Media attention and crude oil volatility: Is there any ‘new’ news in the newspaper?

D Aromi
Facultad de Economía, Universidad de Buenos Aires, IIEP-BAIRES.

A Clements
School of Economics and Finance, Queensland University of Technology

March 20, 2018

Abstract
In recent years there has been a growing interest in the analysis of large volumes of unscheduled news flow. Such news flow has often been used as an exogenous variable for explaining asset returns and or volatility. This paper examines the dynamic relationship between news flow and asset price dynamics from a different perspective. A novel index of media attention is proposed, and in the context of the crude oil market the linkages between media attention and returns and volatility are examined. It is found that media attention reacts strongly to shocks to volatility whereas there is little impact in the opposite direction. As such media attention seems to inherit the persistence in volatility but offers only a little more in terms of information relevant to future volatility. Therefore media attention does not offer a great deal of new news useful for explaining volatility.

Keywords
News, media, linguistic analysis, volatility, crude oil

JEL Classification Numbers
C22,G00
1 Introduction

The Mixture of Distributions Hypothesis suggests that return volatility is related to the flow of information into the market (Clark, 1973, Tauchen and Pitts, 1983). Though theoretically appealing, there has been a great deal of debate surrounding the appropriate measures of information flow. While a wide range of announcements and different sources of information have been considered across different markets, a common theme is that such information is treated as exogenously driving asset prices, returns volatility and or trading volume.

In recent years, there have been rapid advances in the capture, storage and analysis of large volumes of unscheduled news items available to market participants. This has led to the development of a new strand of literature linking electronic news arrivals and market behaviour. Based on ever more sophisticated algorithms to process the contents of unscheduled news, a range of studies consider the impact of news flow on returns, trading activity and volatility, mainly in the context of equities. Tetlock (2007) uses the so-called bag-of-words approach to measure quantifies optimism and pessimism from the Wall Street Journals Abreast of the Market showing that pessimism predict declining market prices. Using a similar technique, Tetlock, Saur-Tsechansky, and Macskassy (2008) show that the fraction of negative words in newspaper articles forecast firm earnings their content captures the hard-to-quantify aspects of fundamentals. Groß-Klußmann and Hautsch (2011) use news data for individual equities traded on the London Stock Exchange sampled at high frequency to examine how news arrivals influence trading activity in individual stocks. Riordan, Storkenmaier, Wagener, and Sarah Zhang (2013) focus on the impact of newswire messages on intraday price discovery, liquidity, and trading intensity of Canadian stocks. Smales (2014) uses news headlines relating to the constituents of the S&P 500 to examine the relationship between aggregate news sentiment and changes in the implied volatility index, the VIX index. Again, these unscheduled news arrivals have been viewed as exogenously influencing prices, volatility and trading activity.

As an alternative measure of information, or more specifically the rate at which investors seek out information (investor attention), internet search volume has also been linked to return volatility. Da, Engleberg, and Gao (2011, 2015), Smith (2012) and Vlastakis and Markellos (2012) consider the impact of Google Search Volume (GSV) on the returns and volatility of a range of assets, individual stocks, foreign exchange and Treasury bonds. More closely related to this paper, Campos, Cortazar, and Reyes (2017) and Afkhami, Cormack, and Hamed Ghodduisi (2017) examine the link from GSV to oil price volatility. Campos et al. (2017) consider a simple measure of GSV related to the term ‘Oil prices’ and oil option implied volatility. Afkhami et al. (2017) employ a complicated multi-stage filtering process to identify the GSV of which energy related terms from an initial set of 90 offer the greatest improvements in volatility prediction.

This paper extends the work of Campos et al. (2017) and Afkhami et al. (2017) in the context of crude oil markets along a number of interesting dimensions. This paper proposes a novel measure of news, or media attention based on newspaper articles which offers an automated
approach to selecting keywords upon which to determine news attention. As discussed in earlier studies, GSV is likely to reflect activity by less sophisticated investors, whereas newswire articles discussed above are only normally accessible to sophisticated institutional traders. Newspaper articles on the other hand are readily available to everyone (at a nominal cost).

