

Volatility persistence and asymmetry under the microscope: The role of information demand for gold and oil

Georgios Bampinas* Theodore Panagiotidis† Christina Rouska‡

January 30, 2018

Abstract

This study explores the relationship between Google search activity and the conditional volatility of oil and gold spot market returns. By aggregating the volume of queries related to the two commodity markets in the spirit of Da *et al.* (2015), we construct a weekly Searching Volume Index (SVI) for each market as proxy of households and investors information demand. We employ a rolling EGARCH framework to reveal how the significance of information demand has evolved through time. We find that higher information demand increases conditional volatility in gold and oil spot market returns. Information flows from Google SVIs reduce the proportion of the significant volatility asymmetry produced by negative shocks in both commodity markets. The latter is more profound in the gold market.

Keywords: Gold, Oil, Google Trends, Volatility, Asymmetry, EGARCH.

JEL Classification: C01, C32, C38, C51, C81, D81, G02, G11

*Department of Economics, University of Macedonia, 156 Egnatia Street. 540 06, Thessaloniki, Greece.
E-mail address: *bampinasg@uom.edu.gr*.

†Department of Economics, University of Macedonia, 156 Egnatia Street. 540 06, Thessaloniki, Greece.
E-mail address: *tpanag@uom.edu.gr*.

‡Department of Economics, University of Macedonia, 156 Egnatia Street. 540 06, Thessaloniki, Greece.
E-mail address: *me145@uom.edu.gr*.

1 Introduction

Historically, investors have utilized commodities as a store of value and a means of exchange. Increased stock market volatility in the period following the global financial crisis has induced investors participating in precious metals and energy markets, mainly due to the shortage of assets and alternative investment vehicles in the world economy (Caballero *et al.*, 2008). Gold and oil, as two of the most widely traded commodities and among the most followed economic indicators, have attracted investors attention not only as 'safe haven' to avoid financial risk, but also as a part of a fundamental investment strategy (Baur and Lucey, 2010; Daskalaki and Skiadopoulos, 2011); this has been further enabled by an increase in the range of products such as actively managed traded funds (hedge funds, ETFs) which offer to investors and traders profit opportunities on both the long- and short-term.¹

Investment incentives of these types differ from traditional ones, while having important consequences for the behaviour of gold and oil price fluctuations. For instance, commodity index traders intensify their investment in commodities to improve portfolio diversification (Daskalaki and Skiadopoulos, 2011; Bessler and Wolff, 2015), to hedge against movements of inflation (Erb and Harvey, 2006; Bampinas and Panagiotidis, 2015; Beckmann *et al.*, 2015), or to 'fly to safety' in times of market turmoil (Baur and McDermott, 2010; Bampinas and Panagiotidis, 2017). Moreover, the vast inflow of commodity index traders has made gold and oil markets more susceptible to the mood of financial markets and economic conditions (Tang and Xiong, 2012; Büyüksahin and Robe, 2014). In this respect, the question is no longer whether investor sentiment affects the gold and oil price fluctuations, but more importantly how to capture investor sentiment in these markets and quantify its effect.

In this study, we gauge investor sentiment for the gold and oil markets by using a direct source of information provided by Google. The sentiment variables, Gold and Oil Search Volume Indexes (SVI's), are compiled from Google Trends. Google Trends provides time

¹See for example <http://www.gold.org/research/gold-demand-trends/gold-demand-trends-full-year-2016> and <https://www.bloomberg.com/view/articles/2016-09-16/the-problem-with-the-oil-etf-it-does-a-great-job>.

series data on the volume a particular keyword search term is entered into the Google search engine. By inputting a keyword search term on Google trends (for example “gold”, “gold price”, “spot gold” and/or “crude oil”, “oil price”, “buy oil”), users can observe the actual flow of worldwide Internet searches for that particular keyword over time. This value is relative to the total number of search on Google in the corresponding time interval. Google Trends allows users to restrict (geographically) SVI results to specific countries. Since both of (dependent) variables of interest are expressed in US dollars, we focus our analysis on the US. Thus, the measures we construct represent the sentiment of the US investors and households towards oil and gold. Figures 1 and 2 demonstrate the gold and crude oil spot returns along the related keywords “gold price” and “oil price”, respectively.

[Insert Figures 1 and 2 about here.]

Much of the extant literature has been interested in studying the link between volatility and the rate at which information flows into the marketplace (e.g. Brailsford, 1996; Lamoureux and Lastrapes, 1990). One of the most appealing explanations for commonly-observed patterns is that volatility is proportional to the rate of information arrival. Formalized via the mixture of distributions hypothesis (MDH), the ARCH and GARCH effects (see Engle, 1982; Bollerslev, 1986) are often attributed to fluctuations in the unobservable rate of information flow (see for example, Clark, 1973; Epps and Epps, 1976; Tauchen and Pitts, 1983). Lamoureux and Lastrapes (1990) find that trading volume is a good proxy for informational arrivals in the US stock market, reducing volatility’s persistence in a conditional variance process. On the contrary, other studies assert that conditional volatility’s persistence remains in the presence of trading volume (Bollerslev and Domowitz, 1993; Bessembinder and Seguin, 1993). Insomuch as the Google Trends can act a barometer for the rate of information flow, it is of interest to question whether there is any link between the information demand captured from Google SVIs and asymmetry in the volatility (usually captured by an EGARCH model) of the two major commodity markets.

Whilst there is an extant literature concerning the relationship between news sentiment and volatility in the stock market, less attention has been given to more specialised benchmarks (for a related study on commodities see e.g., Peri *et al.*, 2014). The primary goal of this paper is to construct a direct measure of investors' information demand from the Internet and quantify its effect on the conditional volatility of gold and oil market returns. This is important because the institutional framework of the commodities markets has distinctive characteristics that differentiate them from the equity markets. In particular, commodity traders face constraints in terms of position limits imposed by the exchange, margin requirements, and the necessity of physical settlement of short positions. Such constraints may explain to some extent why commodity returns behave differently to those witnessed in equity markets (for a discussion, see Gorton and Rouwenhorst, 2006). In order for investors to make rational portfolio optimization choices, it is necessary to have a better understanding of how information demand impact on returns volatility, as well as the degree of commodity market response to asymmetric information flows.

