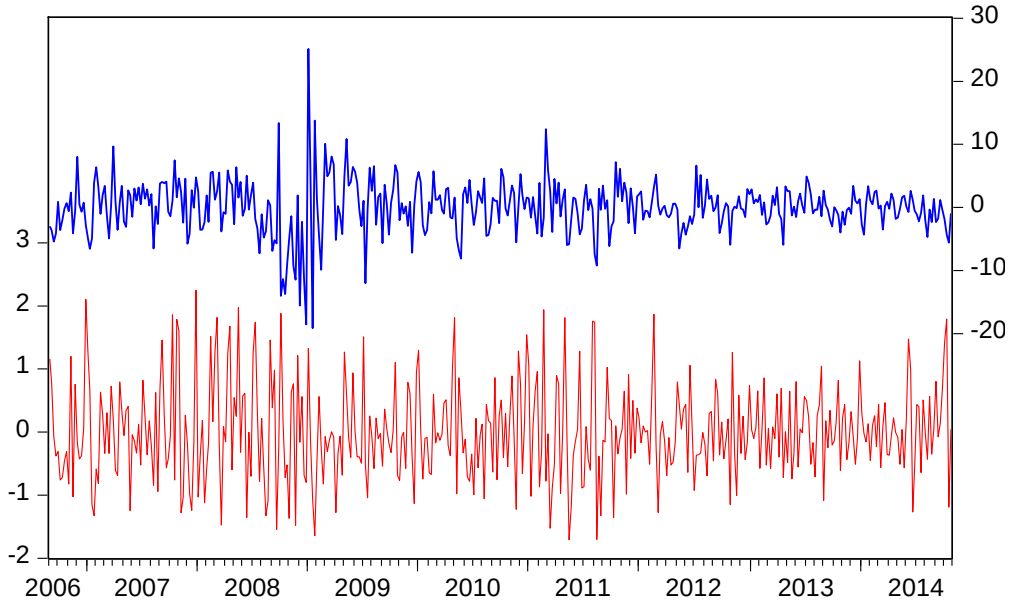


Figure 13.

crude oil returns — Google Index-oil

4.4 The Model

In our empirical application we model the volatility of gold and oil price returns in a GARCH framework. In order to capture conditional volatility and leverage effects in the presence of information impacts, that we will refer to as asymmetric effects in our analysis (Bollerslev 2008); we use an exponential GARCH (EGARCH) assuming normally distributed errors[11] on the full sample, a subsample for the gold price returns and we also estimate the model parameters on a one-year rolling window (50 observations) which is shifted forwards by one week (1 observation).

The EGARCH (1, 1) process can be augmented as follows:

$$y_t = \mathrm{a} + \mathrm{b}y_{t-1} + \varepsilon_t$$

$$\log(\sigma_t^2) = \omega + \alpha\left[\left|\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right|\right] + \beta\log(\sigma_{t-1}^2) + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \lambda \mathrm{x}_t$$

Where σ^2 is the conditional variance, ε_t the error term in the mean equation and α, β, γ, λ are the constant parameters. The α parameter represents the symmetric effect of lagged shocks on the volatility, while the γ coefficient accounts for the asymmetric effect on the volatility; negative values of γ indicate that negative shocks ("bad news") are more destabilizing than positive shocks. If there is a symmetric effect, we expect statistically insignificant γ, indicating that positive and negative return shocks have the same magnitude of impact in volatility. The persistence of shocks in conditional volatility is given by β. Finally, x_t is a variable representing the sentiment indicator. Two alternative representations for x_t can be adopted: the Google Indexes, or the Factors from gold and oil-related keywords, respectively.

[11] The EGARCH(1,1) model on gold returns and the augmented model with the Google Index are estimated under the GED distribution with fixed parameter, as this provides better statistical results compared to the normal distribution (Panagiotidis, Bampinas, and Ladopoulos n.d.).

For the Factors approach in our model we set as x_t variable either the first factor, F1, or combinations of them for the different keyword samples.

5. Empirical Results

This section describes the estimation results of the EGARCH (1, 1) model for the gold and crude oil market, separately. The main findings are presented in Tables 10-19. The tables include the coefficient estimates and the statistic results of the model parameters for the full sample and the rolling estimations as a percentage of significance, for the Google Index and Factors approach. In model selection, Akaike and Schwarz information criteria are used for the models performance evaluation.

5.1 Empirical Results for the Oil market

We begin presenting our results considering the full sample period (Table 10). For both of our models, the benchmark and the augmented EGARCH with the Google Index, we observe that the γ coefficient is negative and statistically significant at the 1 per cent level, indicating asymmetric effect in the crude oil returns. This suggests that shocks have asymmetric effect on the volatility of crude oil prices. The α and β parameters are positive and statistically significant, with β very close to 1, suggesting that shocks to crude oil price volatility tend to persist. Finally, when we include the variable of Google Index as a regressor in the variance equation we notice that is highly significant, with Akaike and Schwarz criteria taking lower values in favor of the augmented model. For the heteroskedasticity test, the ARCH-LM test with 12 lags show that the EGARCH models seem to model the dependence in the conditional volatility.

Table 11 presents the percentage of statistically significant values of γ coefficient for the rolling regression results for each model at 0.01, 0.05 and 0.1 levels of significance. We can see that the proportion of the significant values of γ is decreasing when we include the Google Index

in the variance equation, for all the levels of significance, suggesting that the sentiment variable can overlap the factor of the asymmetric effect in the crude oil price volatility. Figure 14 plots the AIC criterion values and the *p*-values of the γ coefficient for the two rolling models and Figure 15 presents the z-statistic values of the λ coefficient in a scale of 0 to 3 indicating the levels of significance.

EGARCH(1,1) results for the crude oil returns	ω	γ	α	β	λ	Diagnostic test: ARCH-LM
without variance parameters	-0.082800**	-0.110825***	0.171800***	0.980080***		9.052283
	[-2.484]	[-3.896]	[4.846]	[102.126]		(*0.698500*)
oil-Google Index	-0.055797***	-0.100272***	0.059943**	1.004093***	0.275870***	6.304328
	[-2.698]	[-4.422]	[2.147]	[241.798]	[8.868]	(*0.900000*)

Information criteria	AIC	SC
without variance parameters	5.322358	5.379564
oil-Google Index	5.256615	5.323356

Table 10. The parameters of the variance equation for the benchmark and augmented model for the full sample (426 observations). Z-statistics in brackets, prob. Chi-Square (12) in *italics*. *, ** and *** denote statistical significance at 0.10, 0.05 and 0.01 level respectively. Information criteria Akaike and Schwarz.

Models	Benchmark EGARCH (1,1)			Augmented EGARCH (1,1)		
Levels of significance	1%	5%	10%	1%	5%	10%
Number of statistically significant values of γ	39	72	132	16	30	64
Percent %	10.37234	19.14894	35.10638	4.255319	7.978723	17.02128

Table 11. Percentage of the statistically significant values of γ for the 376 rolling outcomes from the rolling benchmark and the augmented EGARCH (1, 1) with the oil-Google Index variance regressor.

