



Testing for the Presence of Asymmetric Information in the Oil Market: A Vector Autoregression Approach

Haytem Ahmed Troug¹, Rashid Sbia^{2*}

¹Department of Economics, University of Exeter, Exeter, UK, ²DULBEA, Solvay Business and Economics School, Ministry of Finance, UAE. *Email: r_sbiam@hotmail.com

ABSTRACT

This paper aims at providing empirical support to claims made by officials in oil-producing countries that investors in the New York Stock Exchange (NYSE) market are involved in the disruption of oil production in some Organization of the Petroleum Exporting Countries countries. The claims state that some investors in the NYSE are financing militias in those countries to close down oilfields and ports, and buy oil before this incident occurs. By doing so, they have access to information that no one else in the market has, and make profits from this information. Using a vector autoregression (VAR) model approach to detect this phenomenon, and being inspired by the asymmetric information theory, we fail to support those claims. We tried to put this theory under investigation by running test on three oil-disruption incidents that occurred in 2013, and all of the results turned out to be insignificant. Nevertheless, this approach was able to detect a period which might involve asymmetric information in the NYSE. In addition, using a VAR model enabled us to measure the duration and magnitude of the effect of a shock in volumes of trade on oil prices in that market.

Keywords: Asymmetry of Information, Stock Market, Oil Market, Organization of the Petroleum Exporting Countries, New York Stock Exchange
JEL Classifications: C50, C58, G02, G14

1. INTRODUCTION

Investors with access to information regarding a company or an industry have the advantage of profiting from prior knowledge of the information before its out to the public or revelation of that information. Insiders are legally defined as senior officials, managers, and direct or indirect owners of 10% or more of the regarded stock, According to paragraph (a) of subdivision 5 of Section 78 of the Insurance Law. Trading by this group is regulated in that any profit from turning over stock in that firm within 6 months must be returned to the corporation. Moral suasion is also used to discourage insider use of private information by publishing their trading activities in the Official Summary of Security Transactions and Holdings. This ongoing phenomenon is defying the whole “market efficiency” hypothesis which was highly regarded during the 60’s of last century.

During 2013 oil disruptions occurred in some developing Organization of the Petroleum Exporting Countries (OPEC) countries. This phenomenon encouraged a lot of senior officials

in the oil industry and related governance institutions to accuse several possible beneficiaries from these disruptions. In this paper we try to focus on one of those suspects and they’re the speculators in the stock market. Ying (1966) was one of the main scholars that highlighted the notion of “asymmetric information” and how to detect it. Our analysis is partly inspired by his approach, while we will try to develop our own approach for this case as well.

A Wall Street motto says “It takes volume to make prices move.” The analysis of trading volume and its relationship with security prices and changes in price is a topic that has been considered for over 40 years. Its roots are generally credited to the work of Osborne (1959). Volume is a calculation of the amount of shares that change owners for a certain security. The size of daily volume on a certain stock can fluctuate on any given day depending on whether there’s new information available about the company, whether the trading day is a full or half a day, and many other probable factors.

Out of the many different factors affecting trading volume, the one which relates the most to the essential valuation of the security,

is the flow of new information on the security. This information can take various forms; it might be a press release or an earnings announcement provided by the institution itself, or it can be news spread from a third party, such as a court ruling or a release by a regulatory agency pertaining to the company, the analysis of trading volume and associated price changes corresponding to information releases has been of much interest to researchers. In this case the rise in the volumes of trade might impose a negative impact on the prices.

The theory suggests certain relationships between trading volume and changes in prices that are consistent with the use of asymmetrically distributed information. Several hypotheses derived from these theories are empirically tested (Karpoff 1998). The first hypothesis, and most spread, is an examination of serial correlation of return (growth rate of prices) residuals during periods of unusually high trading volume. The rejection of independence of return residuals led to an attempt to derive strategies to successfully diagnose these periods of positive serial correlation of residuals and profit accordingly. Because there are potential internal validity problems with the first hypothesis, an alternative test was performed. Trading prior to a period of large price changes was investigated and found to be significantly greater than normal the day prior to a large price change. This is consistent with individuals trading on information the day prior to its public release and subsequent price effect (Sun 2002).

Given the above, we will try to test of the theory of asymmetric information on the price returns of Brent crude oil. We will focus on three incidents that occurred during 2013 where oil supplies were disrupted in three key producing countries:

- Libya: Militias stormed into seaport facilities and blocked exports of crude oil, and stopped an approximate daily production of 650,000 bbl/day at the end of July
- Nigeria: Attacks on pipelines in Nigeria caused production to decline by almost 450,000 bbl/day in June
- Iraq: Crude oil production in Iraq declined by 250,000 bbl/day by the end of August due to continues attacks on exporting pipelines from Iraq to Turkey.

The remainder of the paper is organized as follows. We review the relevant literature on the relationship between trade volumes and prices in Part (2). A detailed exposition of the methodology and a description of the data in our model is presented in Part (3). In Part (4) we present a summary of the data and it's descriptive statistics. Our model of interest will be investigated in Part (5), along with some robustness tests. Section (6) offers concluding remarks.