The focus of this study is to model volatility in the presence of information flow stemming from the Internet (information demand); in particular, we aim to quantify the amount of conditional volatility produced by information demand from Google over time. We employ a rolling EGARCH framework that allows us to examine the evolving influence of information demand on the conditional volatility of the two most important commodities (gold and oil). Following Da *et al.* (2015), we construct a Google Searching Volume Index (SVI), and focus on the rolling estimates of the conditional variance coefficients. A baseline EGARCH model together with a Google SVI-EGARCH augmented one is examined. Two important findings emerge: (i) the significance of β increases in the augmented model that includes investors attention for crude oil, consistent with the nature of information flows coming from the internet, as an expression of noise trading and (ii) the significance of the asymmetry coefficient is reduced in the Google-EGARCH augmented model for gold and oil returns volatility, which is compatible with the intrinsic characteristics of commodities to absorb

efficiently new information in response to negative shocks.

The rest of the paper is organized as follows: Section 2 reviews the literature, section 3 discusses the data and the next one the Methodology. Section 5 presents the results and the last one concludes.

2 Related literature

Merton's (1987) seminal work demonstrates that investor attention matters for security prices and liquidity. This was followed by Sims (2003), Hirshleifer and Teoh (2003), Peng and Xiong (2006), among others. The exact role of information and investor attention in the market equilibrium and efficiency, however, remains elusive. Grossman and Stiglitz (1980) suggest that more information (a greater number of informed investors) leads to more informative prices, resulting in a more efficient market. In this respect, if increased attention leads to more information absorbed by the market, attention should improve market efficiency.

Despite the theoretical justification of the efficient market hypothesis, substantial evidence has been accumulated suggesting that the markets may not be efficient. A strand of the literature in the field of behavioural finance argues that some investors are not fully rational and their demand for risky assets is affected by their beliefs or sentiments that are not fully justified by fundamental news (Forlani, 2002; Kahneman, 2003). Theoretical models used to explain the causes of the violation of market efficiency (see, for instance, Barberis *et al.*, 1998; Daniel *et al.*, 1998; Hong and Stein, 1999) typically investigate some form of investors' reaction to new information.

Empirical investigation of the limited investor attention issues often involves analysis of the news announcements effects (Barber and Odean, 2008; Dellavigna and Pollet, 2009). The investigation of the effect of news and mass media on asset returns exposes several potential issues. For example, Fang and Peress (2009) access attention attracted by a firm by examining a number of published newspaper articles. Nevertheless, there is no consensus as to the extent to which readers of a newspaper pay attention to the specific company. Other

measures of investor attention, such as institutional holdings, analyst coverage, or advertisement expenditures, suffer from similar shortcomings. Huberman and Regev (2001) present an example of how security prices can be artificially inflated by mass media. Therefore, it is difficult to verify empirically a connection between stock returns, market efficiency, and investor attention.

Recent studies introduced a new measure of attention – Internet search query – into the finance literature. As an alternative proxy for investor attention, Da *et al.* (2011) propose the use of information conveyed by search volume on Google. They show that Search Volume Index (SVI) measures attention more consistently than do other well-established attention proxies and primarily captures the retail investors attention. Vlastakis and Markellos (2012) also use data from Google Trends to approximate information demand and public interest at the firm and market level for the 30 largest stocks traded on the NYSE. They show that information demand approximated from the SVI, has a significant impact on individual stock trading volume and returns volatility. Moreover, the authors report a positive link between information demand and investor's risk aversion. In a similar vein, Smith (2012) shows the predictive power of Google internet searches for specific keywords over foreign currency volatility. Dergiades *et al.* (2015) employ both Google and Twitter data to investigate their effect on the eurozone bond markets. Da *et al.* (2015) explore the role of internet search volume as a measure of investors sentiment. Adopting a regression based approach, data for 30 search terms were selected and modified to form the Financial and Economic Attitudes Revealed by Search index (FEARS). Their results reveal that the FEARS index predicts short-term returns reversals, temporary increases in volatility and mutual funds flows, with investors to switch from equity funds to bonds after a spike in the FEARS index. The emerging consensus is that the Google search measures capture the attention of retail investors. An advantage of employing the Google search approach over previously used measures of attention, such as traditional news, is its ability to incorporate actively expressed investor interests.

Numerous articles have explored the use of internet search data as economic indicators. This was initiated by Choi and Varian (2009a) who illustrated its use for predicting US retail sales, automotive sales, home sales and trends in travel destinations. Kholodilin *et al.* (2010) use the monthly US real private consumption, the Consumer Sentiment Indicator from the University of Michigan, the Consumer Confidence Index from the US Conference Board, the 3- month and 10-year US Treasury constant maturity rate, the S&P500 index and Google search data from 220 private consumption related terms. They show that these models, including Google indicators, perform better than the benchmark model (i.e., an AR(1) model), as well as the augmented models with the alternative sentiment indicators. Vosen and Schmidt (2011) obtained year-to-year growth rates for 56 consumption relevant categories from Google Trends. They compare the predictive power of Google indicators to those of two common survey based indicators, namely the University of Michigan's Consumer Sentiment Index and the Conference Board's Consumer Confidence Index. Their results indicate that the performance of the Google indicators is superior to the benchmark and the alternative measures of sentiment, both in-sample and out-of-sample. Carrière-Swallow and Labbè (2011) investigate whether Google activity correlates with consumer purchases in the automobile market of Chile. They show that the models, including Google Trends-based index (GTAI) of automobile related keywords, outperform competing benchmark specifications in terms of nowcasting, both in- and out-of sample.

Labour market studies using internet search data have been carried out in a wide range of countries. For the case of the US, Choi and Varian (2009b) find that unemployment and welfare-related searches can improve predictions of initial jobless benefit claims. Askitas and Zimmermann (2009) employ monthly unemployment rates of Germany from January 2004 to April 2009 and observe strong correlations between specific keywords collected from Google Insights and unemployment rates. D' Amuri and Marcucci (2017) and Suhoy (2009) find similar results for Italy and Israel, respectively. Fondeur and Karamè (2013) test the ability of Google search data to improve predictions of youth unemployment in France.

