PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES
DEPARTMENT OF ECONOMIC & REGIONAL DEVELOPMENT

MSc IN APPLIED ECONOMICS AND ADMINISTRATION

OIL PRICE VOLATILITY WITH ULTRA-HIGH FREQUENCY DATA

By

Pagkalou I. Foteini

Athens, 2016

Preface

This thesis completes my master's degree in Applied Economics and Administration at the Department of Economic and Regional Development (TOPA) at the Panteion University of Social and Political Sciences. The study has been carried out during the winter of 2015 under the supervision of Prof. Stavros Degiannakis.

Foremost, I would like to express my sincere gratitude to my advisor Professor for the continuous support of my study and research and his positive and encouraging attitude. His guidance helped me throughout my research and writing of this thesis.

In addition, I would like to thank the rest of my thesis committee: Prof. Theodosios Palaskas and Prof. Veni Arakelian. My gratitude also goes to the teachers of the Department for the knowledge offered to us throughout our education until graduation.

Foteini I. Pagkalou
Athens
May, 2016

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AIC=Akaike Information Criterion

AR = Autoregressive model

ARCH model = Autoregressive Conditional Heteroscedasticity models

ARIMA model = Autoregressive Integrated Moving Average model

CBOE=Chicago Board Options Exchange

DW= Durbin-Watson statistic

ECM = Error Correction Model

EGARCH = Exponential Generalized Autoregressive Conditional Heteroskedastic model

EUVIX=CBOE/CME FX Euro Volatility Index

GARCH model = Generalized Autoregressive Conditional Heteroscedasticity model.

GDP = Gross Domestic Product

GVZ=CBOE Gold Volatility Index

HAR = Heterogenous Autoregressive Model

HQ=Hannan-Quinn criterion

IV = Implied Volatility

LM = Long Memory

MedRV = Median Realized Variance

OPEC = Organization of the Petroleum Exporting Countries

OPV = Oil Price Volatility

OVX= CBOE Crude Oil ETF Volatility Index

RV = Realized Volatility Measure

S.E.= Standard Error of the Regression

SC=Schwarz Criterion

SV models = Stochastic Volatility models

TYVIX=Treasury Volatility Index U.S

VaR = Value at risk

VIX =CBOE Volatility Index

WTI = West Texas Intermediate

Abstract

This study aims to identify relationships between oil price realized volatility and implied volatilties. Studying the relevant literature I have found that there is a limited material on the estimation and in particular in the predictability of oil price realized volatility. The chief aim of this study is in terms of improving our understanding of dynamic properties of volatility which are key factors for forecasting, especially realized measures that are valuable predictors of future crude oil price volatilities, and most of all offering quantitative studies of new volatility which could aid in facilitating and improving estimation of complex Volatility Models and above all to consider what has been the trends of crude oil price volatility using predictive HAR methods. More specifically, in our research, we created six Heterogenous AutoRegressive (HAR) models. First, we estimate the HAR model for the realized volatility and subsequently we create five HAR models by adding the implied volatilities of another assets, in the effort to identify the best performing model for the estimation of oil price realized volatility. Checks carried out at statistical level and econometric analysis, we found that the combination of crude oil volatility with implied volatilities, results in a better model that improves the estimating accuracy. In particular when there is a correlation between crude oil and the other asset, such as crude oil volatility index (OVX), gives even better results.

1. THESIS INTRODUCTION

1.1 INTRODUCTION TO THE TOPIC

This paper concern itself with identifying, assessing, analyzing, quantifying and investigating the Crude oil price volatility through information gained through ethnographical studies and data gleaned from Energy Forecasting Reasearch Project (Enefor). This study affords front seat view of current trends and movements of oil price volatility estimating with High-Frequency Data in short term and also approximately in long terms. Through appropriate HAR Reporting studies it is possible to quantify oil price volatility estimating and forecasting with High-Frequency Data deployment which measures crude oil price vicissitudes with reasonable degree of accuracy, genuineness and authenticity. There are plethora of studies on topic of fluctuating oil crude prices that offer valuable insights and commentaries but only a few elements over the last 12 year period on specification how oil price realized volatility works, what it affects and how it may best be determined. For these, we needed to isolate the gaps in current literature and fill on this thesis these gaps convincingly and added to existing body of literature on this topic.

The assessment and prediction of variation in the price of oil, is important mainly because of the large variability presents, and the effect that has on economic policy. Indeed, modeling the volatility of oil prices and shaping the criteria of which depends directly can help not only the science of analysis and decision making of enterprises and organizations, but also at state level can be a key tool for improvement of macroeconomic policy responses. This arises from the fact that the price of crude oil and its derivatives, are: a) one of the key variables in the production of macroeconomic projections, b) gives useful information for predicting recession since the unexpected, large and persistent fluctuations real price of oil is harmful to the well-being of both the oil import, and economies that produce oil, c) affect consumer sentiment in a wide range of sectors, including utilities, d) is important in modeling markets durable goods such as vehicles and domestic heating systems and e) can play an active role in the provisioning which depend directly on energy use, such as investment decisions, modeling in the energy sector, the provision of carbon dioxide emissions and climate change and the regulatory policies design. However, despite the importance that the study shows the configuration of the oil price is difficult to model the volatility. This first arises from the fact that the choice of the sample period and the description of models per time series, since the study of the literature shows that other models better suited to long-term period and others in the short and different results are difficult to get as the selection of the sample range. Here, one would expect that the larger the sample, the best predictive power will model. However, many researchers conclude that greater value is to study outliers (eg in times of crisis, war, speculative demand), rather than the data observed for a longer time. Another problem that arises concerns the study of those parameters and items affecting the determination of the value (real or nominal). Many researchers are finding that there is a dependency between both macroeconomic goods, in particular on GDP growth, and the demand and supply conditions. Specifically, the changing circumstances, to reach any reliable conclusions should be taken into account in addition to alternative assumptions and evaluation of sample sensitivity to events socially and historically. From all the above, we conclude that it is useful assessment of and that even this can translate into profits. For this in our work we try to identify those models that best meet in the short run, since it is a piece that has not been investigated largely even, while early studies show that the use of ultra-high frequency data such as intraday it is extremely useful in predicting the volatility of futures of crude oil. We will also attempt to model the fluctuation in oil prices in combination with other goods and in the various stock markets and compare the extent to which the estimate of affected.

1.2 A GLACE AT THE EXISTING LITERATURE

Breeden and Litzenberg (1978) have been long working on deriving probability densities, being consistent with all option prices noted at the same point in time. Britten-Jones and Neuberger (2000) have presented and evaluated various procedures for modeling and forecasting volatility. They tried to shed light on crude oil futures market where oil futures contract is the world's largest trading by volume and the most liquid contract. Given the remarkably high volatility of the market, testing volatility and forecasting models is an enormous challenge.

We will present the following models: GARCH (1.1), exponential GARCH (EGARCH), GJR-GARCH, APARCH, FIGARCH and HAR.

There is no such a significantly negative and asymmetric relationship between changes in volatility and returns. In short, HAR models seem to be the best in out – of- sample performance.

The sample standard deviation of close–to–close return is:

$$\hat{\sigma}^2_{c,t} = \left(ln \frac{c_t}{c_{t-1}} \right)^2$$

Where

c_t stands for the closing price of the trading day *t*

The estimator of volatility has, then, as follows:

$$\hat{\sigma}^2_{P,t=}\frac{1}{4\ln(2)}\left(ln\frac{H_t}{L_t}\right)^2$$

Starting with GARCH(1,1) model:

$$r_t = \mu_t + \varepsilon_t, \varepsilon_t = z_t\sigma_t \sim NID(0,1)$$

$$\sigma_t^2 = \omega_{+\alpha}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2$$

Where

r_t is the daily return
μ_t is the conditional mean
σ_t^2 *is the conditional variance of returns in period t-1*

Nelson (1990) argued that the non-negative constraints in the linear GARCH model are too restrictive. Thus, in EGARCH model specifications there are no restrictions regarding the parameters α and β. It is given as:

$$\text{Log } (\sigma_t^2) = \omega + \alpha(|z_{t-1}|) - E(z_{t-1}) + \gamma z_{t-1} + \beta log(\sigma_{t-1}^2)$$

Where

$$z_{t=}\sigma_t^{-1}\varepsilon_t$$

The estimated conditional variance is strictly positive and it does not require the non-negativity constraints used in the estimation of GARCH models. By including both the standardized value and its absolute value, the variance equation is allowed to capture any asymmetry in the relation between market returns and conditional volatility.
Fulvio Corsi, Francesco Audrino and Roberto Reno (2012) point out that "volatility dynamics have been modeled in order to take into account their most salient features: clustering, slowly decaying auto-correlation, and non-linear responses to previous market information of a different type". Volatility persistence seems to be the result of the aggregation and collection of the heterogeneous components present in financial markets. Heterogeneity may be of various differences in locations, risk profiles, information etc. Traders, market makers, institutional investors or insurance companies trade with different frequencies. Traders trade every day, having an intraday horizon, whereas companies trade less frequently. Different types of traders provoke different types of volatility components.

It has been noticed that volatility over the long-run horizon has more robust influence on volatility over the short-run horizon. This is because short-term (intraday) traders' reaction is influenced by the expected future size of trends and risk. The heterogeneous volatility components lead to a simple AR model that considers volatilities realized over different time horizons and hence is called HAR type model. "The combination of the ease of implementation with a very accurate fit of financial volatility time series has made the HAR models very popular in the financial econometrics community" (Fulvio Corsi, Francesco Audrino, Roberto Reno (2012)).

The most important features of the financial markets' volatility is the long-range dependence (realized volatility provides auto-correlations at very long lags), the leverage effect (returns are negatively correlated with volatility) and the jumps (extreme price variations).

"In practice, the HAR model provides a simple and flexible method to fit the partial auto-correlation function of the empirical data with a step function which has a predefined tread depth and an estimated rise height….it fits the persistence properties of the financial data, as well as (and potentially better than) true long memory models, such as the fractionally integrated one, which, however, are much more complicated to estimate and deal with. For these reasons, the HAR model has been employed in several applications in the literature, of which we provide an admittedly incomplete list", (Fulvio Corsi, Francesco Audrino, Roberto Reno (2012).

Thus, high frequency data are more appropriate in measuring, modeling and forecasting volatility (Barndorff-Nielsen and Shephard, 2007). This is also confirmed by Hansen and Lunde (2010) and Sevi (2014). Intraday high frequency data provide a more accurate measure of the current volatility, which, in turn, is rather useful in forecasting future volatility. Research in the University of Panteion (2015) extends Haugom (2014), Sevi (2014) and Prokopczuk (2015) research, which focus on oil price realized volatility and forecast using the HAR-RV model by considering 14 exogenous variables (such as stocks, foreign exchange, macroeconomics, etc) to examine whether their realized volatilities influence oil volatility forecasts. In Panteion's study it is reported that Phan (2015) examined whether the S&P500 volatility improves the oil price volatility forecasts, by proving actually that cross-market volatility interaction improves the forecasts for the oil price volatility.

1.3 RESEARCH QUESTION

The major volatility that is currently observed in crude oil has defied major quantification and it is now crucial that major methods are set in a place where they could afford quantification of the elusive, oil price management over time. The research has as a target to address the model so that we can estimate and predict oil price volatility. More specifically, the main emphasis is given on the nature and capturing of the realized volatility of Brent Crude Oil, as well as references that other assets, such as the macroeconomic indicators and indexes of various economic assets, can play a key role on that. Furthermore, the research aims at defining the basic gains of participants – indicators, which are referred to the enhancement of their overall contribution and risk mitigation from embarking on wrong estimates. Thus, triggered by the problem statement, the main research question can be expressed as follows:

"Do the HAR-X models of oil price realized volatility and other assets implied volatilities and services provide a comparative advantage to the calculations on variations, against those who have been performed only on the basis of oil prices volatility?"

Some of the aforementioned problems concern - apart from the importance of reclaiming HAR models on estimations and predictions of oil price volatility (Corsi, 2009 and Sevi, 2014) - the necessity for higher utilization of different assets on forecasting the realized volatility of oil (Degiannakis and Fillis, 2015). Thus, after taking into consideration these challenges, our analysis and interest behind this research question is to examine whether or not the assets that we have selected to investigate, could be an attractive option for our performance.

1.4 RESEARCH METHODOLOGY

Our methodology was based essentially on the fundamental character and nature of the previously discussed research question. Having as an aim to introduce a model that could offer an adequate quantity of significant evidence and provide answers related to the research question, we selected the ordinary least squared method in our study. Data analysis was performed with the OLS regression technique, which, according to Rutherford (2001), constitutes the basis of several other relevant techniques. The specific technique in combination with the dummy variable coding, may also be used to incorporate grouped explanatory variables and data transformation methods. An advantage of OLS regression technique is that it allows us to control the model assumption, linearity, constant variance and the effect of outliers by using common graphical methods (Hutcheson and Sofroniou, 1999). In addition, secondary data, such as previous reports, essays, published articles, etc, were utilized to obtain useful answers for the sub-question of our research.

Secondary studies using HAR methods, which offers asymmetric propagation of volatility between the long and short term horizons, and offers addictive cascades model of different volatility components designed to imitate actions of diverse types of market players. The Model HAR is short for Heterogenous Autoregressive Model of realized variance (HAR-RV) and is essentially predictive model for daily integrated volatility.