2. LITERATURE REVIEW

Seven years after Osborne (1959) initiated the discussion on how the securities prices could be fitted as a log-normal distribution with the variance term dependent on the trading volume, Ying produced a paper in 1966 highlighting the importance of volumes of trade and criticized the lack of attention on the topic during that period. He also highlighted that prices and volumes of sales in the stock market are joint products of a single market mechanism, and any study that tries to separate prices

from volumes or vice versa will certainly yield inadequate if not false results. The paper applied a series of statistical tests on a 6-years daily series of turnover as volumes¹, and Standard and Poor's 500 index returns from January 1957 to December 1962 for price data.

Ying applied different formations of the two variables from levels of the two variables to the growth rate of the log term of these variables. The main results that were found are:

1. Small volumes are usually accompanied by a fall in price
2. Large volume are usually accompanied by a rise in price
3. A large increase in volume could be accompanied by either a large rise in price or, in rare cases, a large fall in price
4. A large volume is usually followed by an increase in the price
5. If the volume keeps on decreasing consecutively for a period of 5 trading days, then prices tend to fall over the next four trading days
6. If the volume keeps on increasing consecutively for a period of 5 trading days, then prices tend to rise over the next four trading days.

Karpoff (1987), in addition, conducted a survey on the literature that took place on the relationship between volumes and returns. In his paper, Karpoff was able to summarize the following results from the literature at that time:

- o Volumes are positively correlated with absolute price changes. In this remark, Karpoff noted that the change in prices could take any direction. The movement depends on the magnitude of the volume change. This comes in line with the results published by Ying, and will be further tested in this paper.
- o The market normalizes when the information is fully revealed. This result will be very crucial in our analysis, where we expect that the adjustment of prices will fully take place after information of the closure of oil field will take place.
- o Volume is heavy in bull markets, light in bear markets. This result comes very intuitive as we expect investors to hold back on their investment whenever their unconfident about the market conditions, and vice versa.
- o Studies that use a mixture of price and volume data to draw inferences need to be aware of this relationship. For example, trading volume is often used to verify whether or not a price change was due to any informational content, and also whether investor interpretations of information are consistent or differing.

A paper by Lamoureux and Lastrapes (1990) tried to present empirical support for the above claim. By using daily return data from 20 actively traded stocks, they were able to find empirical support to the hypothesis that if we include volumes of trade as an alternate for the arrival of information to the market of individual securities in a generalized autoregressive conditional heteroskedasticity (GARCH) model. The variance of daily price increments shows a positive relationship to the rate of daily information arrival. Rashes (2001) also described an example

¹ Volume divided by shares outstanding is called turnover. This indicator is often preferred in most papers to avoid any bias. Of course, this procedure will not be needed in our study since we're only using a univariate model.

where information releases had an effect on the trade volumes and the prices. His paper examined the effect of information spread of one company on the volumes of trade and prices of another company that was linked to be taken over by the former. In his paper Rashes found that the days with the highest trading volumes during the period 11/1/1996 to 11/13/1997 all happened on periods when there was merger news on MCI Communications, showing that Massmutual's volume was correlated with MCI Communications' trading volume. Using an ordinary least squares (OLS) model, Rashes assigns a dummy variable to the period when there was the merger news. The usage of dummy variables proved in various studies its efficiency in testing the significance of the effect of certain events.

In his theory of the permeation of private information into the stock price, Morse (1980) took daily price and volume data from 1973 to 1976 for 50 securities from different stock markets in the US and used Capital Asset Pricing Model (CAPM) model to test the null hypothesis that there is zero serial correlation of returns during periods of abnormally high trading. Obviously, rejection of this hypothesis would be consistent with the use of asymmetric information and partially adjusting prices.

The desire to use residuals in the CAPM model is mainly due to the fact that private information is most probably specific to the individual firm or portfolio. That's why removing wide market non-informational factors will help isolate the effects of the private information. When the privately held information finally becomes public, the prices would complete their adjustment and the informed traders could harvest their profits. The null hypothesis is no difference between trading prior to a large price change and a small price change.

Morse in 1980 concluded that periods of abnormally large volume usually had positive autocorrelation of returns. He concluded that his findings were a result of the existence of asymmetrical information in the market. In particular, once investors are aware of an event that might cause higher earnings in the near future these investors will trade heavily on the issue until the price reflects the valuation of the security if the private information became public, and this persistent, upward, price movement should be reflected to a positive autocorrelation of price growth, as in our case.

In a different, but supporting approach, Copeland (1976) conducted an experiment on the behavior of asset prices and volumes of trade, while controlling for the arrival and spread of information. He compared the behavior of the bid-ask spreads and the fluctuations of volumes of trade under three scenarios: Strong-form market efficiency; semi-strong form efficiency; and private information, where traders use no market information. The paper concluded that the volumes of trade fluctuate the most as information tends to spread asymmetrically and, thus, the bid-ask spread grows as we move away from the case of strong-form market efficiency.