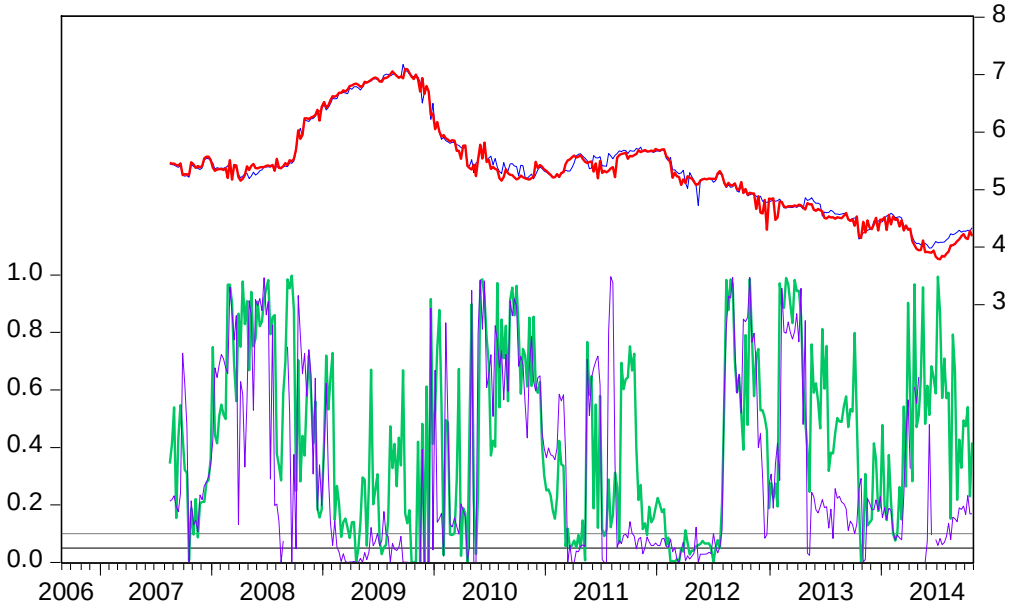


Figure 14. AIC values and γ p-values for the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the oil-Google Index. Horizontal lines indicate significance at 5% and 10%

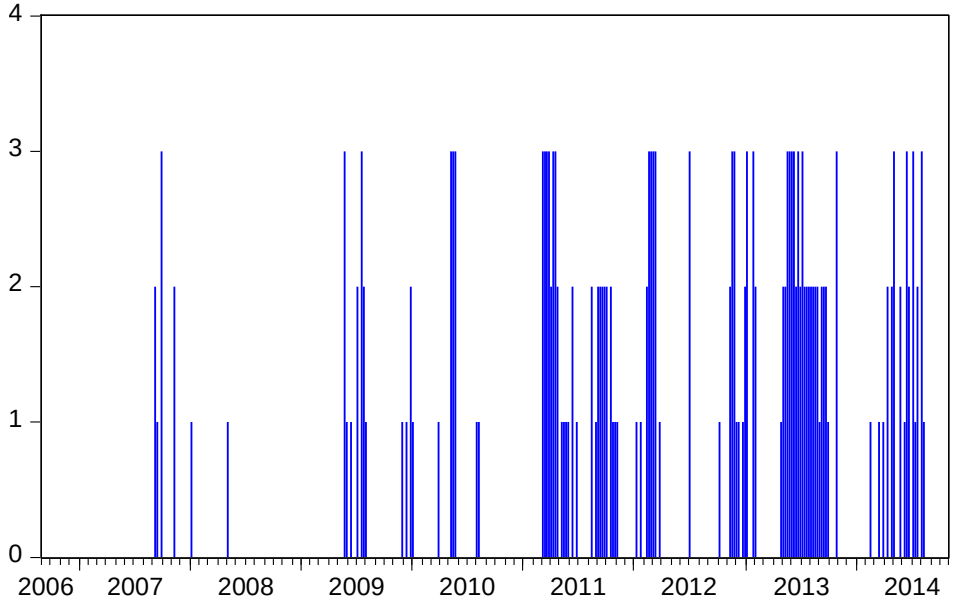


Figure 15. Rescaled z-statistic values of the λ coefficient including the oil-Google Index.The values indicate the level of significance; 3 for 1%, 2 for 5%, 1 for 10% and 0 for insignificance.

At next we illustrate the results for the Factors approach (Table 12). Again, for the full sample and all of the alternative EGARCH models, we have strong evidence of asymmetric effect on volatility shocks and persistence. The augmented EGARCH models include either pairs of the first two factors, the first two factors separately or all of them together for every keyword sample. The results support that when we comprise Google information variables in the variance equation the coefficient β declines; especially when we consider the model with the full number of factors for the full oil-related keyword sample, F1 (49), F2 (49), F3 (49), parameter β decreases from 0.98 to 0.45, contributing to explain 54% of the volatility. Also, the coefficient α, which expresses the symmetric effect, declines and becomes insignificant or less significant in the augmented models. Table 12 also includes the values of the Akaike and Schwarz criteria. The model with factors F1 (20), F2 (20) has the best performance according to AIC, but the model with F1 (20) is superior in SC, an expected result as SC imports a larger penalty term for additional parameters in the model. Again, the ARCH-LM test imply that the benchmark and augmented EGARCH models adequately captures the ARCH effect.

In Table 13 we can see the percentage decline of the number of statistical significant values of γ parameter for the rolling EGARCH models including the factors variables, with only exception the models containing F2(20) or F2(49). That is, Google Trends factors attain to capture more accurately the volatility clustering in crude oil prices. Figures 16-18 plot the AIC values and the *p*-values of γ parameter of the rolling EGARCH models in compare to the benchmark.

EGARCH(1,1) results for the crude oil returns	ω	γ	α	β	λ 1	λ 2	λ 3	Diagnostic test: ARCH-LM
without variance parameters	-0.074795**	-0.105932***	0.158024***	0.980291***				10.71827
	[-2.403]	[-4.221]	[4.854]	[107.277]				(*0.553200*)
F1(20), F2(20)	0.087966*	-0.084959***	0.056957	0.945146***	0.041358***	-0.016371*		9.352511
	[1.677]	[-2.865]	[1.528]	[65.364]	[3.364]	[-1.920]		(*0.672600*)
F1(20)	0.037336	-0.069535***	0.081348**	0.957987***	0.035983***			9.974456
	[0.813]	[-2.574]	[2.179]	[78.636]	[3.335]			(*0.618200*)
F2(20)	-0.049899	-0.116927***	0.146203***	0.974116***	-0.011051			11.45061
	[-1.438]	[-4.309]	[4.402]	[93.437]	[-1.315]			(*0.490700*)
F1(49), F2(49), F3(49)	1.346590***	-0.186539***	-0.019374	0.452364**	0.348635***	-0.155585**	-0.190865**	9.770451
	[2.646]	[-2.778]	[-0.185]	[2.175]	[2.690]	[-2.283]	[-2.100]	(*0.636100*)
F1(49), F2(49)	0.058599	-0.078499***	0.069051*	0.953290***	0.036141***	-0.014806**		9.753981
	[1.224]	[-2.768]	[1.904]	[74.124]	[3.354]	[-2.174]		(*0.637500*)
F1(49)	0.005831	-0.075127***	0.107642***	0.962454***	0.030399***			9.640937
	[0.138]	[-2.702]	[2.849]	[81.295]	[3.150]			(*0.647400*)
F2(49)	-0.059164*	-0.111096***	0.148691***	0.977029***	-0.006985			11.11010
	[-1.773]	[-4.260]	[4.508]	[102.946]	[-1.066]			(*0.519500*)

	Benchmark	F1(20), F2(20)	F1(20)	F2(20)	F1(49), F2(49), F3(49)	F1(49), F2(49)	F1(49)	F2(49)
AIC	5.327169	**5.295637**	5.300789	5.327022	5.301391	5.300612	5.307853	5.328464
SC	5.375964	5.360698	**5.357717**	5.383951	5.374584	5.365673	5.364781	5.385392

Table 12. The parameters of the variance equation for the benchmark and augmented models for the full sample (526 observations). Z-statistics in brackets, prob. Chi-Square (12) in *italics*. *, ** and *** denote statistical significance at 0.10, 0.05 and 0.01 level respectively.