In the context of the crude oil market, this paper considers the link between the proposed index of media attention and volatility using more up-to-date volatility measures. In doing so, it identifies how much news (or information) do newspaper articles actually capture in relation to the evolution of volatility. As discussed earlier, the previous research discussed here treats new flow, or media attention in the form of GSV as exogenous, but there is anecdotal evidence to show this may not be an entirely valid assumption to make. Take for instance, the period of significant turmoil in equity markets at the start of 2018 where the Dow Jones fell about 5% in one week during February. Subsequent to these price falls and heightened volatility, headlines in the Markets section of the Wall Street Journal on February 9 were dominated by the events of the previous week:

*Largest Dow Industrials Declines by points*

*Dow Industrials plunge into correction territory*

*The obscure trade that might have triggered the sell off,*

indicating that some news flow or media attention clearly reacts to market conditions themselves. Such anecdotal evidence is motivation to begin this analysis from a different viewpoint and investigate the dynamic relationship between media attention and volatility and not simply treat news flow as exogenous. Next, informational content of news attention is considered relative to two benchmark approaches, option implied volatilities and the Heterogeneous Auto-Regressive (HAR) model of Corsi (2009).

It is found that media attention surrounding the crude oil market is predominantly driven by past volatility of oil market returns. Shocks to volatility lead to significantly higher rates of news attention for out to two weeks into the future. While there is very little impact of shocks to news attention on future volatility. Given that news reacts to movements in volatility, it likely inherits the well-known persistence exhibited by volatility itself. It is found that news attention offers small improvements in in-sample fit and forecast accuracy beyond option implied volatilities and the benchmark HAR model. These small improvements are evident when using longer lags of news attention over longer forecast horizons. These results, in conjunction with the causal links from volatility to news indicate that news attention mostly reflects the underlying persistence in volatility and does not offer a great deal of new news in relation to crude oil markets.

While more the more general links between news and volatility are the focus here, and not simply forecasting oil volatility, these results along with those of Campos et al. (2017) and Afkhami et al. (2017) suggest that a fruitful avenue of research is to compare the different sources of
news sources together in the long-term component of more complex models of volatility for forecasting purposes.

Section 2 outlines the measures of crude oil volatility used here, along with the methodology used to create the index of media attention. Section 3 reports the empirical results relating to both the dynamic links between volatility and news attention, and the informational content of oil IV. Section 4 provides concluding comments.

2 Data

Sample period spans 10 May 2007 to 2 May 2017 containing 2513 daily observations. Both daily, and 5-minute returns on the United States Oil (USO) exchange-traded fund were collected from Thomson Reuters Tick History. The daily returns (RETS) are shown in the middle panel of Figure 1. These show the familiar pattern of mostly negative returns associated with the collapse in oil prices from historical highs in 2008-2009, and again in 2014-2016. Intraday 5-minute prices (9:30am to 4pm NY time) were collected and used to construct daily estimates of realized volatility (RV). The daily RV estimates are shown in the lower panel of Figure 1. Volatility increased markedly during the 2008-2009 periods prices fell, and rose again (but to lower levels) during the falls of 2014-2016.

OVX_t is the implied volatility index derived from options on the USO oil fund. It is calculated in the same manner as the VIX index based on the S& P 500 and is also published by the CBOE and reflects the market’s expectation of crude oil volatility over the following 22 trading days.

An index of media attention on crude oil markets is constructed using the Global Vectors (GloVe) approach of Pennington, Socher, and Manning (2014). GloVe is based on a word vector representation which has been shown to efficiently summarize semantic information corresponding to each word and can be used to measure relatedness between different words. Following Pennington et al. (2014), word vector representations as a linear structure of meaning based on a term co-occurrence matrix. Let W denote a dictionary and X_{ij} denote the number of times word i occurs in the context of word j. Then, word vector representations \{v_i\}_{i \in W} solve:

\[
\min_{\{v_i\}_{i \in W}} \sum_i \sum_j f(X_{ij}) \left[ v_i' v_j + b_i + b_j - \log(X_{ij}) \right]^2
\]

where \( f(X_{ij}) \) is an increasing and concave weighting function and \( b_i \) is word i’s bias.