3. DATA AND METHODOLOGY

Our analysis is going to be on daily prices of Brent-crude oil and their trade volumes in the New York Stock Exchange (NYSE). The

data ranges from 11/23/2009 till 9/12/2014 and was downloaded from the Bloomberg database. In addition, given that previous studies have concluded that prices suffer from a unit root, we calculated the returns on prices using the following formula:

$$R(t) = \frac{P(t) - p(t-1)}{P(t-1)}$$

We will employ a vector autoregression (VAR) model to perform our analysis. VAR models were first introduced to econometrics by Sims (1980). VARs have often been advocated as an alternative to large-scale simultaneous structural models. One of the benefits of this model is that we don't need to specify which variables are endogenous or exogenous—they are all endogenous, unless there are some variables that are desired to be treated as exogenous based on economic theory. Also, given the generality of the model, the VAR model could be easily extended to other models i.e., vector error correction model, VAR moving average. Plus the model was able to overcome some of the old problems in econometrics like the identification problem in structural models. The VAR model assumes that the relationship can be described by the following structural form equation:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix},$$

We're expecting the signs of all coefficients of the model to be positive. Also, we note that this is a general model, and extensions may be applied further on depending on the tests that we will conduct.

Implementing a VAR model will enable us to conduct one of the most prominent features of the model and that is the impulse response function (IRF). The IRF captures the effect of an exogenous one-unit shock in the error term on the variables of the model. We note here that we use the Cholesky ordering in the basic form of IRF's.

Our next step will be to regress the error term in the second equation $e_{2,t}$ on $y_{1,t}$ and include a dummy variable that corresponds with the periods of interest in our study.

$$e_{2,t} = C_0 + C_1 Y_{1,t} + C_2 D_n + u_t$$

The aim of this step is to test for heteroskedasticity, and test for the presence asymmetric information in those period explained by abnormal behaviour of the error term in those periods.

4. EMPIRICAL ANALYSIS OF THE DATA

We first start by plotting the two series of our model to get an indication of the evolution of the series in our model. As we can see from Figure 1, the prices series doesn't look mean-reverting at the level. The evolution of the prices in Figure 1 tells us that the data might not be stationary at I(0). This preliminary result comes in line with previous literature that was cited earlier (Morse, 1980).

To get a more concrete indication, we show in Figure 2 the evolution of price growth (returns) and compare the results. We can clearly see that prices at the first-difference look more stationary than at the level. Both graphs show that prices are stationary at I(1), but we will confirm this result later on in this section once we perform the unit-root tests.

We now turn our attention to the evolution of trade volumes. We can see in Figure 3 that volumes of trade (vol) are mean-reverting at the level I(0). This result also comes consistent with stationarity tests that were conducted on volumes of trade in previous literature.

4.1. Normality

We test for normality for both variables of the model. If our data comes from a normal distribution, we consider the JB statistic has a Chi-squared distribution asymptotically. This statistic can be used to test whether the data comes from a normal distribution, or not. The null hypothesis in this test is a joint hypothesis of the

skewness and the Kurtosis. The null hypothesis assumes that under a normal hypothesis the expected skewness is zero and the expected value kurtosis is 3. The JB test takes the following form:

$$JB = \frac{n}{6} (S^2 + \frac{1}{4}(K - 3)^2)$$

S is the skewness, and K the kurtosis. As the definition of JB shows, any deviation from this increases the JB statistic.

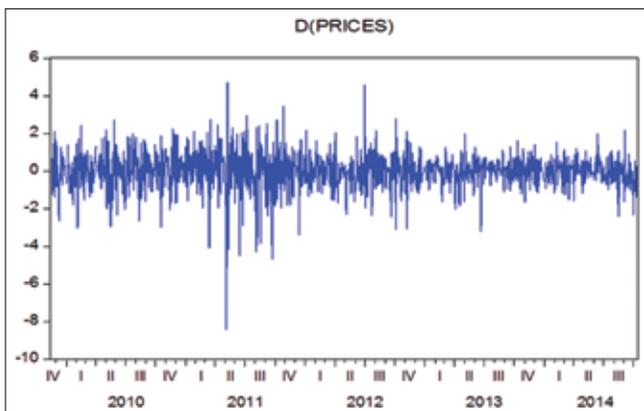
Figures 4 and 5 show descriptive statistics on our two variables. While the outcome of the volumes statistics comes close to our

Figure 1: The evolution of prices in the sample period



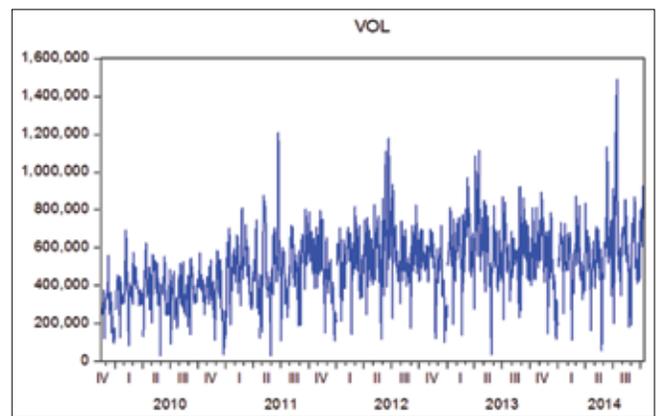
Source: Author’s calculation based on data collected from the New York Stock Exchange

Figure 2: The evolution of the first-difference of prices during the sample period