Percentage of the statistically significant values of γ			
Levels of significance	1%	5%	10%
Benchmark EGARCH (1,1)	10.5042	17.85714	27.73109
F1(20), F2(20)	9.87395	13.23529	16.59664
F1(20)	4.411765	8.193277	15.96639
F2(20)	17.64706	25	31.51261
F1(49), F2(49), F3(49)	9.243697	15.96639	20.58824
F1(49), F2(49)	10.71429	16.80672	22.05882
F1(49)	6.512605	11.34454	19.32773
F2(49)	13.86555	22.68908	30.2521

Table 13. Percentage of the statistically significant values of γ for the 476 rolling outcomes from the rolling benchmark and the augmented EGARCH (1, 1) models with the oil Factors variance regressors.

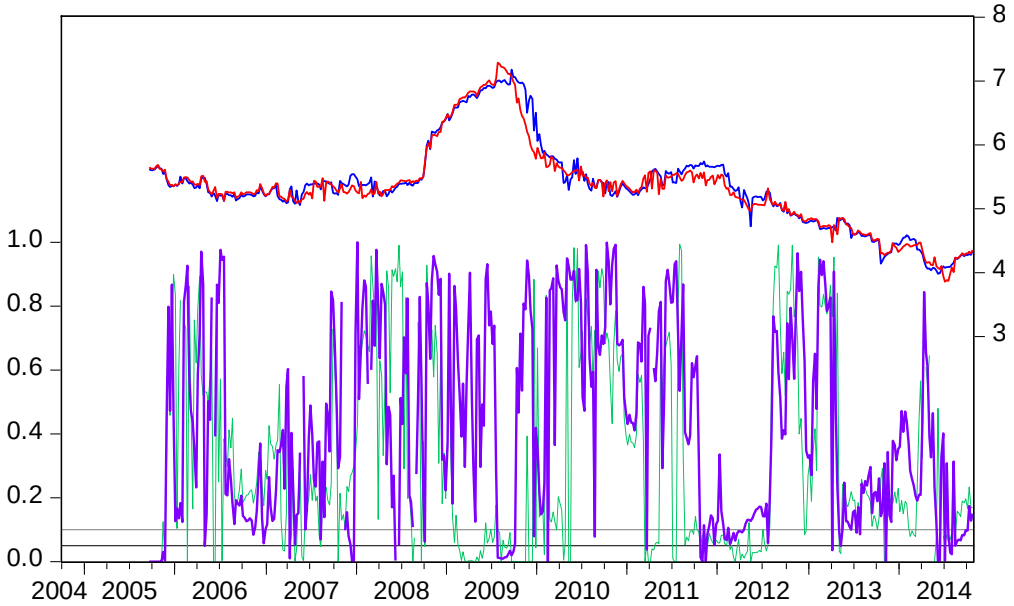
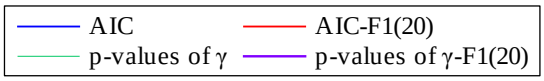


Figure 16. AIC and γ p-values for the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the oil factor F1(20). Horizontal lines indicate significance at 5% and 10%



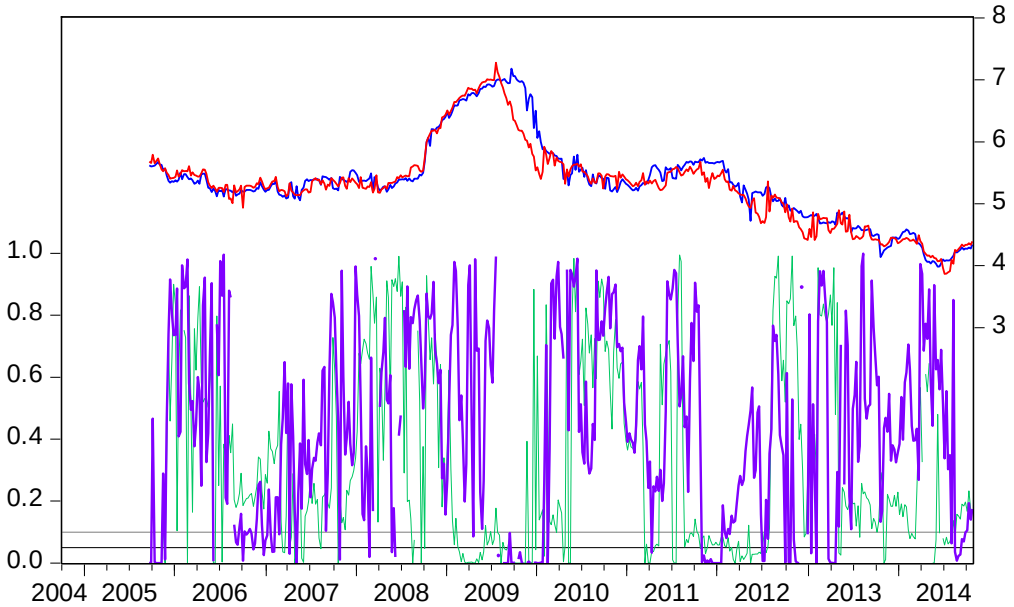
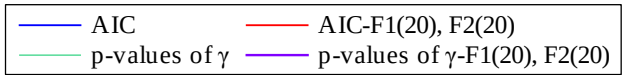


Figure 17. AIC values and γ p-values for the benchmark and the augmented rolling EGARCH(1,1) with the oil factors F1(20), F2(20). Horizontal lines indicate significance at 5% and 10%



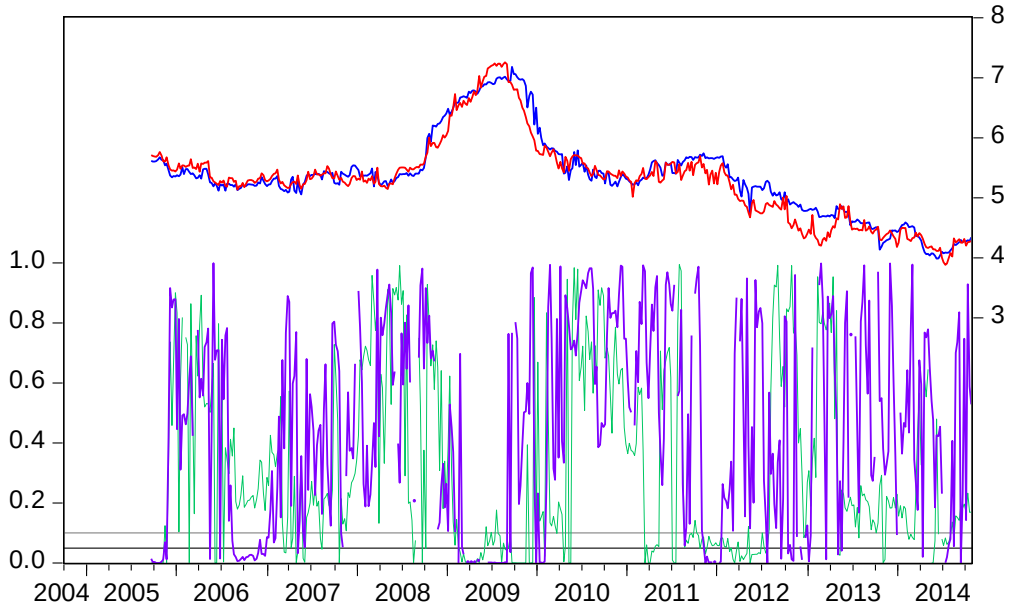


Figure 18. AIC values and γ p-values for the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the oil factors F1(49), F2(49), F3(49). Horizontal lines indicate significance at 5% and 10%

