Is the Internet Search Driving Oil Market? A Revisit through Time-Frequency approaches

Bouoiyour, Jamal and Selmi, Refk

CATT, University of Pau, France, Tunis Business School, Tunisia

August 2015

Is the Internet Search Driving Oil Market? A Revisit through Time-Frequency approaches

Jamal BOUOIYOUR[1]
Refk SELMI[2]*

Version August 2015

Abstract: The present research seeks to address whether internet search drives oil market. For this purpose, we perform two analyses to empirically gauge the relevance of Google search Index as a measure of investors' attention. Firstly, we test if extracting public moods oriented to crude oil using web contents, can help to predict crude oil. Secondly, we analyze the informational content of three oil events (OPEC cuts, 2008 global financial crisis and Libya war) in terms of their effects on the behavior of the crude oil. To achieve this goal, we intend to decompose the causality between attention and oil price into different time scales and frequencies using frequency domain causality test and nonlinear causality test-based wavelet. To ascertain the robustness of our results, we replicate the same testing procedure using another attention proxy which is the number of tweets. The paper decisively confirms that there is a short-run relationship between attention and crude oil. In addition, we show that world crude oil responding to oil events display sharp differentiation. If OPEC cuts had short- and medium-run causality and Libya war exhibits a short-term causality, the attention to global financial collapse had a longer time interval and a wider scale of influence. The first finding implies that internet search is a very practical way to compute investors' attention that can help in predicting short-run fluctuations in the oil market. For the second outcome, different shock origins and distinct properties of oil events may be advanced as possible element of explanation that may exhibit different effects on crude oil.

Keywords: Crude oil; oil events; investors' attention; Google Trends; Twitter; time-frequency approaches.

JEL Classification Numbers: E31, E32, Q43.

[1] CATT, University of Pau, France; E-mail: jamal.bouoiyour@univ-pau.fr

[2] Tunis Business School, Tunisia; E-mail: s.refk@yahoo.fr * *Corresponding author*

1. Introduction

The global oil market was largely threatened by short-run disturbances and sudden shocks, causing crude oil to exhibit swelling volatility, deepening the boost of oil futures market. The latter has widely benefitted from the sharp increase of oil marketisation and the development of electronic carriers. Due to the ever increasing diffusion of high-speed Internet access, Internet-based services are available to more people in the world than ever. The internet has played a substantial role in changing the landscape of oil market. It enables to provide unknown immediacy of news coverage from a huge number of sources. In a sense, Internet search may allow us to find information about unexpected events in real time over the whole world (see for instance, Salaverria 2005 and Mueller and Lemstra 2011).

In fact, the excessive crude oil volatility has exceeded the extent that can be explained by the interplay between supply and demand fundamental factors. As the crude oil varies intensely with events around the world, this study discusses the global financial collapse, the OPEC announcements and the Libyan war. Each of these oil events has its own characteristics. The global financial crisis of 2008 and the Libyan war belong to the category of emergency events, while OPEC conferences seem traditional events that usually arise at an expected time. These different shock origins of these oil-related events cause them to exert a different impact on oil price. Nowadays, searching on Google, social networking on Facebook or Twitter, and video streaming on YouTube are the better ways to spread and receive such news. Via the Internet, information related to oil-related events can be transmitted rapidly, causing large crude oil changes by adjusting the traders' market expectations. Based on these considerations, the Internet search becomes day-to-day a potential tie helping to better predict market prices. Some studies have been done on the effects of retrievable data from Web content or Web usage. Behavioral finance research considers that traders' investment decisions are highly driven by emotion (Damasio 1999 and Dolan 2002). For example, some research show how online information predicts book sales, Grexit, crypto-market, among others (Mishne and Glance 2006, Mitchell et al. 2012; Bouoiyour and Selmi 2015). Moreover, one can argue that information pass-through has heavily influenced the market-pricing mechanism; specifically, crude oil seems potentially sensitive to information disclosure that impacts the formulation of investors' decisions. Indeed, supplementary market factors should be added into the analytical framework to effectively forecast oil prices (Fan and Xu 2011; Guo and Ji 2013; Li et al. 2015).

Intuitively, financial attributes of the oil market sharply stand out, and the impact of external information on crude oil needs further research to gain better insights into a new issue, these concerns underscore the need to pursue a different approach to obtain more reliable measures of the attention to crude oil. This will allow us to effectively address whether public concern on the Web has a great influence on investors' decision making in terms of the oil market. In this paper, based on data series constructed from Internet platform through search query volume on Google and Twitter, the public concerns in the oil market are determined via search queries and the number of tweet backs related to specific search terms. This research provides better paths about how Internet-based data may be used to predict crude oil over different time scales/frequencies.

This study has two primary objectives. The first objective is to test whether extracting public moods oriented to crude oil and major oil events (OPEC cuts, 2008 global financial crisis and Libya war), using web contents, may predict crude oil. Instead of measuring the impact of an event by using dummy variables as it usually done, the present work introduces the concept of Internet concern as a quantitative measure. In computational viewpoint, a dummy variable is restricted to 1 over the period at which the event emerges and thus cannot fully reflect the changing mode of the event. A proper climax of the event can be determined by the degree of the Internet concern for such events. In that context, web search engines are binding tightly bound with the public concerns over the real world. The second aim is to explore if the attention' effects on crude oil seem transitory or structural by disentangling the short-, the medium- and the long-run causality between the focal variables. Unlike the causal standard approaches that consider the direction of causality between economic variables across the full studied period, the time-frequency causality enables to evaluate causality among different time-frequencies. Basically, the standard measurement of the causality for various periods may be performed by standard models by subdividing the sample period to well specific sub-periods. This procedure seems vulnerable since it is based on a relatively small number of data that may threaten the robustness of the results. To avoid this problem, the time-frequency analysis seems able to keep all the observations over the period of investigation in each of the frequencies involved. To our best knowledge, there is only one study focusing on the examination of the causality between attention to oil-related events and oil price in two main periods using co-integration model (short-run and long-run). Guo and Ji (2013) consider the causality between attention to oil events and oil price without accounting for additional control variables that may explain potentially the focal relationship, which may be ineffective.

Studying the bivariate linkage may not be robust when some relevant explanatory variables are not included[3]. The investigation of bivariate dependence-based unconditional data analysis may lead to confusing outcomes since the occurrence of noise cannot be heavily neglected, disrupting then the linkage investigated (Aguiar-Conraria and Soares 2011 and Bouoiyour et al. 2015). This highlights the importance of the consideration of potential control variables (multivariate-based conditional data analysis) to gain clearer and conclusive insights into a new topic.

By applying frequency domain causality test and scale-by-scale nonlinear causality test (based on Taylor approximation within wavelets) , our results indicate that attention, which is measured by Google search queries related to "crude oil", Granger-cause the oil price in the higher frequencies, highlighting their ability to serve as oil market short-run predictors. Besides, our findings reveal that crude oil responding to oil events display sharp differentiation, conditioning upon potential control variables. If OPEC announcements had a short- and medium-run causality and Libya war exhibits a short-term causality (, the attention to global financial collapse had a persistent (structural) effect. Differences in shock origins and distinct properties of oil events are advanced as elements of explanations that may exhibit different impacts on crude oil. To ensure the robustness of our results, we replicate the same testing procedure using another attention proxy (Twitter). These findings appear fairly robust across the different methods employed and the distinct attention proxies used.

The structure of the article is as follows: Section 2 presents a brief background about the crude oil and the major oil-related events. Section 3 describes the data and presents the empirical strategy. Section 4 reports our main findings and discusses them. This section also offers a robustness check. Section 5 concludes.

2. Crude oil and oil-related events: Background

The main aim of this article is to assess the traders' investment behaviors oriented to oil events on crude oil. As all commodities, the oil price is determined by the interplay between

[3] When we consider only two variables, we generally fall on the problem of simple regression without control variable which is unable to capture appropriately proper results.

supply and demand; hence crude oil may be highly influenced by supply-side shocks, aggregate demand shocks and precautionary demand shocks[4].

As we know, the supply of oil depends potentially on geological discovery, the legal and tax framework for oil extraction, the cost of extraction, and the political situation in oil-producing countries. The political instability in oil producing countries as well as conflicts with other countries can threaten heavily energy market by destabilizing the oil price. However, this research focuses three oil events: the global financial collapse of 2008 which is representative of aggregate demand shock, the Libyan war after the Arab uprisings and OPEC announcements (precautionary demand shocks). Each of these oil-related events has its unique features. The global financial crisis and the Libyan war represent emergency events, while OPEC decisions correspond to traditional events that usually arise at a specific time.

The release of the Organization of Petroleum Exporting Countries (OPEC) decisions usually exerts a substantial impact on oil price. OPEC's announcements affect potentially the expectations and may lead to swelling volatility of the crude oil markets. The announcements may be in the form of "cut", "maintain", and "increase" decisions regarding changes in oil production levels. This research focuses on OPEC cutbacks that have played a pulling role in explaining current expectations in the markets and the different investment and speculative strategies. While OPEC decisions are important for better understanding oil markets, it is worth noting that few studies devote attention to these conferences (Mensi et al. 2014).