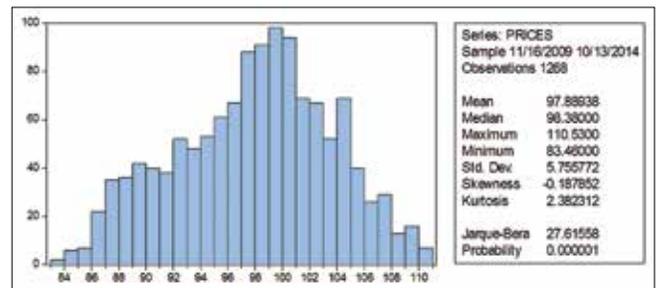
Source: Author’s calculation based on data collected from the New York Stock Exchange

Figure 3: The evolution of trade volumes in the sample period



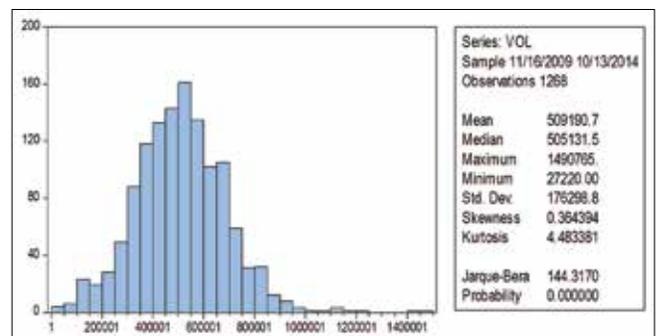
Source: Author’s calculation based on data collected from the New York Stock Exchange

Figure 4: Descriptive statistics of prices



Source: Author’s calculation based on data collected from the New York Stock Exchange

Figure 5: Descriptive statistics of trade volumes



Source: Author’s calculation based on data collected from the New York Stock Exchange

expectations, the results on prices, shown in Figure 4, comes as surprise to us. Although it's slightly skewed to the left and its kurtosis is near the normal value, the JB tests highly rejects the null hypothesis of normality. A similar test was conducted on the first-difference of prices, and the null hypothesis was rejected as well².

Figure 5 depicts the results that we got for the descriptive statistics of the trade volumes variable. From the Figure 5 we notice that there are a significant amount of outliers to consider the trade volumes series a normal. So we can confirm a shift from normality in the trade volumes. Its kurtosis 4.48 which is different from that of standard normal (which is 3) and so does its skewness. As a result, we highly reject the null hypothesis of normality.

4.2. Stationarity

As mentioned above, we test here for the presence of unit root in our two series. We have discussed earlier that prices, according to our plots and previous literature, will not be stationary at the level of the series. We reached this conclusion after testing for this using the Dickey–Fuller test. Now we go over three tests that are widely used to detect the presence of unit root in a series (Davidson, 2000).

4.2.1. Augmented Dickey–Fuller (ADF) test

The ADF test is an extension of the DF test, and it follows the same procedure but it is applied to the model, where we add lags of first difference:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t$$

Where α is a constant, β the coefficient on a time trend and P the lag order of the autoregressive process. Noting that this is the general formula, and the test could be implemented without a constant (drift) or without a constant and a trend. The number of lags to include in the model could be easily determined by using the information criterion models.

The null hypothesis of the test could be tested to verify if $\gamma = 0$, against the alternative hypothesis of $\gamma < 0$. Of course the value of τ could be in some rare cases larger than zero. But that only occurs when the data sample is relatively small. Our statistic will take the following form:

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$$

This τ then is compared to the relevant critical value for the Dickey–Fuller test. If the test statistic is bigger in absolute value than the critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present.

4.2.2. Phillips–Perron (PP) test

$$\nabla y_t = \rho y_{t-1} + u_t$$

∇ is the first difference of the variable at period t . similar to the ADF test, the PP test deals with data for y_t that might have a

higher order of autocorrelation than is formally reported in the test equation. The main virtue of the PP tests is that it makes adjustment for any serial correlation and heteroskedasticity in the errors term with a non-parametrical approach by modifying the Dickey–Fuller test statistics. PP's test statistics can be viewed as Dickey–Fuller statistics that have been made robust to serial correlation by using the Newey–West adjustment. We note here that if there was no serial correlation in the series the test will transform to a standard ADF test.

4.2.3. Kwiatkowski, Phillips, Schmidt and Shin (KPSS) test

The KPSS test has been widely used in the last two decades, the KPSS test, is due to Kwiatkowski et al. (1991). They derive their test by starting with the model:

$$y_t = \beta_0 Dt + \mu_t + u_t$$

Dt contains deterministic components which can be constant or constant with a time trend, the latter is more used in practice. The KPSS test was a twist to a long series of tests that tested the null hypothesis of the series being non-stationary. Some literature (Davidson) reported that existing tests might tend to believe that the series suffered from a unit root while it didn't. On the contrary, the null hypothesis of the KPSS test is that the series is stationary.

We conclude this argument by suggesting that all of the above test should be conduct to the same series for concreteness.

Table 1 clearly indicates that the two variable of the model (trade volumes, change in prices) are stationary at I(0).