AIC — AIC-F1(49),F2(49),F3(49)
p-values of γ — p-values of γ-F1(49),F2(49),F3(49)

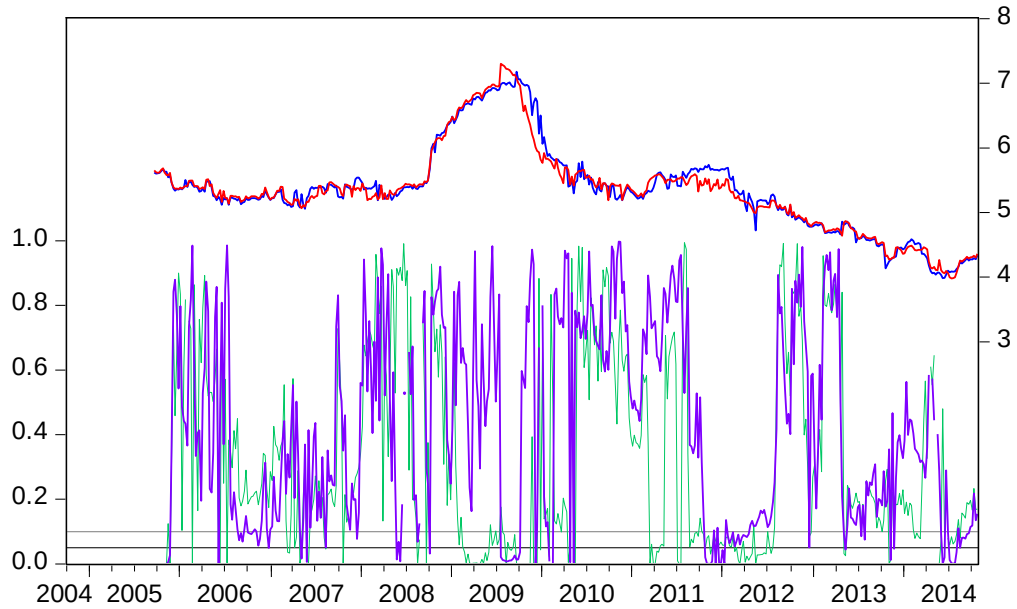


Figure 19. AIC values and γ p-values of the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the oil factor F1(49). Horizontal lines indicate significance at 5% and 10%

AIC — AIC-F1(49)
p-values of γ — p-values of γ-F1(49)

In short, the asymmetric effect is present in the crude oil market and, in a certain degree, overlapped by the sentiment variables derived from Google Trends information. This can be interpreted as follows; 'bad news' and crucial events, such as the Iraq war, the OPEC's supply and the relatively recent US recession are mainly responsible for the higher fluctuations in crude oil prices for this time period. At the same time, the 'bad news' can trigger excess internet search activity of less informed, noise traders and, by extension, irrational investment decisions out of panic and fear (Da, Engelberg, and Gao 2011) leading to an amplified volatility.

5.2 Empirical Results for the Gold market

At first, we present the results for the full time period. In Table 14 we can see the statistics and the values of the information criteria for the models with and without the Google Index. For both of our models, the α parameter accounting for the symmetric effect is positive and statistically significant at the 1 per cent level. We also observe that the λ coefficient is highly significant and the persistence in volatility is strong for the two models. No asymmetric effect is observed for the full sample and the information criteria point the augmented model as superior, with the ARCH-LM test with 12 lags of the squared residuals capturing the 'ARCH effect' properly.

Nevertheless, when we consider a subsample of the period 9/17/2006-10/21/2012, we notice that the asymmetry parameter is positive and significant, indicating that positive shocks have larger impact than negative shocks in gold price volatility, noted as inverted asymmetric reaction to shocks (Baur 2011). For the augmented model, the parameters γ and β become insignificant, suggesting that the Google Index manage to overlap a part of the asymmetry and

explain the persistence in the volatility. However, this augmented EGARCH fails to adequate model heteroskedasticity and is not preferred according to the information criteria.

For the rolling approach, Table 16 shows the proportion of statistically significant values of the γ coefficient for both of our models. As we can see, including the sentiment index in the variance leads to a reduction of the significance of the γ parameter. Figure 20 presents the AIC values and the γ parameter *p*-values for both of the rolling EGARCH models and Figure 21 plots the z- statistic values of the λ coefficient.

EGARCH(1,1) results for the gold returns	ω	γ	α	β	λ	Diagnostic test: ARCH-LM
without variance parameters	-0.086461*	0.020023	0.229657***	0.936640***		7.055134
	[-1.763]	[0.495]	[2.845]	[31.315]		(*0.8539*)
gold-Google Index	-0.102090**	-0.005673	0.182797***	0.971956***	0.202045***	6.614289
	[-2.261]	[-0.166]	[2.578]	[45.165]	[2.879]	(*0.8820*)
Information criteria		AIC			SC	
without variance parameters		4.333432			4.390740	
gold-Google Index		4.323114			4.389973	

Table 14. The parameters of the variance equation for the benchmark and augmented model for the full sample (426 observations). Z-statistics in brackets, prob. Chi-Square (12) in *italics*. *, ** and *** denote statistical significance at 0.10, 0.05 and 0.01 level respectively. Information criteria Akaike and Schwarz.

EGARCH(1,1) results for the gold returns-Subsample	ω	γ	α	β	λ	Diagnostic test: ARCH-LM
without variance parameters	-0.132510**	0.121429**	0.243976***	0.965400***		14.56026
	[-2.075]	[2.088]	[3.250]	[32.259]		(*0.2664*)
gold-Google Index	1.110822***	0.069435	0.499398***	0.053041	-0.366786***	44.59012
	[3.227]	[0.556]	[2.701]	[0.282]	[-3.680]	(*0.0000*)
Information criteria		AIC			SC	
without variance parameters		4.393725			4.464543	
gold-Google Index		4.439608			4.522230	

Table 15. The parameters of the variance equation for the benchmark and augmented model for the subsample (319 observations). Z-statistics in brackets, prob. Chi-Square (12) in *italics*. *, ** and *** denote statistical significance at 0.10, 0.05 and 0.01 level respectively. Information criteria Akaike and Schwarz.

Models	Benchmark EGARCH (1,1)			Augmented EGARCH (1,1)		
Levels of significance	1%	5%	10%	1%	5%	10%
Number of statistically significant values of γ	21	65	84	5	10	25
Percent %	5.585106	17.28723	22.34043	1.329787	2.659574	6.648936

Table 16. Percentage of the statistically significant values of γ for the 376 rolling outcomes from the rolling benchmark and the augmented EGARCH (1, 1) with the gold-Google Index variance regressor.