The 2008 financial collapse is perceived as the worst economic crisis since the Great depression of the 1930s. The 2008 crisis prompted financial institutions collapsing and countless companies. As immediate consequence, oil prices plummeted owing to low external demand to fall on December 23, 2008, to US$30.28 a barrel. The demand from the Organisation for Economic Co-operation and Development (OECD) countries came to a sudden halt and recession loomed as the financial crisis skewed the balance between oil supply and demand. In response and in an essay to curb the drop of crude oil price, the OPEC introduced a series of cuts. As a result, oil prices have begun to stabilize at levels ranging in the mid 2009 US$ 40 per barrel. Given these considerations, the market internet concern related to the global financial crisis jumped remarkably in 2008. However, this heavy increase

[4] Based on Kilian (2009)'s study, the supply-side shocks are shocks mainly owing to the current physical availability of crude oil; the aggregate demand shocks are shocks to the demand itself generated by excessive fluctuations in the global business cycle; and precautionary demand shocks occur from the uncertainty about shortfalls in the expected supply relative to the expected demand.

in investors' attention notably dissipated since in 2009 as the panic alleviated and the world economy began to recover slowly.

On 2011, the Brent price reaches $100 a barrel for the first time since the end of 2008, due to the "Arab Spring" events. The increase in oil prices that coincided with Arab uprisings has great attention about the main causes for oil price dynamic fluctuations. The reasons behind this increase in oil price may include the possibility of delays in the transit of oil tankers through the Suez Canal during the unrest in Egypt and the shut-off of oil from Libya (which had been exporting 1.3 million barrels of oil per day before the aftermath of "Arab Spring"). The Libyan war led to noticeable cuts in oil production (by approximately 90%). Although Libya represents only 2% of the global oil supply, and the disruption of crude oil exports from Libya can be effectively avoided by a surplus production from Saudi Arabia, the Libya's war impacted substantially Brent prices, which increased by almost 20%. It must be stressed here that unrest in Syria, Yemen, Bahrain, as well as Saudi Arabia, the world's largest oil producer have largely influenced the World oil market. Given this last consideration, OPEC members announced their commitment to meet any shortfalls in demand due to the sudden uprisings and increased deeply the production. Nevertheless, Saudi Arabia declared that, due to the decrease of global demand, the oil production dropped by approximately 800,000 barrels per day, and it blamed speculative trading for a sharp increase in the crude oil price.

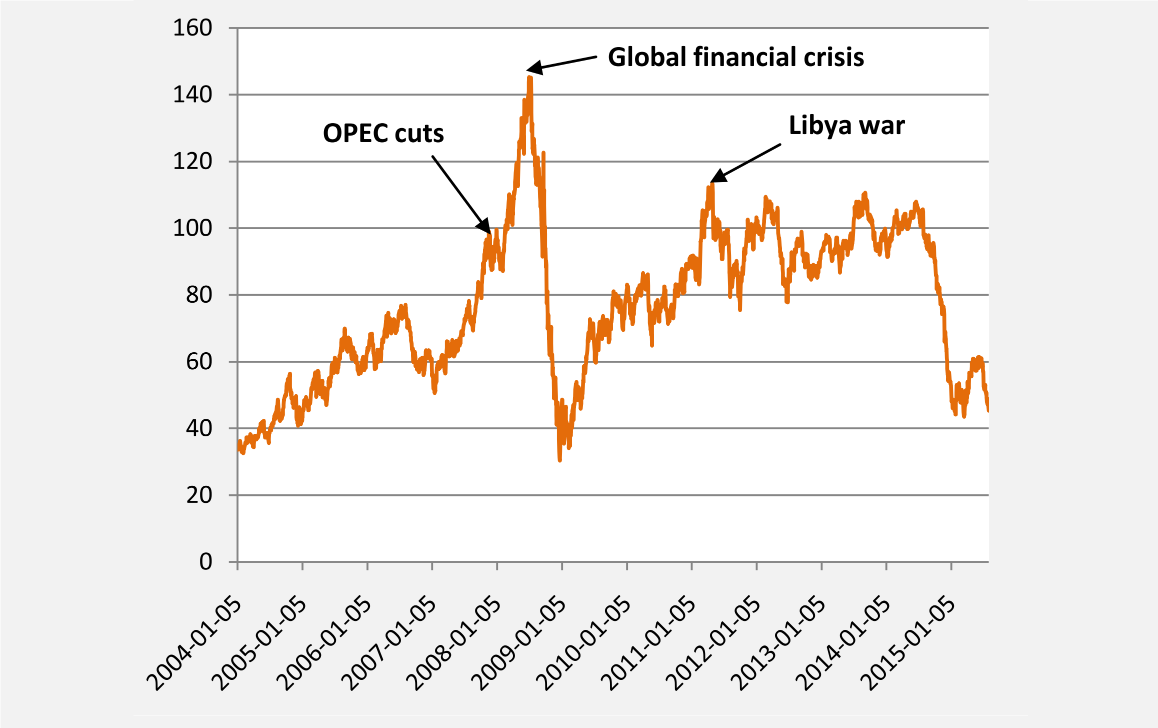
Figure 1 clearly indicates that oil price experienced several jumps and excessive swings over the period spanning between 2004 to 2015[5]. The figure depicts the greater appreciations and precipitous depreciations in oil price due essentially to the events that undergoes World oil market. The sizable volatility of oil price, the differences of origins of oil-related events and their possible dramatic consequences highlight insightful hypotheses to be tested throughout the rest of this study:

***H_1*: Does the extraction of investors' moods oriented to crude oil may help to properly predict crude oil?**

H2: Do different oil events may exhibit distinct effects on crude oil?

[5] For details about the reasons behind our choice of this period, please see data section.

Figure 1. Crude oil evolution and oil-related events



Source: US Energy administration.

3. Methodology and data

Many economists are aware that there is actually more time scales in between the short run and the long run; hence, the decomposition of the evolution of economic phenomena in distinct scales or frequencies help better understand their structure and also to see more accurately their behavior through different periods (short-, medium- and long-run). The application of frequency approach in studying macroeconomic time series appears late compared to other disciplines, greatly motivated by the hope to gain full insights about the structure of economic relationships. Considering that a series can have repetitive or cyclical behavior seems highly important because it allows us to distinguish between the spectral analysis and standard investigations through conventional methods which assume that time series have an independent behavior over time.

The causality analysis is largely stands on the seminal work by Granger (1969) and his followers. This study considers that the Granger causality test may yield different results for distinct time scales/ horizons of the data. Building on, we use frequency domain causality test and scale-by-scale nonlinear causality test based on Taylor approximation within discrete wavelet framework.

3.1. Frequency domain causality

The Granger causality can be analysed under different horizons. For example, Geweke (1982) and Hsoya (1991) proposed measures for Granger causality under a frequency domain framework. Given their usefulness, testing frequency-by-frequency causality has been widely and extensively examined (Breitung and Candelon 2006) and then has been employed in several researches (Bodart and Candelon 2009; Bouoiyour et al. 2015, etc…). The Breitung and Candelon (2006)'s test disentangles the short-, medium- and long-run relationship between the variables studied.

To meticulously present the Breitung and Candelon (2006)'s contribution, let start by defining $z_t = [x_t, y_t]$ as a two-dimensional time series vector with t = 1,…, T. It is supposed that z_t has a finite-order VAR representation $\theta(L)z_t = \varepsilon_t$

where $\theta(L)z_t = 1-\theta_1 L - ... - \theta_p L^p$ is a 2 × 2 lag polynomial with $L^k z_t = z_{t \to k}$. It is assumed that the vector ε_t is white noise with $E(\varepsilon_t) = 0$ and $E\ (\varepsilon_t\varepsilon_t') = \Sigma$, $E(\varepsilon_t \varepsilon_t') = \sum$ where $\sum$ is a positive definite matrix. Next, let G be the lower triangular matrix of the Cholesky decomposition $GG' = \sum -1$, such that $E(\eta_t \eta_t') = 1$ and $\eta_t = G\varepsilon_t$ ηt = Gεt. The system is stationary expressed as follows:

$$z_t = \phi(L)\varepsilon_t = \begin{pmatrix} \phi_{11}(L)\phi_{12}(L) \\ \phi_{21}(L)\phi_{22}(L) \end{pmatrix} \begin{pmatrix} \varepsilon_{12} \\ \varepsilon_{21} \end{pmatrix} \tag{1}$$

$$= \Psi(L)\upsilon_t = \begin{pmatrix} \psi_{11}(L)\psi_{12}(L) \\ \psi_{21}(L)\psi_{22}(L) \end{pmatrix} \begin{pmatrix} \eta_{12} \\ \eta_{21} \end{pmatrix}$$

Then, the spectral density can be derived from the previous matrix and written as following:

$$f_x(w) = \frac{1}{2\pi} \left\{ \left|\psi_{11}(e^{-iw})\right|^2 + \left|\psi_{12}(e^{-iw})\right|^2 \right\} \tag{2}$$

The measure of causality proposed by Geweke (1982) and extended by Hosoya (1991) is represented as follows:

$$M_{y\to x}(w) = \log\left[1+\frac{\left|\psi_{12}(e^{-iw})\right|^2}{\left|\psi_{11}(e^{-iw})\right|^2}\right] \tag{3}$$

As $\left|\psi_{12}(e^{-iw})\right|^2$ seems a "complex" nonlinear function of the VAR parameters[6], Breitung and Candelon (2006) and in order to resolve this drawback argue that the hypothesis $M_{y\to x}(\omega) = 0$ correspond to a linear restriction on the VAR coefficients

$$H_0 : R(\omega)\phi(L) = 0 \tag{4}$$

where $R(\omega) = \begin{bmatrix} \cos(\omega)\cos(2\omega)\ldots\cos(p\omega) \\ \sin(\omega)\sin(2\omega)\ldots\sin(p\omega) \end{bmatrix}$

Thus, its significance can be tested by a standard F-test. Based on their evidence, to assess the significance of the causal relationship, we compare the causality measure for $\omega \in [0,\pi]$ with the critical value of a χ^2 distribution with 2 degrees of freedom, which is 5.99.