1.5 THESIS OVERVIEW

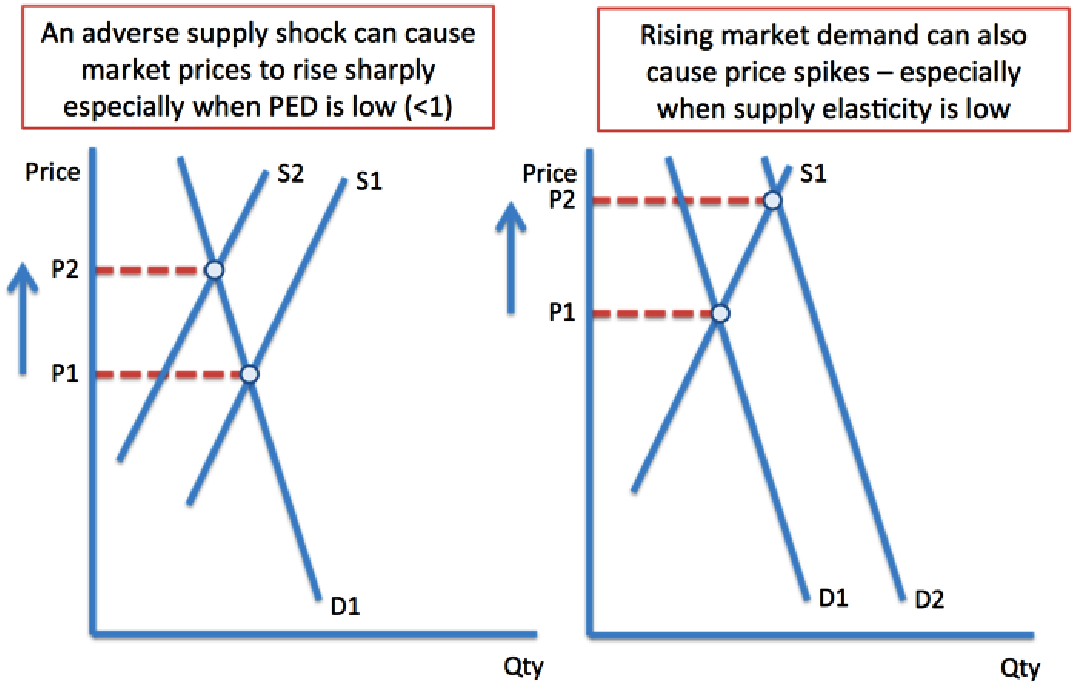
In this first section, we provide a short introduction to the topic, in careful apposition, as a presentation of the main objectives and the scope of this study. In the second section of our study, the structure and the progress of oil price volatility is presented, in order to understand under which conditions it can be estimated. After data selection and model configuration, the statistical software of choice is adopted and the methodologies on interpreting and assessing their fit are compared. The results are recorded and analyzed on the “Empirical Results and Discussion” Chapter and final quantitative studies of the HAR model are presented, based on the previous Chapters. Finally, in this part we present the gleaned findings on the usefulness, accuracy and genuineness of the HAR Models, their constraints - as much as this paper is concerned - and most of all, the major benefits it augurs for study purposes. Final conclusions are provided at the end of this study with recommendations for further research.

2. OIL PRICE VOLATILITY ANALYSIS

2.1 BASIC ECONOMIC THEORY ABOUT OIL PRICE

The following figure will illustrate the theory about the price and the quantity of oil using some basic microeconomic theories. When a fine, in our case, crude oil, is rare and has very few subsidies the demand is meant to be inelastic. This means that even if the price rises up, the quantity demanded will slightly change. Whereas if the demand of a product is elastic (e.g. the branded athletic shoes “Nike”), then an increase in the price of goods may induce a large change in the quantity demanded, since consumers will turn to other products (sneakers, cheap athletic shoes with no brand name, etc). In our case oil has inelastic demand and that makes the whole situation much more interesting. The importance of finding evidence, so as to be able in the future to control its performance is a real bet for econometricians, politicians and researchers.

Figure 2.1: Causes of price volatility – Supply and Demands Analysis



Source: www.tutor2u.net

If supply is shrinked from s1 to s2 in the case of an inelastic demand, a rise in price will be induced, decreasing at the same time the quantity of the supplied good, as shown in the above figure on the left.

On the other hand, an increase in the demand of the good from D1 to D2 will provoke an increase in the price of the good.

This provides a pure view of the nature of the good of crude oil. It is a rare good, with inelastic demand, few producers, large demand and few subsidies.

VIX measures the expected volatility in the S&P 500 Index over the following 30 days. Similarly in oil, the Oil VIX measures the degree of risk in oil prices and it is an indicator of close-time-horizon price movements. When the Oil VIX rises, oil prices tend to fall and vice versa.

Figure 2.2: Relationship between the Oil VIX and oil prices

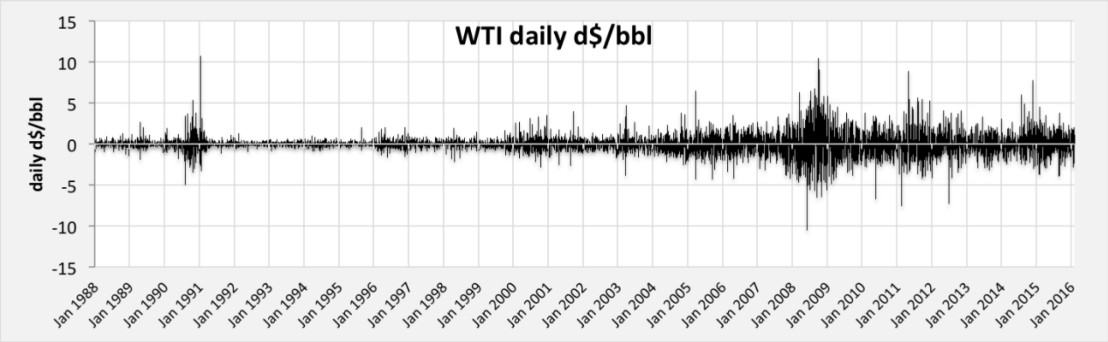
Source: EIA CBOE

The above figure presents the relationship between the Oil VIX and oil prices. The arrows show how oil prices have fallen when the Oil VIX rises. An increase in the Oil VIX implements a larger volatility (i.e. risk) and a larger possibility for additional price decreases.

It is difficult to make a good prediction where oil prices will go from there. This is because of the exogenous factors of supply and demand that have an impact on its price. Therefore, in the lack of war, terrorist attacks, natural disasters or other unforeseen events, the price of oil may remain range-bound and possibly fall even further.

The following figure outlines much better the increase of oil price volatility over the years.

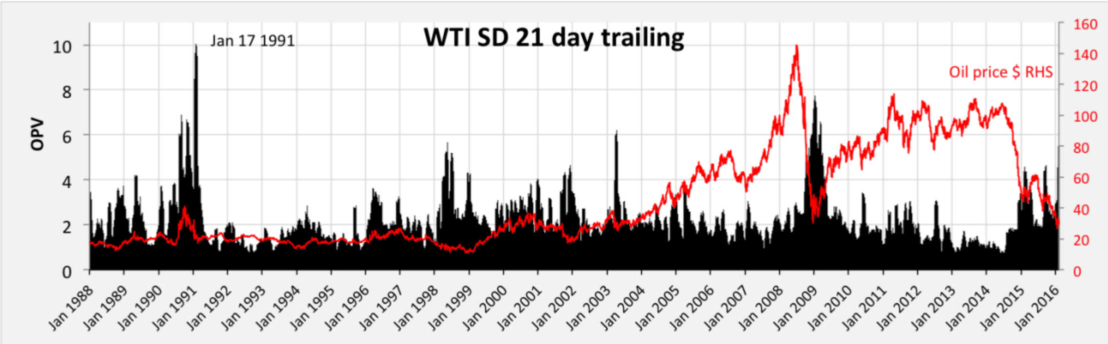
Figure 2.3: The increase of oil price volatility over the years



Source: www.oilprice.com

Oil has become more volatile over the years. Periods of variability are evidently illustrated with an obvious upward trend towards the variability of a larger extent as the years pass by. This implicates the augmentation in oil price.

Figure 2.4: Periods of variability



Source: www.oilprice.com

The red line represents the oil price fluctuation and the black fluctuations indicate the price volatility. As it can been seen, volatility is not always distinctive and distinguishable in oil price volatility. When oil price volatility (OPV) is greater than 4, volatility is supposed to be high. When OPV is less than 2, volatility is supposed to be low. Four major periods of high volatility may be distinguished from the aforementioned elements:

a) 1990 - Iraq's invasion in Kuwait

b) 1991- Initiation of Desert Storm Operation

c) 1998/9 - Oil price crisis (price decrease)

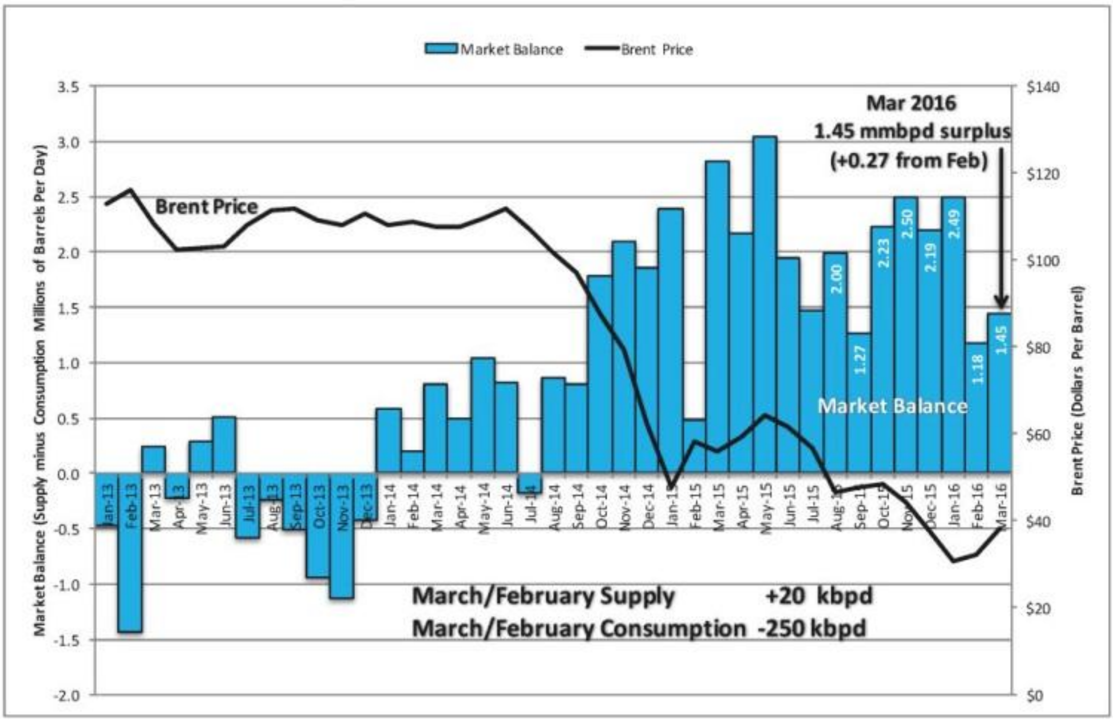
d) 2008 - Financial crash

Once, again the price does not always display the actual volatility and the occurring changes.

We may say thought that when price falls, traders put their bets, since it is considered to be the “gold” period to trade and hence frequency at short-run volatility is rather high.

Oil price volatility may provide and indicator of the future trend of the oil price and that’s the main reason why people in research focus that much on oil price volatility. After periods of high volatility, prices tend to rise.

Figure 2.5: EIA World Liquids Market Balance (Supply minus Consumption)



Source: www.oilprice.com (EIA data)

In March 2016 oil reached the highest price so far, indicating simultaneously a supply dive of crude oil.

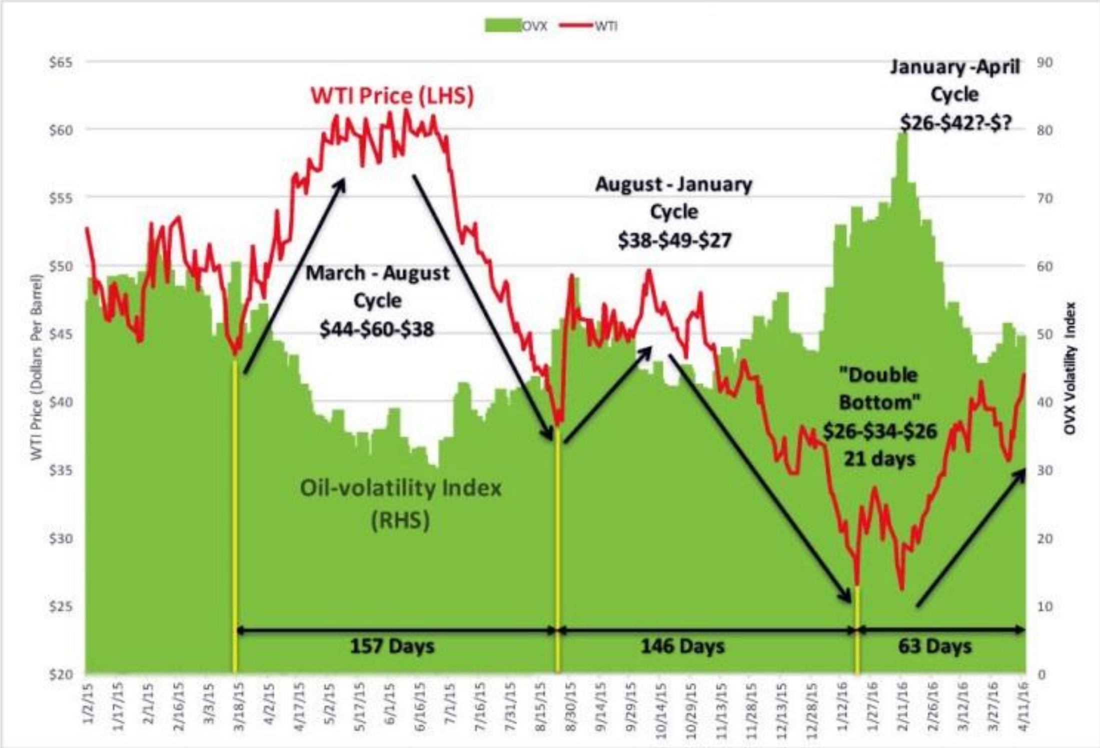
The net surplus is defined as follows:

NET SURPLUS = SUPPLY – CONSUMPTION

In other words the net surplus is the difference between supply and consumption. The surplus in March 2016 has risen to 1.45 million barrels per day in comparison with the net surplus in February, which was 270.000 barrels per day. The high price with the high surplus is an indicator of the difficult and relatively durable recovery period that will follow, in order to reach the level of a lower price (jump recovery).

“Meanwhile, on Wednesday, April 12th, Brent futures closed at almost $45 and WTI futures at more than $42 per barrel, the highest oil prices since early December 2015, (www.oilprice.com)”.

Figure 2.6: EIA World Liquids Market Balance (Supply minus Consumption)



Source: www.oilprice.com

The price of oil per barrel increased from $26 to $45 and people in the market believe than even a freeze of oil production from OPEC, in cooperation with Russia, will ambiguously make the difference.

In 2015, we had two price cycles. In March, from $44 per barrel the price has raised to $60 per barrel and that fluctuation endured for 50 days.

The second price cycle was in August of 2015 when the oil price had raised from $38 to $49 per barrel in a short period of seven days due to the China's and US storage withdrawals.

In December 2015, the oil price dropped to $26.55 per barrel

ECONOMETRIC MODELS ON OIL PRICE FORECASTING ARE:

- Time series models
- Financial models
- Structural models

Time series again is said to use historical data.

Oil prices can follow an AR process:

$$S_t = \Phi_1 S_{t-1} + \cdots . + \Phi_q S_{t-p} + \varepsilon_t = \Phi_p(L) S_t + \varepsilon_t$$

Where p is the order of the AR(p) process

$\Phi_p(L)$ is the polynomial in the lag operator L of order p

ε_t is a white noise error

In this model (auto-regressive model) prices are not driven by random fluctuations but by historical data.