4.3. Cointegration

In this study, we embark on examining the long run relationships between our two variables in the model. Among the cointegration techniques employed are the VAR-based multivariate Johansen. This will be our model of interest in conducting the test:

$$\Delta xi = n + \sum_{i=0}^{p-1} \phi_i \Delta xi + a\beta' X_t - 1 + \varepsilon$$

Where n is the number of variables in our model, ε is white noise, is our long-term relationship vector, and (a) and ϕ capture the fluctuation in the short-term.

In the Table 2, both methods used in the Johansen coninetgration test show that the two variables have more than one cointegration equilibria. The notation of having more than one cointegration equilibria comes as counterintuitive, where it demolishes the idea of the two variables converging on a unique long-term path. Nevertheless, as the test is only generated from the sample data, our main benefit from this test is to conclude that the two variables are cointegrated and have a long-term relationship, no matter how many cointegration equilibrias the test give us^{3/4}.

² The results are available upon request.

³ As shown in the equation, the Johansson cointegration technique relies heavily on the number of tested variables; in example, if we have a model that contains up to 8 variables, the Johansen test will test up to 8 possible cointegration equilibriums.

⁴ The results are available upon request.

Table 1: Unit root tests

Variable	ADF test	PP Zt test	KPSS test	Order of integration
D (prices)				
Statistic	-38.22967	-38.37276	0.16865	I (0)
P*	0.0000	0.0000	1%<t>5%	
Volumes				
Statistic	-3.65705	-20.94601	0.33760	I (0)
P*	0.0257	0.0000	1%>t	

*MacKinnon (1996) one-sided P values. **The probabilities in the ADF and PP are calculated by the Mackinnon one sided P values which are based on the df critical value. While the critical values for the KPSS test are based on critical values that were created by the authors of the test. In our sample case the critical values are: 0.216 for the 1% level, 0.146 for the 5% level, and 0.119 for the 10% level⁵. ADF: Augmented Dickey-Fuller, PP: Phillips-Perron, KPSS: Kwiatkowski, Phillips, Schmidt and Shin

Table 2: Johansson cointegration test

Number of cointegrations	Calculated statistic	Critical value	P**
Trace test			
None*	79.01136	15.49471	0.0000
At most 1*	5.154784	3.841466	0.0232
Maximum-Eigen value test			
None*	73.85657	14.26460	0.0000
At most 1*	5.154784	3.841466	0.0232

*Rejection of the hypothesis at the 0.05 level. **MacKinnon-Haug-Michelis (1999) P values

4.4. Granger Casuality

The last test that we perform in this section before moving on to the model estimation is the Granger casuality test. Performing the test will enable us to determine which direction the relationship between the variables takes. Also, in many practical works, it’s used as a tool to construct the ordering of variables in a model when conducting an IFR in a VAR model (Enders 2010). In addition, once knowing the direction of the relationship between variables, that will help us in projecting future values of the variable of interest: If the volume variable affects the growth in prices variable, the former should help improving the predictions of the latter variable.

The Table 3 shows that we can reject the first null hypothesis that says “trade volumes does not Granger cause prices.” This result comes as expected, and very intuitively as we expect more volumes of trade to be a strong indication of strong demand that causes prices to rise. The second null hypothesis that “prices doesn’t Granger cause trade volumes” is rejected at the 1% and 5% levels, but it’s accepted at the 10% level. In the latter case, we always expect that prices will have an effect on the volumes of trade.

5. THE MODEL

5.1. Lag Selection Criteria

The first step we take, in constructing our model, is to determine the number of optimal lags to include in our model. The procedure that we follow is quite counterintuitive but more practical; where we first run the VAR model for a “test-run” and then apply the the selection criterion tests to determine the optimal number of lags. Our final step is to run the VAR model

Table 3: Granger causality tests

Null hypothesis	Obs.	F-statistics	P
VOL does not Granger cause PRICES	1202	4.74613	0.0003
PRICES does not Granger cause VOL		1.91081	0.0898

Table 4: Selection criteria tests

AIC	$-2 \frac{\log k}{T} + \frac{k}{T} 2$
SC	$-2 \frac{\log k}{T} + \frac{k}{T} \log(T)$
HQ	$-2 \frac{\log k}{T} + \frac{k}{T} 2 \ln(\ln(T))$

Source: Davidson (2000). AIC: Akaike info criterion, SC: Schwarz criterion, HQ: Hannan-Quinn criterion

again with the number of optimal lags that were suggested by the tests. In this regard, we present our main selection criterion tests in Table 4.

In the Table 5 represents the number of observations and k represents the number of parameters. Here we note that the Akiake information criterion is considered biased towards high order of lags, while the Schwarz criterion gives more weight to less lags (Davidson, 2000). Hannin–Quinn (HQ) test is considered to be the most relevant criterion according to the literature.

As shown in Table 5, we have premitted the Eviews package to test for the optimal lag up to 90 period. Although this might seem extreme, but given the rich structure of the data, giving the package more lags to operate on wil give more accurate results. The results from Table 5 were similar to our predictions on the behaviour of the selection criterions. We conclude with choosing the 5th lag as suggested by the HQ test. Given that this is working-daily data, the 5th lag is a reasonable choice where it reflects that corresponding days of the week have similar patterns.