To begin here, the word ‘oil’ is used, with a set of S of the most closely words selected from the dictionary W of all words by proximity of their vector (100 dimension) representations.\(^1\) Let \( n_{wt} \) denote the number of times word w is observed in period t. Then, the news attention index

---

\(^1\) The list of the 50 most closely related words is reported in the Appendix.
for period $t$ is given by:

$$ NEWS_t = \frac{\sum_{w \in S} n_{wt}}{\sum_{w \in W} n_{wt}} $$

That is, the index is given by the number of occurrences of words in $S$ as a fraction of the total number of occurrences of words in the dictionary $W$.

The texts used in this exercise are extracted from a publicly available subset of content published in the Wall Street Journal. The data can be found at http://pqasb.pqarchiver.com/djreprints/. For each article published in the newspaper, this website provides access to the headline, the lead and a fraction of the body. Two collections of texts are used in this exercise. The training corpus is given by content published between 1970 and 2006 upon which the 50 words in $S$ are chosen.\(^2\) The news attention index, $NEWS_t$, is then computed for each day using a second corpus of material published during the sample period. It is important to note that the information contained in $NEWS_t$ can be viewed as being available to the market on day $t$ as the articles are published early in the morning prior to the day’s trading. Vector representations of words were computed using package text2vec in platform R.

The news attention index, $NEWS_t$, is shown in the top panel of Figure 1. A surge is $NEWS_t$ is evident around the peak, start of the fall in prices and onset of higher volatility in 2008 appearing to slowly subside as prices continues to fall. Periods of greater attention are also evident around the periods of higher volatility in 2014-2016.

The subsequent empirical analysis will reveal whether media attention simply reflects past and at best current market conditions, or also provides useful information relating to future market conditions.

3 Empirical Analysis

3.1 Dynamic link between news attention and volatility

As news articles and hence the variable $NEWS_t$ is available before trading on day $t$, there is a natural calendar structure underlying the relationship between news attention and volatility. Therefore a Structural Vector AutoRegression (SVAR) framework is used to examine the linkages between news attention and volatility.

$$ A_0 Y_t = \mu + \sum_{i=1}^{k} \theta_k Y_{t-k} + v_t $$

\(^2\)As a robustness check, the subsequent analysis has been conducted on different indices of attention based on different assumptions for $S$ and the dimension for the word vector representations. The results are qualitatively the same leading to the same conclusions drawn here.
where $Y_t = [RV_t; NEWS_t]'$ is a vector containing the news and volatility measures. The matrix $A_0$ reflects the short-run calendar restriction and is defined as follows

$$A_0 = \begin{pmatrix} 1 & \alpha_{12} \\ 0 & 1 \end{pmatrix} \tag{2}$$

where the coefficient $\alpha_{12}$ captures the same-day contemporaneous link from news attention to volatility. Optimal lag structures varied widely across different information and decision criteria, and as such a mid-range lag structure of 5 was chosen for all models. To examine the link between news attention and market conditions more deeply, $RV_t$ is replaced in two further specifications, $Y_t = [RETS_t^+; NEWS_t]$ and $Y_t = [RETS_t^-; NEWS_t]$ where $RETS_t^+$ and $RETS_t^-$ are dummy variables to indicate the occurrence of more extreme positive and negative returns above 2% in absolute terms (capturing approximately the top and bottom decile of returns). Such an analysis will reveal whether there are significant asymmetries in the relationship between news attention and the sign of extreme returns which are closely related to volatility itself but also reflect the information in the sign of the returns (oil price increase or decrease). Impulse response functions (IRFs) from the model using the three definitions of $Y_t$ will be used examine the links between news and volatility.
IRFs from model given the main specification, \( Y_t = [RV_t; NEWS_t]' \) are shown in Figure 2. These show that while shocks to either RV or NEWS have a significant positive impact on their own future paths. Shocks to RV have a lasting positive impact on NEWS meaning that news attention increases for most of the next weeks into the future subsequent to shocks to volatility. On the other hand, a shock to news attention has a little discernible impact on future RV.

Figure 2: Impulse response functions for the \([RV_t; NEWS_t]\) model. 95% confidence intervals are given by the dashed lines.