Under the assumption that in financial markets prices tend to go back to their average level after the shock, given a long run equilibrium level S_t^* of the oil price and a mean reversion rate *a* we obtain the following:

$$S_{t+1} - S_t = a(\, S_t^* - S_t) + \varepsilon_t$$

Where future price variations depend on the difference between actual and long run price levels

Error correction models (ECM) are designed to capture movements towards the equilibrium. Suppose we have two variables X and Y with an equilibrium level between them Y=aX

$$Y_t = a + \lambda_1 (Y_{t-1} - aX_{t-1}) + \varepsilon_t$$

Abosedra (2005) proposes to accommodate the one-month-ahead price of oil for every day by using the previous day's spot price and by using the monthly average price to obtain a monthly predictor for the future oil price X

$$S_t = a + \beta X_{t-1} + \varepsilon_t$$

Since the co-integration between S and X leads to biased estimates for α and β a non- linear estimation is suggested:

$$S_t = \ a + \beta X_{t-1} + \sum_{i=1}^{n} \rho_i (S_{t-1} - \alpha - \beta X_{t-i-1}) + \sum_{j=m}^{m} \Phi_j \Delta X_{t-j} + \varepsilon_t$$

Bopp and Lady (1991) prove that this model has a very poor forecasting ability. Their model takes into consideration the consequences of the reduction of the OPEC, using a leverage variable and a dummy variable, which capture the effects of the twin towers attack in 2001:

$$S_t = a + \beta_1 S_{t-1+}\beta_2 S_{t-12+} \sum_{j=0}^{5} \gamma_j D01_j + S99 + \varepsilon_t$$

Pindyck (1999) uses data from a range of 127 years, trying to see whether time series models are helpful in forecasting long horizons, under the assumption that nominal oil prices deflated by wholesale prices p:

$$P_t = \rho p_{t-1} + \beta_1 + \beta_2 t + \beta_3 t^2 + \Phi_{1t} + \Phi_{2t} + \varepsilon_t$$

$$\Phi_{1t} = a_1 \Phi_{1,t-1} + \upsilon_{1t}$$

$$\Phi_{2t} = a_2 \Phi_{2,t-1} + \upsilon_{2t}$$

which is a more accurate explanation of oil price fluctuations

With a deterministic trend:

$$P_t = \rho p_{t-1} + \beta_1 + \beta_2 t + \varepsilon_t$$

which performs better in forecasting oil prices.

2.2 GENERAL ASPECTS OF OIL PRICE VOLATILITY

Oil shocks influence macroeconomy (Hui Guo and Kevin L. Kliensen, 2006). Oil volatility is asymmetric, provoking asymmetric affects and influences. In particular, extreme oil price declines aggregate output since they slim down investments by increasing uncertainty and they create resource reallocation. Authors' survey indicate that volatility's constructed measurement, using daily crude oil future prices, underline a negative effect on future gross domestic product for the time horizon between 1984 and 2004.

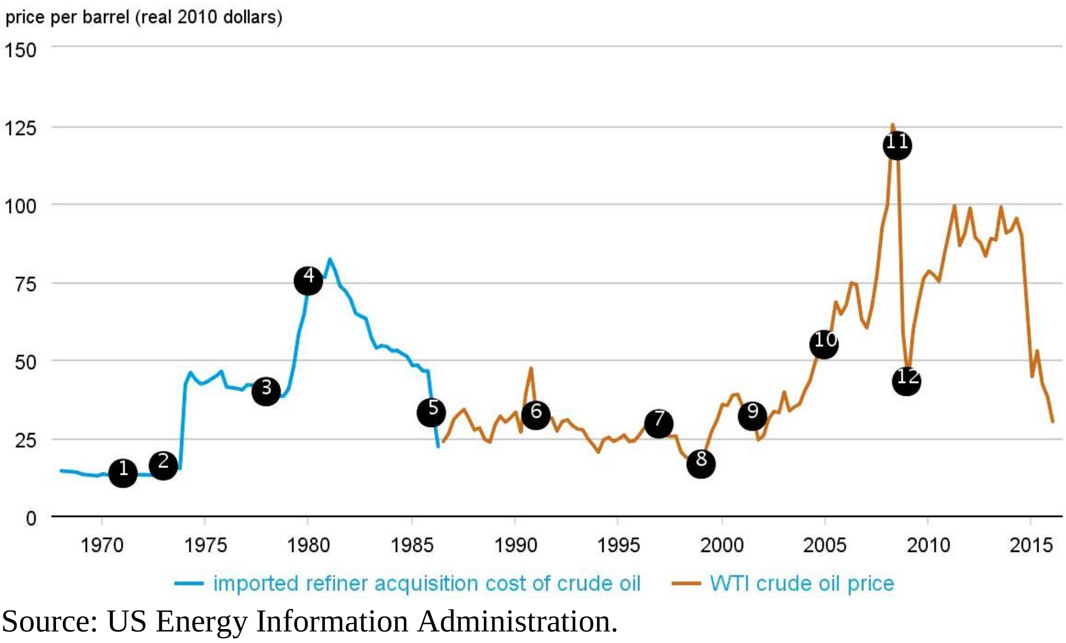
Monetary policymakers are really interested in oil price shocks and oil price volatility due to its significant effects over the economy. US recessions - and more particularly that one of the 2001 - provoke a sharp oil price increase. The indication is that an increase in crude oil's price increases simultaneously the production costs, which, in turn, decrease the future gross domestic products. We should seriously take into account that even though countries all over the world have started producing oil subsidiaries during the last years, they are not in a full position to be unaffected from oil price volatility and to be independent, as regards the various sources of production. In other words, oil hasn't become yet fully subsided and, hence, oil price volatility, jumps and fluctuations do matter.

Additionally, oil price volatility influences investments. Oil shocks affect aggregate output adversely, in a sense that they increase uncertainty for the future; they increase the risk of investment because of the ambiguous future oil volatility. The potential of being higher may either delay the investment (until things become a little bit clearer) or make harder and longer the decision-making process. Alternatively, this short of uncertainty provoked by oil price volatility may induce rather costly infrastructure resource allocation, which actually may double the cost of an already placed investment, under the strong alternative of shutting down the business of a whole company (after an oil shock increase). The evidence, though, of an oil price decrease is ambiguous.

We should not neglect the fact that many oil shocks are not controlled and cannot be easily explained by standard macroeconomic variables, simply because they are a result of exogenous events, like wars and political instability.

In the next figure, we may see historically the fluctuations and the volatility of oil price over the years due to global historic events, proving that oil price volatility is a unique and rare case of a sometimes unexplainable and unpredictable change that is interconnected with many products and assets. Economists and politicians face such a big challenge in their attempt to analyze volatility, the causes of volatility, the extent of volatility, the factors by which this volatility is determined and finally the extent to which one can intervene to these factors in order to reduce uncertainty and potential risk.

Figure 2.7: Crude oil prices and key geopolitical and economic events



Source: US Energy Information Administration.

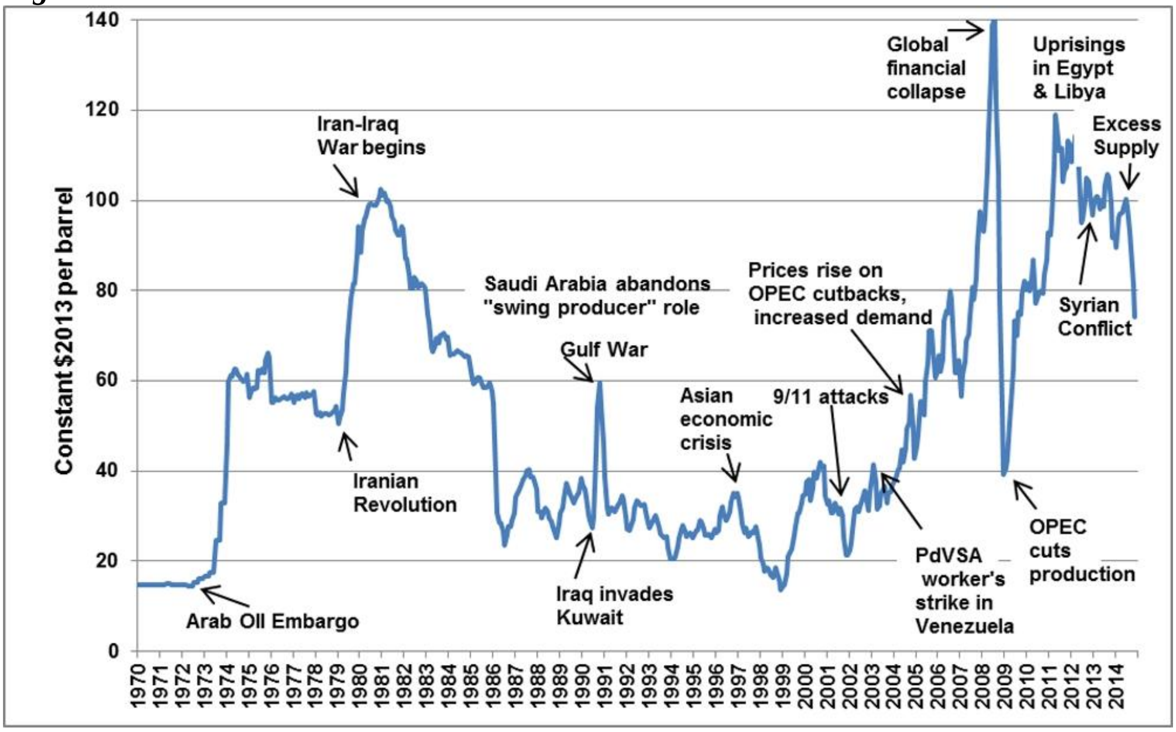
1. US spare capacity exhausted
2: Arab Oil Embargo
3: Iranian Revolution
4: Iran-Iraq War

5: Saudis abandon swing producer role
6: Iraq invades Kuwait
7: Asian financial crisis
8: OPEC cuts production targets
9: 9-11 attacks
10: Low spare capacity
11: Global financial collapse
12: OPEC cuts production targets

As we can see from the above figure, Iran – Iraq War boosted oil price, inducing an increase in oil price volatility, followed by a sharp decrease that took place when Saudis abandoned swing producer role. The global financial collapse also affected oil price volatility.

WORLD CRUDE OIL PRICE AND ASSOCIATED EVENTS, 1970-2014

Figure 2.8: "What Drives Crude Oil Prices?"



Source: Energy information Administration (January 2015)

Exogenous events are again evident along with their effects in oil price volatility. A representative example that is illustrated clearly in the above figure is the Syrian Conflict that caused high oil price volatility.

The above evidence reinforce the argument that oil price volatility is a big challenge to investigate, analyze, try to predict, shed light both in the research field and in the market field, as well as to provide a useful tool for oil price volatility predictions. Such a tool may produce better political strategies, monetary strategies, decision making and investment decisions.

Volatility does matter since the frequent price fluctuations provoke an increase in the value of an asset or a sharp fall. Moreover and as a whole the year-to-year total costs may strongly affect the entire economy. As regards the households' basket, if oil's price increases, then the cost of living will also rise, while less disposable income will be left for the rest of the market. Most countries do not produce oil. They import it, which, in turn, means an outflow income.

EIA's investigation around oil volatility is rather interesting. It examined the extent to which the growth of the economy is affected and whether energy prices are assumed to be stable or predictable. The survey had a time horizon of four years. Other things equal, under the assumption of steady or predictable energy prices, the economy growth is estimated to reach the level of a 0,2% increase in the GDP (gross domestic product). This clearly indicates that, other things equal, economy will perform better if energy prices are stable or predictable. The potential of energy prices to be stable is very small. Therefore, the only hope for the economy left is to be able to predict oil price volatility to its possible and achievable extent.

Fluctuations in global aggregate demand and global growth are components of price volatility. If the change in the world's aggregate demand for oil affects oil's volatility there would be the same volatility performance for other goods as well. But this actually is not the case, indicating that oil volatility is something different, unique and complicated in comparison with the market performance of other goods.

A key explanation of oil price volatility is the increase in the foreign value of a US dollar. There is evidence that an increase in the US dollar foreign exchange value is accompanied with a fall in oil prices. We should not forget that oil is priced is US dollars. Thus, a decline in the foreign exchange value of the US dollar makes oil more expensive for non-US consumers.

Consequently, non-US consumers may reduce demand, which, in turn, may smooth oil price. Sometimes though, a strong US currency may not affect oil price, reflecting, thus, the result of other changing conditions (like global economy).

Correlations over the years between oil price and the foreign exchange value of the US dollar seem to vary.

We should make clear that oil volatility is not a result of supply. If oil was stored and released in special times at predefined quantities then such an argument would have a base. But this is not the case.

Oil is not and it cannot be stored. Changes in oil inventories are one parameter behind oil's volatility. But it cannot simply and without special evidence be considered the main and vital factor of oil's price fluctuations.

2.3 ECONOMETRIC APPROACHES HISTORICALLY USED IN FORECASTING OIL PRICE VOLATILITY

Accurate and considerable forecasts of volatility are rather crucial inputs and components for a range of finance applications like derivatives pricing, portfolio selections, asset allocation etc. As we have analyzed it before, volatility is not always observable and researchers are trying to extract volatility from other variables. Volatility models are being used to measure and predict - to the possible extent - future volatility.

Niaz Bashiri Behmiri and Jose R. Pires Manso outline that forecasting oil price volatility can be approached by two methods: the quantitative method, which in turn is divided in econometrics (time series models, financial models and structural models) and non-standard methods, and the qualitative method (the mechanism under which wars, natural disasters etc may affect oil price volatility). Time series analysis is based on historical data. Time series analysis involves naïve models, exponential smoothing models and auto-regressive models like ARIMA, ARCH, GARCH. Niaz Bashiri Behmiri and Jose R. Pires Manso have presented results of various studies that indicate that "a) time series models are adequate for forecasting oil prices in the short run, but they have limited forecasting ability in the medium and long-term, b) time-series models provide accurate forecasts of oil price volatility, but a single model cannot be used in every case, c) oil prices and their volatility display significant nonlinearity, which indicates that small shocks to the economy could have large and unpredictable implications for oil prices and their volatility". More specifically Wang (2005) uses the ARIMA model to examine crude oil data from January 1970 to December 2003. The output of sample forecasts gives evidence that the linear ARIMA model provides poor forecasting compared with a non-linear artificial neutral network. Xie (2006) also applies an ARIMA model to examine prices in the same period, indicating once again that the ARIMA model provides poor forecasting. Fernandez (2010) does an out-of-sample both for short- and long-run horizons, using daily natural gas and Dubai crude oil prices for the period 1994-2005 in the ARIMA model. For the short-run forecasts the ARIMA model has superior performance, but for long-horizon forecasts the ARIMA model doesn't perform well. The ARIMA is linear and hence is inappropriate to describe the non-linear components of oil price time series.