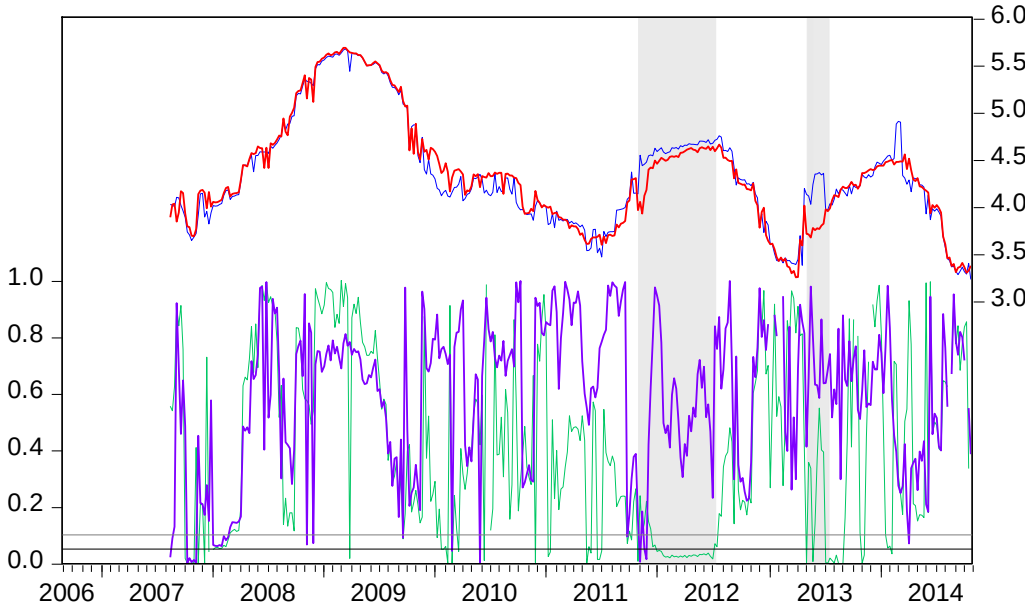
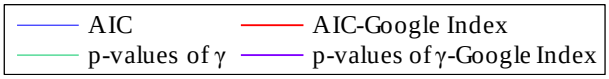


Figure 20. AIC values and γ p-values for the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the gold-Google Index. Horizontal lines indicate significance at 5% and 10%.



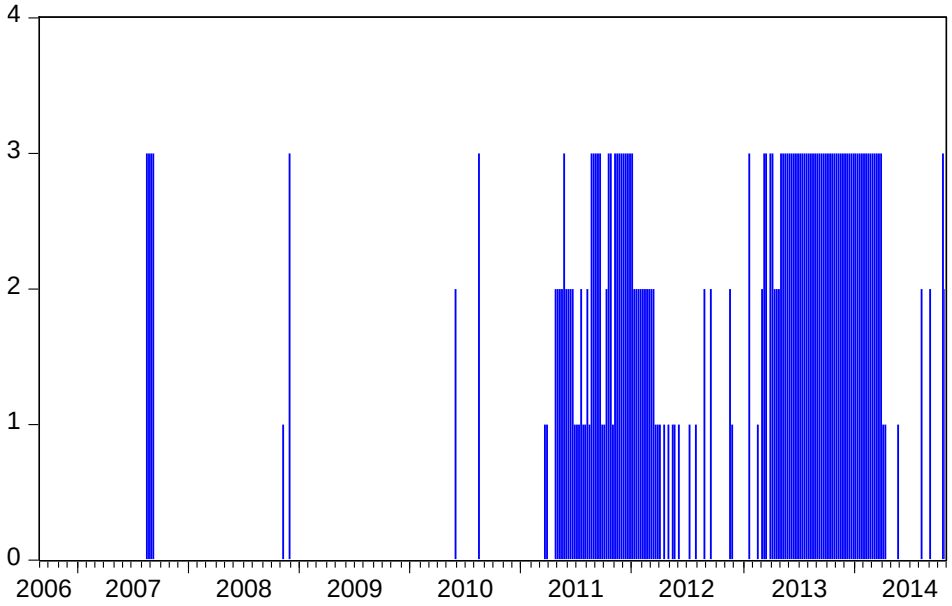


Figure 21. Rescaled z-statistic values of the λ coefficient including the gold-Google Index.The values indicate the level of significance; 3 for 1%, 2 for 5% 1 for 10% and 0 for insignificance

Finally, we illustrate the results for the factors approach. Table 17 includes the results for the full sample. We observe that the β coefficient is positive and significant at the 1 per cent level, and its value declines when we include the factors variables in the variance equation. So, the factors generated from Google information explain a part of the volatility persistence. Again, we observe a positive and highly significant α parameter and insignificant γ parameter. For the augmented models the outcomes for asymmetry are mixed presenting a pattern of significance either for the parameter α, or the parameter γ for each of the alternative EGARCH models, denoting that asymmetry is a time varying attribute. Considering the information criteria and the ARCH-LM test, the models with better performance are the ones which include either the factors F1 (61), F2 (61), F3 (61), F4 (61), F5 (61), or the first factors F1 for each of the gold-related keyword samples, namely F1 (20), F1 (61). For these models we notice that the γ coefficient is insignificant and α, β significant and positive with only exception the model including the five

factors, where the parameter α is negative. Also, the reduction of the value of β is about 5 to 10 percent.

In Table 18 we consider a subsample of the data during the period 10/17/2004-10/21/2012 for the factor models combining the first two factors or the full number of factors for each keyword sample and we compare them to the benchmark. As we can see, strong persistence in the volatility is present and partially explained from the factors. Also, the asymmetry parameter γ is positive and highly significant, implying that 'good news' increase volatility more than 'bad news'. Lastly, when we comprise the factors in the model, the α, γ parameters decline and become insignificant, except for the model augmented with the five factors, where the symmetry parameter is negative and statistically significant at the 5 per cent level, indicating that the significant factors capture the information effects, too. For the information criteria results, the Akaike value is lower for the augmented EGARCH with the five factors and Schwarz criterion points to the model with the factors F1 (20), F2 (20), which is expected as Schwarz prefers more parsimonious models.

For the rolling EGARCH approach, Figures 22-25 show the AIC values and the γ *p*-values of the models which include the first factors and the full number of factors for every keyword sample. Table 19 summarizes the proportion of the statistically significant values of the γ parameter for all the levels of significance, for all of our models.

EGARCH(1,1) results for the gold returns	ω	γ	α	β	λ 1	λ 2	λ 3	λ 4	λ 5
without variance parameters	-0.078976**	0.003954	0.217343***	0.941353***					
	[-2.361]	[0.151]	[3.840]	[48.117]					
F1(20), F2(20)	0.442210***	-0.116680**	-0.024883	0.685088***	0.149909***	0.199231***			
	[4.962]	[-2.336]	[-0.313]	[11.402]	[5.804]	[4.463]			
F1(20)	-0.039460	0.031805	0.219485***	0.909697***	0.048238***				
	[-0.990]	[0.900]	[3.365]	[38.227]	[4.437]				
F2(20)	0.127568**	-0.102966***	0.117789*	0.847704***	0.091128***				
	[2.148]	[-3.111]	[1.742]	[23.391]	[3.667]				
F1(61), F2(61), F3(61), F4(61), F5(61)	0.336306***	-0.031347	-0.184756***	0.853069***	0.100297***	-0.085600***	0.0442***	-0.0574***	0.06***
	[6.489]	[-0.885]	[-3.684]	[46.010]	[8.160]	[-5.771]	[5.114]	[-6.253]	[5.582]
F1(61), F2(61)	0.293457***	-0.089759	0.042578	0.763519***	0.124848***	-0.105330***			
	[3.207]	[-1.621]	[0.543]	[12.515]	[4.149]	[-2.936]			
F1(61)	-0.005979	0.034444	0.213232***	0.888330***	0.063496***				
	[-0.134]	[0.886]	[3.029]	[34.343]	[5.169]				
F2(61)	-0.035904	-0.030172	0.183885***	0.929121***	-0.018881				
	[-0.792]	[-0.841]	[3.432]	[32.063]	[-1.277]				