To consider how news attention interacts with the information in extreme price movements as opposed to total RV, Figures 3 and 4 show the IRFs given \( Y_t = [RETS_t^+; NEWS_t] \) and \( Y_t = [RETS_t^-; NEWS_t] \) respectively. Shocks to extreme positive (\( RETS_t^+ \)) or negative (\( RETS_t^- \)) returns have broadly similar impacts on future news attention, in both cases a positive effect (slightly more significant for negative returns) most of the two weeks into the future. Though the effect is not as strong as that evident with shocks to volatility. As opposed to RV, shocks to news attention are associated with a greater occurrence of extreme price movements in the future, with the effect being stronger for extreme price falls relative to price increases.

Figure 3: Impulse response functions for the \([RETS_t^+; NEWS_t]\) model. 95% confidence intervals are given by the dashed lines.
In summary, these results show that shocks to volatility (and to a slightly lesser degree extreme price changes) have a substantial impact on news attention. However, on the other hand, shocks to news attention have little effect total volatility but do contain some information about the occurrence of extreme price movements, showing that there is a great deal of asymmetry in the links in both directions. Given the higher frequency of news attention relative to GSV, a richer more complex set of dynamics can be revealed. Overall, these results also give some empirical support to the anecdotal evidence presented earlier that media attention is at least in some part driven by recent market conditions. While news attention reacts strongly to volatility, the question of whether it contains any useful information for explaining volatility dynamics is addressed in the following section.

3.2 Information in news attention

First the informational content of news attention relative to option implied volatility (IV) OVX index is considered. While earlier studies have linked information in GSV to volatility, none of these studies considered information in GSV relative to other forecasts such as IV, often shown to be the most accurate forecast of volatility, Poon and Granger (2003). This is an important question as OVX should reflect all relevant conditioning information and represent the market’s best prediction of crude oil volatility (see Poon and Granger 2003 for a broad overview of this issue across a wide range of markets). While there has been a great deal of work in this area, little attention has been paid to oil markets, and none to the question of whether offer news attention offers incremental information over IV. This analysis here differs from that of Campos et al. (2017) which considers the role of GSV in predicting OVX. Here the central question is whether media attention contains more information regarding future volatility than that contained in OVX.

This issue is addressed within forecast regression framework of Clements and Hendry (1998)
Table 1 reports the estimation results for the IV models in equation in 3 at a forecast horizon of lagged news attention are of interest. Both the fit of these models and the significance of the coefficients on attention. All regressions are estimated in logarithmic form, a commonly used approach in the forecasting literature. Both the fit of these models and the significance of the coefficients on lagged news attention are of interest.

Table 1 reports the estimation results for the IV models in equation in 3 at a forecast horizon of $k = 22$, therefore the dependent variable is $ln(RV_{t+1→t+22})$. News attention does not offer any benefits at the shortest horizon $k = 1$, with the results at $k = 5$ being broadly consistent with the case reported here. Newey-West standard errors are reported in parentheses to control for overlapping nature of the dependant variables. Unsurprisingly, the coefficients on $OVX_t$, are highly significant in all cases indicating that IV in its own right is useful for explaining future RV. Only the coefficient on $N_{1,t}$ is significant at 10% when it is included in the regression as the sole explanatory variable. Overall, there is little evidence that past media attention offers any incremental information beyond O VX for the purposes of explaining future volatility and hence O VX is informationally efficient with respect to (encompasses) past media attention. Such a result is consistent with the findings of many previous studies where it is found that IV in general is efficient with respect to most information and offers a high quality forecast. In terms of the current context, given the strong reaction of media attention to past volatility, while it may contain information about the history of volatility it contains little new news (information) about future volatility that is not already compounded into option markets.

Next, the news attention will considered in the context of the Heterogeneous Auto-Regressive (HAR) framework of Corsi (2009). HAR style regression are a very simple tool to capture much of the long-term persistence in RV (and related measures) and has become somewhat of a benchmark in the financial econometrics literature. This analysis extends that of Afkhami et al. (2017) in using both higher frequency RV estimates and higher frequency measures of news attention both available daily thus being able to rely on the HAR framework as opposed to GARCH models. A range of HAR regressions that include the lags of news attention are estimated using the same lagged moving averages in news attention:

$$\begin{align*}
\ln(RV_{t+1→t+k}) &= \beta_0 + \beta_1 \ln(OVX_t) + \beta_2 \ln(RV_{5,t}) + \beta_3 \ln(RV_{22,t}) + \epsilon_t \\
\ln(RV_{t+1→t+k}) &= \beta_0 + \beta_1 \ln(OVX_t) + \beta_2 \ln(N_{1,t}) + \epsilon_t \\
\ln(RV_{t+1→t+k}) &= \beta_0 + \beta_1 \ln(N_{1,t}) + \beta_2 \ln(N_{5,t}) + \beta_3 \ln(N_{22,t}) + \epsilon_t
\end{align*}$$ (3)
Dependent variable: $\ln(RV_{t-22})$