They find that Google information enhance prediction models in terms of both level and accuracy.

Nevertheless, very few applications of Google search query data focus on commodity markets. Peri *et al.* (2014) use weekly data for internet search volumes of the keyword 'corn price(s)', newspapers information for the same keyword and corn futures prices, from January 2004 to July 2011. Employing an EGARCH model, they attempt to capture the effect of information flows from internet and newspapers on the conditional volatility of corn futures prices. Their results indicate that the internet search activity enhances the volatility produced by negative shocks, consistent with the notion that internet search mostly reflects the noise traders' activity. Smales (2014) examines the relationship between news sentiment and returns in the gold futures market over the period 2003-2012 and finds that negative news sentiment invokes a greater contemporaneous response in returns of gold futures.

Volatility of oil and gold is not only an important factor in investors and consumers behavior but it also has significant consequences for the financial markets and the economy. Factors driving oil price volatility include market fundamentals (fluctuations in supply, demand, and market power), some of them related to expectations of future production, consumption and traders behavior. Oil is also considered a driver of inflation and in turn, inflation is a driver of gold (O'Connor *et al.*, 2015). Inflationary expectations may lead investors to purchase gold, either to hedge against the expected decline in the value of money (see Jaffe, 1989) or to speculate on the associated increase in the price of gold (Bampinas and Panagiotidis, 2015). Cai *et al.* (2001) use intraday data for the US and find that prices of gold futures have time varying volatility and that announcements concerning GDP and inflation have a strong impact on gold return volatility. Batten *et al.* (2010) examine the volatility of four precious metals and show that gold volatility is explained by monetary variables.

3 Data

Our study employs weekly time series on crude oil and spot prices, over the period October 2004 to October 2014. The data for gold were obtained from the Federal Reserve Economic Data (FRED) that is quoted in US dollar per ounce, while crude oil price data (West Texas Intermediate) are obtained from the US Energy Information Agency, quoted in US dollar per barrel.² Weekly rates of return are calculated by taking the weekly first difference log-returns, i.e. $r_t = 100 \times (\ln(P_t) - \ln(P_{t-1}))$, where P_t is the gold or the crude oil price at time t . Table 1 reports the descriptive statistics for the price returns data; the excess kurtosis value and the Jarque-Bera test for normality both indicate that the returns series exhibit significant departure from normality. The challenge in this section is how to quantify information demand (we also use the term investor attention/sentiment index as equivalent terms). We will gauge the attention that oil and gold receive by employing data from Google search queries. The next section presents the methodological details for this.

[Insert Tables 1 about here.]

3.1 Construction of the Google Index

We start by constructing a list of sentiment-reveal search terms that could influence the two commodities under examination, namely gold and oil.³ We initiate the analysis with primary keywords such as “gold”, “gold price”, “gold rate”, “gold stock”, “spot gold”, “COMEX GOLD”, “oil price”, “oil stock”, etc. To get a better understanding on how these terms are searched in Google, we have tried each one in Google Trends and got the 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

²Gold prices are available at: <https://fred.stlouisfed.org/> and crude oil prices are available at: <https://www.eia.gov/>.

³Note that recent text analysis uses Harvard IV-4 Dictionary to categorize words as “positive”, “negative”, “strong”, “weak” and so on. Therein the word “gold” is classified as an economic word with positive sentiment. See http://www.wjh.harvard.edu/~inquirer/spreadsheet_guide.htm.

used the top 30 or 40 related searches from Google Correlate.⁴ This generates approximately 260 related words for each commodity.

Next, we remove duplicates, terms with inadequate data (query series with zero search volume) and terms with no economic meaning. For example, a search for “gold” results in related searches “how much gold”, “grams of gold”, “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.

We downloaded the weekly Search Volume Index (SVI) for each of these terms over the sample period of 10/3/2004-10/26/2014 from Google Trends. Over the last 10 years Google was the most popular search engine in the US with 247 million unique visitors and a market share of 63.9 percent among all search engines. Thus, we focus our analysis to the US.

An important issue with the data from Google Trends is the sampling method. For every query, a measurement error is present in the generated series. Requests (to Google) at different days for the same query produce (slightly) different results. To identify any deviations in the data, we downloaded daily the SVI series for each keyword, for 60 different days, following Carrière-Swallow and Labbè (2011). The correlation coefficient of the samples for each term is found close to 0.9. 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. We set the weekly change in search term i as:

$$DLSVI_{i,t} = \ln(SVI_{i,t}) - \ln(SVI_{i,t-1})$$

Following Da *et al.* (2015), to mitigate any concerns about outliers, difference in variance across terms and to address the issues of slow-moving time trends and seasonality, we make a number of adjustments to the SVI measures. At first, we winsorize each of the $DLSVI_{i,t}$ at the 5% level. Then, we seasonally adjust the series by regressing $DLSVI_{i,t}$

⁴Google 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. For more on Google Correlate see <https://www.google.com/trends/correlate>.

on monthly dummies and keep the residuals. Finally, to make each time-series comparable, we standardize them by subtracting the mean and divide the difference by the standard deviation. This procedure results in an adjusted (winsorized, standardized and seasonally adjusted) weekly change of SVI, adjDLSVI (adjusted DLSVI), for each search-term. We keep the $\text{adjDLSVI}'s$ with non-zero observations for the same time span. The final list of adjusted DLSVI terms along the unit root tests are given in Tables 2. ADF and PP unit root tests confirm that the adjusted DLSVI series are stationary.

[Insert Tables 2 about here.]