	Benchmark	F1(20), F2(20)	F1(20)	F2(20)	F1(61), F2(61), F3(61), F4(61), F5(61)	F1(61), F2(61)	F1(61)	F2(61)
AIC	4.312979	4.201913	4.287760	4.280750	**4.131274**	4.244043	4.277463	4.312930
SC	4.361774	4.266974	4.344689	4.337678	**4.220733**	4.309104	4.334392	4.369858
Diagnostic test: ARCH-LM	6.914106 (*0.8632*)	23.32699 (*0.0251*)	11.84107 (*0.4585*)	14.99138 (*0.2419*)	8.594908 (*0.7371*)	22.20916 (*0.0352*)	15.21430 (*0.2299*)	7.815168 (*0.7994*)

Table 17. The parameters of the variance equation for the benchmark and augmented models for the full sample (526 observations). Z-statistics in brackets, prob. Chi-Square (12) in *italics*. *, ** and *** denote statistical significance at 0.10, 0.05 and 0.01 level respectively.

EGARCH(1,1) results for the gold returns- Subsample	ω	γ	α	β	λ 1	λ 2	λ 3	λ 4	λ 5
without variance parameters	-0.129105***	0.104355***	0.218676***	0.972670***					
	[-2.878]	[2.700]	[4.260]	[68.388]					
F1(20), F2(20)	0.215568**	-0.042892	-0.053862	0.846658***	0.061598***	0.134258***			
	[2.421]	[-0.818]	[-0.680]	[19.492]	[3.251]	[3.766]			
F1(61), F2(61), F3(61), F4(61), F5(61)	0.292977***	-0.025442	-0.154407**	0.871397***	0.088047***	-0.063333***	0.04576*	-0.050***	0.0597***
	[3.453]	[-0.651]	[-2.274]	[31.622]	[4.472]	[-3.539]	[1.938]	[-4.384]	[3.226]
F1(61), F2(61)	0.095902	-0.005562	0.085249	0.869552***	0.055304***	-0.076725***			
	[1.064]	[-0.095]	[1.062]	[18.293]	[2.667]	[-2.627]			

	Benchmark	F1(20), F2(20)	F1(61), F2(61), F3(61), F4(61), F5(61)	F1(61), F2(61)
AIC	4.299602	4.235787	**4.219980**	4.282685
SC	4.357423	**4.312882**	4.325986	4.359780
Diagnostic test: ARCH-LM	14.79931 (0.1396)	10.46770 (*0.4005*)	7.731976 (*0.6550*)	13.19073 (*0.2132*)

Table 18. The parameters of the variance equation for the benchmark and augmented models for the subsample (419 observations). Z-statistics in brackets, prob. Chi-Square (10) in *italics*. *, ** and *** denote statistical significance at 0.10, 0.05 and 0.01 level respectively.

Percentage of the statistically significant values of γ			
Levels of significance	1%	5%	10%
Benchmark EGARCH (1,1)	3.151261	14.4958	21.0084
F1(20), F2(20)	6.932773	10.29412	14.28571
F1(20)	5.252101	11.13445	16.80672
F2(20)	5.252101	13.44538	20.37815
F1(61), F2(61), F3(61), F4(61), F5(61)	1.680672	4.621849	12.39496
F1(61), F2(61)	8.613445	14.28571	19.11765
F1(61)	6.302521	11.97479	19.32773
F2(61)	9.87395	16.59664	20.58824

Table 19. Percentage of the statistically significant values of γ for the 476 rolling outcomes from the rolling benchmark and the augmented EGARCH (1, 1) models with the gold Factors variance regressors.

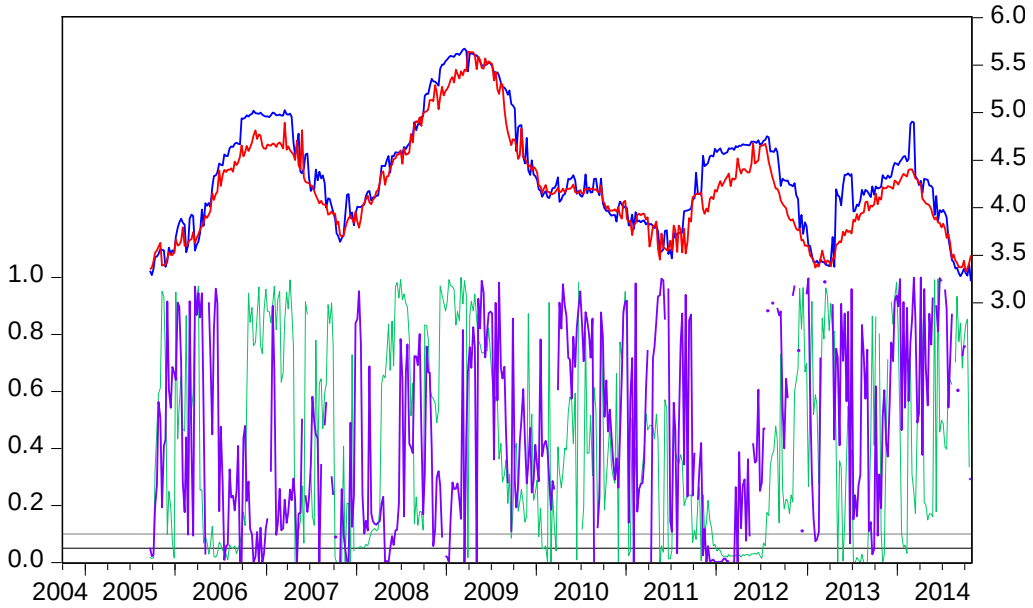


Figure 22. AIC values and γ p-values for the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the factors F1(20), F2(20). Horizontal lines indicate significance at 5% and 10%

AIC — AIC-F1(20), F2(20)
p-values of γ — p-values of γ-F1(20), F2(20)

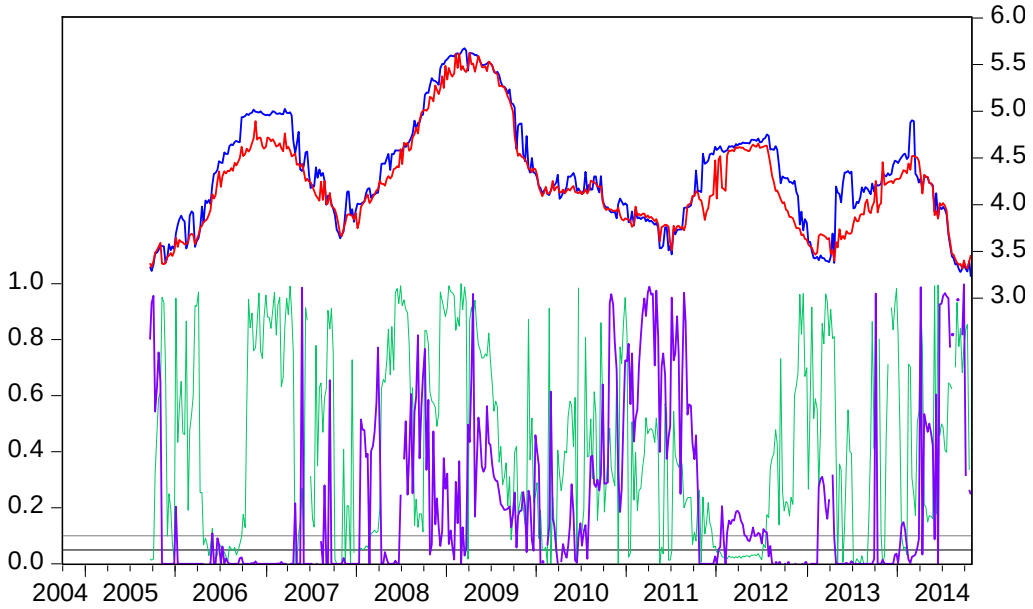


Figure 23. AIC values and γ p-values of the benchmark EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the factor F1(20). Horizontal lines indicate significance at 5% and 10%