<table>
<thead>
<tr>
<th></th>
<th>Column 1</th>
<th>Column 2</th>
<th>Column 3</th>
<th>Column 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1$</td>
<td>1.0330***</td>
<td>1.0292***</td>
<td>1.0030***</td>
<td>0.9864***</td>
</tr>
<tr>
<td></td>
<td>(0.0111)</td>
<td>(0.0115)</td>
<td>(0.0123)</td>
<td>(0.0133)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.0693*</td>
<td>0.0020</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0408)</td>
<td>(0.0157)</td>
<td>(0.01428)</td>
<td></td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.2090</td>
<td>0.0756</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.1540)</td>
<td>(0.1306)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.2483</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2847)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Regression results for equation 3. Newey-West standard errors are reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Table 2 reports the estimation results for the HAR models in equation 4 again based on $\ln(RV_{t+1-22})$ as the dependent variable and Newey-West standard errors. The coefficients on lagged RV, $\beta_1$, $\beta_2$ and $\beta_3$ are all highly significant, a standard result in HAR models. In relation to the inclusion of news attention, again broadly similar results to the previous IV models are found. When $\ln(N_{1,t})$ is first added to the regression in isolation, $\beta_4$ is significant at 10%. Again, as the longer lags are included they dominate and drive out the importance of $\ln(N_{1,t})$ with the coefficient on $\ln(N_{22,t})$ significant at 5%. The model with all three lags of news attention offers a modest increase in explanatory power relative to the HAR benchmark.

The OVX regression results seems to indicate that there is not a great deal of new news in the rate of media attention, in the sense that is encompassed by forward looking market volatility predictions. However, the HAR results show that news attention reflect longer term persistence in volatility as longer term lags in news are important in regression explaining volatility over longer horizons.

It has been found that news attention is not particularly important at short forecast horizons, however long term averages of news seems to be important for longer term forecasts. Given the strong reaction of news attention to volatility, it seems as though news flow reflects the persistence exhibited by volatility and longer term averages in news attention help explain volatility over longer horizons. This observation leads to a potentially interesting avenue for future research to consider the role played by news in component models for volatility where it may be useful for helping to explain the slower moving longer-term component of volatility.

While news attention offers some benefits in terms of in-sample fit, the final question considered
Dependent variable: $\ln(RV_{t\rightarrow t+22})$

<table>
<thead>
<tr>
<th>$\beta_1$</th>
<th>0.1399***</th>
<th>0.1397***</th>
<th>0.1379***</th>
<th>0.1363***</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.0174)</td>
<td>(0.0176)</td>
<td>(0.0176)</td>
<td>(0.0170)</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.1184*</td>
<td>0.1188***</td>
<td>0.1243*</td>
<td>0.1303**</td>
</tr>
<tr>
<td></td>
<td>(0.0630)</td>
<td>(0.0628)</td>
<td>(0.0649)</td>
<td>(0.0652)</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.5989***</td>
<td>0.5890***</td>
<td>0.5664*</td>
<td>0.5464***</td>
</tr>
<tr>
<td></td>
<td>(0.0832)</td>
<td>(0.0807)</td>
<td>(0.0861)</td>
<td>(0.0908)</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.0799*</td>
<td>0.0007</td>
<td>0.0002</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.0440)</td>
<td>(0.0166)</td>
<td>(0.0157)</td>
<td></td>
</tr>
<tr>
<td>$\beta_5$</td>
<td></td>
<td>0.2447</td>
<td>-0.1269</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.1495)</td>
<td>(0.0962)</td>
<td></td>
</tr>
<tr>
<td>$\beta_6$</td>
<td></td>
<td></td>
<td>0.5120**</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.2571)</td>
<td></td>
</tr>
</tbody>
</table>

$R^2$ | 0.7795 | 0.7806 | 0.7830 | 0.7861 |
Adj. $R^2$ | 0.7792 | 0.7803 | 0.7826 | 0.7855 |

Table 2: Regression results for equation 4. Newey-West standard errors are reported in parentheses. *p<0.1; **p<0.05; ***p<0.01

is whether this translates to out-of-sample improvements in forecast accuracy.