Cheong (2009) uses GARCH type models to compare the volatility forecasting ability. He uses WTI, the Brent crude oil spot prices for the period 1993-2008. As regards the out-of-sample accuracy estimated 5-, 20-, 60- and 100-day horizons, results show that "the standard short memory GARCH normal and student-*t* models outperform…and GARCH models that account to asymmetric reaction of oil volatility to price changes perform better at longer horizons". Kang (2009) executes a similar test, the results of which also indicate that the out-of-sample forecasts using a single GARCH model does not outperform the other models for both Brent and WTI.

Giliola Frey, Matteo Manera, Anil Markandya, Elisa Scarpa (2009) outline the importance and significance of studying oil prices. As they have outlined in 2007 the WTI (West Texas Intermediate) reported a price of 72$/b whereas in 2008 the price has been boosted to 100$/b (a 38% increase). They also support that despite the fact that up to 2030 people all around the world will try to switch from liquid fuels to other energy subsidies, oil will still be a significant fuel option and the total increase from 2005 to 2030 in fuel consumption is expected to rise by 1,2%. Generally speaking, oil is important due to its contribution to the generation of electricity. Going a little bit back in history they remind us that by the end of the 19th century United States was the basic consumer of oil and its North Eastern region was by then the main source of oil. Oil became famous shortly in Europe. The increasing consumption caused depletion of the US reserves and other sources of oil were discovered in Iraq, Saudi Arabia. By the end of WWII, oil was the predominant source of energy. The cooperation between the United States and the Saudi Arabia created a strong alliance that could control the production and the price of oil. In 1950, oil sources were discovered in Middle East and OPEC was established, that is, a strong cartel. In 1970, US were out of oil and started importing it. In other words, it became an oil dependent country. In 1973, the support of the United States to the country of Israel provoked the embargo of the Arab countries and as a result the price of oil increased by 400%. Since then the stability of oil price was vanished and a period of oil price's fluctuations started.

The second uncertainty phase was in the period 1979-1980 where the Iran–Iraq relationships were under war, pushing oil price to double and proving the inability of OPEC to control the price. Non-OPEC countries increased oil production to meet supply and demand needs. Saudi's Arabia oil production caused a huge price decrease.

It was only in the 1990s, after the Gulf War and the invasion of the Kuwait, that the oil consumption started to increase aided by the growth of the Asian economies. The decrease of the oil price in 1997-1998 was due to the OPEC's increase of production, accompanied by a stability of growth of the Asian economies, leaving OPEC with oversupplies.

A common approach to forecast volatility is time series models using asset prices along with historical data and information. Volatility models are being divided into two main categories: the auto-regressive conditional ARCH models and the stochastic volatility models (SV). The latter ones include unobserved jumps to the return variance, which make things rather difficult.

ARCH models take into account the conditional variance as if it was observable. Researchers have shown a preference in the ARCH models and their extensions since their results over the time-varying volatility are quite decent. Despite the success of the ARCH models some studies assert that they have poor performance in out-of-sample forecasts (Figlewski, 1997).

Other researchers (Andersen and Bollerslev, 1998) opposite the reason that make ARCH models less dynamic, as regards the out-of-sample performance, by asserting that it is not the failure of the models, but the failure to specify the volatility proxy. Log absolute or squared returns are very noisy to the measurement errors. Thus, inefficient inferences concerning latent volatility are being provided.

The alternative solution that Andersen and Bollerslev suggest is to be estimated by high frequency intraday returns. As they argue, under appropriate conditions the use of high frequency data to predict volatility is an unbiased, highly efficient estimator of return volatility, and it gives the chance of a more accurate forecast for future volatility.

High frequency data, though, have been criticized for susceptibility to microstructure effects, for discrete observations, bid ask bounce and for the fact that not all financial markets really have available high frequency data to proceed.

The extreme value or the range-based estimator is another solution to discover the latent volatility. Daily opening, daily closing, high and low prices data commodities and currencies are necessary to produce a worthy estimation of the volatility.

Alizadeh, Brandt and Diebold (2002) underline that the log range (which is the high daily log price minus the log daily low price) is an efficient and reliable volatility estimator.

One way to forecast volatility involves building time series models for volatility, using past asset prices in combination with other relevant historical information. These are divided in two distinctive categories, the auto-regressive conditional heteroskedastic (ARCH) family Engle (1982), and the stochastic volatility (SV) (Clark, 1973). SV models include an unobserved shock to the return variance into representation of the volatility dynamics. In contrast, an advantage of the ARCH models is that they treat the conditional variance, given past information, as observable, and maximum-likelihood methods can be used to estimate the model parameters. The ARCH model is the mostly used one, as well as its various extensions, since they have been proven to capture the time-varying volatility observed in financial data.

Despite the fact that ARCH models have been widely used, they provide poor out-of-sample forecasts and don’t shed much light on the variability of ex post realized volatility. Andersen and Bollerslev (1998) assert that this not a failure of the ARCH models, but a failure to define accurately the volatility proxy. Absolute or squared returns are noisy due to the measurement errors. Therefore inefficient inferences are remarked, regarding underlying latent volatility and its dynamics. They proposed an alternative measure for an ex-post latent volatility, estimated by high-frequency intraday returns. Under appropriate conditions, it is an unbiased highly efficient estimator. Nevertheless, advocates of evil blame it for susceptibility to microstructure due to non-synchronous trading, discrete price observations bid-ask bounce, intraday periodic volatility patterns and due to the fact that high frequency data is not available for all financial markets.

Alternatively, there is also the extreme value or range-based estimator. They require daily opening, closing, high and low prices. Alizadeh, Brandt and Diebold (2002) showed that the log range (difference between logarithms of daily high and low prices) is a rather efficient volatility estimator.

Another forecasting approach is option prices. Options have been used for managing price risk. Their hedging function is the inference of information about markets assessment of specific assets' future volatility. Research and practically frequent tests proved that the option prices approach has superior predictive ability. Studies though show that the implied volatility is higher than the realized volatility. The bias in implied volatility may be the presence of a volatility risk premium.

Another volatility approach is based on the information reached in option prices. Options are being used to reduce price risk since they are robust against unfavorable market moves. Options are also useful to transfer information of an asset's future volatility.

The ability of the realized volatility by option prices to predict future volatility is said to provide sufficient evidence that it is efficient. Nevertheless surveys prove that realized volatility is an upward biased predictor and that most of the times it is higher than the future-realized volatility. This is partly explained because investors seem to prefer to pay a premium to hedge against the upward market-realized volatility.

Another aspect of the observed bias in the realized volatility is that the realized volatility is based on the money Black – Scholes model on implied volatility, one of whose assumptions is that volatility is considered to be constant. Once constant volatility is used as an assumption to analyze options and price is being determined under conditions of stochastic volatility measurement errors occur.

Britten – Jones and Neuberger (2000) form a model-free implied volatility, which includes whole cross-section option prices rather than money prices alone. In that way, the measurement error is being reduced to an adequate extent.

GARCH models have been widely used to model the return volatility. Much research has been done and still carries on about crude oil's forecasting volatility.

Kang (2009) argued that, as regards oil volatility CGARCH and FIGARCH models are performing much better compared to GARCH and IGARCH models, as a method to capture long memory volatility.

Cheong (2009) in his investigation ends up to the point where he argues that the GARCH model performs better than others he tested (namely APARCH, FIGARCH, FIAPARCH) in the Brent crude oil market. FIAPARCH model, in turn, has a superior performance in relation with other models in the WTI crude oil market.

Mohammadi and Su (2010) used four GARCH models: GARCH, EGARCH, APARCH and FIAGARCH in eleven crude oil markets. EGARCH and APARCH models, according to their research and point of view based on their survey, are doing better in an out-of-sample forecast evaluation.

Wei (2010) edited a larger number of linear and non-linear GARCH models, using more loss functions. He comes up with the conclusion that the long-run non-linear models outperform the linear ones for crude oil's volatility forecasting.

Sadorsky (2006) uses both univariate and multivariate models to forecast daily volatility in petroleum futures price returns for the time horizon between 1988 and 2003. GARCH models seem to perform efficiently.
Aggolucci (2009), on the other hand, argues than GARCH models CGARCH and TGARCH models following a GED error distribution perform better.

Researchers like Parkinson (1980), Garman and Klass (1980), Ball and Tours (1984), Rogers and Satchell (1991), Kunimoto (1992), Yang and Zhang (2000) advocate that range data is available in most financial assets over long time horizons, making them a more efficient estimator of volatility than a variance estimator based purely on closing prices.

Schwert (1990) argues that range data are not really helpful to forecast stock returns. However, they actually do important work in predicting volatility. Bali (2003) provides empirical evidence that the estimation of the value at risk is superior to that of a standard approach. Daily frequent extreme value estimators are significantly more efficient than the traditional estimators.

Alizadeh, Brandt and Diebold (2002) investigate log range volatility's proxy properties. They claim that the log range is more efficient because the log range is one quarter of the error of the standard volatility proxies. In other words, they proved that range based volatility estimators provide a much better and more accurate volatility forecast.

Shu and Zhang (2006) reinforce Alizadeh, Brandt and Diebold's evidence by developing respectively a range-based covariance, thus showing that this measure is much more efficient.

The value of range-based estimators is also proven and is demonstrated in more complex volatility models, such as that of Chou (2005), who created the conditional auto-regressive range CARR model. He asserts that the CARR model outperforms the GARCH model.

Generally speaking, when market prices often change over a short period of time the market is considered to have a high volatility. When prices over a short period of time are relatively stable with only small fluctuations, the market is supposed to have low volatility.

In energy markets with assets like oil, gas, etc prices and their volatility is rather important, simply because the ability to increase their value, which is typically counted in millions or billions of dollars, depends on the ability to buy or sell at a profitable price.

Walid Chkili, Shawkat Hammoudeh, Duc Khuong Nguyen examine in their survey the suitability of GARCH-class models "in modeling conditional volatility and market risk (VaR) out of four most widely traded commodities (crude oil, natural gas, gold and silver), in the presence of long memory and asymmetric effects". There is not much research and literature on the choice of the "right" volatility model to forecast future volatility. They prove in their research that non-linear volatility models are best for the estimation of the VaR forecasts, both in short- and long-trading positions. Aloui and Mabrouk, 2010, as well as Cheong, 2009 and Wei 2010 assert that the long-run memory and asymmetry properties are important and rather significant in the improvement of the accuracy of the VaR estimates. Most papers separate long memory from asymmetry and usually focus on the long memory. Their survey takes both into account. Moreover, past surveys approach the volatility forecasting by forecasting conditional return and not conditional volatility. It is concluded that the best model is the FIAPARCH one, which provides the best VaR estimates and forecasts. The number of violations of that model is the lowest. Acknowledging the importance of volatility asymmetry of commodities and its importance in forecasting, asset valuation, hedging and risk management several surveys try to model commodity volatility behavior. Choi and Hammoudeh (2009) argue that long memory (LM) univariate GARCH models are better than standard GARCH models in forecasting commodity volatility. More specifically they argue that "The FIGARCH model provides strong evidence of long memory (LM) for most of oil and products price returns". Cheong (2009) discovers that simple GARCH model process oil data much better than GARCH models, which accommodate asymmetry and LM. Moreover, he declares in his research that "the non-parametric GARCH model yields superior performance compared to an extensive class of parametric GARCH models, based on loss functions and Hansen's superior predictive ability test. Walid Chkili presented a number of GARCH type models to cover spot and future prices of crude oil. Namely, these are three non-linear eGARCH models: FIGARCH, FIAPARCH and HYGARCH. The whole concept of choosing non-linear GARCH models is that commodities like oil are vulnerable and various shocks in markets affect its volatility. Their basic difference from linear GARCH–class models is that they take into account LM and volatility asymmetry.

Starting with the FIGARCH model, we have the opportunity via this model to distinguish the long memory from the short memory in the conditional variance. FIGARCH (1,d,1) model has as follows:

$$\sigma_t^2=\omega+\beta\sigma_{t-1}^2+[1-(1-\beta L^{-1})(1-\lambda L)(1-L)^d]\varepsilon_t^2$$

Where $\omega>0$, $\beta<1$ and $\lambda<1$, d is the fractional integration parameter and satisfies $0\leq d\leq 1$. If d=0 in short memory, we have GARCH (1,1) , if d=1 we have IGARCH (1,1)

The second model used is the FIAPARCH one, which accommodates both LM and asymmetry effects in the conditional volatility. FIAPARCH (1,D,1) is:

$$\sigma_t^{\delta}=\omega(1-\beta L)^{-1}+[1-(1-\beta L)^{-1}(1-\lambda L)(1-L)^d](|\varepsilon_t-\gamma\varepsilon_t)^{\delta}$$

Where again $\omega>0$, $\delta>0$, $\beta<1$ and $\lambda<1$. The parameter γ refers to the asymmetry under the condition $-1<\gamma<1$. When $\gamma>0$, negative market shocks will have more impact on the product's volatility return than positive shocks will have. The fractional integration parameter d under the condition $0\leq d\leq 1$ demonstrates LM in the conditional variance process. When $\gamma=0$ and $\delta=2$ we have the GIGARCH model. When d=0 we have the APARCH model.

Finally, we have the HYGARCH hyperbolic GARCH model which is demonstrated as follows:

$$\sigma_t^2=\omega+[1-(1-\beta L)^{-1}\lambda L\{1+\alpha((1-L)^d-1)\}]\varepsilon_t^2$$

Where $\omega>0$, $\alpha\geq 0$, $\beta<1$, $\lambda<1$ and $0\leq d\leq 1$. Davinson (2004) indicates that the HYGARCH model «permits the existence of the second moment at more extreme amplitudes than the simple IGARCH and FIGARCH models do".

The crucial element for investors is to find out the best model for predicting the VaR of their asset portfolios. VaR estimates and forecasts are produced from two GARCH models, the EGARCH and the FIAPARCH model. In Walid Chkili's research, VaR forecasts both in the long- and the short-run are produced with α ranging from 5% to 0,25%. The accuracy is tested using the Kupiec likelihood ratio (LR) test. In-sample results show that the GARCH model is not doing that well in the short-run and long-run positions. The EGARCH model is not better. The FIAPARCH model is much better than the others. Mabrouk (2010) indicates that models taking into account LM and asymmetry develop the quality and accuracy of the VaR estimations. Out-sample VaR estimations indicate the same pre-mentioned results. FIAPARCH model's superior performance is evident.

A realized volatility model is Corsi's (2009) model named HAR-RV model, as it is explained in a Henrik Soyland Langeland paper (2013). Traders have different time horizons. In the short term, the trader will be influenced both by short- and long-term volatility, whereas the long-term trader is not influenced by short-term volatility. Andersen (2007) indicates that HAR-RV model in forecasting performs much better than the long memory ARFIMA model.