The next step to specify which search terms are more significant for the gold and oil returns. For this, we run recursive rolling regressions of the adjusted DLSVI on the contemporaneous oil and gold market returns for every term separately. We use an expanding window of 100 observations (2 years) and a step of 50 observations. We then rank the t -statistics from every rolling outcome from the smallest to the 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 October 3, 2004 –August 27, 2006 gives us three significant search terms for gold: “gold prices”, “gold price” and “price gold”. We use these observations for the period September 3, 2006 –August 12, 2007. The generated Google Index for week t during this period is simply the average $\text{adjDLSVI}_{i,t}$ of these three terms on week t and is defined as:

$$\text{GoogleIndex}_t = \sum_{j=1}^n R^j(\text{adjDLSVI}_t)$$

Given the need for an initial window, our Google Index begins in September 3, 2006.⁵

4 Methodology

This section provides the econometric approach adopted to develop an understanding of the relationship between information demand and volatility in the gold and oil spot markets. The

⁵The full data sample for the index approach includes 426 observations for gold and crude oil data.

baseline model utilises a regression specification for the returns equation of the form:

$$r_t = c + \varepsilon_t \quad (1)$$

where $\varepsilon_t = z_t h_t^{1/2}$, r_t is the daily stock returns, h_t is the conditional variance and c the constant term. $\{z_t\}$ is a sequence of independent, *iid* random variables with zero means and unit variance. This implies that $\varepsilon_t | \mathcal{F}_{t-1} \stackrel{d}{\sim} (0, h_t)$ where d stand for the distribution. Nelson (1991) indicates that a Generalized Error Distribution (GED) is preferred for GARCH models. Thus, we assume that the error term follows the GED distribution.⁶

In the empirical application, we analyse the conditional volatility of spot oil and gold price returns by controlling for the information flows. We model the two commodity price return volatilities in an asymmetric GARCH framework. In particular, we employ an augmented EGARCH model to capture conditional volatility and possible leverage effects:

$$\log(h_t) = \omega + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right] + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta \log(h_{t-1}) + \lambda DLSVI_t \quad (2)$$

where h_t is the conditional variance. The α parameter represents the symmetric effect of the model whereas β measures the persistence in the conditional volatility. The asymmetric effect of the volatility is expressed by the γ coefficient. If $\gamma > 0$ “good news” (or positive shocks) are more influential than “bad news” (or negative shocks), whereas if $\gamma < 0$ “bad news” generate a larger impact on volatility. The information demand variable $DLSVI_t$ for gold and oil is included in the conditional variance equation as exogenous (this is the adjusted variable that was explained in the previous section). The estimated λ would be the focus of our estimation ($\lambda=0$ would be the baseline model and $\lambda \neq 0$ the augmented one). As noted in Kalev *et al.* (2004), when a proxy for information flows is included into a conditional variance equation most of the observed volatility persistence would diminish. In Eq. (2) parameter restrictions are not necessary to ensure the non-negativity of h_t .

⁶In a similar vein, Bampinas *et al.* (2018) conclude that the GED is the most preferable compared to the normal and the Student’s *t* in accomplishing the non-negativity and stationarity constraints of the EGARCH specification, within the S&P500 stocks universe.

Our final step involves rolling estimation (sub-sample analysis) to gauge the evolution of the coefficients and their significance through time. For this purpose, we perform a rolling-regression estimation for Eq.(2) at a fixed window of 50 observations, which is shifted forward by one week (a rolling sample of about 1 year of data is taken with a step size of 1 week). The first estimate is based on a regression using observations 1-50, the second, observations 2 to 51, the third, observations 3 to 52, and so on.

5 Results

Our analysis starts with an EGARCH model for the full sample for the logarithmic returns of crude oil. The results for both the baseline and the augmented model (the one presented in Equation 2 that includes investor attention in the conditional variance) are presented in Table 3. For both estimated models (Panel A in Table 3), we observe that the γ coefficient is negative and statistically significant at the 1% level, indicating an asymmetric effect on the volatility of crude oil returns. That is, volatility is likely to be higher when lagged shocks are negative. 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 Google Index (the adjusted DLSVI) as an additional regressor in the variance equation we notice that is highly significant. The volatility persistence parameter β increases while the asymmetry parameter γ decreases but remains significant. Akaike and Schwarz criteria are taking lower values in favor of the augmented model.

Panel B in Table 3 presents the percentage of statistically significant values of β and γ coefficients for the rolling regressions results for each model at all conventional levels of significance. We observe that the proportion of the significant β values increases by 27% on average when we consider the augmented model. On the contrary, the γ coefficients decrease by more than half when we include the Google Index in the variance equation, for all the levels of significance. This reduction of asymmetry when information demand is included suggests that the sentiment variable reflects information that was previously incorporated in

the asymmetric factor of the crude oil price volatility.

[Insert Tables 3 and 4 about here.]

Table 4 presents the empirical results for the gold market. For both the baseline and the augmented model (Panel A in Table 4), the α parameter accounting for the symmetric effect is positive and statistically significant at the 1 % 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 to the augmented model as superior, with the ARCH-LM test with 12 lags of the squared residuals capturing the ‘ARCH effect’ satisfactory.

Panel B in Table 4 shows the proportion of statistically significant values of the β and γ coefficients for models when we adopt the rolling regression approach. As we can see, including the sentiment index in the variance equation leads to a reduction of the proportion significance of the β and γ parameters by 36% and 77% (at the 1% significance level), respectively. That is, the Google SVI for gold explains much of the volatility persistence while absorbs most of the volatility in the presence of negative shocks.

Overall our results regarding the crude oil market are against the MDH, for both static and rolling regression analysis. The inclusion of information demand variables amplifies the volatility persistence. Another interesting result is the reduction of the significance of the leverage effect. Specifically, the internet-information flow variable reduces by more than half the volatility produced by negative shocks. On the contrary, our results for gold are in line with the MDH considering the rolling regression analysis. Moreover, Google information flows decrease much of the volatility persistence (36%) as well as a large proportion of negative shocks (77%).

Our results may be interpreted in accordance with the nature of information flows coming from the internet, as an expression of the perception of retail investors. We show that the increased searching volume in Google can amplify volatility persistence in crude oil but not

in the gold market. One plausible explanation might be the increasing importance of the gold market in the portfolio choice of investors; both retail and institutional. This has been further enabled by an increase in alternative investment vehicles such as exchange traded funds and the physical gold demand (coins, bars, jewellery) for investment purposes. Retail investors might still consider gold as an investment opportunity, to mitigate some of the risks faced by holding other assets. On the other hand, oil lacks the characteristics of gold. Actions of retail investors which are often considered as uniformed can lead to increase in volatility (for a discussion see Dimpfl and Jank, 2016). In our case, an increase in information demand after a negative shock for oil may generate noise which induces volatility to persist.