Forecasts are generated from four models, denoted below as $O VX$, $O VX+NEWS$, $H AR$ and $H AR+NEWS$. $O VX$ and $H AR$ are simply first models in equations 3 and 4. The remaining two models contain the monthly lagged moving average of news attention as it is found to be the most important in-sample. Model coefficients were estimated under a rolling scheme based on 1000 observations. Forecasts were generated over the 1,5 and 22 day horizons. Model coefficients were re-estimated each day as the rolling window moved through time. Forecast accuracy for each of the four models is measured using the QLIKE loss function,

$$QLIKE = \left(\frac{RV_{t+1\rightarrow t+k}}{F_t}\right) - \log \left(\frac{RV_{t+1\rightarrow t+k}}{F_t}\right) - 1.$$ (5)

where $F_t$ denotes the relevant forecast formed at time $t$. For each forecast horizon, differences in forecast accuracy of the four models will be assessed using the Model Confidence Set (MCS) introduced by Hansen, Lunde, and Nason (2011). The MCS approach avoids the specification of a benchmark model, and starts with a full set of candidate models $M_0 = \{1,...,m_0\}$. All loss differentials, $d_{ij,t}$ (based on $QLIKE$) between models $i$ and $j$ are computed and the null
Table 3: MCS results comparing the four forecasting models. P-values reported are based on the range statistic. A p-value $> 0.05$ indicates that a model is included in the MCS.

The focus of this paper is not directly on developing the most accurate forecasting model, more on a broader investigation of the dynamics between news or media attention and volatility. However, these forecasting results in conjunction with the earlier in-sample results indicate that the rate of media attention offers some useful information regarding longer-term persistence in volatility. This appears to be a direct result of the impact that past volatility has on media attention. These results do lead to a possibly interesting avenue for future research. It may be interesting to consider the role of news flow in long-term volatility movements within the context of components models for volatility.

4 Conclusion

Traditionally, measures of unscheduled news flow have been treated as exogenous drivers of returns, volatility or trading activity. However, there is little understanding of the dynamic
relationship between market conditions and media attention. Here a novel index of media attention to crude oil markets was proposed which can be interpreted as a rate of information flow available to a wide range of investors. The empirical analysis here examined the links between news attention and volatility, along with the role for news attention within volatility models.

In terms of the link between news attention and volatility, it was found that news attention reacts strongly to shocks to volatility, whereas relatively little effect is observed in the other direction, though there is some impact on future extreme price movements. This provides some support to the anecdotal evidence (in the context of equity markets) provided here that media attention is least in part drive by recent market conditions, and hence would reflect the persistence exhibited by volatility itself. This result provides a deeper understanding of the dynamics between news flow and volatility and suggests that news attention may not be entirely exogenous and may contain less new news than commonly assumed.

By incorporating past news attention into volatility models, small increases in in-sample fit and forecast performance (though not statistically significant improvements) are observed. These improvements are only seen using longer lags of news attention and at longer forecast horizons. This indicates that news mostly contains information about the longer-term persistence in volatility which is a direct result of the strong reaction of news attention to past volatility. As news attention is found to react to past volatility and inherits it persistence, and possibly little more information, one could conclude there is not a great deal of new news in media attention relating to crude oil.

5 Appendix: words related to oil

50 words selected by proximity of vector representation: alaskan, alberta, amoco, arabia, ashland, barrel, barrels, calgary, chevron, crude, crude-oil, discovered, discovery, drilling, energy, exploration, exporting, exxon, field, gas, gasoline, glut, gulf, houston, husky, indonesia, leases, marathon, mexico, mobil, natural, naturalgas, offshore, oil, oils, opec, petroleum, pipeline, producing, properties, refineries, refinery, refining, saudi, sea, shell, tankers, texaco, well and wells.

References