5.2. Regression

In our VAR model we tried to fit the best structure for our data, and our representation of the model where the change in prices is the dependent variable is:

$$D(\text{PRICES}) = C(1)*D(\text{PRICES}(-1)) + C(2)*D(\text{PRICES}(-2)) + C(3)*D(\text{PRICES}(-3)) + C(4)*D(\text{PRICES}(-4)) + C(5)*D(\text{PRICES}(-5)) + C(6)*\text{VOL}(-1) + C(7)*\text{VOL}(-2) + C(8)*\text{VOL}(-3) + C(9)*\text{VOL}(-4) + C(10)*\text{VOL}(-5) + C(11)$$

The C’s in the above equation represent the coefficients. The coefficients that we obtained from the VAR models were:

$$D(\text{PRICES}) = -0.0857642158517*D(\text{PRICES}(-1)) - 0.0586659993177*D(\text{PRICES}(-2)) + 0.0184358326867*D(\text{PRICES}(-3)) + 0.00123809476594*D(\text{PRICES}(-4)) - 0.0360604002967*D(\text{PRICES}(-5)) - 2.76281732617e-07*\text{VOL}(-1) + 1.19382756347e-07*\text{VOL}(-2)$$

5 The detailed results are available upon request

$$- 1.98872978363e-08 * VOL(-3) + 5.76915362094e-07 * VOL(-4) - 3.26437450876e-07 * VOL(-5) - 0.0462414453954$$

We note that the coefficients of the lags in the model turned out with varying signs. This might be a result of the volatility of trade and prices in the NYSE. We also note that after running a serial correlation on the residuals of the model, we found that there was no presence of serial correlation (Table 6) in all the lags except the 4th lag. We make a remark here that the 4th lag of the trade volumes was the most significant out of the covariates of the model.

The next step that we did was to test for the forecasting ability of the model in hand by conducting an in-sample forecast. In our model we truncate the last 21 days of the series (from September 15th to October 13th 2014), and try to forecast those days with our obtained regression results. The next step is to compare the forecasts of those 21 observations with the actual values by plotting both of them in the same graph (Figure 6).

In the Figure 6, the blue line represent that actual values of the prices (PRICES) and the red line represents the forecasts of those values (PRICES [Scenario 1]). We can see that there is a small deviation in our forecasts from the original values. For concreteness, we also compute the ratio of the two standard deviations. By computing this ratio we're determining how much our forecasts are capturing the deviation in our model of interest. The closer we're to one the more accurate the forecasting power of our model is. The calculated ratio in our model was about 93%, which means that our forecast is missing 7% of the deviation in the original series. Nevertheless, given the complexities of the determinants of the oil prices, this model turns out to produce decent forecasts.

5.3. Detecting the Presence of Asymmetric Information

After obtaining the regression results we would like to investigate the impulse response relationship between two variables in a higher dimensional system. Lamoureux and Lastrapes (1990), concluded in their study that the sum of the coefficients on the ARCH and GARCH specification would be a measure of the persistence of the shocks. As the sum of the coefficients of the GARCH is 0.826381 close to unity we can say there is a great persistence of shocks to volatility. We will study this type of

causality by tracing out the effect of an exogenous shock or innovation in the error term of the trade volume variable on the return variable. As noted above, according to the Cholesky approach, the order of the variables plays a crucial role in determining the magnitude and length of the shock. Many scholars have lately criticized this approach because of its limitations in multivariate regressions. But in our case that will not impose a problem in our case since the relationship is already defined, and the causality effect has already been proven by the Granger causality test above.

The Figure 7 shows that a one standard deviation shock in the volume variable will lead the growth of returns to go into negative territories until it reaches -0.03 in the second period of the shock. Then it goes back up until it reaches 0.06 in the fifth period. The effect of the shock dies out after 10-11 periods. The changing direction of the return series after the shocks comes in line with the empirical studies that were conducted on the relationship between the two variables of interest. These studies confirm that the direction of the return variable is mainly affected by the timing and persistence of the shock. Nevertheless, our main goal from this procedure is to get an estimate of the time length we should assign to our dummy variables further on.

We now regress the residuals obtained from the second model, where the trade volumes variable was the dependent variable, on the prices variable to control for heteroskedasticity and also to control for symmetric information, we also include the dummy variables for the periods that correspond to 10 days

Figure 6: In-sample forecasting



Figure 7: Impulse response function

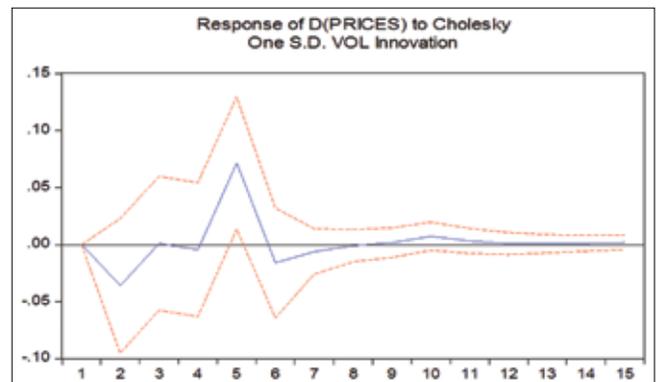


Table 5: VAR lag order selection criteria

Lag	AIC	SC	HQ
0	32.84426	32.86104	32.85085
1	29.13610	29.18645*	29.15585
2	29.12870	29.21261	29.16162
3	29.11723	29.23470	29.16331
4	29.10877	29.25980	29.16802
5	29.07957	29.26416	29.15198*
6	29.07756*	29.29572	29.16314
7	29.08081	29.33254	29.17956
.			
.			
.			
90	29.27483	32.31231	30.46643