3.2. Scale-by-scale nonlinear Granger causality test

The wavelet transform has been defined by Daubechies (1992) as "a tool that cuts up data or functions into different frequency components, and then studies each component with a resolution matched to its scale." Thus, wavelet analysis enables to separate each variable into its constituent frequency components. This method is of interest as it relates the relationship studied to the frequency at which activity in the time series occurs. There are at least two types of wavelets: father wavelet ϕ which represents the low frequency and mother wavelet ψ which captures high-frequency components expressed, respectively, as follows:

$$\phi(x) = \sqrt{2}\sum_k l_k \phi(2x-k) \tag{5}$$

$$\psi(x) = \sqrt{2}\sum_k h_k \phi(2x-k). \tag{6}$$

where l_k and h_k are respectively the low-pass and high-pass filter coefficients. $l_k = \frac{1}{\sqrt{2}}\int \phi(t)\phi(2t-k)dt$; $h_k = \frac{1}{\sqrt{2}}\int \psi(t)\phi(2t-k)dt.$

[6] For details, you can refer to Gradojevic and Lento (2015).

A wavelet decomposition of a function $f(t)$ can be defined as a sequence of projections into father and mother wavelets $s_{J,k}, d_{J,k}, \ldots, d_{1,k}$, which can be written as follows:

$$s_{J,k} \approx \int \phi_{J,k}(t) f(t) dt \tag{7}$$

$$d_{j,k} \approx \int \psi_{j,k}(t) f(t) dt \text{ ,j=1,2.... } J. \tag{8}$$

where $s_{J,k}$ is the smooth behavior of the signal at a specific time scale. The coefficients $d_{j,k}$ represent deviations from the trend.

At this stage, the wavelet decomposition is expressed as following:

$$f(t) = \sum_k s_{J,k}\phi_{J,k}(t) + \sum_k d_{J,k}\psi_{J,k}\ (t) + \sum_k d_{J-1,k}\psi_{J-1,k}(t) + \ldots\ldots + \sum_k d_{1,k}\psi_{1,k}(t) \tag{9}$$

where J is the number of multi-resolution levels.

After decomposing the variables studied into different scales, we test the two hypotheses mentioned in the previous section through a nonlinear causality test proposed by Péguin-Feissolle et al. (2008). Unlike standard linear Granger which loses a great problem when the nexus between the variables of interest seems nonlinear, the test of Péguin-Feissolle et al. (2008) is based on a Taylor expansion of the nonlinear model y_t around a well specific point expressed as follows:

$$y_t = f^*(y_{t-1}, \ldots, y_{t-q}, x_{t-1}, \ldots, x_{t-n}, \theta^*) + \varepsilon_t \tag{10}$$

where θ^* is a parameter vector and ε_t ~ $nid(0,\sigma^2)$; the sequences x_t and x_t are weakly stationary. The functional form of f^* is unknown but we assume that is adequately represents the causal relationship between x_t and y_t. While trying to test noncausality hypothesis, we start by the fact that x_t does not cause y_t if the past values of x_t does not contain any information about y_t that is already contained in the past values of y_t itself. Following Péguin-Feissolle and Teräsvirta (1999), we linearize f^* in (10) by expanding the function into a k-order Taylor series around an arbitrary fixed point in the sample space.

It must be stressed here that the unconditional causality (bivariate causality) is vulnerable. . Accurately, when we account only for two variables, it will be difficult to capture solid and unambiguous findings with regard to the focal interaction dynamic since it may distort the estimate. Ignoring relevant control variables when investigating causal links between two time series may beget to overestimation of the causality structure among other statistical properties; hence the usefulness of employing causality-based conditional data analysis (Bouoiyour et al. 2015) by adding relevant explanatory variables including World oil demand and supply.

3.3. Data

For empirical purpose and while trying to test if attention through Google search Granger-cause crude oil fluctuations under specific horizons, we use quarterly time-series data related to crude oil, the interest to crude oil and the attention to oil related events (OPEC's cuts, 2008 global financial crisis and Libya war) for the period spanning between 2004 and 2015. The short time range of our time series data is due to search engine data availability. To measure trader attention to crude oil price and then to the aforementioned major oil events, we obtain Internet search activity data from Google search queries[7] (http://www.google.com/trends), a tool we selected that can ensure the internationality of the basic data in the present article. This Google service provides accurate and proper data, indicating how frequently search terms have been used in the Google Search engine. It seems obvious that market participants looking for information about oil price use many possible search terms. For the present study, we use Google Trends that display the number of searches for the search terms "crude oil", "OPEC cuts", "global financial crisis" and "Libya war". We should mention here that the daily data related to Google search for the period between 01/12/2004 and 01/07/2015 are converted to quarterly data (quarter)[8]. The crude oil price and the additional control variables (World oil demand and World oil supply[9]) are collected from quandl website

[7] Google search queries on the Web are subsets of the public concerns in the real world, and have enough large representative samples from about 100 countries. The searches performed in Google each day amounted approximately 4.7 billion until 2011 (Guo and Ji 2013). Web search engines are becoming the more effective tool for reaching clearer information on the Web. While attempting to compare the search data, Google Trends findings are normalized. More precisely, search data are divided by a common variable, which is here the total searches, while avoiding the variable's effect on the data (Google 2014). It is important to mention here that the google search algorithm has been edited countless times by Google Inc in order to optimize the google search Index.

[8] We have used day_covert Matlab code (http://www.aoni.waseda.jp/motegi/day_convert.txt). The data were converted to quarterly frequency, since one of the main contributions of our research is to see whether the nexus between trader' investments attention and oil market seems conditioning upon potential control variables. However, the relevant fundamentals of crude oil used (World oil demand and World oil supply) are available on quarter-by-quarter basis.

[9] It is well known that when all the factors that could affect the price of oil are considered, the most influential remain supply and demand; hence the relevance of the choice of these explanatory variables.

(http://www.quandl.com). We transform all the considered variables by taking natural logarithms to correct for potential heteroskedasticity and dimensional differences between the series. Then, we first-difference the time series studied to generate quarter-on-quarter time series and standardize them to exhibit a zero mean and variance of one.

For robustness check, we have used social networking data (Twitter). Note that for twitter, we use the tweet backs related to the keywords "crude oil" and "global financial crisis"[10].

4. Findings discussion

4.1. Main frequency-domain causality results

4.1.1. The nexus between the attention to "crude oil" and crude oil

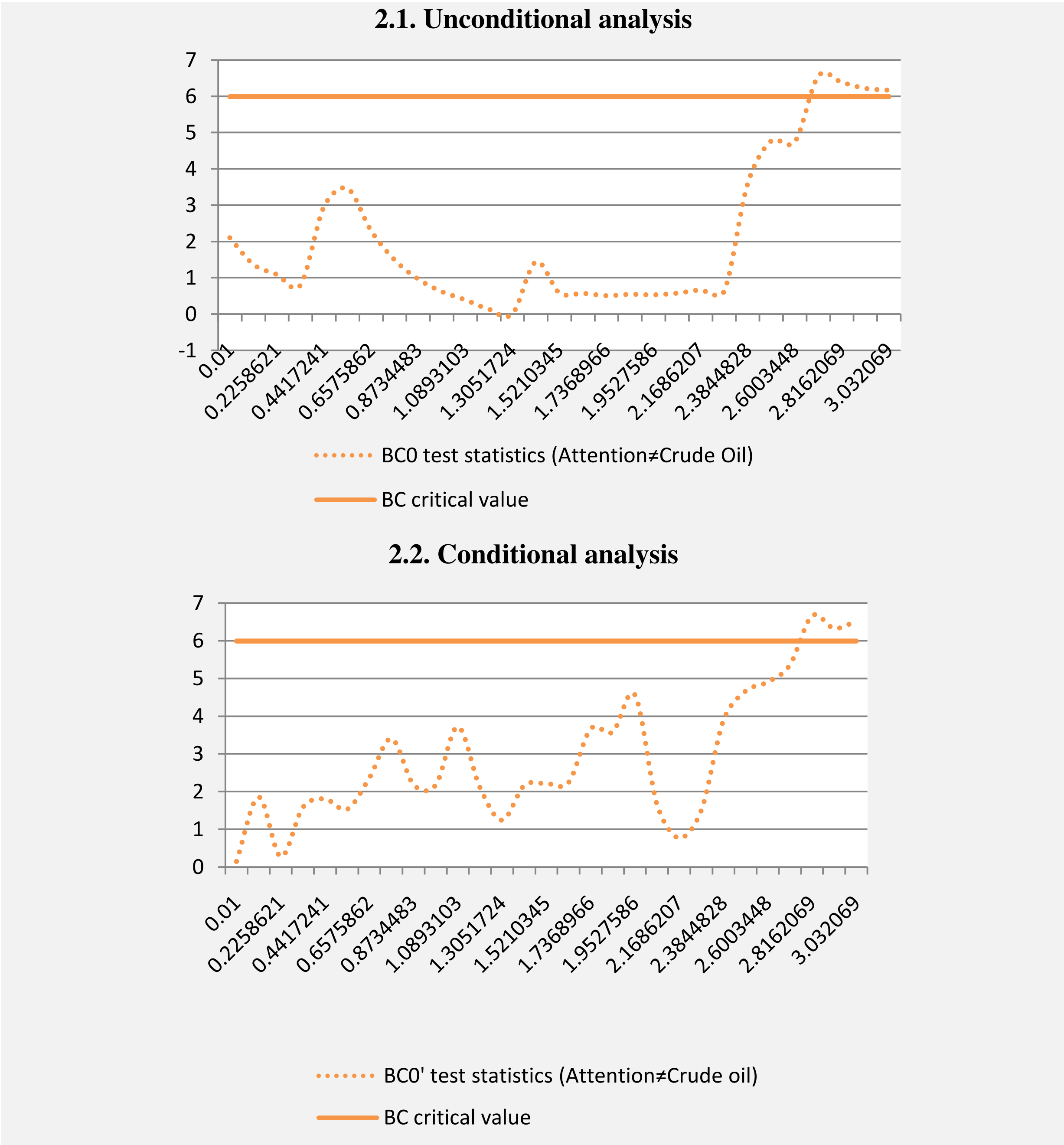
Frequency domain causality was exploited here in the form of a forecast combination in order to get accurate prediction of crude oil price across well specified horizons for the period 2004q1-2015q2. The figure contains the test statistics with their 5 percent critical values (dashed line) over the interval [0, π]. We assess if investors' attention Granger-cause crude oil among different frequencies involved. The results reveal that the null hypothesis of Google queries no Granger-cause crude oil price is rejected for ω less important than 2.49 (less than 2.50 quarters) or when $\omega \in [0.79\pi; 0.96\pi]$[11] (graph2.1, Figure 2). When oil demand and oil supply are accounted for as potential control variables, the results change slightly; we note that the cycle length becomes shorter. More precisely, the null hypothesis is verified for ω moving between 0.89π and 0.96π (less than 2.2 quarters, graph 2.2, Figure 2). The results obtained in this study clearly reveal that Google search query is a reliable proxy to measure attention of investors to oil market in the short-run. Accurately, the search volume data from Google Trends contains valuable information that may help investors in predicting crude oil changes in shortest time horizons. The observed outcomes also indicate that the relationship between traders' attention and crude oil seems deeply conditional to relevant control variables including World oil demand and supply.