The realized volatility is defined as:

$$RV_{t+1d}^{d} = c + \beta^{d} RV_{t}^{d} + \beta^{w} RV_{t}^{w} + \beta_{m} RV_{t}^{m} + \tilde{\omega}_{t=1d}^{d}$$

Where d stands for day
W stands for week
And m stands for month

The above equation is the HAR(3)-RV with a simple auto-regressive structure, which can easily be extended by adding additional regressors.

The HAR (3)-RV can then be extended to HAR-RV-IV by adding additional regressors, such as the implied volatility IV. We then come up with the following model:

$$RV_{t+1} = \beta_0 + \beta_1 RV_t + \beta_2 RV_t^{w} + \beta_3 RV_t^{m} + \beta_4 IV_t + \epsilon_{t+1}$$

After that, we can take into account the HAR-RV-EX model with RTN_t^{+} defined as the max $(RTN_{t,}0)$ and RTN_t^{-} as the min $(RTN_{t,}0)$ with RTN representing the percentage change in price from market close at t-1 to market close at t.

$$RV_{t+1} = \beta_0 + \beta_1 RV_t + \beta_2 RV_t^w + \beta_3 RV_t^m + \beta_4 SIZE_t + \beta_5 NTR_t + \beta_6 OI_t + \beta_7 RTN_t^+ + \beta_8 RTN_t^- + \beta_9 BAS_t + \beta_{10} SL_{t+}^+ + \beta_{11} SL_t^- + \epsilon_{t+1}$$

If $SIZE_t$ is the daily average size of trades
NTR_t is the average number of contracts traded during day t
And OI_t is the number of open interests at day t for the monthly cycle
RTN_t^+ is max ($RTN_{t,}0$)
RTN_t^- is min ($RTN_{t,}0$)

The conclusions that can be made from the above modes are shortly displayed below. The three coefficients of the HAR-RV model are highly significant. One day's volatility has the strongest effect on the following day's volatility level.

When IV is included, a reduction happens in all three RV coefficients, but the largest reduction is observed in the monthly measure. IV loses its statistical significance and hence it performs better in shorter horizons.

By adding the exogenous variables, the estimated coefficient of the short-term component decreases. Thus, the information parameter of the Ex variable overlaps the information in the daily RV measure. The R squared values have increased.

Henrik Soyland Langeland, using Corsi's (2009) HAR-RV model, combined the realized volatility RV with the implied volatility IV and other exogenous EX variables by using high frequency data and found that the HAR-RV model fits the RV time series significantly better when both IV and EX are added to the model. IV is statistically significant in the short-run rather than in the long-run. "Implied volatility improves predictions most significantly for short-term predictions, whereas other market variables and particularly the bid ask spread had a more significant effect than implied volatility for long term forecasts".

2.4 LITERATURE REVIEW ON HAR MODELS

Benoit Sevi, 2014, using intraday data for a period of 66 days asserts that the decomposition of continuous and discontinuous (jumps) from negative or positive intraday returns gives a forecasting result competing to the models that include different components of the realized variance. Sevi's analysis provides "a large-scale empirical analysis of the forecasting accuracy of various time series models derived from the innovative HAR (Heterogeneous Autoregressive). There is much literature concerning the predicting procedure using intraday data. Fleming, Kirby, and Ostdiek, 2003, as well as Corsi, Fusari and La Vecchia, 2013, reinforce the idea of the aforementioned Benoit Sevi process, regarding the forecasting accuracy of time series models, using intraday prices. Asymptotic theory is used to detect jumps. Realized semi-variances are the high-frequency semi-variances about daily data. Results are then compared to forecasts from implied volatility. Sevi completed his study firstly because forecasting the most traded good was really exciting, secondly because he was aware that the results of such a research affect risk management and portfolio selection and thirdly because in general Sevi believes that shedding light on forecasting crude oil prices, using intraday data, will also shed light on jumps in stock index.

HAR – type models are used in combination with realized semi-variances and detected jumps (Patton and Sheppard,2011). Crude oil's returns volatility has been processed and studied using the GARCH-type models. Sevi declares via his paper that power auto-regressive models are superior in the short-run in predicting volatility while GARCH type models are better in the long-run. Ding, Granger, and Eangle (1993) and Mohammadi and Su (2010) have the same point of view, demonstrating the better performance of the APARCH model compared to the GARCH or EGARCH models. The basic difference between the GARCH models and the HAR models is that the formers used daily data, while the latter use up all the information contained in intraday information/data.

To process the volatility and jump detections some equations should be introduced:

Starting with the realized variance (RV)

$$RV_{t,M} = \sum_{j=1}^{M} r_{tj}^2$$

where frequency tends to infinity

$$RV_{t,M} = \int_{t-1}^{t} \sigma_s^2 ds + \sum_{j=1}^{J(t)} K^2(t_j)$$

To disentangle continuous jumps the bi-power variation is introduced:

$$BVP_{j,M} = \xi_1 \sum_{j=1}^{M-1} |r_{tj}||r_{tj+1}|$$

The BVP has certain drawbacks though. The adjacent return doesn't tend to zero and "the large "jump" intraday return results in an upward bias of the BPV". This zero return provokes a downward bias of the BVP. The median realized variance (MedRV) seems more promising. It is presented as follows:

$$MedRV_{t,M} = \frac{\pi}{6-4\sqrt{3}+\pi}\left(\frac{M}{M-2}\right)\sum_{j=2}^{M-1} med\left(|r_{tj-1}|,|r_{tj}|,|r_{tj+1}|\right)^2$$

MedRV eradicates the impact of jumps. The estimator is stronger to zero returns.

To detect jumps in our study we use the adjusted jump ration statistic. The test statistic for day t is defined as:

$$ZJ_{BPV}(T,M) = \sqrt{M}\frac{\left(RV_{t,M} - RPV_{t,M}\right)RV_{t,M}^{-1}}{\left(\left(\xi_1^{-4}+2\xi_1^{-2}-5\right)max\left\{1,TQ_{t,M}BPV_{t,M}^{-2}\right\}\right)^{1/2}}$$

Which is extended for the research's purpose to the MedRV estimator:

$$ZJ_{MedRV}(t,M) = \sqrt{M}\frac{\left(RV_{t,M} - MedRV_{t,M}\right)RV_{t,M}^{-1}}{\left(0.96max\left\{1,MedRQ_{t,M}MedRV_{t,M}^{-2}\right\}\right)^{1/2}}$$

"When the ZJ statistic is significant, then the difference between the RV and the BPV or MedRV is too large" which means we have a "jump". So the jump component has as follows:

$$J_{t,a}(M) = \left[RV_{t,M} - BPV_{t,M}\right]xI[ZJ_{BPV}(t,M) > \Phi_\alpha]$$

The RV-BPV gives us the squared jump component and the continuous component equals the BPV. When there is no jump the continuous component is similar to the RV.
Patton and Sheppard (2011) introduced the significance of the realized semi-variance defined as follows:

$$RSV_{t,M}^{-} = \sum_{j=1}^{M} r_{tj}^2 xI_{[r_{tj<0}]}$$

Sevi uses the HAR model (Corsi, 2009), which is estimated using standard ordinary least squares.
The average realized variance over the period [t+1, t+h] is:

$$RV_{t+1,t+h} = h^{-1}[RV_{t+1} + RV_{t+2} + \cdots + RV_{t+h}]$$

The eleven time series HAR models can now be illustrated:

The HAR-RV-J, by Andersen et al (2007), where a jump component is completed using one day lagged squared jump, has as follows:

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \beta_{SQJ} J_t + \varepsilon_t$$

The same authors have also introduced the HAR-CJ model in 2007, with the following equation:

$$RV_{t+1,t+h} = \beta_0 + \beta_{C1} C_t + \beta_{SQJ1} J_t + \beta_{C5} C_{t-1,t-4} + \beta_{SQJ} J_{t-1,t-4} + \beta_{C22} C_{t-5,t-21} + \beta_{SQJ22} J_{t-5,t-21} + \varepsilon_t$$

Two years later, in 2009, Corsi presents the HAR-RV model, which is expressed as:

$$RV_{t+1,t+h} = \beta_0 + \beta_1 RV_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t$$

In 2011, Patton and Sheppard introduced several models and more specifically the PS model, using the realized semi-variances decomposition of one day so that lagged positive and negative components are distinguished:

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t$$

The PSlev model, which examines whether the superior significance of the negative realized semi-variance does not come from a leverage effect:

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \gamma RV_t I_{|r_t<0|} + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t$$

the HAR-RSV model that assumes that positive and negative realized semi-variances can have a different forecasting power at different tags:

$$RV_{t+1,t+h} = \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \beta_5^- RSV_{t-1,t-4}^- + \beta_5^+ RSV_{t-1,t-4}^+ + \beta_{22}^- RSV_{t-5,t-21}^- + \beta_{22}^+ RSV_{t-5,t-21}^+ + \varepsilon_\tau$$

the HAR-RV-SJ model:

$$RV_{t+1,t+h} = \beta_0 + \beta_J \Delta J_t + \beta_C C_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t$$

and finally the HAR RV-SJd model, with a discrimination between positive and negative jumps:

$$RV_{t+1,t+h} = \beta_0 + \beta_J^- \Delta J_t I_{[\Delta J<0]} + \beta_J^+ \Delta J_t I_{[\Delta J>0]} + \beta_C C_t + \beta_5 RV_{t-1,t-4} + \beta_{22} RV_{t-5,t-21} + \varepsilon_t$$

During the same year, Chen and Ghysels provided the CG model, where one day lagged squared jump is considered:

$$\begin{aligned} RV_{t+1,t+h} = {} & \beta_0 + \beta_1^- RSV_t^- + \beta_1^+ RSV_t^+ + \beta_5^- RSV_{t-1,t-4}^- + \beta_5^+ RSV_{t-1,t-4}^+ + \beta_{22}^- RSV_{t-5,t-21}^- \\ & + \beta_{22}^+ RSV_{t-5,t-21}^+ + \beta_{SQJ} J_t + \varepsilon_\tau \end{aligned}$$

Finally, Sevi presents another two models, that is, the HAR-CSJ model, where signs of jumps are taken into account over a time period larger than one day:

$$\begin{aligned} RV_{t+1,t+h} = {} & \beta_0 + \beta_{1J} \Delta J_t + \beta_{1c} C_t + \beta_{5J} \Delta J_{t-1,t-4} + \beta_{5C} C_{t-1,t-4} + \beta_{J22} \Delta J_{t-5,t-21} \\ & + \beta_{22C} C_{t-5,t-21} + \varepsilon_t \end{aligned}$$

and the HAR-CSJd model, where there is a discrimination between negative and positive signed jumps:

$$\begin{aligned} RV_{t+1,t+h} = {} & \beta_0 + \beta_{1J}^- \Delta J_t I_{[\Delta J_t<0]} + \beta_{1J}^+ \Delta J_t I_{[\Delta J_t>0]} + \beta_{1c} C_t \\ & + \beta_{5J}^- \Delta J_{t-1,t-4} I_{[\Delta J_{t-1,t-4} t<0]} + \beta_{5J}^- \Delta J_{t-1,t-4} I_{[\Delta J_{t-1,t-4}>0]} + \beta_{5C} C_{t-1,t-4} \\ & + \beta_{22J}^- \Delta J_{t-5,t-22} I\left[\Delta J_{t-5,t-21} t < 0\right] + \beta_{22J}^+ \Delta J_{t-5,t-22} I\left[\Delta J_{t-5,t-21} t > 0\right] \\ & + \beta_{22} RV_{t-5,t-21} + \varepsilon_t \end{aligned}$$

Sevi's research provides us with important results. It is proved that the fit of the predictive regressions may be improved by taking into account components like the squared jump component, the continuous component, the detected jumps and the realized semi-variances. Nevertheless in an out-sample case study the results are not much better than sophisticated models' results.

The survey of Stavros Degiannakis and George Filis, 2015, is also of great interest. They made research also in the oil price volatility forecasting. They have admitted that oil prices monopolize media interest and it is a real challenge for a researcher to contribute in that field of research. According to their survey, the involvement of hedge funds in the oil market, which provoked the financialisation of the market, makes research really interesting. Accurate and authoritative forecasts of oil price volatility are rather crucial for policy makers and oil traders. Sadorsky (2006) was the first to deal with the oil volatility forecasting using GARCH, TGARCH models, followed by Sadorsky and McKenzie (2008) who showed that GARCH-type models provide accurate forecasts. Kang (2009) using daily oil prices proved that CGARCH and FIGARCH models are more useful. Nomikos and Pouliasis (2011) using one day ahead oil price volatility proved that both MIX-GARCH and MRS-GARCH models are better than simple GARCH models. Kang and Yoon (2013) assert that ARFIMA-FIGARCH models have better performance. Most researchers argue that univariate GARCH models give more accurate results. Efimova's and Serletis's(2014) survey is the first to introduce the S&P500 daily returns in their univariate GARCH-type and multivariate models. Their findings indicate that univariate models produce better forecasting results and that the involvement of the S&P500 did not improve the results. Andersen and Bollerslev (1998) "introduce an alternative measure of daily volatility, which considers intraday data, namely the Realized Volatility. The realized volatility is based on the idea of using the sum of squared returns to generate more accurate daily volatility measures. Numerous studies have shown that the intraday data are able to produce better forecasts, compared to the daily data (Hansen and Lunde, 2005, Engle and Sun, 2007). From 2014 and on surveys on forecasting oil volatility have started using high frequency data. Prokopczuk (2015) made his survey using intraday data and a HAR model showing like Sevi that modeling jumps doesn't help the accuracy of forecasting of the HAR-RV model. Stavros Degiannakis' and George Filis' research approaches the oil price realized volatility forecasting using the HAR-RV model and considering 14 exogeneous variables that concern four different asset classes (i.e. stocks, foreign exchange, commodities and macroeconomics) and explain whether their realized volatilities affect oil volatility forecasts, They provided a solution on how to deal with exogenous variables. The forecasting accuracy is based on individual asset class, their combined forecasts and the forecast-averaging. By using the period of crisis 2007-2009, the model confidence set and the direction of change (DoC) accuracy can be evaluated by applying the Median Absolute Error and the Median Squared Error in forecasts during a 66-day period.