*Lag order selected by the criterion. AIC: Akaike info criterion, SC: Schwarz criterion, HQ: Hannan-Quinn criterion

before the disruption occurred. We assign a dummy variable to each period:

- D1: The last 10 days of July 2013, corresponding to the oil disruption that occurred in Libya.
- D2: The last 10 days of May 2013, corresponding to the oil disruption that occurred in Nigeria.
- D3: The last 10 days of August 2013, corresponding to the oil disruption that occurred in Iraq.

We run the following regression 3 times interchanging the dummy variables in each regression.

$$e_{2,t} = C_0 + C_1 Y_{1,t} + C_2 D_n + u_t$$

We note here that we also tried running the regression while including all of the dummy variables at once, but the results were pretty much similar to the ones we got when we regressed the residual term on each dummy variable on its own.

The results shown in Table 7 clearly indicate that all of the dummy variables coefficients were insignificant. The OLS results above indicate that prices have a significant but negligible effect on the error terms in all three models. Also, all of our three models, although significant, showed a very low explanation power that didn't exceed 1% in all three variables. This result might also be attributed to the size of the disrupted quantities compared to world supply of oil. All three disruptions combined only represent about <2% of world supply of crude oil. Details of all three regressions are found in Appendix (5).

5.4. Testing for Structural Breaks

As our last robustness test, we try to detect the presence of a structural break in our VAR model, if there's one. We argue here that if detect a structural break in the model, we might be able to find a period where there was asymmetric information prior and during that structural break. In addition, most of the tests conducted to test for a structural break in the data are mainly focused on the error term as the main indicator of a change in the relationship between the variables of the model.

We turn our attention to two widely used tests that are used in testing for structural breaks in econometric model:

The Chow structural break-point test (known break-point):

Assuming that the estimated regression is:

$$Y_t = a + b_1 x_1 + b_2 x_2 + e$$

We split the time series based on a chosen structural point:

$$Y_{1t} = a_1 + b_2 x_1^1 + b_2 x_2^1 + e_1$$

$$Y_{2t} = a_2 + b_2 x_1^2 + b_2 x_2^2 + e_2$$

Under the following null and alternative hypothesis:

Table 6: VAR residual serial correlation

Lags	LM-statistics	P
1	3.328580	0.5044
2	4.925860	0.2950
3	2.832056	0.5863
4	11.42412	0.0222
5	3.448623	0.4857
6	4.929834	0.2946

P value from Chi-square with 4 df. VAR: Vector autoregression

Table 7: Regression results of the residuals on selected dummy variables

Covariates	Model 1	Model 2	Model 3
Covariates			
Constant	0.259747 (2.800310)***	0.279802 (2.733567)***	0.275208 (2.691567)***
Prices	-5.07E-07 (-2.967157)***	-5.27E-07 (-2.728254)***	-5.25E-07 (-2.713649)***
D1	0.162607 (0.452267)		
D2		-0.344068 (-0.931970)	
D3			0.161359 (0.436122)
F	4.533558***	4.050290**	3.708829**
R ²	0.007568	0.007472	0.006847
Adjusted R ²	0.005899	0.005627	0.005001
D.W	2.009316	1.996088	1.995907

The terms in the brackets represent the t-statistic of the variables. *10% significance, **5% significance, and ***1% significance

Table 8: Quandt-Andrews and Chow test results

Statistic	Value	P*
Quandt-Andrews		
Maximum LR F-statistic (5/09/2011)	5.917216	0.0025
Maximum Wald F-statistic (5/09/2011)	23.66887	0.0025
Exp LR F-statistic	1.559357	0.0182
Exp Wald F-statistic	8.160222	0.0029
Ave LR F-statistic	2.685281	0.0055
Ave Wald F-statistic	10.74112	0.0055
Chow breakpoint test 5/09/2011		
Null hypothesis: No breaks at specified breakpoints		PF (4,1232)
F-statistic	5.917216	0.0001

*P values calculated using Hansen's (1997) method

$$H_0 : a_1 = a_2, b_1^1 = b_1^2, b_2^1 = b_2^2$$

$$H_1 : a_1 \neq a_2, b_1^1 \neq b_1^2, b_2^1 \neq b_2^2$$

The test takes the following expression:

$$f = \frac{(\sum e^2 - (\sum e_1^2 + \sum e_2^2) / (k))}{(\sum e_1^2 + \sum e_2^2) / (m1 + n2 + 2k)}$$

Thus, if our calculated coefficient is significant, we conclude that our model suffers from a structural break.