Moreover, the inclusion of information demand variable reduces much of the volatility asymmetry generated by negative shocks in both markets. This is in line with the findings of Baur and Dimpfl (2016) for the gold market, who argue that in the case of a negative shock in the gold market investors use gold search queries as a substitute for an immediate reaction. This may reduce the likelihood of overreaction and thus reduce volatility asymmetry. In our case, this holds not only for the gold market but also for the oil market. The latter may also be attributed to the intrinsic characteristics of commodity markets (especially gold) to act as safe haven in times of market turmoil, irrational investors behavior or panic reactions.

6 Conclusions

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 spot returns. Our information demand variables were extracted from a list of searching keywords, following the methodology of Da *et al.* (2015). Our empirical analysis supports the view that Google indicators intensify volatility in both oil and gold markets. This result can be explained by the content of noise trading activity in which internet search may constitute an additional source of volatility. The latter does not hold for gold market when we consider the rolling regression analysis. Moreover, Google information demand on gold and crude oil markets by noise

traders or uniformed investors is found to reduce volatility especially in the presence of a negative shock, when there is an intense need for alternative investment vehicles or “safe havens”. Our analysis allows as to better understand the intrinsic characteristics of gold and oil which can be interpreted in light of behavioural finance and contribute to the discussion on the role households and investors sentiment in the two major commodity markets.

References

- Askitas, N. and Zimmermann, K. F. (2009) Google econometrics and unemployment forecasting, *Applied Economics Quarterly*, 55, 107-120.
- Bampinas, G. and Panagiotidis, T. (2015) Are gold and silver a hedge against inflation? A two century perspective, *International Review of Financial Analysis*, 41, 267–276.
- Bampinas, G. and Panagiotidis, T. (2017) Oil and stock markets before and after financial crises : a local Gaussian correlation approach, Forthcoming in the *Journal of Futures Markets*.
- Bampinas, G., Ladopoulos, K. and Panagiotidis, T. (2018) A note on the estimated GARCH coefficients from the S&P1500 universe, Forthcoming in the *Applied Economics*.
- Barber, B.M. and Odean, T. (2008). All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies*, 21, 785818.
- Barberis, N., Shleifer, A. and Vishny, R. (1998) A model of investor sentiment, *Journal of Financial Economics*, 49, 307343.
- Batten, Jonathan A., Ciner, C., and Lucey, B.M. (2010). The macroeconomic determinants of volatility in precious metals markets. *Resources Policy*, 35, 65–71.
- Baur, D.G. and Dimpfl, T. (2016). Googling gold and mining bad news, *Resources Policy*, 50, 306–311.
- Baur, D.G. and Lucey, B.M. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold, *Financial Review*, 45, 217–229.
- Baur, D.G. and McDermott, T.K. (2010). Is gold a safe haven? International evidence, *Journal of Banking & Finance*, 34, 1886–1898.
- Beckmann, J., Berger, T. and Czudaj, R., (2015), Does gold act as a hedge or a safe haven for stocks? A smooth transition approach, *Economic Modelling*, 48, 16-24.
- Bessembinder, H. and P. Seguin (1993) Price Volatility, Trading Volume, and Market Depth: Evidence from the Futures Market, *Journal of Financial and Quantitative Analysis*, 28, 21–39.
- Bessler, W. and Wolff, D. (2015) Do commodities add value in multi-asset portfolios? An out-of-sample analysis for different investment strategies, *Journal of Banking & Finance* 60, 1–20.
- Bollerslev, T. (1986) Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 31, 307327.
- Bollerslev, T. and I. Domowitz (1993) Trading Patterns and Prices in the Inter-bank Foreign Exchange Market, *Journal of Finance*, 48, 14211443.
- Brailsford, T. J. (1996) The Empirical Relationship Between Trading Volume, Returns, and Volatility, *Accounting and Finance*, 35, 89–111.
- Büyüksahin, B. and Robe, M.A., 2014. Speculators, commodities and cross-market linkages. *Journal of International Money and Finance*. 42, 38–70.
- Caballero, R.J., Farhi, E., Gourinchas, P-O. (2008) Financial Crash, Commodity Prices and Global Imbalances. NBER Working Paper No. 14521.
- Cai, J., Cheung, Y. L., and Wong, M. C. S. (2001). What moves the gold market? *Journal of Futures Markets*, 21, 257278.
- Choi H, Varian H. (2009a). Predicting the present with Google Trends. Economics Research Group, Google.
- Choi H, Varian H. (2009b). Predicting initial claims for unemployment benefits. Economics Research Group, Google.