[10] We test the robustness for only one oil-related event (global financial collapse) due to Twitter data availability for the rest of search terms (OPEC cuts and Libya war).

[11] Recall that the frequency (ω) on the horizontal axis can be translated into a cycle or periodicity of T quarters by $T = (2\pi / \omega)$, where T is the period.

Summing up, either unconditionally or conditionally, the great anxiety of investors about crude oil uncertainty exerts a powerful role in explaining the crude oil, but in the higher frequencies (short-run). This means that the first hypothesis to be tested (***H1***) is well supported. This result (short-run relationship) seems intuitive since oil price is obviously determined by its fundamentals. In the long-run, the demand and supply play the major roles in explaining the changes in oil price.

Figure 2. The frequency-domain causality between attention to crude oil and crude oil



Note: The horizontal line represents the 5% critical value of the null hypothesis test of no Granger causality at frequency *w*.

4.1.2. *The nexus between the attention to oil-related events and crude oil*

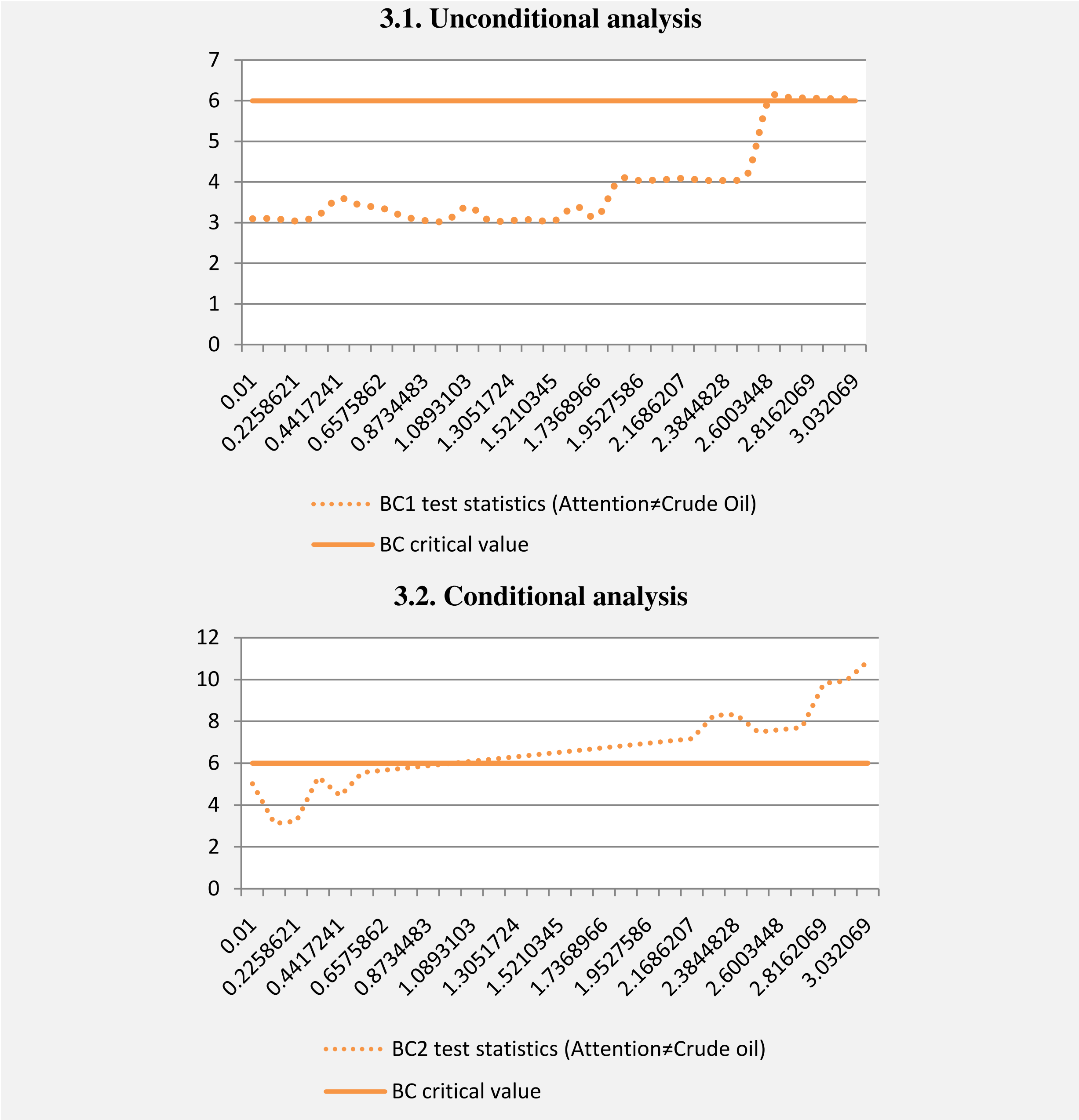
Figures 3, 4 and 5 depict the frequency causality between three oil-related events and crude oil. From the causality measures over frequencies from zero to π, we emphasize three main findings.

Figure 3 shows the predictive power of Google search trends related to the search term "OPEC cuts" for crude oil over the period 2004q1-2015q2. We worthy note that the null hypothesis of attention to OPEC production cuts no Granger-cause crude oil is rejected at the 5% significance level when $w \in [0.82\pi;0.96\pi]$. This implies that high frequencies of Google trends (short-run cyclical components), with wave length of less than 2.3 quarters (graph 3.1), is that which causes crude oil price. These results change remarkably when incorporating World oil demand and supply as control variables. The cycle length becomes much longer; a causality from attention to oil is verified in the short and medium terms, when $w \in [0.34\pi;0.96\pi]$ (graph 3.2). Notably, the consideration of control variables changes the cycle at which the causality is well supported. The cycle length reaches 5.8 quarters. Expectedly, the announcement of a production cut has a positive effect on oil market (Mensi et al. 2014). This study provides evidence that, under the production cut scenario, there is a medium- and short-run causality running from these OPEC announcements and crude oil when demand and supply are considered, implying that traders or investors need time to effectively digest the information that is released, and that the influence of that information will arise and persist over a short period of time. Accordingly, Mensi et al. (2014) show that the volatility of crude oil price increases remarkably in the OPEC's post-announcement period of the cut decisions and add that the options traders may formulate good decisions before the meetings when they see the oil market is in glut. In that context, the OPEC announcements may serve as exploitable opportunities, but may also prompt great speculation behavior.

Unlike the other oil-related event (OPEC cuts), the 2008 global financial crisis had a longer time interval and a wider scale of influence. We find, from Figure 4 (unconditional analysis), a significant causality at lowest frequencies (above 28.3 quarters). Precisely, the null hypothesis of attention to global financial crisis no Granger-cause crude oil is rejected at the 5% significance level when $w \in [0.01;0.22]$. . Conditioning upon additional control variables, the null hypothesis is well rejected at the cycle between 0.01 and 0.65, and thus a long-run causality seems supported when w is above 9.6 quarters. Normally, oil events trigger short term influences on oil market. However, the 2008 crisis is seemingly a structural crisis (long term). This crisis has affected intensely the whole world prompting a massive de-leveraging by financial institutions and then to a collapse in the price of oil (Watts, 2009). These dramatic consequences made central banks around the world to cut interest rates and several governments act by implementing economic stimulus packages to promote economic growth and to enhance confidence in the financial markets. Oil exporters are not an exception. According to the Economist Intelligence Unit, the demand fell by 0.2 per cent in 2008 and by 0.4 per cent in 2009. Plummeting world demand is largely driven by the great drop in consumption among European countries. There is also a decline by approximately 2.9 per cent in oil demand in OECD countries in 2008. The financial crisis contributes widely in deepening the uncertainty in the demand for oil, accentuating the uncertainty about the crude of oil price and leading to disequilibrium between oil demand and supply. As it is very hard to predict whether policies pursued did or did not succeed in delaying the dramatic consequences of global financial collapse and in arresting contagion among Eurozone, the internet concern persisted after the 2008 financial crisis broke out.