AIC — AIC-F1(20)
p-values of γ — p-values of γ-F1(20)

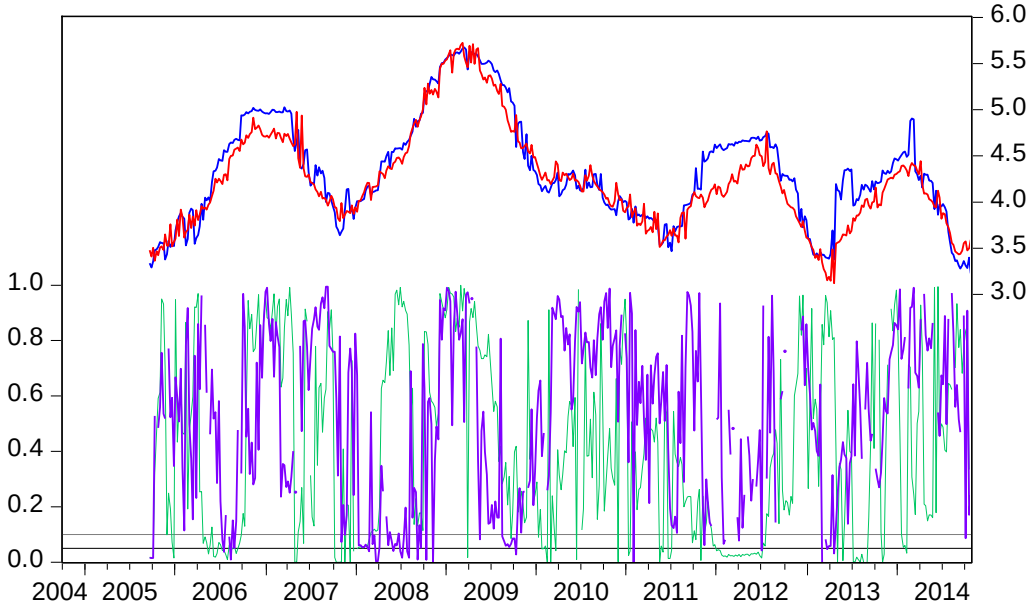


Figure 24. AIC values and γ p-values of the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the factors F1(61), F2(61), F3(61), F4(61), F5(61). Horizontal lines indicate significance at 5% and 10%

AIC — AIC- F1(61),F2(61),F3(61),F4(61),F5(61)
p-values of γ — p-values of γ-F1(61),F2(61),F3(61),F4(61),F5(61)

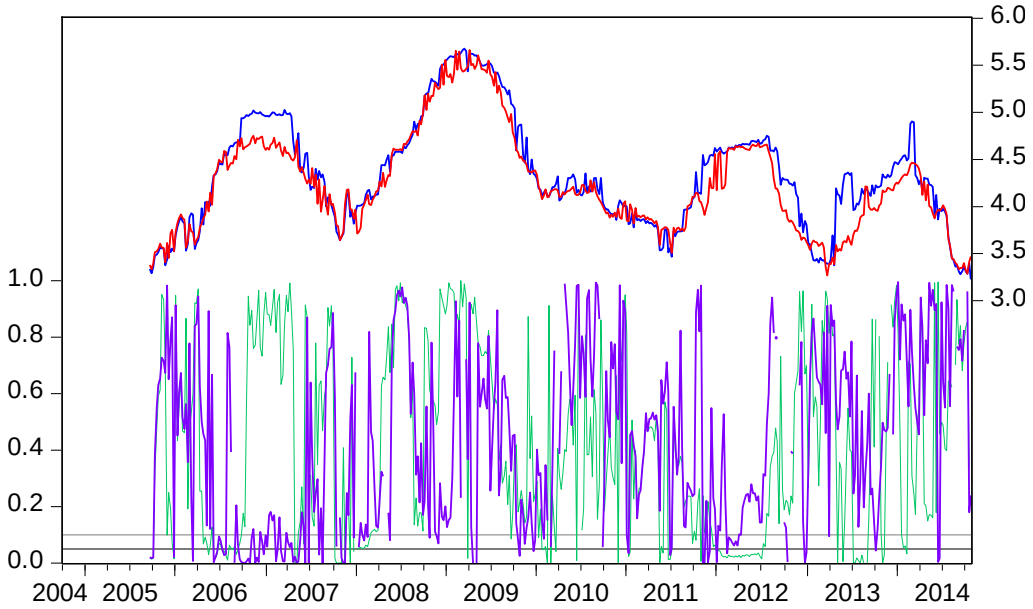


Figure 25. AIC values and γ p-values for the benchmark rolling EGARCH(1,1) and the augmented rolling EGARCH(1,1) with the factor F1(61). Horizontal lines indicate significance at 5% and 10%

AIC — AIC-F1(61)
p-values of γ — p-values of γ-F1(61)

This contrary effect of the shocks in the volatility of gold price returns can be described considering the 'safe heaven' property of gold. In times of crisis and financial or macroeconomic uncertainty, investors transmit the volatility from the other markets, where negative news have occurred, to the gold market as a result of increased hedging positions (Baur 2011).

The results of our analysis point out that the significance of the asymmetry parameter declines when we include our information demand variables. So, a part of the uncertainty can be explained by the sentiment indicators generated from Google Trends.

6. Conclusion

In our analysis we use an EGARCH model to capture the effects of Google Trends information on the conditional volatility of crude oil and gold price returns. Our information demand variables were generated either as common factors, extracted from a list of searching keywords, or as a Google Index, constructed using the methodology of Da, Engelberg, and Gao (2010), as a direct measure of sentiment. Our findings support that there is a significant impact of the Google indicators on the conditional volatility of returns. In general, we found evidence of asymmetry and persistence of shocks across the data samples, overlapped by the sentiment indicators in certain time periods. Especially, for the gold price returns and over the subsample period this asymmetric effect can be characterized as inverted, with 'good news' have greater impact on the volatility than 'bad news'. The results are consistent with sentiment-induced investment decisions; internet search activity can reflect the financial uncertainty in times of turmoil that can induce an amplified volatility in the markets.

7. References

Agnolucci, Paolo
2009 Volatility in Crude Oil Futures: A Comparison of the Predictive Ability of GARCH and Implied Volatility Models. Energy Economics 31(2): 316–321.

Artola, Concha, and Enrique Galan
2012 Tracking the Future on the Web: Construction of Leading Indicators Using Internet Searches. Banco de Espana Occasional Paper(1203). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2027861, accessed December 12, 2014.

Askitas, Nikolaos, and Klaus F Zimmermann
2009 Google Econometrics and Unemployment Forecasting. Applied Economics Quarterly 55(2): 107–120.

Baur, Dirk G.
2011 Asymmetric Volatility in the Gold Market. Available at SSRN 1526389. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1526389, accessed February 22, 2016.