- Dimpfl, T, Jank, S. (2016). Can internet search queries help to predict stock market volatility?, *European Financial Management*, 22, 171–192.
- Kholodilin, K.A., M. Podstawska, and B. Siliverstovs (2010) Do Google Searches Help in Nowcasting Private Consumption? A Real-Time Evidence for the US, KOF Swiss Economic Institute, *Working Paper* (256).
- Vosen, S. and Schmidt, T. (2011) Forecasting private consumption: survey-based indicators vs. Google Trends. *Journal of Forecasting*, 30, 565578.
- Carrière-Swallow, Y., and Labbè, F. (2013). Nowcasting with Google Trends in an emerging market, *Journal of Forecasting*, 32, 289–298.
- Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41, 135–155.
- Da, Z., P. Gao, and J. Engelberg. (2011) In search of attention, *Journal of Finance*, 66, 146199.
- Da, Z., J. Engelberg, and P. Gao. (2015) The sum of all FEARS investor sentiment and asset prices. *Review of Financial Studies*, 28, 1–32.
- DAmuri, F., and Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment, *International Journal of Forecasting*, 33, 801–816.
- Daniel, K., Hirshleifer, D. and Subrahmanyam, A. (1998) Investors, psychology and security market under- and overreactions. *Journal of Finance*, 53, 1839–1885.
- Daskalaki, C. and Skiadopoulos, G. (2011). Should investors include commodities in their portfolios after all?, *Journal of Banking & Finance*, 35, 2606–2626.
- Dellavigna, S., Pollet, J.M., (2009) Investor inattention and Friday earnings announcements, *Journal of Finance*, 64, 709749.
- Dergiades, T., Milas, C. and Panagiotidis, T., (2015), Tweets, Google trends, and sovereign spreads in the GIIPS, *Oxford Economic Papers*, 67, 406–432.
- Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50, 987–1007.
- Epps, T. W., and Epps, M. L. (1976). The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis, *Econometrica*, 44, 305–321.
- Erb, C.B. and Harvey, C.R., (2006). The strategic and tactical value of commodity futures markets, *Financial Analysts Journal*, 62, 69–97.
- Engle, R.F and Patton, A.J. (2001) What good is a volatility model? *Quantitative Finance*, 1, 237–245.
- Fang, L. and Peress, J. (2009) Media coverage and the cross-section of stock returns. *Journal of Finance*, 64, 2023-2052.
- Fondeur, Y., and F. Karamè (2013) Can Google Data Help Predict French Youth Unemployment?, *Economic Modelling*, 30, 117125.
- Forlani, D. (2002). Risk and rationality: The influence of decision making and perceived outcome control on managers' high-risk decisions. *Journal of Behavioural Decision Making*, 15(2), 125140.
- Gorton, G., Rouwenhorst, K.G., 2006. Facts and Fantasies about Commodity Futures. *Financial Analysts Journal*. 62, 47–68.
- Grossman, S.J. and Stiglitz, J.E. (1980) On the impossibility of informationally efficient markets, *The American Economic Review*, 70, 393-408.
- Hirshleifer, D. and Teoh, S.W. (2003) Limited attention, information disclosure, and financial reporting, *Journal of Accounting and Economics*, 36, 337386.
- Hong, H. and Stein, J. (1999) A unified theory of underreaction, momentum trading and

- overreaction in asset markets, *Journal of Finance*. 54, 2143-2184.
- Huberman, G. and Regev, T. (2001) Contagious speculation and a cure for cancer: a nonevent that made stock prices soar, *Journal of Finance*, 56, 387396.
- Jaffe, J. F. (1989). Gold and gold stocks as investments for institutional portfolios. *Financial Analysts Journal*, 45, 53–59.
- Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioural economics. *American Economic Review*, 93(5), 14491475.
- Kalev, P., Liu, W., Pham, P. K., and Jarnebic, E. (2004). Public information arrival and volatility of intraday stock returns. *Journal of Banking and Finance*, 28, 14411467.
- Lamoureux, C. G. and W. D. Lastrapes (1990) Heteroskedasticity in Stock Return Data: Volume Versus GARCH Effects, *Journal of Finance*, 45, 221–230.
- Merton, R.C. (1987). A simple model of capital market equilibrium with incomplete information, *Journal of Finance*, 42, 483510.
- Nelson, D.B., (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica*, 59, 347–370.
- O'Connor, F., Lucey, B., Batten, J. and Baur, D. (2015). The financial economics of gold-A survey. *International Review of Financial Analysis*, 41, 186–205.
- Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning, *Journal of Financial Economics*, 80, 563602.
- Peri, M., Vandone, D., and Baldi, L. (2014). Internet, noise trading and commodity futures prices, *International Review of Economics & Finance*, 33, 82-89.
- Sims, C.A. (2003). Implications of rational inattention, *Journal of Monetary Economics*, 50, 665–690.
- Smales, L.A. (2014). News sentiment in the gold futures market. *Journal of Banking & Finance*, 49, 275–286.
- Smith, G. P. (2012) Google Internet Search Activity and Volatility Prediction in the Market for Foreign Currency, *Finance Research Letters*, 9, 103–110.
- Suhoy T. (2009) Query indices and a 2008 downturn: Israeli data. Discussion Paper No. 2009-06, Bank of Israel.
- Tang, K. and Xiong, W., 2012. Index investment and the financialization of commodities. *Financial Analysts Journal*. 68, 54–74.
- Tauchen, G. E. and M. Pitts (1983) The Price Variability-volume Relationship on Speculative Markets, *Econometrica*, 5, 485–550.
- Vlastakis, N., and R. N. Markellos (2012) Information Demand and Stock Market Volatility, *Journal of Banking & Finance*, 36, 18081821.

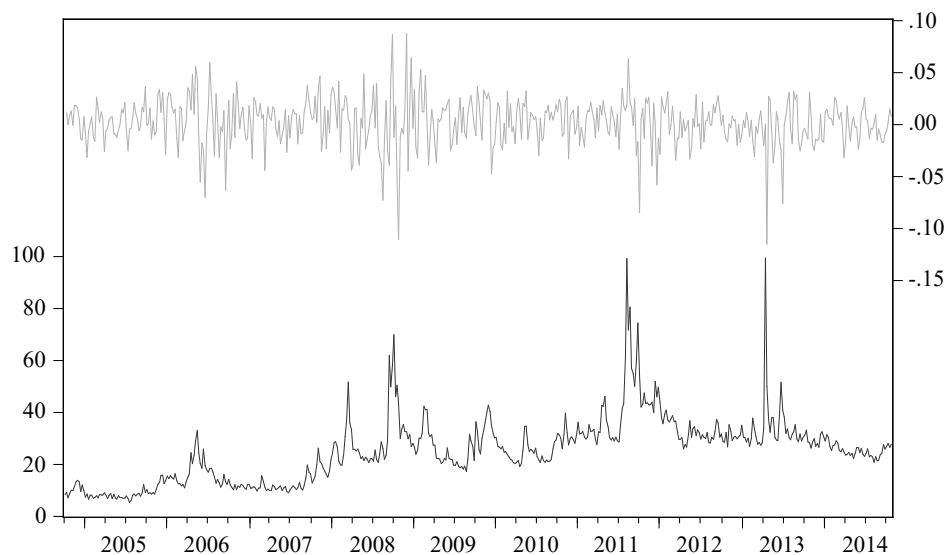


Figure 1: This figure shows the gold spot returns (right axis) and the 60 days Google searching volume cross-sectional average of the keyword “gold price” (left axis).