Realized volatility DRV is demonstrated as follows:

$$DRV_t^{\tau} = \sqrt{\sum_{j=1}^{\tau} \left(logP_{tj} - logP_{tj-1}\right)^2}$$

Hansen and Lunde, 2005, argue that the realized volatility should include the fact that even when markets are closed, information keeps on flowing and, hence, they proposed to adjust/replace the intraday volatility with the close-to-open inter-day volatility as follows:

$$DRV^{(\tau)}_{(HL),t} = \sqrt{\omega_1\left(logP_{t_1} - logP_{t-1_\tau}\right)^2 + \omega_2 \sum_{j=2}^{\tau}\left(logP_{t_j} - logP_{t_{j-1}}\right)^2}$$

The annual realized volatility series is:

$$RV^{(\tau)}_{(HL),t} = \sqrt{252}xDRV^{(\tau)}_{(HL),t'}$$

Let us introduce you the HAR models. The first type is the HAR-RV:

$$\begin{aligned} log\left(RV^{\tau}_{HL,OIL,t}\right) \\ &= w_0^{(t)} + w_1^{(t)} log\left(RV^{(\tau)}_{(HL),OIL,t-1}\right) + w_2^{(t)}\left(5^{-1}\sum_{k=1}^{5} log\left(RV^{(\tau)}_{(HL),OIL,t-k}\right)\right) \\ &+ w_3^{(t)}\left(22^{-1}\sum_{k=1}^{22} log\left(RV^{(\tau)}_{(HL),OIL,t-k}\right)\right) + \varepsilon_t \end{aligned}$$

Following the HAR-RV-X model

$$log\left(RV^{\tau}_{HL,OIL,t}\right) = w_0^{(t)} + w_1^{(t)} log\left(RV^{(\tau)}_{(HL),OIL,t-1}\right) + w_2^{(t)}\left(5^{-1}\sum_{k=1}^{5} log\left(RV^{(\tau)}_{(HL),OIL,t-k}\right)\right) + w_3^{(t)}\left(22^{-1}\sum_{k=1}^{22} log\left(RV^{(\tau)}_{(HL),OIL,t-k}\right)\right) + w_{(\alpha),4}^{(t)} log\left(RV^{(\tau)}_{(HL),X_{(\alpha)},t-1}\right) + w_{(\alpha),5}^{(t)}\left(5^{-1}\sum_{k=1}^{5} log\left(RV^{(\tau)}_{(HL),X_{(\alpha)},t-k}\right)\right) + w_{(\alpha),6}^{(t)}\left(22^{-1}\sum_{k=1}^{22} log\left(RV^{(\tau)}_{(HL),X_{(\alpha)},t-k}\right)\right)$$

$$+\varepsilon_t$$

Using these data they come up with the one day ahead forecasting realized volatility.

The conclusion of the article is that the HAR-RV model outperforms the forecasting accuracy at all forecasting horizons. The HAR-RV-X models should be used from stakeholders, while as regards the oil price volatility the HAR-RV model should ideally be used. In addition, it is obvious that using the Median Absolute Error is better than using the Median Squared Error. “The fact that HAR-RV-X models, which combine multiple asset classes’ volatilities are the best performing models, provides strong support to our argument that different asset classes’ volatilities provide important information for the forecast of oil price volatility, given the fact that there are different channels through which every asset class could have an impact on the oil price volatility”.

2.5 INTERACTION OF OIL PRICE VOLATILITY WITH STOCK MARKETS' VOLATILITY

Din Hoang Phan, Susan Sunila Sharma, Paresh kumar Narayan (2015) shed light on another significant aspect of crude oil's forecasting price volatility by examining the interaction between crude oil and the equity markets for the period between 2009 and 2012. According to their survey the integration of the bid – ask spread and trading volume components leads to a better prediction of the oil's price volatility. Trading information also improves the oil price's volatility predictability (both for in-sample and out-of–sample cases). The trading strategy based on predictive regression models yields utility gains to mean-variance investors. There is such strong evidence regarding the impact of the oil price on stock markets. Dispreong (2008), Park and Ratti (2008) and Miller and Ratti (2009) assert that "there is a negative effect of crude oil price returns on stock returns" whereas some other researchers like Chen (1986), Huang (1996) and Wei (2003) do not indentify a statically significant effect. Narayan and Sharma (2011) argue there is a mixed affection. Sectors related to oil, such as transportations, are positively affected by oil price changes, while the rest of the sectors are negatively affected. Another point of view about the relationship between oil price's volatility and stock markets' volatility is the cross – market volatility transmission. Agren (2006) finds significant interaction in a sample of countries (US, UK, Japan, Norway and Sweden) with the exception of Sweden. Hammoudeh (2004) analyses the volatility interaction among five S&P oil sector stock indices and oil prices from the US market using GARCH models. It is proved that there exists a bidirectional interaction between the return volatility of oil futures and oil sector indices. Malik and Edwing (2009) have also found a significant volatility interaction between oil prices and five US equity sector indices using the GARCH model. Arouri (2011) underlies the fact that there is a uni-direction transmission of volatility in the European market with oil markets affecting stock markets, while in the US market the volatility transmission is bidirectional.

D.H.B. Phan, S.S. Sharma, P.K. Narayan (2015) include for the first time information on trading volume and bid-ask spread in testing cross market volatility interactions between crude oil and equity markets. They discover significant evidence that the involvement of this parameter (information on trading volume and bid-ask spread) lead to more accurate forecasts of the volatility in both markets. "Crude oil and equity markets are heavily traded and studies based on low-frequency data, such as daily, weekly or monthly data, may fail to capture information contained in intraday price movements. As volatility is a key input for market risk evaluation and derivatives in pricing, intraday volatility modeling and forecasting are important to market participants who are involved in intraday trading, such as day traders, high – frequency portfolio managers, and program traders". Wang and Wang (2010), as being presented in D.H.B. Phan, S.S. Sharma, P.K. Narayan (2015) analysis.

2.6 THE IMPORTANCE OF JUMPS

Marcel Prokpczuk (2014) defines jump intensity “as the number of jump days in that month over the total number of trading days in the particular month”. The dynamics of jumps are crucially important in predicting crude oil’s volatility forecasting. HAR type models, according to their survey, capture the dynamics of jumps. Especially for oil, Marcel Prokpczuk (2014) asserts that the portion of positive jumps (4.1%) for the period of 2008-2009 was twice as high compared to that of the negative jumps. Generally speaking, the intensity of crude oil displays a large variation compared with other goods like gasoline, natural gas, heating oil. Noticeable is that the mean of the positive jumps of oil is almost the same with the mean of the negative jumps. The results of the survey indicate that volatility increases after a jump event. The coefficient of oil (approximately 0.098), which is higher than that of gasoline, etc, indicates that a negative jump enters the regression with a negative loading implying a prediction of increase of the future volatility. In modern finance, we focus on volatility rather than on variance, as modern portfolio theories indicate. Nevertheless, tests on jump detection identify jumps in variance and not in volatility. Modeling of jumps, in order to improve the accuracy of variance forecasting, have led us to similar conclusions. To sum up, the analysis indicates that jumps are rare, with a varying intensity, and the results from testing whether modeling of jumps help us to predict future volatility are depressingly showing that jumps do not significantly improve the accuracy of volatility forecasting in energy (oil) markets.

2.7 THE IMPORTANCE OF USING HIGH FREQUENCY DATA

The availability of high frequency data has boosted activity in economics. The basic reason that high frequency data can improve forecasting accuracy is that volatility is highly persistent. A more accurate measurement of the current volatility, using high frequency data, is precious and significant for forecasting future volatility.

Another reason that high frequency data are considered useful for an improved volatility forecasting is that it improves the understanding of the dynamic properties of volatility. Future volatility may be predicted using realized measures. In addition, realized measures have contributed in the development of new volatility models that produce more accurate forecasts. The evaluation of volatility forecasts has been also improved by high frequency data. The estimation of complex volatility models, like the continuous time volatility models, can be improved and developed by realized measures. The understanding of the factors and the driving forces that affect volatility and their importance has been improved by the use of high frequency data. Summing up, the squared intraday returns give us the realized variance. The realized variance shows that volatility models deliver accurate forecasts (Andersen and Bollerslev, 1998).

Using high frequency data, the volatility forecasting procedure may be divided into the reduced form volatility forecasting and the model-based forecasting. The reduced

form volatility model is a time series model (e.g. ARIMA) with realized measures used to produce volatility forecasts.
The model-based approach specifies the distribution of returns by using a model for returns (e.g. GARCH).

In conclusion, high frequency data are used to produce volatility forecasts. High frequency data help us understand better the components and factors that determine volatility.

2.8 CHAPTER HIGHLIGHTS

In this part are summarized the most important information of the chapter. First of all were presented some basic economy theory over oil price in order to increase the intelligibility the causes of oil price volatility and then presented background information about the mechanism of oil price volatility and the models with which we can estimate her, focusing on oil price realized volatility by using HAR models. Two other issues that considered, was the importance of using high frequency data in our research that define better the volatility and the notability of using specific other assets on estimating and predicting better the volatility of oil price. Thus, our attempts are heading in the direction of providing enough evidence to support whether exogenous variables' implied volatilies contribute in oil price volatility estimation significally, by using ultra-high frequency data.

3. MODEL SPECIFICATION AND DATA SELECTION

3.1 INTRODUCTION

This part of our study presents the method we used to build the model, that is, the identification of the dependent variable, the data collection, the selection of the appropriate statistical software for model calculations and finally the presentation of our outputs. Furthermore, we define the criteria used for the selection of the implied volatilities of other indicators, which were included in our sample as regards oil price realized volatility. The results from the estimation of each HAR Model, as well as the key conclusions of this study, are quoted in the next Chapter.

3.2 BUILDING THE MODEL

Realized Volatility Measure

The realized volatility can be calculated with the following mathematical equation:

$$RV_t^{(\tau)} = \sum_{j=1}^{\tau} (\log P_{tj} - \log P_{tj-1})^2$$

It should be mentioned that as the time interval (τ) extends towards infinity, the realized volatility coincides with the integrated volatility. Microstructure frictions that are correlated with high sampling frequency have as a consequence the addition of a greater noise volume to the estimated volatility. Therefore, microstructure frictions may influence the balance between accuracy and potential bias.

In order to solve this problem, Andersen et al. proposed the configuration of the volatility signature plot, where the average realized volatility is set against the sampling frequency. Given the fact that the bias due to microstructure frictions increases, as the number of samples gets higher, the signature plot can be used so as to determine the highest frequency where the average realized volatility remains as steady as possible.

Optimal sampling frequency

Towards this direction, in our attempt to define the highest frequency with the stabilized realized volatility, the inter-day variance (y_t^2) is decomposed into the intra-day variance ($RV_t^{(\tau)}$) and the intra-day auto-covariances (y_{ti}, y_{ti-j}), as follows:

$$y_t^2 = RV_t^{(\tau)} = \sum_{j=1}^{\tau-1} \sum_{i=j+1}^{\tau} y_{t_{i-j}}$$

The measurement error is expressed by the latter part, assuming that $E(y_{ti}, y_{ti-j}) = 0$ for $j \neq 0$. Based on the aforementioned, when the auto-covariance bias is eliminated, then the sampling frequency is optimal.

Inter-day adjustments

Intraday volatility is influenced even when markets are closed, since the flow of information never stops. In order to enhance accuracy and address this problem, the close-to-open inter-day volatility was proposed (Hansen & Lunde, 2005), which is estimated as follows:

$$RV_{t(HL)}^{(\tau)} = w_1(\log P_{t_1} - \log P_{t_1 - 1_\tau})^2 + w_2 \sum\nolimits_{j=2}^{\tau} (\log P_{t_j} - \log P_{t_j - 1_\tau})^2$$

where the values of ω_1 and ω_2 are such, so as to minimize the difference between realized volatility and integrated volatility, that is, to minimize the variance of RV.

Heterogenous Auto-Regressive model

Corsi et al. proposed in 2009 the HAR model, which is estimated as:

$$\log\left(RV_{OIL,t}^{\tau}\right) = w_0^{(\tau)} + w_1^{(\tau)} \log\left(RV_{,OIL,t-1}^{(\tau)}\right) + w_2^{(\tau)}\left(5^{-1}\sum_{j=1}^{5} \log\left(RV_{OIL,t-k}^{(\tau)}\right)\right) + w_3^{(\tau)}\left(22^{-1}\sum_{j=1}^{22} \log\left(RV_{OIL,t-j}^{(\tau)}\right)\right) + \varepsilon_t$$

Heterogenous Auto-Regressive model with eXogenous variables

After incorporating exogenous variables in the HAR model, the HAR-X model has as follows:

$$log\left(RV_{OIL,t}^{\tau}\right) = w_0^{(\tau)} + w_1^{(\tau)} log\left(RV_{OIL,t-1}^{(\tau)}\right) + w_2^{(\tau)}\left(5^{-1}\sum_{j=1}^{5} log\left(RV_{OIL,t-j}^{(\tau)}\right)\right) + w_3^{(\tau)}\left(22^{-1}\sum_{j=1}^{22} log\left(RV_{OIL,t-j}^{(\tau)}\right)\right) + w_{(m),4}^{(\tau)}(IV_{X_{(m)},t-1}^{\tau}) + w_{(m),5}^{(\tau)}\left(5^{-1}\sum_{j=1}^{5} log\left(IV_{X_{(m)},t-1}^{(\tau)}\right)\right) + w_{(m),6}^{(\tau)}\left(22^{-1}\sum_{j=1}^{22} log\left(IV_{X_{(m)},t-1}^{(\tau)}\right)\right) + \varepsilon_t$$

In the present study, we provide information regarding the 30-day futures contracts for the following indexes: CBOE Volatility Index (VIX), CBOE Crude Oil ETF Volatility Index (OVX), CBOE/CME FX Euro Volatility Index, CBOE Gold Volatility Index and Treasury Volatility Index U.S (TYVIX). The choice of the aforementioned indexes is justified by the growing literature that confirms the cross-market transmission effects, either of returns or volatilities, between the Crude Oil Volatility Index and the other four indexes, as macroeconomic indicators. Based on these interactions, we support strongly that these four indexes provide crucial information for the future movements of the oil price volatility. In addition, we posit that the specific variables are among the most tradable futures contracts worldwide and they are ideal as their combined trading spans across the full day and they represent the stock market indexes of the largest economies in the world. Furthernore, we maintain that the EUR/USD is the main currency that exercises an impact on oil fluctuations, while recent studies have shown that the US 10-year T-bill futures, which are also included in our study, are responsive to change in the economic conditions. We treat the US 10-year T-bill as a variable that approximates global economic developments, given the importance of the US in the global economy. More specifically, the CBOE indexes of interest are the following ones:

The CBOE Volatility Index, with the trademarked ticker symbol VIX, is an index of the Chicago Board Options Exchange, which is used to calculate the implied volatility of the S&P500 index options; The VIX is designed to measure the market's expectation of stock market volatility over the next 30-day period. The volatility indexes used by the CBOE are basic tools for the measurement of market expectations of volatility conveyed by option prices. The VIX is calculated and disseminated in real-time by the CBOE and theoretically it is comprised by a weighted blend of prices for a range of options on the S&P500 index. It is quoted in percentage points, in the same way the standard deviation of a rate of return is expressed, e.g. 15.32.