Figure 5 describes the predictive power of the attention to Libya war for crude oil. Unconditionally, we clearly show a short-run causality. Notably, the null hypothesis that Google search queries related to "Libya war" does not Granger-cause crude oil is rejected for the frequency, ω, lower than 2.60 corresponding to a cycle of 2.4 quarters (graph 5.1). Our evidence remains meaningful, even after incorporating potential control variables. Accurately, we support a causality at highest frequencies (short-run), i.e., when ω is lower than 2.92 corresponding to a cycle length 2.1 quarters (graph 5.2). When the Libyan war broke out, the Internet concern exerts an impact on crude oil but only in the short-run. Unsurprisingly, the investors searching for information on the Internet (in particular, Google) as a reaction to Libya war news will not persist since well cognizant by traders in oil market that Libya represents only 2% of the global oil supply, and thus a drop of oil exports from Libya can be easily cancelled out by a surplus production from major oil exporting countries like Saudi Arabia. From these outcomes, we can confirm the hypothesis (***H2***) thereby the different oil-related events exert different effects on crude oil price.

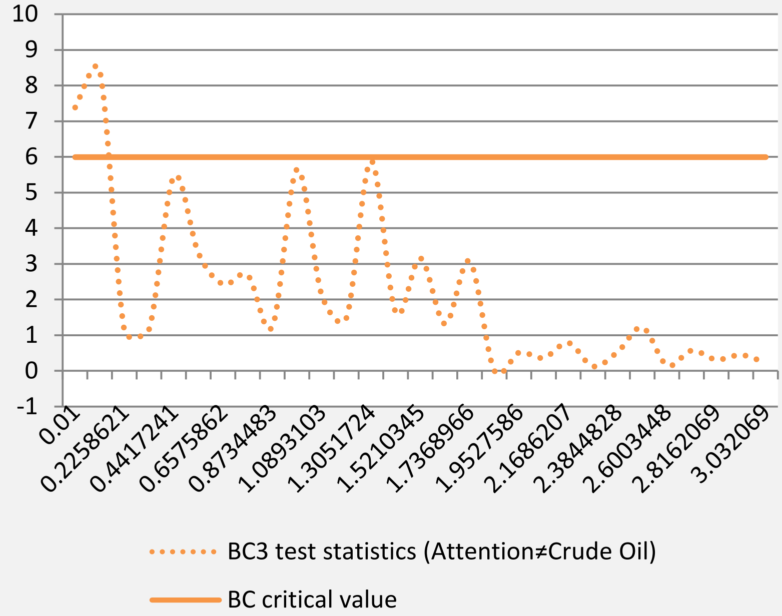
Figure 3. The frequency-domain causality between attention to OPEC cuts and crude oil



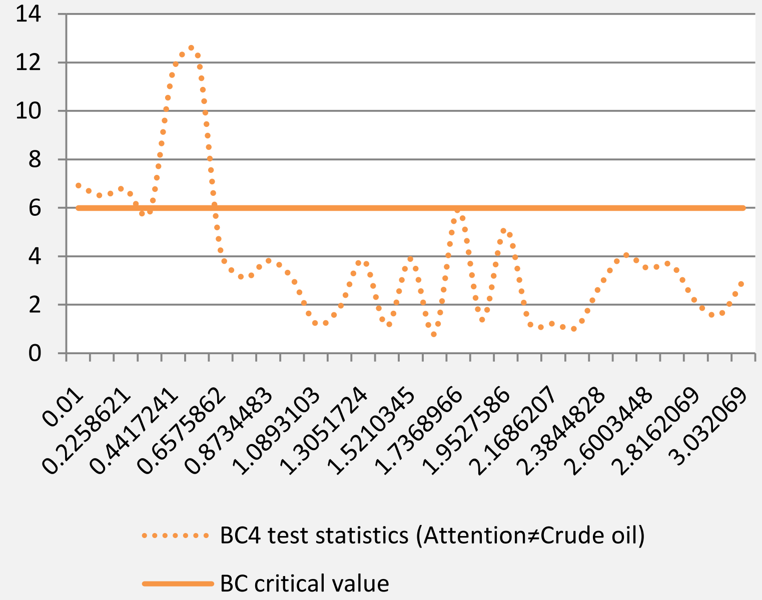
Note: The horizontal line represents the 5% critical value of the null hypothesis test of no Granger causality at frequency *w*.

Figure 4. The frequency-domain causality between attention to global financial crisis and crude oil

4.1. Unconditional analysis



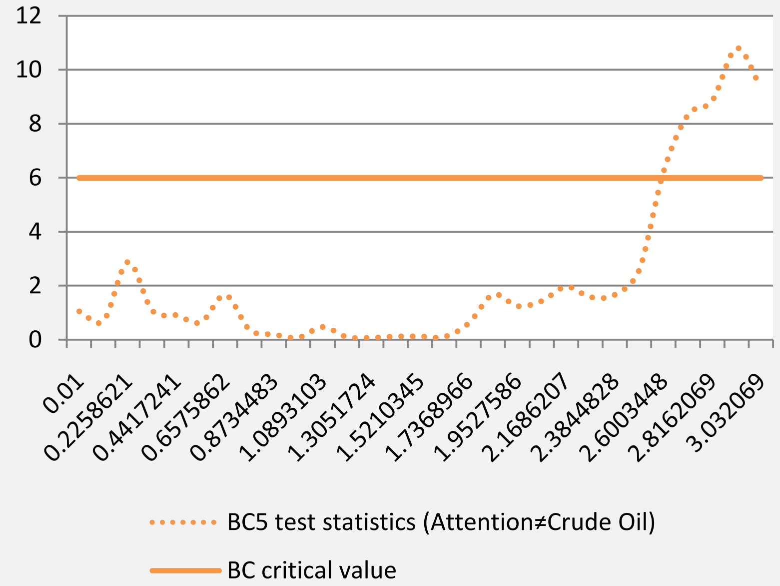
4.2. Conditional analysis



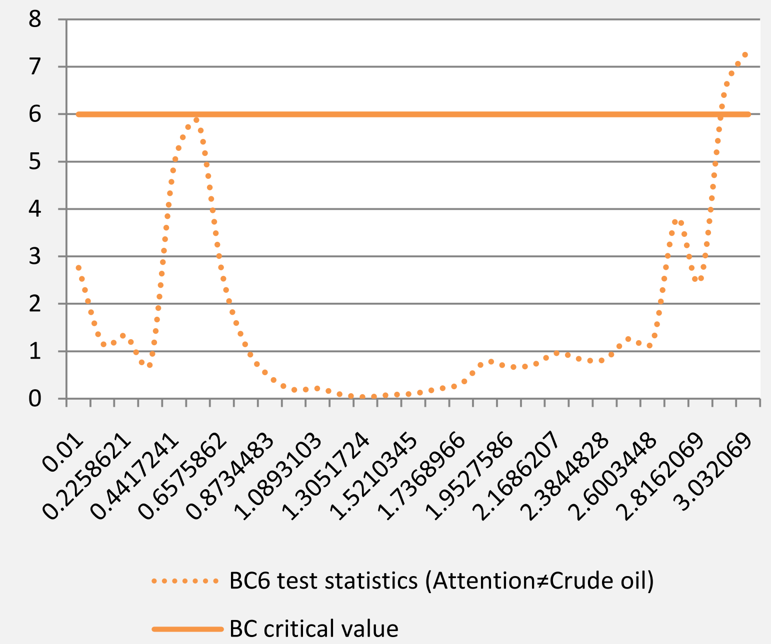
Note: The horizontal line represents the 5% critical value of the null hypothesis test of no Granger causality at frequency *w*.

Figure 5. The frequency-domain causality between attention to Libya war and crude oil

5.1. Unconditional analysis



5.2. Conditional analysis



Note: The horizontal line represents the 5% critical value of the null hypothesis test of no Granger causality at frequency *w*.

4.2.Main discrete wavelet-based nonlinear causality test results

4.2.1. Preliminary analysis

A nonlinear causality test based on Taylor approximation within different time scales through discrete wavelet decomposition has been applied. This application consists, in first step, to decompose the variables of interest and the additional control variables into different scales. Specifically, in order to perform a wavelet decomposition of attention to "crude oil", attention to the central oil-related events, crude oil (in level), oil demand and oil supply in a set of five orthogonal components D1, D2, D3, D4, D5, that stand for different dealing frequencies over the period 2004q1-2015q2 (from low to medium and high frequencies), we choose the Symmlet basis. The latter is orthogonal, near symmetric and have a compact support and good smoothness properties. Table 1 presents the time scale interpretation of discrete wavelet multiresolution assessment; each time scale corresponds to a well specified time-scale of a category of traders at the oil market.