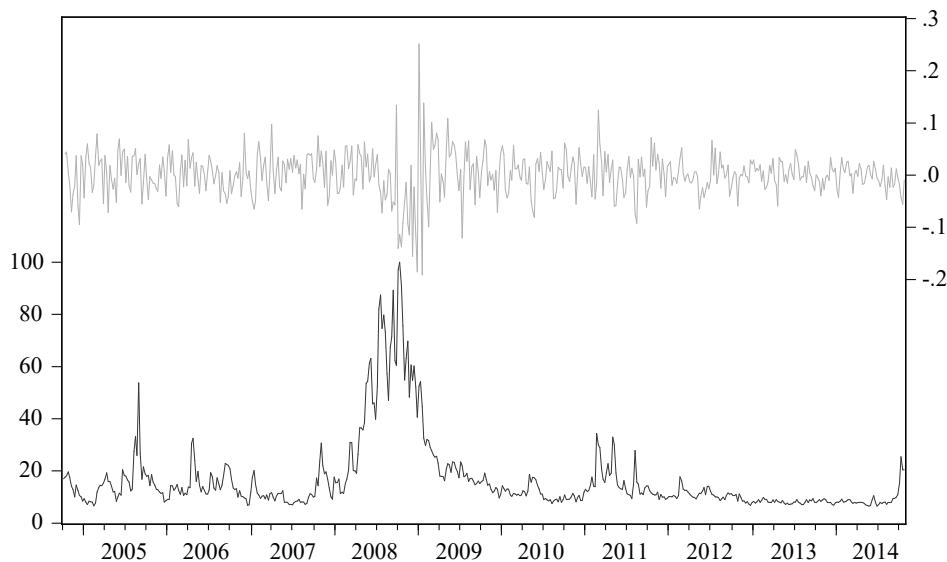


Figure 2: This figure shows the crude oil spot returns (right axis) and the 60 days Google searching volume cross-sectional average of the keyword “oil price” (left axis).

Table 1: This table reports the descriptive statistics of gold and oil weekly returns. J-B denotes the Jarque-Bera test statistic which tests the null hypothesis of normality. The period of study is October 2004 to October 2014.

| | gold price returns | crude price returns |
|-----------|--------------------|---------------------|
| Mean | 0.210 | 0.095 |
| Median | 0.413 | 0.247 |
| Maximum | 8.745 | 25.124 |
| Minimum | -11.497 | -19.099 |
| Std. Dev. | 2.255 | 4.142 |
| Skewness | -0.637 | -0.189 |
| Kurtosis | 6.384 | 7.540 |
| J-B | 286.13 | 454.08 |
| Prob. | 0.000 | 0.000 |
| Obs. | 426 | 426 |

Table 2: This table reports the gold and oil *adjDLSVI* search terms that were used for the Google index construction. The Augmented Dickey-Fuller (ADF) and the Phillips-Perron (PP) unit root tests are applied with constant and with constant and a time trend. The critical values for ADF and PP statistics are taken from MacKinnon (1996). Lag lengths are determined via AIC for ADF test. PP was conducted using Bartlett kernel (Newey-West automatic). *, ** and *** denote statistical significance at the 10%, 5% and 1% level respectively.

| | ADF constant | ADF constant and trend | PP constant | PP constant and trend |
|-----------------------------|-------------------|---------------------------|-------------------|--------------------------|
| <i>Panel A : Gold terms</i> | | | | |
| best gold | -15.261***[0.000] | -15.258***[0.000] | -80.333***[0.000] | -86.371***[0.000] |
| buy gold | -16.363***[0.000] | -16.445***[0.000] | -41.447***[0.000] | -43.856***[0.000] |
| gold rate | -17.218***[0.000] | -17.220***[0.000] | -42.124***[0.000] | -42.960***[0.000] |
| gold bar | -16.803***[0.000] | -16.789***[0.000] | -58.136***[0.000] | -58.206***[0.000] |
| gold market | -16.992***[0.000] | -17.024***[0.000] | -38.173***[0.000] | -38.907***[0.000] |
| gold ounce | -29.988***[0.000] | -29.964***[0.000] | -40.307***[0.000] | -40.592***[0.000] |
| gold price | -19.690***[0.000] | -19.706***[0.000] | -28.464***[0.000] | -29.067***[0.000] |
| gold prices | -26.089***[0.000] | -26.092***[0.000] | -29.492***[0.000] | -29.897***[0.000] |
| gold | -15.012***[0.000] | -15.000***[0.000] | -30.628***[0.000] | -30.682***[0.000] |
| gold spot | -20.926***[0.000] | -20.923***[0.000] | -33.083***[0.000] | -33.280***[0.000] |
| gold bullion | -16.212***[0.000] | -16.223***[0.000] | -47.738***[0.000] | -49.851***[0.000] |
| gold value | -18.789***[0.000] | -18.785***[0.000] | -47.471***[0.000] | -48.672***[0.000] |
| kitco | -19.237***[0.000] | -19.273***[0.000] | -32.270***[0.000] | -33.788***[0.000] |
| of gold | -17.557***[0.000] | -17.548***[0.000] | -35.007***[0.000] | -35.078***[0.000] |
| ounce gold | -29.989***[0.000] | -29.965***[0.000] | -40.497***[0.000] | -40.771***[0.000] |
| precious metals | -24.258***[0.000] | -24.236***[0.000] | -45.107***[0.000] | -45.059***[0.000] |
| price gold | -19.688***[0.000] | -19.704***[0.000] | -28.770***[0.000] | -29.073***[0.000] |
| price of gold | -30.791***[0.000] | -30.787***[0.000] | -38.611***[0.000] | -39.305***[0.000] |
| rose gold | -16.219***[0.000] | -16.209***[0.000] | -98.721***[0.000] | -117.431***[0.000] |
| spot gold | -20.893***[0.000] | -20.891***[0.000] | -33.061***[0.000] | -33.273***[0.000] |
| <i>Panel B : Oil terms</i> | | | | |
| barrel oil | -19.878***[0.000] | -19.899***[0.000] | -28.948***[0.000] | -29.433***[0.000] |
| barrel price | -19.322***[0.000] | -19.304***[0.000] | -28.943***[0.000] | -28.889***[0.000] |
| buy oil | -18.181***[0.000] | -18.208***[0.000] | -48.470***[0.000] | -52.144***[0.000] |
| crude oil price | -20.696***[0.000] | -20.680***[0.000] | -34.379***[0.000] | -34.327***[0.000] |
| crude oil prices | -20.089***[0.000] | -20.070***[0.000] | -30.144***[0.000] | -30.114***[0.000] |
| crude oil | -19.933***[0.000] | -19.973***[0.000] | -28.751***[0.000] | -28.848***[0.000] |
| crude price | -20.331***[0.000] | -20.314***[0.000] | -34.456***[0.000] | -34.383***[0.000] |
| crude prices | -20.298***[0.000] | -20.278***[0.000] | -29.982***[0.000] | -29.955***[0.000] |
| crude | -20.081***[0.000] | -16.975***[0.000] | -29.524***[0.000] | -29.736***[0.000] |
| current oil | -20.367***[0.000] | -20.347***[0.000] | -31.374***[0.000] | -31.338***[0.000] |
| gas price | -19.094***[0.000] | -19.082***[0.000] | -25.870***[0.000] | -25.855***[0.000] |
| oil barrel | -19.868***[0.000] | -19.887***[0.000] | -28.917***[0.000] | -29.387***[0.000] |
| oil chart | -18.054***[0.000] | -18.038***[0.000] | -44.725***[0.000] | -44.838***[0.000] |
| oil futures | -17.636***[0.000] | -17.640***[0.000] | -37.538***[0.000] | -37.593***[0.000] |
| oil price | -20.182***[0.000] | -20.183***[0.000] | -26.231***[0.000] | -26.249***[0.000] |
| oil prices | -20.048***[0.000] | -20.030***[0.000] | -26.557***[0.000] | -26.520***[0.000] |
| oil | -19.816***[0.000] | -19.805***[0.000] | -25.794***[0.000] | -25.792***[0.000] |
| oil stock | -21.249***[0.000] | -21.229***[0.000] | -44.345***[0.000] | -44.321***[0.000] |
| price of oil | -21.041***[0.000] | -21.030***[0.000] | -30.218***[0.000] | -30.226***[0.000] |
| price oil | -20.161***[0.000] | -20.160***[0.000] | -26.084***[0.000] | -26.134***[0.000] |