The idea of a volatility index and financial instruments based on such an index was first developed and presented by Professor Menachem Brenner and Professor Dan Galai in 1986. They published their research in their academic article with the title "New Financial Instruments for Hedging Changes in Volatility", which appeared in the Financial Analyst Journal in 1989. Several years later, at a January 1993 news conference, Professor Robert Whaley announced the development of a tradable stock market volatility index based on index option prices and thus CBOE had computed

VIX on real-time basis. The VIX translates roughly the expected movement in the S&P index, over the upcoming 30-day period, which is then annualized. CBOE disseminates the index values continuously during trading hours. These indexes constitute major barometers of investor views and market fluctuations relating to listed options.

The Gold Volatility Index is used by the Chicago Board Options Exchange (CBOE) to calculate the market's expectation of 30-day fluctuation of gold prices, using the VIX methodology to options on SPDR Gold Shares. The trademarked ticker for the specific index is the GVZ and like other indexes on which the VIX methodology is applied, GVZ uses options spanning a wide range of strike prices. The SPDR Gold Shares (GLD), also known as SPDR Gold Trust, is part of the family of exchange-tranded funds that represent fractional, undivided interest managed and marketed by State Street Global Advisors. For a few years, the fund constituted the second-largest exchange-traded fund in the world, but as of the close of 2014 it has dropped out of the top ten.

CBOE has GVZ data going back to 2008. Comparing daily changes in GLD to GVZ shows that just under 60% of trading days the two will move in the opposite direction. So today was in the about 40% camp where GVZ moved in sync with GLD. So unlike equity market index volatility (VIX, VXN, VXEEM, VXEWZ) volatility on GLD can (and often does) rise when the price of GLD moves up. This piece of knowledge by itself explains why volatility traders may want to explore GVZ futures or options as a different way to play macro market movements.

The CBOE Crude Oil ETF Volatility Index, with the trademarked ticker "OVX", known also as Oil VIX, is a measurement tool used to calculate the market's expectation of the 30-day fluctuation of crude oil prices, using the VIX methodology to United States Oil Fund options spanning a wide range of strike prices. The United States Oil Fund has as a role to track market changes in crude oil prices and constitutes an exchange-traded security. The performance of the Fund is directed towards the reflection, as closely as possible, of the spot price of West Texas intermediate light, sweet crude oil, minus USO expenses, and this is achieved by holding near-term future contracts and cash.

The Treasury Volatility Index of the U.S., with the ticker symbol TYVIX, represents the index with the former title "CBOE/CBOT 10-year U.S. Treasury Note Volatiliy Index" and ticker VXTYN. This index utilizes the CBOE's VIX methodology to estimate a constant 30-day expected volatility of 10-year Treasury Note futures prices and is measured according to transparent pricing from CBOT's actively traded options on the T-Note futures. The TYVIX is the first exchange-traded volatility index for U.S. Treasuries and calculates the expected percentage changes in its CBOT futures on 10-year Treasury Notes over a one-month period.

Historically, the TYVIX is characterized by upward spikes, when 10-year Treasury note and future prices present wide fluctuations, especially large downswings. The properties of this index provide an innovative mechanism for core instruments of the U.S. fixed income market, including mortgage backed securities and corporate, municipal and government bonds.
The CBOE/CME FX Euro Volatility Index, with the trademarked ticker EUVIX, is one of the four volatility indexes used by the CBOE to measure the market's expectation of 30-day currency-related fluctuation, using once again the VIX methodology to options on currency-related measurement tools. Besides the CBOE/CME FX Euro Volatility Index, in the same category the following indexes are also included: the CBOE/CME FX Yen Volatility Index (Ticker: JYVIX), the CBOE/CME FX British Pound Volatility Index (Ticker: BPVIX) and the CBOE EuroCurrency Volatility Index (Ticker: EVZ).

The logic of the model is to examine whether the dependent variable (oil price realized volatility) and the adjusted implied volatility of different indexes' prices are linearly related, while at the same time there is enough evidence that all variables are statistically significant. It should be highlighted that the data we obtained were converted into their logarithm form with an aim to avoid different measurement tools and incorporate all the prolific characteristics of the time series. Another significant reason to use the logarithm form was that we tried to address any potential heteroscedasticity problems.

3.3 DATA SELECTION

The realized volatility of oil is constructed at a 23 minutes sampling frequency according to Degiannakis and Filis (2016). The data for the exogenous variables OVX, VIX, TYVIX, EUVIX and GVZ, are available for the period 3rd of January 2012 to 31 December 2015.

4. EMPIRICAL METHODOLOGY AND ECONOMETRIC ISSUES

4.1 EQUATIONS OUTPUTS ANALYSIS

We worked with the statistical package "EViews" (Econometric Views) and in this section we present the method we analyze our results with the outputs of our equations (See Appendices).

4.1.1 Description

Eviews displays the equation window, which provides the estimation output view. Besides the introductory elements presented at the beginning, such as the dependent variable, the method, the sample and the observations, in the second section of the window we see four columns that correspond to the estimates of the parameters, the standard error, the t-test performance of the parameters and their p-values. Finally, the third section of the window displays a number of statistical quantities that can be used to draw conclusions, as regards the suitability of the model, as well as to explain the relationship between our variables.

The standard regression, when utilizing matrix notation, can be estimated as follows:

$$y = X\beta + \varepsilon$$

where y stands for a T-dimensional vector that incorporates observations on the dependent variable, X stands for a T x k set of independent variables (where T is the number of observations and k the number of right-hand side regressors)

β expresses a k-vector of coefficients and

ε is a T-vector of disturbances.

4.1.2 Coefficient Results

REGRESSION COEFFICIENTS

The linear models utilized in this study include coefficients that estimate the marginal contribution of the independent variable to the dependent variable, while other variables remain fixed. The least squares regression coefficients *b* are calculated with the standard OLS formula:

$$b = (X'X)^{-1} X' y$$

"C" regressor corresponds to a constant or intercept coefficient in the regression that, when all of the other independent variables are zero, it stands for the base level of the prediction. If all other variables are kept fixed, then the remaining coefficients express the slope of the relations between the relevant independent variable and the dependent variable.

Standard Errors

The standard errors are an expression of the statistical reliability of the coefficient estimates. More specifically, a high standard error implies greater statistical noise in the estimations. The normal distribution of errors means that there are approximately 2 out of 3 chances that the true regression coefficient lies within one standard error of the reported coefficient, as well as 95 chances out of 100 that it lies within two standard deviations.

T-statistics

The t-statistic is the ratio of an estimated coefficient to its standard error. It can be utilized to examine whether a coefficient is equal to zero or not. In the interpretation of the t-statistic, it should be secured that the observations of the t-statistic take place while assuming that the coefficient is equal to zero.

Probability

While assuming that the errors are normally distributed or that the estimated coefficients are asymptomatically normally distributed, the probability of drawing a t-statistic or a z-statistic is as the observed one is presented in the last column of the output. Probability is also expressed as the p-value or the marginal significance level. P-value is can be used in order to reject or accept the hypothesis that the true coefficient is zero against a two-sided alternative other than zero. A t-distribution with T-k degrees of freedom is used to calculate the p-values for t-statistics, while a standard normal distribution can provide the p-value for z-statistics.

4.1.3 Summary Statistics

R-squared

The R-squared (R^2) statistic expresses the ability of the regression to predict the values of the dependent variable within the sample. It stands for the fraction of the variance of the dependent variable that is interpreted by the independent variables. Eviews calculates the R-squared (R^2) as follows:

$$R^2 = 1 - \frac{\hat{\epsilon}'\hat{\epsilon}}{(y-\bar{y})'(y-\bar{y})}; \qquad \bar{y} = \sum_{t=1}^{T} y_t / T$$

where *y* represents the mean value of the dependent variable.

Adjusted R-squared

The adjusted R^2, which penalizes the R^2 for the addition of regressors that do not contribute to the explanatory properties of the model, is expressed as follows:

$$\bar{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-k}$$

The adjusted R^2 decreases every time a regressor is added and is always lower than the R-squared or even negative in poorly fitting models.

Standard Error of the Regression

The S.E. of the regression depends on the estimated variance of the residuals and is calculated as follows:

$$S = \sqrt{\frac{\hat{\epsilon}'\hat{\epsilon}}{(T-k)}}$$

Sum-of-Squared Residuals

The sum-of-squared residuals, which is useful is several statistical calculations, can be estimated with the following formula:

$$\hat{\epsilon}'\hat{\epsilon} = \sum_{i=1}^{T}(y_i - X_i'b)^2$$

Log Likelihood

Log likelihood values are calculated at the estimated values of the coefficients and may be used to perform likelihood ratio tests, examining the difference between the log likelihood values of the restricted and unrestricted versions of an equation. The following mathematical equation is used to calculate log likelihood values:

$$l = -\frac{T}{2}(1 + \log 2\pi) + \log(\hat{\epsilon}'\hat{\epsilon}/T)$$

It should be mentioned that Eviews includes constant terms in the log likelihood calculations.

Durbin-Watson Statistic

The serial correlation in the residuals is calculated using the Durbin-Watson statistic, which is estimated as follows:

$$DW = \sum_{i=2}^{T} (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2 / \sum_{i=1}^{T} \hat{\varepsilon}_t^2$$

It is generally accepted that when DW is less than 2, then there is evidence of positive serial correlation.

Mean and Standard Deviation of the Dependent Variable

The mean and standard deviation of the dependent variable are calculated as follows:

$$\bar{y} = \sum_{i=1}^{T} y_t / T\,, \qquad S_y = \sqrt{\sum_{i=1}^{T} (y_t - \bar{y}^2)/(T-1)}$$

Akaike Information Criterion

The Akaike Information Criterion (AIC) is used in model selection for non-nested alternatives and is calculated as:

$$AIC = -\frac{2l}{T} + 2k/T$$

The length of a lag distribution can be chosen by selecting the specification with the lowest value of the AIC.

Schwartz Criterion

Instead of the AIC that imposes a larger penalty for additional coefficients, the Schwarz Criterion (SC) can be used:

$$SC = -\frac{2l}{T} + (k \log T)/T$$

F-Statistic

The test of the hypothesis that the total amount of slope coefficients (the constant or intercept ones are not included) in a regression are zero provides the F-statistic, which is estimated as:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(T-k)}$$

The aforementioned statistic is characterized by an F-distribution with k-1 numerator degrees of freedom and T-k denominator degrees of freedom. It should be mentioned that since the F-statistic is depended only on the sums-of-squared residuals of the estimated equations, it is poor in heterogeneity or serial correlation.

4.2 SELECTION OF BEST MODEL ON THE BASIS OF CRITERIA

Several times in the model selection procedure the information criteria can be used for guidance. The distance from the “real” model is expressed with the Kullback-Leibler quantity of information and is calculated by the log likelihood function. The role of the information criterion is the measurement of information that strikes a balance between this measure of goodness of fit and parsimonious specification of the model. The basic information criteria are provided by the following formulas:

Akaike info criterion (AIC)	-2(l/T) + 2(k/ T)
Schwarz criterion (SC)	-2(l/ T) + klog(T)/T
Hannan-Quinn criterion (HQ)	-2(1/ T) + 2klog(log(T))/T

where *l* is the value of the log of the likelihood function and *T* observations are used to estimate *k* values. It should be mentioned that the information criteria are based on -2 times the average log likelihood function with a penalty function adaptation.
Eviews re-centers the criteria by subtracting off the value for the saturated model, a factor analysis form of the information criteria, provided by the following formulas:

Akaike info criterion (AIC)	(T-k)D / T - (2/T)df
Sfhwarj: criterion (SC)	(T-k)D / T - (log(T)/T)df
Hannan-Quinn criterion (HQ)	(T-k)D / T - (2 In(log (T)) /T)df

where *D* is the discrepancy function and *df* the number of degrees-of-freedom in the estimated dispersion matrix.

The aforementioned information criteria can be used as a model selection guide. They have been used in time series analysis in order to define the proper length of the distributed lag. It should be mentioned, though, that the criteria rely heavily on the unit of measurement of the dependent variable y. Therefore, they cannot be used to choose between a model with dependent variable y and one with log(y).

If the models we have available were characterized by a restricted-unrestricted relationship, we could use one of the trinity tests and see what rates are statistically significant and what are not. But since we have to compare models, which are not directly related to each other, we use various selection criteria to provide a better model. We will follow the normal and regular approach by using the adjusted R^2, the standard error of regression, as well as other various criteria based on squares of residuals. A good choice is the Schwartz criterion, which actually is an improved version of the Akaike's Information Criterion, since, usually, the Akaike Criterion is selected as the best model with the largest number of parameters. We should also note that both of these criteria are the most appropriate models, based on adaptation data, and not the predictive power of the model with respect to the dependent variable, since they are based on the function of models' maximum likelihood. Finally, we should mention that the statistical package Eviews that we used is considered as the best model for the adaptation of the data that is the lowest in the Akaike and Schwartz criteria for out-of-sample comparison of models, as applied in our case.

4.3 TESTING MODELS RESIDUALS

For each estimated model, as regards some matters on residuals, it should be noted that they follow normal distribution to have constant fluctuation (homoskedasticity) and they are uncorrelated with each other. Then, we must check where these matters apply. To control regularity, we have selected the *Jarque Bera Test* that has been built in Eviews, to control homoskedasticity we used the *ARCH LM Test* - the relevant literature is the best way to estimate the variation - as well as the *White Test (non-cross terms)* and finally to control auto-correlation we have chosen the *Serial Correlation LM Test.*
More specifically :

4.3.1 The Jarque-Bera goodness-of-fit test

The Jarque-Bera goodness-of-fit test verifies the hypothesis that the data come from a normal distribution and is calculated as follows:

$$JB=(N-k/6)(S^2+(K-3)^2/4)$$

where N is the number of observations, S is the asymmetry, K is the curvature and k is the number of parameters that were evaluated to create the time series. If the value of the statistic test is greater than the Chi-Square distribution with two degrees of freedom, then the hypothesis is rejected and the corresponding p-value is displayed. We worked with a = 5% and this means in practice that if the p-value is less than 5%, then the initial hypothesis will be rejected for any significance level greater than 5%. We could, in order to increase the probability that we are in, diminish the value of a.

4.3.2 White Heteroskedasticity Test

White (1980) constructed a control test where the initial case of homoskedasticity of the residuals is tested against the alternative hypothesis that is characterized by heteroskedasticity. The White's statistic test, calculated as the number of observations on the determinative model coefficient, is a model that has as a dependent variable the squares of the residuals of the model and as explanatory variables, the independent variables of the model, their squares and their product by two (White Heteroskedasticity-cross terms). But we have worked with the variation of the statistic test of

White, the White Heteroskedasticity-non cross terms, which does not include what is happening now in pairs of explanatory variables.