Table 1. Frequency interpretation of scales based on discrete wavelet decomposition

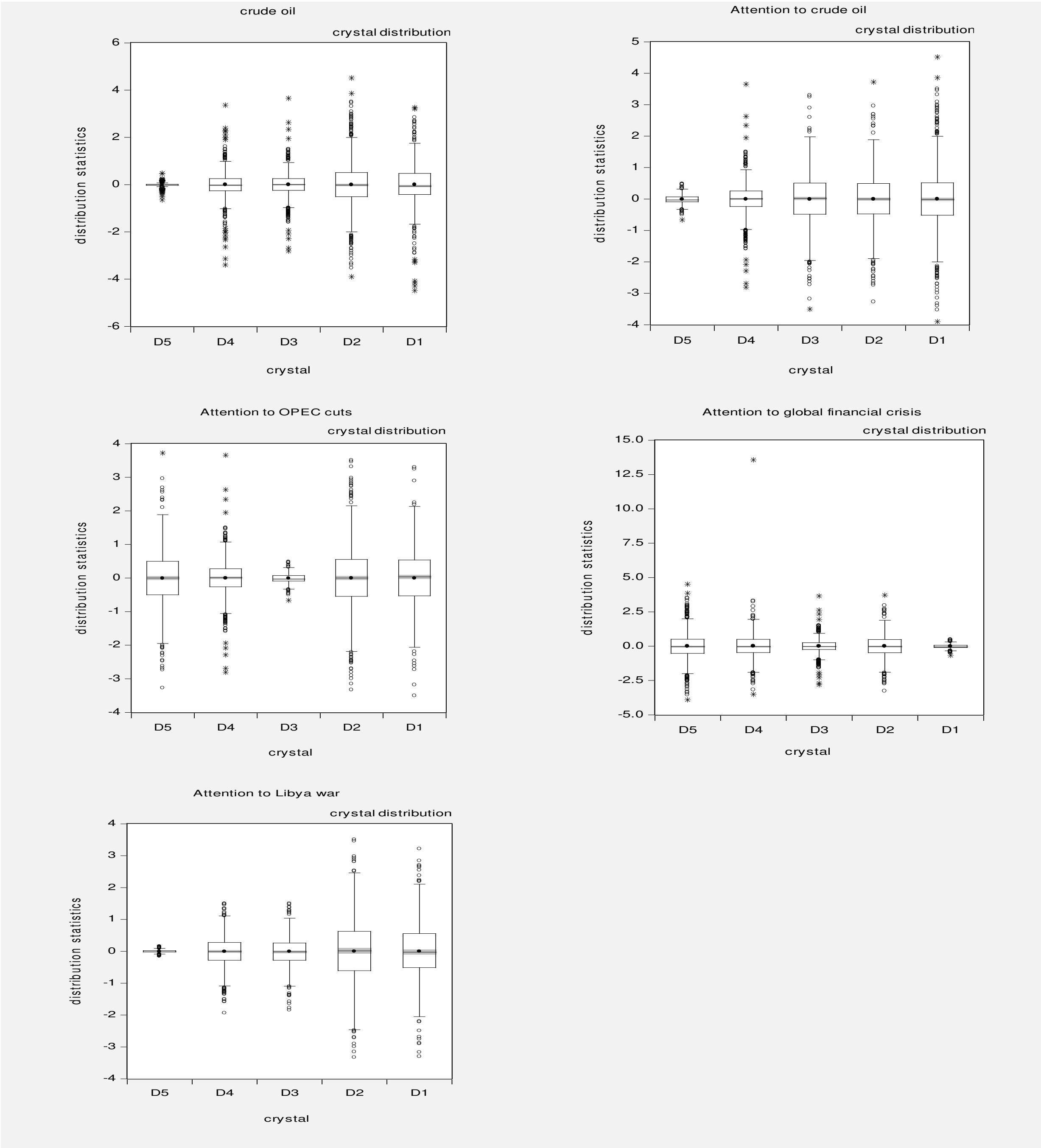
Wavelet scales	Trading frequency (quarters (q))
D1	2-4q
D2	4-8q
D3	8-16q
D4	16-32q
D5	above 32q

A second step seeks to separate the contribution of energy in the time-series due to changes at a specific scale. Table 2 reports the energy of each scale as percentage of the overall energy for the variables of interest, ie crude oil price, attention to crude oil and attention to oil-related events. This is done in order to lose much information about frequencies' contribution. We discuss three main periods: short-run (D1+D2), medium-run (D3) and long-run (D4+D5). For almost all series, the short-run dominates the other scales, except the attention to global crisis of 2008 where the longest time scales seems preponderant. Figure 6 depicts the box plot for each of the focal variables in order to show the crystal energy distribution, as summarized in Table 2. The crystal energy distribution reveals that crystals D1 and D2 contain most of the series' energy, except for attention to global financial collapse. The wavelet decomposed variables investigated are presented in Figure A.1 (Appendices), showing that the time series studied vary sharply over time.

Table 2. Energy decomposition for the variables of interest

Wavelet scales	Crude oil	Attention to crude oil	Attention to OPEC cuts	Attention to global financial crisis	Attention to Libya war
D1 (2-4q)	38.19%	46.14%	40.36%	0.73%	31.55%
D2 (4-8q)	33.38%	28.76%	42.15%	11.41%	50.01%
D3 (8-16q)	13.15%	10.98%	1.34%	14.52%	9.37%
D4 (16-32q)	14.04%	12.00%	8.23%	34.56%	8.86%
D5 (above 32q)	1.24%	2.11%	7.92%	38.77%	0.21%

Figure 6. Crystal energy distribution for the variables of interest



4.2.2. The nexus between the attention to "crude oil" and crude oil

After decomposing the series under consideration, we test the null hypothesis that attention to crude oil due to its great uncertainty does not Granger-cause crude oil over distinct time scales. The results of wavelet-based nonlinear causality test are reported in Table 3. We find that the null hypothesis of attention, measured through Google Trends by searching the term "crude oil", does not Granger-cause crude oil is rejected in lowest time scales (short-run), ie when the scale is between 4 and 8 quarters (D2). These results change slightly when relevant explanatory variables are accounted for. Our findings often suggest the existence of causality from attention to crude oil in the short-run but in larger scale corresponding to the scales D1 (2-4q) and D2 (4-8q). These results confirm those of frequency domain causality. Both techniques, unconditionally and conditioning upon further control variables, indicate that investors moods about crude oil uncertainty, measured by the Google search queries, may be perceived as short-term predictors of crude oil. Thus, the hypothesis (***H1***) seems well verified. This highlights the consistency of the empirical strategy pursued throughout this research.

Table 3. Conditional vs unconditional scale-by-scale nonlinear causality test: Attention to crude oil-crude oil nexus

D1	D2	D3	D4	D5
H_0: Attention to crude oil does not-Granger cause crude oil				
Unconditional analysis				
0.6381 (0.1765)	2.3351** (0.0099)	0.41294 (0.2401)	0.2159 (0.1165)	0.8261 (0.4058)
Conditional analysis				
2.5312* (0.0261)	3.6214** (0.0057)	0.6378 (0.7554)	0.7498 (0.6567)	1.1457 (0.8849)

Notes: (.): the p-value; p-value<0.01: ***; p-value<0.05: **; p-value<0.1

4.2.3. The nexus between the attention to oil-related events and crude oil

A scale-by-scale causality has been re-tested while attempting to address if investors moods may exhibit different impacts on crude oil among different oil-related events. From a first examination of the outcomes reported in Table 4, we note that the results seem sensitive to the oil event studied and the inclusion of additional control variables.

From Table 4.1, we notice that the null hypothesis of attention to OPEC's announcements (production cuts) no-Granger cause crude oil is rejected at lowest time scales (D1). When accounting for oil demand and supply, the findings often reveal a short-run linkage but in a wider scale (D1 and D2). The results do not change considerably when testing the scale-by-scale causality running from attention to Libya war to oil market. We find usually a short term nexus (D1 and D2, Table 4.3), sustaining the evidence that oil shocks trigger shortest effects, which dissipate in longest horizons (Guo and Ji 2013). Unlike OPEC cuts and Libya war, the influence of attention to global crisis on crude oil appears persistent (structural). Specifically, the null hypothesis that the investors' interest to 2008 financial collapse does not Granger-cause crude oil is supported at longer time scales either unconditionally (D4 and D5, Table 4.2) or conditionally to World oil demand and supply (D3, D4 and D5).This means that anxiety about the possible detrimental effects of global financial crisis continues to expand up to now and investors seem unable to digest clearer information about crisis in the short-run. We confirm therefore the second hypothesis (***H2***) to be tested in this study.

Table 4. Conditional vs unconditional scale-by-scale nonlinear causality test: Attention to oil-related events-crude oil nexus

D1	D2	D3	D4	D5
4.1. H_0: Attention to OPEC's cuts does not-Granger cause crude oil				
Unconditional analysis				
5.0832* (0.0105)	1.1342 (0.9561)	0.5484 (0.5814)	0.2297 (0.1489)	0.1827 (0.2625)
Conditional analysis				
3.7652** (0.0028)	3.2183** (0.0110)	0.6273 (0.4108)	0.5274 (0.5581)	0.6178 (0.5532)
4.2. H_0: Attention to global financial crisis does not-Granger cause crude oil				
Unconditional analysis				
0.3154 (0.2231)	0.4521 (0.1964)	0.6251 (0.2296)	3.6891** (0.0064)	4.1271* (0.0128)
Conditional analysis				
0.1922 (0.1053)	0.3258 (0.1273)	4.1213* (0.0222)	3.1857* (0.0181)	3.6932** (0.0037)
4.3. H_0: Attention to Libya war does not-Granger cause crude oil				
Unconditional analysis				
4.4615** (0.0099)	3.9621*** (0.0005)	2.2562** (0.0034)	0.8954 (0.1633)	0.9124 (0.6617)
Conditional analysis				
3.2569*** (0.0008)	4.3784*** (0.0010)	0.6128 (0.5554)	1.2459 (0.7637)	0.6754 (0.3986)

Notes: (.): the p-value; p-value<0.01: ***; p-value<0.05: **; p-value<0.

4.3. Robustness check

The robustness check part present the results of frequency domain causality (Figure A.2 and Figure A.3, Appendices) and scale-by scale nonlinear causality test (Table A.1 and Table A.2, Appendices) for testing the causality between the attention to crude oil, measured via the number of tweets related to the search term "crude oil", and the crude oil price on the one hand and the nexus between the investors moods related to the global financial crisis (Twitter) and crude oil. The relationships between the rest of oil events and crude oil have not been tested due to the availability of data (the number of tweets related to the two keywords "OPEC cuts" and "Libya war" are unavailable over the period 2004q1-2015q1 We report only the results of conditional data analysis, while the findings of unconditional assessment are available upon request.