Black, Angela J., Olga Klinkowska, David G. McMillan, and Fiona J. McMillan
2014 Forecasting Stock Returns: Do Commodity Prices Help?: Forecasting Stock Returns: Do Commodity Prices Help? Journal of Forecasting 33(8): 627–639.

Bollen, Johan, Huina Mao, and Xiaojun Zeng
2011 Twitter Mood Predicts the Stock Market. Journal of Computational Science 2(1): 1–8.

Bollerslev, Tim
2008 Glossary to Arch (garch). CREATES Research Paper 49. http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=1263250, accessed February 24, 2016.

Carrière-Swallow, Yan, and Felipe Labbé
2013 Nowcasting with Google Trends in an Emerging Market: Nowcasting with Google Trends in an Emerging Market. Journal of Forecasting 32(4): 289–298.

Choi, Hyunyoung, and Hal Varian
2012 Predicting the Present with Google Trends: PREDICTING THE PRESENT WITH GOOGLE TRENDS. Economic Record 88: 2–9.

Da, Zhi, Joseph Engelberg, and Pengjie Gao
2010 The Sum of All Fears: Investor Sentiment and Asset Prices. SSRN eLibrary. https://www3.nd.edu/~pgao/papers/FEARS_20131007.pdf, accessed December 12, 2014.
2011 In Search of Attention. The Journal of Finance 66(5): 1461–1499.

Fondeur, Y., and F. Karamé

2013 Can Google Data Help Predict French Youth Unemployment? Economic Modelling 30: 117–125.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, et al.
2009 Detecting Influenza Epidemics Using Search Engine Query Data. Nature 457(7232): 1012–1014.

Goel, S., J. M. Hofman, S. Lahaie, D. M. Pennock, and D. J. Watts
2010 Predicting Consumer Behavior with Web Search. Proceedings of the National Academy of Sciences 107(41): 17486–17490.

Kholodilin, Konstantin A., Maximilian Podstawski, and Boriss Siliverstovs
2010 Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US. KOF Swiss Economic Institute Working Paper(256). http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1615453, accessed February 11, 2016.

McLaren, Nick, and Rachana Shanbhogue
2011 Using Internet Search Data as Economic Indicators. Bank of England publications, Research and analysis. http://www.bankofengland.co.uk/publications/quarterlybulletin/qb110206.pdf.

Panagiotidis, Theodore, Georgios Bampinas, and Konstantinos Ladopoulos
N.d. A Note on the Estimated GARCH Coefficients from the S&P 1500 Universe.

Peri, Massimo, Daniela Vandone, and Lucia Baldi
2014 Internet, Noise Trading and Commodity Futures Prices. International Review of Economics & Finance 33: 82–89.

Preis, Tobias, Helen Susannah Moat, and H. Eugene Stanley
2013 Quantifying Trading Behavior in Financial Markets Using Google Trends. Scientific Reports 3. http://www.nature.com/doifinder/10.1038/srep01684, accessed December 12, 2014.

Smith, Geoffrey Peter
2012 Google Internet Search Activity and Volatility Prediction in the Market for Foreign Currency. Finance Research Letters 9(2): 103–110.

Vlastakis, Nikolaos, and Raphael N. Markellos
2012 Information Demand and Stock Market Volatility. Journal of Banking & Finance 36(6): 1808–1821.

Vosen, Simeon, and Torsten Schmidt
2011 Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends. Journal of Forecasting 30(6): 565–578.

Yang, Yan
2011 Learning Stock Volatility Using Keyword Search Volume.

8. Appendix

Table. 20 gold-related keywords with non-zero searching volume(non-zero observations)

BEST GOLD
BUY GOLD
GOLD RATE
GOLD
GOLD BAR
GOLD BULLION
GOLD MARKET
GOLD OUNCE
GOLD PRICE
GOLD PRICES
GOLD SPOT
GOLD VALUE
KITCO
OF GOLD
OUNCE GOLD
PRECIOUS METALS
PRICE GOLD
PRICE OF GOLD
ROSE GOLD
SPOT GOLD

Table. 20 oil-related keywords with non-zero searching volume

PRICE OIL
PRICE OF OIL
OIL STOCK
OIL PRICES
OIL PRICE
OIL FUTURES
OIL CHART
OIL BARREL
OIL
GAS PRICE
CURRENT OIL
CRUDE PRICES
CRUDE PRICE
CRUDE OIL PRICES
CRUDE OIL PRICE
CRUDE OIL
CRUDE
BUY OIL
BARREL PRICE
BARREL OIL

Table. 61 gold-related keywords (full sample)

BEST GOLD	GOLD TRADING
BUY GOLD	GOLD VALUE
BUYING GOLD	GOLD PRICE CHART
CURRENT GOLD	HISTORICAL GOLD
CURRENT GOLD PRICE	INDIA GOLD PRICE
GLD	KITCO
GOLD	KITCO.COM
GOLD BAR	KITCO GOLD
GOLD BULLION	LIVE GOLD PRICE
GOLD CHART	MONEX
GOLD CLOSE	OF GOLD
GOLD COMMODITY	OUNCE GOLD
GOLD DEALER	PRECIOUS METALS
GOLD DEALERS	PRICE GOLD
GOLD FUTURES	PRICE OF GOLD
GOLD INDEX	PURCHASE GOLD
GOLD INVESTMENT	ROSE GOLD
GOLD JEWELLERY	SPOT GOLD
GOLD MAPLE	SPOT GOLD PRICE
GOLD MARKET	SPOT PRICE GOLD
GOLD NUGGET	WHERE TO BUY GOLD
GOLD OUNCE	GOLD STOCK
GOLD PRICE	GOLD STOCK MARKET
GOLD PRICE CHART	GOLD STOCK PRICE
GOLD PRICE HISTORY	GOLD STOCKS
GOLD PRICE IN	GOLD TODAY
GOLD PRICE IN INDIA	
GOLD PRICE INDIA	
GOLD PRICE LIVE	
GOLD PRICES	
GOLD QUOTE	
GOLD RATE	
GOLD SILVER PRICES	
GOLD SPOT	
GOLD SPOT PRICE	

Table. 49 oil-related keywords (full sample)

BARREL OF GAS	PRICE OF CRUDE
BARREL OF OIL	PRICE OF CRUDE OIL
BARREL OIL	PRICE OF OIL
BARREL OIL PRICE	PRICE OF OIL BARREL
BARREL PRICE	PRICE OIL
BARREL OF CRUDE OIL	PRICE OIL BARREL
BLOOMBERG ENERGY	SPOT OIL
BLOOMBERG OIL	SWEET CRUDE
BUY OIL	THE PRICE OF OIL
COST OF OIL	OIL STOCK
CRUDE	OIL STOCK PRICE
CRUDE OIL	OIL STOCK PRICES
CRUDE OIL CHART	OIL STOCKS
CRUDE OIL PRICE	OIL TODAY
CRUDE OIL PRICES	PRICE OF BARREL OF OIL
CRUDE PRICE	
CRUDE PRICES	
CURRENT OIL	
CURRENT OIL PRICE	
CURRENT OIL PRICES	
GAS PRICE	
NYMEX OIL	
OIL	
OIL BARREL	
OIL BARREL PRICE	
OIL CHART	
OIL FUTURES	
OIL PER BARREL	
OIL PRICE	
OIL PRICE CHART	
OIL PRICE HISTORY	
OIL PRICES	
OIL QUOTE	
OIL SPOT	