Table 3: This table reports the EGARCH(1,1) results for the crude oil spot returns. *, ** and *** denote statistical significance at the 10%, 5% and 1% level respectively. ARCH-LM test examines the null hypothesis of no ARCH effects in the standardised residuals. In Panel B, the proportion of statistically significant β and γ coefficient values are calculated from the 376 rolling subsamples of the EGARCH(1,1) baseline and the augmented models

| | ω | γ | α | β | λ | ARCH-LM | AIC | SIC |
|--|-----------------------|-----------------------|---------------------|----------------------|---------------------|------------------|---------|---------|
| <i>Panel A : EGARCH(1, 1) model (Full sample)</i> | | | | | | | | |
| Baseline model | -0.082** [-2.484] | -0.110*** [-3.896] | 0.171*** [4.846] | 0.980*** [102.12] | | 9.052 [0.698] | 5.322 | 5.379 |
| Augmented model | -0.055*** [-2.698] | -0.100*** [-4.422] | 0.059** [-2.147] | 1.004*** [241.79] | 0.275*** [8.868] | 6.304 [0.900] | 5.256 | 5.323 |
| <i>Panel B : EGARCH(1, 1) model (Rolling sample)</i> | | | | | | | | |
| Baseline model | | | | | | | | |
| Significance levels | 1% | 5% | 10% | | | 1% | 5% | 10% |
| Number of significant β coefficients | 138 | 153 | 172 | | | 188 | 209 | 213 |
| Proportion (%) | (36.70) | (40.69) | (45.74) | | | (50) | (53.98) | (56.64) |
| Number of significant γ coefficients | 39 | 72 | 132 | | | 16 | 30 | 64 |
| Proportion (%) | (10.37) | (19.14) | (35.10) | | | (4.25) | (7.97) | (17.02) |

Table 4: This table reports the EGARCH(1,1) results for the gold spot returns. *, ** and *** denote statistical significance at the 10%, 5% and 1% level respectively. ARCH-LM test examines the null hypothesis of no ARCH effects in the standardised residuals. In Panel B, the percentage of statistically significant β and γ coefficient values are calculated from the 376 rolling subsamples of the EGARCH(1,1) baseline and the augmented models.

| | ω | γ | α | β | λ | ARCH-LM | AIC | SIC |
|--|-----------------------|--------------------|---------------------|----------------------|---------------------|------------------|---------|---------|
| <i>Panel A : EGARCH(1, 1) model (Full sample)</i> | | | | | | | | |
| Baseline model | -0.086* [-1.763] | 0.020 [0.495] | 0.229*** [2.845] | 0.936*** [31.315] | | 7.055 [0.853] | 4.333 | 4.390 |
| Augmented model | -0.102*** [-2.261] | -0.005 [-0.166] | 0.182** [2.578] | 0.971*** [45.165] | 0.202*** [2.879] | 6.614 [0.882] | 4.323 | 4.389 |
| <i>Panel B : EGARCH(1, 1) model (Rolling sample)</i> | | | | | | | | |
| Baseline model | | | | | | | | |
| Significance levels | 1% | 5% | 10% | | | 1% | 5% | 10% |
| Number of significant β coefficients | 155 | 171 | 173 | | | 100 | 110 | 125 |
| Proportion (%) | (41.22) | (45.47) | (46.01) | | | (26.59) | (29.25) | (33.24) |
| Number of significant γ coefficients | 21 | 65 | 84 | | | 5 | 10 | 25 |
| Proportion (%) | (5.58) | (17.28) | (22.34) | | | (1.32) | (2.65) | (6.64) |