Figure A.2 depicts the predictive power of the attention to crude oil for crude oil. We show that the null hypothesis that the attention does not Granger-cause crude oil is rejected even if we incorporate the main oil fundamentals (demand and supply) for the frequency, ω, lower than 2.49 corresponding to a cycle length less than 2.5 quarters (short-run). By replicating the same testing procedure to the Twitter data related to global financial collapse of 2008 in order to assess its impact on oil price (Figure A.3), we note that the null hypothesis of the number of tweets no Granger-cause crude oil price is rejected at the 5% significance level for ω above 0.65 corresponding to more than 9.6 quarters (long-run). The results appear also fairly robust when using nonlinear causality test-based wavelet approach. Our findings summarized Table A.1 (Appendices) reveal that the null hypothesis of attention to crude oil (Twitter) no-Granger cause crude oil price is rejected under low time scale (D1). When focusing on oil-related events and in particular global financial crisis, we support a long-run causality at higher time scales (D4 and D5). We ascertain therefore the robustness of our results among the different techniques carried out and the proxies used: (i) The two hypotheses tested throughout this research were well verified; (ii) The effect of the attention to global financial crisis on crude oil appears structural (long-run causality).

The internet search -either Google trends or Twitter- only reflects the traders' investments moods under specific time. The investors are highly influenced by sudden events in both national and international contexts. Obviously, the bad news tend to trigger more powerful effect on the investor. Generally speaking, the mass media plays a pulling role in dramatizing the situation to reach its main aim which is "the buz". So the investors' behaviors in response to each event cannot effectively and properly reflect the oil price or accurately supplant their main fundamentals (supply, demand, economic growth, discoveries of new deposits, etc.). This is the reason behind finding short-run causality between the attention of oil-related events and crude oil, except for the 2008 financial crisis. It is true that crisis seems structural and therefore is still felt to this day. A country like Spain has not yet returned to its wealth level of 2007. Greece, another member of the European Union which is sinking into crisis; but this remains a very special case.

We have attempted, from this research, to use some proxies able to depict investors' sentiment (Google search queries and tweet backs number) to appropriately predict the crude oil. Despite their usefulness, these variables do not fully reflect on their own sense the feeling of investors. In other words, it seems absolutely true that the use of these attention proxies appear fruitful and meaningful in economic and financial applications. However, it is insufficient to rely on only the traders' investment attention to forecast oil price as it is done in some papers including the Guo and Ji (2013)'s study. An effective analysis should account for the oil price fundamentals based on conditional data analysis.

5. Concluding remarks

With the sharp increase of oil marketisation and the development of electronic information carriers over the current period, the question "Is the internet search driving oil market" has devoted tremendous attention. The present paper revisits this question while attempting to reach better ways over the period 2004-2015. This study has two main aims. The first one is to test if an extraction of investors' moods oriented to crude oil and major oil events (OPEC cuts, global financial crisis and Libya war), using web contents, may help to properly predict crude oil. The second one is to explore if the influences of these events transitory or structural. For this purpose, different models were employed. Instead of analyzing the time series at their original level, as it is usually done, the present study deals with the relationship between attention and crude on a frequency-by-frequency and scale-by-scale basis, even if we incorporate potential control variables (conditional data analysis). As a result, some new conclusions on oil market concerns are disclosed. We emphasize two main findings.

Firstly, a short-run relationship between Google search queries and crude oil is highly supported in highest frequencies (short-run). Secondly, oil events play a pulling role in deepening uncertainty and complexity of the worldwide oil market. However, the crude oil responds differently to the attention of these events. If OPEC production cuts exert short and medium term causality and the attention to Libya war plays o potential role on crude oil in the short-run, the interest to 2008 global financial crisis exhibits a longer scale effect. This sharp distinction among oil events' effects highlight deeper difference in their origins and properties. These evidences remain strong even when a number of control variables are accounted for. To ascertain the robustness of our results, we replicate the same testing procedure by using as proxy of investors attention the number of tweets related to the search terms "crude oil" and "global financial crisis". By doing so, the results remain fairly solid. We confirm the first hypothesis that the extraction of investors' moods oriented to crude oil help to effectively predict crude oil but in the short-run. Unfortunately, we cannot verify whether different oil may exert different effects on crude oil due to the lack of data (Twitter) for OPEC cuts and Libya war. Despite this lacuna, we unambiguously check that the attention to financial collapse on oil market is persistent (long-run).

Beyond testing relevance of attention to oil-related events to gain clearer information about oil market, this paper proves the usefulness of search query data (Google Trends and Twitter) in forecasting short-run fluctuations in the oil market. Search volume may be perceived as a very practical way to reveal and compute the traders' investments attention that can help in predicting short run fluctuations in oil market.

Before ending, it must be stressed that we interpret the observed outcomes as providing ample support for the notion that extracting information from oil-related events is helpful in guarding against the swelling crude oil volatility. Nevertheless, these findings are still preliminary and several extensions appear warranted. It is recommended thus to conduct further research by employing other measures investors' attention with other Internet-based data in the Big Data Era to confirm our results.

References

Aguiar-Conraria, L. and Soares, M-J. (2011), "Oil and the macroeconomy: using wavelets to analyze old issues." Empirical Economics, 40, pp. 645-655.

Bodart, V., et Candelon, B. (2009), "Evidence of interdependence and contagion using a frequency domain framework." Emerging Markets Review, 10(2), pp. 140-150.

Bouoiyour, J. and Selmi, R. (2015), "Greece withdraws from Euro and runs on Bitcoin; April Fools Prank or Serious Possibility?" Working paper, CATT, University of Pau.

Bouoiyour,J., Selmi, R., Tiwari, A-K. and Shahbacz, M., (2015), "The Nexus between Oil price and Russia's Real Exchange rate: Better Paths via Unconditional vs Conditional Analysis", Energy Economics, In press.

Breitung, J., and Candelon, B., (2006), "Testing for short and long-run causality: a frequency domain approach." Journal of Econometrics 132, pp. 363-378.

Damasio, A., (1999), "The Feeling of What Happens." Harcourt Brace, New York.

Daubechies, I., (1992), "Ten Lectures on Wavelets." CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA, vol.61.

Dolan, R-J., (2002), "Emotion, Cognition, and Behavior." Science, 298 (5596), pp. 1191-1194. Fan, Y. and Xu, J-H., (2011), "What has driven oil prices since 2000? A structural change perspective." Energy Economics, 33(6), pp. 1082-1094.

Geweke, J. (1982), "Measurement of linear dependence and feedback between multiple time series." Journal of American Statistical Association 77, pp. 304-324.

Gradojevic, N. and Lento, C., (2015), "Multiscale analysis of foreign exchange order flows and technical trading profitability." Economic Modelling, 47, pp. 156–165

Granger, C.W.J. (1969), "Investigation causal relations by econometric models and cross spectral methods." Econometrica, 37, pp. 424-438.

Guo, J-F., and Ji, Q., (2013), "Oil price volatility and oil-related events: An Internet concern study perspective." Applied Energy, 137, pp. 256-264.

Kilian, L. (2009), “Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market.” Am Econ Rev, 99(3), pp. 1053-1069.

Li , X., Jian M. , Wang, S. and Zhang , X., (2015), “How does Google search affect trader positions and crude oil prices?” Economic Modelling, 49, pp. 162-171.

Mensi, W., Hammoudeh, S. and Yoon, S-M., (2014), “How do OPEC news and structural breaks impact returns and volatility in crude oil markets? Further evidence from a long memory process.” Energy Economics, 42, pp. 343-354.

Mishne, G., and Glance, N., (2006), “Predicting Movie Sales from Blogger Sentiment.” Working paper, University of Amsterdam.

Mitchell, A., T. Rosenstiel and L. Christian (2012), “What Facebook and Twitter Mean for News.” Pew Research Center, Annual Report on American Journalism: The State of the News Media 2012.

Mueller, M. L. and Lemstra, W. (2011), “Liberalization and the Internet”. In: Eds. Künneke, R. W. and Finger, M., International handbook of network industries. Cheltenham, UK: Edward Elgar. Pp. 144-161.