Quandt-Andrews test (unknown break-point):

Table 9: Regression results with structural break tests

Covariates	Model 4					
	Constant		Prices		D4	
	0.247425	(0.247425)**	-4.41E-07	(-2.333810)**	-3.589884	(-7.037580)***
Supplementary statistics	R ²	Adjusted R ²	F-statistic	D.W		
	0.050381	0.048616	28.54316***	2.054487		

The terms in the brackets represent the *t*-statistic of the variables. *10% significance, **5% significance, and ***1% significance

The Quandt–Andrews test is conducted on the basis that the structural break point(s) are unknown. The three tests that are developed by Quandt–Andrews are based on the Chow test:

- o The maximum statistics. This test chooses the point which has highest probability of being a structural breakpoint, and it takes the following expression:

$$\text{Max } F = \max_{t_1 \leq t \leq t_2} \{F(t)\}$$

- o The exponential test. It takes the following expression:

$$\text{Exp } F = \ln \left[\frac{1}{K} \sum_{t=t_1}^{t_2} \left(\frac{1}{2} F(t) \right) \right]$$

- o The average test, this test takes average of all the iterated tests that are made on all the possible structural breakpoints:

$$\text{Ave } F = \frac{1}{k} \sum_{t=t_1}^{t_2} F(T)$$

In our case it’s more convenient to start the procedure with the Quandt–Andrews test, and then confirm the results by the Chow test, as shown in Table 8.

All of the Quandt–Andrews results shown in Table 8 clearly indicate that our model suffers from a structural break. Furthermore, the maximum statistics test proposes that the most likely structural-break point is 05/09/2001. We verify that by conducting the Chow structural-break test on that same date.

We go back to our residuals model and regress the residuals on the new dummy variable (D4) detected from the structural break tests, and we get:

The results shown in Table 9 indicate that the dummy variable characterizing the period prior to the structural-break point is significant at the 1% level. We also notice that there’s a general improvement in the model; R² increased to 5% but it’s still in the bottom levels, and the *F* static is more significant than before. The structural break that occurred in May 2011 was attributed to bearish market sentiment, which triggered an outflow of investment from the paper oil market, according to OPEC’s monthly oil report. Nevertheless, there was a spreading rumor during that time that there will be a releasing of large volumes of the IEA’s Strategic Petroleum Reserve oil World supply-demand to restore balances, which helped mitigate further market price increases mid-2011. The effect of this rumor can also be attributed to the quantities of supply in this matter, where the IEA has significant amounts of reserves that might actually affect the demand-Supply balances, unlike the first three incidents.

6. CONCLUSION

Using a VAR model approach to detect the presence of asymmetric information in the NYSE regarding oil disruption in some oil producing countries, we fail to support those claims made by several officials in those countries. We tried to put those claims under investigation by running test on three oil disruption incidents that occurred in 2013, and all of the results turned out to be insignificant. Nevertheless, the model detects a period which might involve asymmetric information in the NYSE (05/09/2011). We link the significance of our results to the quantities of oil relative to the total world supply of oil, where the amount of oil evolved in each incident plays a role in determining the significance of our tests.

The IRFs that we employed in this paper indicates that a shock in the trade volumes would have a varying moderate impact on oil prices that would last to about 10 days.

Lastly, previous papers pointed out that asymmetry in information may be squandered within a few hours. In that case, daily data will not be sufficient to detect any serial correlation in returns. So we suggest the use of other models to detect this phenomenon more precisely. The Markov switching model might be a good candidate to fit this kind of relationship that probably has more than one structural-break point.

REFERENCES

Chow, G. (1960), Tests of equality between sets of coefficients in two linear regressions. *Econometrica*, 28(3), 591-605.

Copeland, T. (1976), A model of asset trading under the assumption of sequential information arrival. *Journal of Finance*, 31(4), 1149-1168.

Davidson, J. (2000), *Econometric Theory*. 1st ed. United Kingdom: Blackwell Publishing Limited.

Enders, W. (2010), *Applied Econometric Time Series*. 3rd ed. United Kingdom: Wiley, John and Sons, Incorporated.

Karpoff, J. (1987), The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22(1),109-126.

Karpoff, J. (1998), Costly short sales and the correlation of returns with volume. *Journal of Financial Research*, 11(3),173-188.

Kwiatkowski, D., Phillips, P., Schmidt, P. (1991), Testing the null hypothesis of stationarity against the alternative of a unit root, how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54, 159-178.

Lamoureux, C., Lastrapes, W. (1990), Heteroskedasticity in stock return data: Volume versus GARCH effects. *The Journal of Finance*, 45(1), 221-229.

Morse, D. (1980), Asymmetrical information in securities markets and trading volume. *Journal of Financial and Quantitative Analysis*,

- 15(5), 1129-1148.
OPEC, Monthly Oil Report, June 2011.
- Osborne, M. (1959), Brownian motion in the stock market. *Operations Research*, 7(2),145-173.
- Rashes, M. (2001), Massively confused investors making conspicuously ignorant choices (MCIMCIC). *Journal of Finance*, 56(5), 1911-1927.
- Sims, C.A. (1980), Macroeconomics and reality. *Econometrica*, 48(1), 1-48.
- Sun, W. (2002), Relationship between trading volume and security prices and returns. MIT Laboratory for Information and Decision Systems Technical Report; p2638.
- Ying, C. (1966), Market prices and volumes of sales. *Econometrica*, 34(3),676-685.