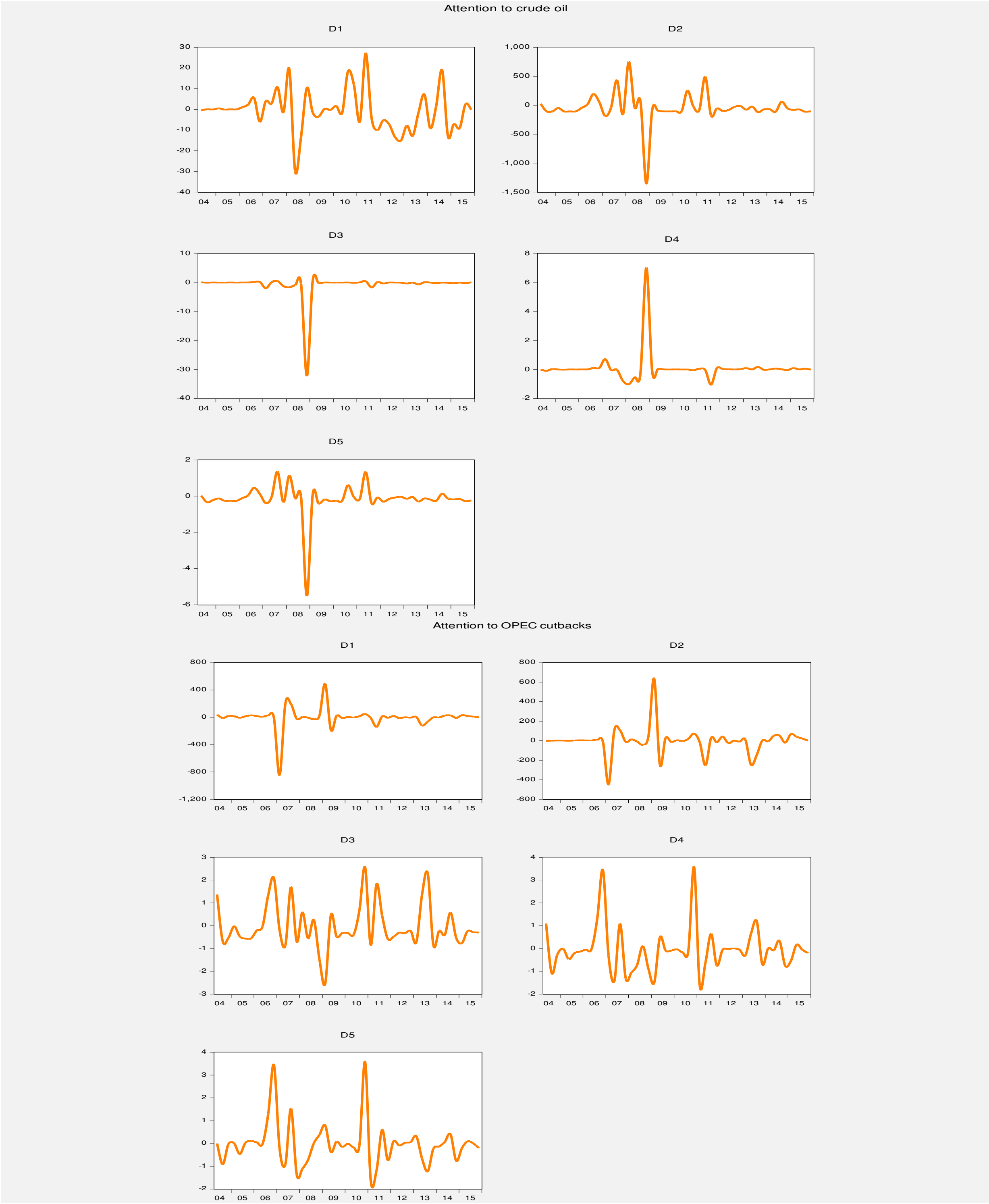
Péguin-Feissolle, A., Terasvirta, T., (1999), “A general framework for testing the Granger non-causality hypothesis.” Stockholm School of Economics Working Paper Series in Economics and Finance n° 343.

Salaverría, R., (2005), “El impacto de Internet en los medios de comunicación en España.” Sevilla: Comunicación Social Ediciones y Publicaciones.

Watts, L., (2009), “The Financial Crisis and its Impact on the Oil & Gas Industry.” Worldbank working paper, March.

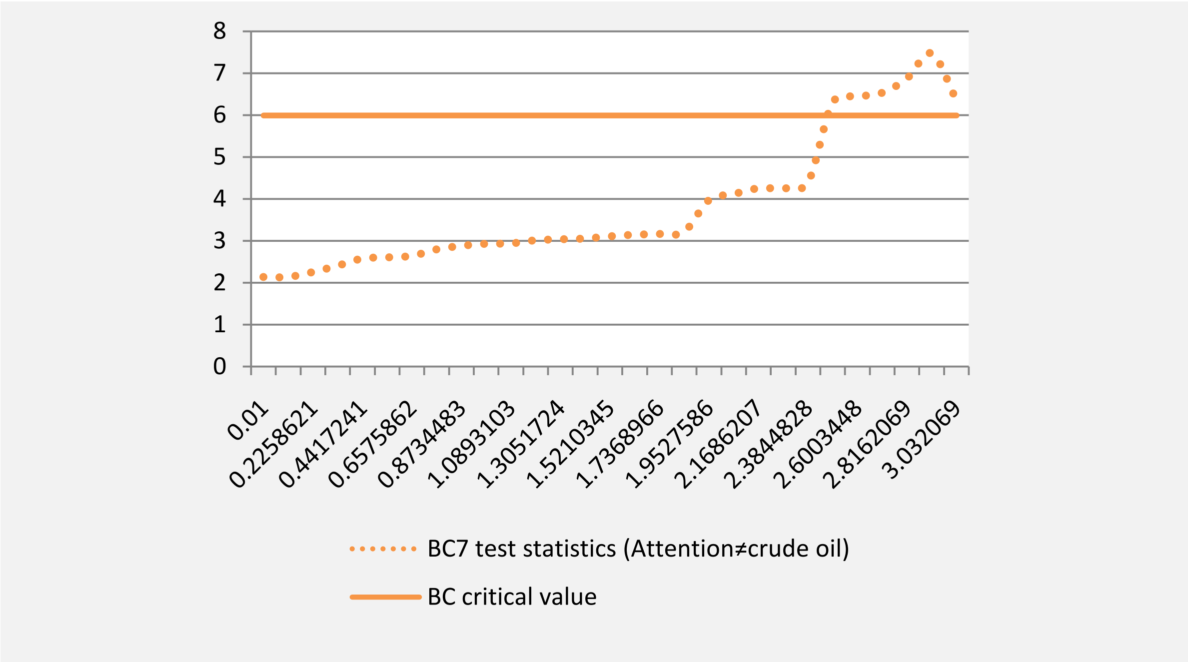
Appendices

Figure A. 1. Discrete wavelet decomposition for the variables studied



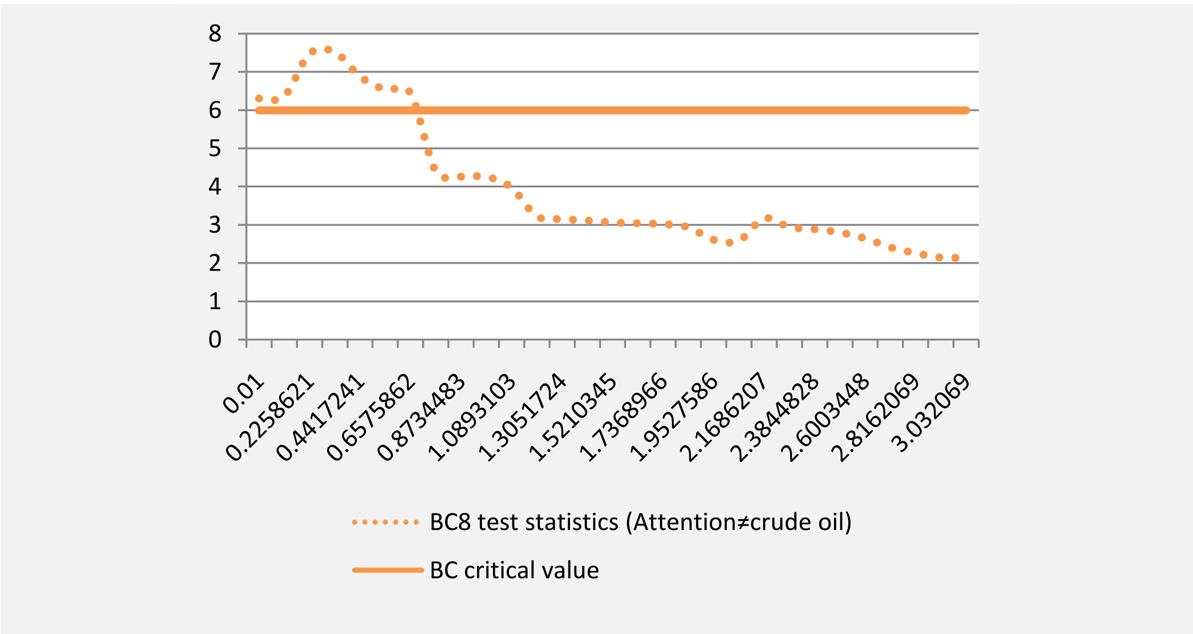
Notes : D1 :2-4q; D2 : 4-8q ; D3 : 8-16q ; D4: 16-32q; D5 : above 32q.

Figure A.2. The frequency-domain causality between attention to crude oil (Twitter) and crude oil



Note: The horizontal line represents the 5% critical value of the null hypothesis test of no Granger causality at frequency *w*.

Figure A.3. The frequency-domain causality between attention to global financial crisis (Twitter) and crude oil



Note: The horizontal line represents the 5% critical value of the null hypothesis test of no Granger causality at frequency *w*.

Table A.1. Conditional scale-by-scale nonlinear causality test: Attention to crude oil (Twitter)-crude oil nexus

D1	D2	D3	D4	D5
H_0: Attention to crude oil does not-Granger cause crude oil				
6.1455***	0.6245	1.1045	0.7213	0.5437
(0.0000)	(0.3873)	(0.9219)	(0.6085)	(0.5021)

Notes: (.): the p-value; p-value<0.01: ***; p-value<0.05: **; p-value<0.1

Table A.2. Conditional scale-by-scale nonlinear causality test: Attention to global financial crisis (Twitter)-crude oil nexus

D1	D2	D3	D4	D5
H_0: Attention to global financial crisis does not-Granger cause crude oil				
0.7615	0.9010	0.6381	4.2415*	3.7629**
(0.2318)	(0.8761)	(0.5513)	(0.0963)	(0.0038)

Notes: (.): the p-value; p-value<0.01: ***; p-value<0.05: **; p-value<0.1