4.3.3 Arch LM Test

Engle (1982) defined a new modeling method of heteroskedasticity in time series, the AutoRegressive Conditional Heteroscedasticity models and at the same time he proposed a LM test on the hypothesis of the residuals of a Regression Model, characterized by this form of heteroskedasticity. The LM statistic test of Engle, follows the Chi-Square distribution with i degrees of freedom (for i lag) and is calculated as the number of observations on the model's determinative factor in determining the model. In fact, the model on which the above calculation is based has as a dependent variable the square of the residuals of the model and as explanatory variables the time lags of the residuals' squares. The disadvantage of this method is that we cannot suppose that the statistic test follows the F-distribution, since the results also offer a statistical quantity referred as the F-statistic.

4.3.4 Serial Correlation LM Test

The test of Breusch (1978) and Godfrey (1978) is one of the Lagrange Multiplier Tests (LM Tests). This test can be used to establish the serial auto-correlation of any class, without assumptions that there are no lags of the dependent variable as explanatory variables. The LM statistic test of Breusch and Godfrey follows the Chi-Square distribution with i degrees of freedom and is calculated as the number of observations on the determinative factor of the model. The model on which the calculation of LM Test is based has as dependent variable the residuals of the model and as explanatory variables, the time lags of residuals. Just as in the LM statistic test of Engle, the results also provide a statistical quantity, referred as F-statistic, about which we cannot assume that it follows the F-distribution and therefore it should not be used.

5. EMPIRICAL RESULTS AND DISCUSSION

This chapter is based on the deliberations and results of the methods.
Choosing as a dependent variable the oil price realized volatility (RV) each time, we built the first HAR model with independent variable the logarithmic values of RV for 1- day, 5-days and 22-days ahead ("reference model"). Then we built five HAR models that had the same dependent variable and as independent variables they had the implied volatilities of a different indicator each time (OVX, VIX, TYVIX, EUVIX, GVZ) for the respective jumps. Finally, we compared these five models with each other and with the reference model.

The six models that were created are presented in the in the following table.

Table 5.1: Created Models

SN	Designation	Dependent Variable	Independent Variables
1	**HAR-RV-1**	RV_{oil}	RV_{oil}
2	**HAR-RV-X-1**	RV_{oil}	IV_{VIX}
3	**HAR-RV-X-2**	RV_{oil}	IV_{TYVIX}
4	**HAR-RV-X-3**	RV_{oil}	IV_{OVX}
5	**HAR-RV-X-4**	RV_{oil}	IV_{GVZ}
6	**HAR-RV-X-5**	RV_{oil}	IV_{EUVIX}

Source: Modified by the author

Correlation Analysis

In our attempt to determine the extent at which the aforementioned variables are related to each other, we utilized correlation analysis. As a rule, the correlation coefficient's values always range from -1 to +1. The +1 value means that the variables examined are absolutely related in a positive linear sense, while the -1 value that they are not related in a positive linear sense. Values close to 0 imply that there is no linear relationship between the variables. In the case of our model, there is neither perfect relationship between the examined variables, nor no linear relationship at all.

The illustration of our outcomes is presented in the table below (Table 5.2).

Table 5.2: Summary

	HAR-RV-1	HAR-RV-X-1	HAR-RV-X-2	HAR-RV-X-3	HAR-RV-X-4	HAR-RV-X-5
w_0	0.098534	0.085610	0.084039	0.032057	0.088447	0.105334
w_1	0.369806	0.319072	0.363972	0.230258	0.331069	0.349188
w_2	0.267783	0.257292	0.266390	0.194576	0.266630	0.245639
w_3	0.329937	0.392880	0.332640	0.436039	0.371476	0.349019
w_4		0.502536	0.163309	1.289146	0.485157	0.364659
w_5		-0.295323	-0.074915	-0.847824	-0.304207	-0.085995
w_6		-0.204176	-0.071928	-0.326692	-0.179023	-0.250254

Note: Predictive regression between oil price realized volatilities and implied volatilities of other predictors at various lags following models 1 to 6 are listed in Table 1. Estimation were made by WLS with fitted values of an OLS regression for weights. The values indicate statistical significance at least at the 5% level.
VIX = Volatility Index, TYVIX = U.S Treasury Volatility Index, OVX = Crude Oil Volatility Index, GVZ = Gold Volatility Index, EUVIX = Euro Volatility Index

Assess the model's fit

There are several techniques on evaluating data's fit.
The standard error of estimations ($S.E. of\ regression$), the coefficient of determination ($R - squared$), and the F-test of the analysis of variance ($F - statistic$) are three main statistical methods used to achieve this purpose.

Table 5.3. Consolidated values of R-squared, F-statistic and S.E. statistics

	R-Squared	Adjusted R-squared	S.E.	F-statistic and Prob(F-statistic)	
HAR-RV-1	0.734805	0.734006	0.228216	919.9087	0.0000
HAR-RV-X-1	0.743694	0.742146	0.224697	480.2134	0.0000
HAR-RV-X-2	0.735451	0.733852	0.228282	460.0919	0.0000
HAR-RV-X-3	0.759138	0.757683	0.217822	521.6164	0.0000
HAR-RV-X-4	0.740269	0.738700	0.226194	471.6976	0.0000
HAR-RV-X-5	0.738149	0.736567	0.227115	466.5381	0.0000

Source: Modified by the author

Table 3 illustrates all the values of these three statistics, as result of regression analysis in order to conclude whether the model fits or not.

Diagnose violations of required conditions

In the case of time series, like our models, first of all we must examine if the residuals are independent. The Durbin-Watson statistical test is an initial step on testing the null hypothesis of no evidence of autocorrelation. A value near 2 indicates no autocorrelation, while values range between 0 and 4. More specifically, values lower than 2 indicate positive first-order auto-correlation and up to 2 negative first-order auto-correlation. In the table below (Table 5.4) the DW statistic values for each model are presented.

Table 5.4. Consolidated Durbin-Watson statistic values

Models	Durbin-Watson statistic
HAR-RV-1	2.021286
HAR-RV-X-1	2.008678
HAR-RV-X-2	2.020224
HAR-RV-X-3	1.974770
HAR-RV-X-4	2.007205
HAR-RV-X-5	2.009644

Source: Modified by the author

After conducting the Durbin-Watson test we carried out three additional assessments on the residuals (White Test, ARCH LM Test, Serial Correlation LM Test) and demonstrated that there is no auto-correlation and heteroscedasticity. The relevant diagrams and tables are presented in Appendices.

The table below summarizes the main features of the models' assessment.

Table 5.5. Consolidated values of Akaike info criterion, Schwarz criterion and Hannan - Quinn criterion

	HAR-RV-1	HAR-RV-X-1	HAR-RV-X-2	HAR-RV-X-3	HAR-RV-X-4	HAR-RV-X-5
AIC	-0.113055	-0.141150	-0.109492	**-0.203298**	-0.127874	-0.119744
SC	-0.093424	-0.106795	-0.075138	**-0.168943**	-0.093519	-0.085389
HQ	-0.105594	-0.128093	-0.096435	**-0.190241**	-0.114817	-0.106687

As mentioned in the previous chapter, the basic Criterion in the selection of the best model is the Schwartz Criterion, because it can be used to compare any models, regardless of the number of parameters, with the only restriction to have the same pendent variable.

6. CONCLUSION

In this thesis we studied the methodology that has been developed on estimating and forecasting oil price volatility. Our limited research was based on recent surveys, which proved that the correlation between HAR models and exogenous volatilities from different asset classes improves the forecasting accuracy (see, Sévi, 2014; Degiannakis and Fillis, 2015). We have presented the results from an empirical analysis of six HAR-type time-series models for 1000 observations, the aim of which was to estimate the realized volatility of oil, in support of the information of the implied volatilities of other economical indicators (exogenous variables). We obtained very strong evidence towards the right direction that supported these studies. More specifically, in our study, where six models with a common dependent variable were compared, that is, the realized volatility of oil price, we concluded that the best model for the assessment of RV_{OIL}, on the basis of the SC criterion, is the HAR-X-3 Model that has as independent variables the implied volatilities of the OVX Index. After that, the HAR-X-1 (with independent variables the implied volatilities of the VIX Index), HAR-X-4 (with independent variables the implied volatilities of the GVZ Index), HAR-RV-1 (without exogenous variables), HAR-X-5 (with independent variables the implied volatilities of the EUVIX Index) models follow and the HAR-X-2 (with independent variables the implied volatilities of the TYVIX Index) model is the last one. This makes it clear that in most cases the exogenous variables bearing information from other markets offer more information on estimating the oil price realized volatility, but require thorough examination, in which exogenous variables will be selected to be integrated in our model.

There are several directions for future research and the challenge is to improve the forecastability on oil price volatility. One relates to the investigation of the optimum sampling frequency to measure the volatility of crude oil prices and the pursuit of periods with relatively similar variability in their dynamic to the weigh optimally, such as the semi-parametric approach. One could also attempt to model the volatility of oil prices in combination with other goods and various stock markets and compare the extent to which the estimation is affected. Given the use of different models in our econometric analysis, we will be able to compare the diffusion achieved by the realized volatility of oil price in relation to: stocks, exchange rates, the real economy and other commodities, exploring accurate quantitative economic relationships linking them, and conversely create a tool for the predictability of variation of other economic indicators, related or indirectly related, with oil-market.

7. REFERENCES

Abosedra, S. (2005). Futures versus univariate forecast of crude oil prices, OPEC Energy Review, 29, Issue 4

Agnolucci, P. (2009). Volatility in crude oil futures: a comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics*, 31, 316-321

Agren, M., (2006). Does oil price uncertainty transmit to stock markets? (No. 2006: 23). Working Paper, Department of Economics, Uppsala University.

Alizadeh, S., Brandt, M.W., Diebold, F.X. (2002), Range-Based Estimation of Stochastic Volatility Models, *The journal of FINANCE*, 57, Issue 3, 1047–1091

Aloui, C., Mabrouk, S. (2010). Value –at – risk estimations of energy commodities via long memory, asymmetry and fat-tailed GARCH models. *Energy Policy* 38, 2326-2339.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, 885-905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, 89(4), 701-720.

Arouri, M. E. H., Jouini, J., & Nguyen, D. K. (2011). Volatility spillovers between prices and stock sector returns: implications for portfolio management. *Journal of International Money and Finance*, 30(7), 1387-1405.

Bali, T.G. (2003). An Extreme Value Approach to Estimating Volatility and Value at Risk, *The Journal of Business* 76, 1, 83-108

Ball, C. A., Torous, W. N. (1984). The maximum likelihood estimation of security price volatility: Theory, evidence, and application to option pricing. *Journal of Busi-*

ness, 57, 97-112.

Barndorff-Nielsen, O., Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4, 1-30.

Bopp, A.E, Lady, G.M. (1991). A comparison of petroleum futures versus spot prices as predictors of prices in the future, *Energy Economics*, 13,4, 274-282

Breeden, D., T., Litzenberger, R.H. (1978). Prices of State-Contingent Claims Implicit in Option Prices, *The Journal of Business*, 51, 4, 621-651

Breusch, T.S.(1978). Testing for Autocorrelation in Dynamic Linear Models. Australian Economic Papers, 17, 334-355

Britten – Jones, M., Neuberger, A. (2000) Option Prices, Implied Price Processes, and Stochastic Volatility, *The journal of FINANCE*, 55, 2, 839–866

Chen, N., Roll, R., Ross, S. (1986). Economic Forces and the Stock Market, *The Journal of Business*, 59, 3, 383-403

Chen, X., Ghysels, E, (2011). News – good or bad – and its impact on volatility predictions overmultiple horizons. Review of Financial Studies 24, 46-81.

Cheong, C. W. (2009). Modelling and forecasting crude oil markets using ARCH – type models. *Energy Policy* 37, 2346-2355

Choi, K., Hammoudeh, S. (2009). Long memory in oil and refined products markets. *Energy* 30, 57-76.

Chou, R. (2005). Forecasting _nancial volatilities with extreme values: the conditional autoregressive range (CARR) model, *Journal of Money, Credit and Banking 37, 561-582.*

Clark, P. K. (1973), A subordinated stochastic process model with finite variance for speculative prices, Econometrica: *Journal of the Econometric Society*, 135-155.

Corsi, F. (2009). A simple approximate longmemorymodel of realized volatility. *Journal of Financial Econometrics* 7, 174-196.

Corsi, F., Fusari, N., La Vecchia, D. (2013). Realizing smiles: Options pricing with realized volatility. *Journal of Financial Economics* 107, 284-304.

Davinson, J. (2004). Moment and memory proprieties of linear conditional heteroskedasticity models and a new model. J. Bus. Econ. Stat 22, 16-29

Degiannakis, S., Filis, G. (2015). Forecasting oil price realized volatility: A new approach, ENEFOR, Panteion University.

Ding, Z., Granger, C., Engle. (1993). A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, 1, 1, 83-106

Efimova, O., & Serletis, A. (2014). Energy markets volatility modelling using GARCH. *Energy Economics*, 43, 264-273.

Engle, E., & Sun, Z. (2007). When is noise not noise–a microstructure estimate of realized volatility.

Engle, R.F. (1982). Autoregressive Conditional Hetroskedasticity with Estimates of the Variance of U.K. Inflation. *Econometrica,* 50, 987-1008

Figlewski, S. (1981). Futures trading and volatility in the gnma market, *The Journal of Finance*, 36, 445–456.

Fleming, J., Kirby, C., Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility, *Journal of Financial Economics*, 67, 473–509.

Garman, M., Klass, M. (1980). On the estimation of security price volatilities from historical data, *The Journal of Business*, 53, 67-78.

Godfrey, L.G. (1978). Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables. *Econometrica*, 46, 1293-1302

Hammoudeh, S., Dibooglu, S., & Aleisa, E. (2004). Relationships among US oil prices and oil industry equity indices. *International Review of Economics & Finance*, 13(4), 427-453.

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH. *Journal of Applied Econometrics*, 20, 873- 889.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2), 453-497.

Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking & Finance*, 47, 1-14.

Huang, R. D., Masulis, R. W., & Stoll, H. R. (1996). Energy shocks and financial markets. *Journal of Futures Markets*, 16, 1, 1-27.

Kang, S.-H., & Yoon, S.-M. (2013). Modelling and forecasting the volatility of petroleum futures prices. *Energy Economics*, 36, 354–362.

Kang, S.H., Kang, S.-M., Yoon, S.-M. (2009). Forecasting volatility of crude oil-markets. *Energy Economics* 31, 119-125.

Kunitomo, N. (1992). Improving the Parkinson method of estimating security price volatilities, *Journal of Business* 65, 292-302.

Malik, F., & Ewing, B. T. (2009). Volatility transmission between oil prices and equity sector returns. *International Review of Financial Analysis*, 18(3), 95-100.

Miller, J., Ratti, R. (2009). Crude oil and stock markets: Stability, instability, and bubbles, *Energy Economics*, 31, 559-568.

Mohammadi, H., Su, L. (2010). International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models. *Energy Economics* 32, 1001-1008.

Narayan, P., Narayan, S. (2007). Modelling oil price volatility. *Energy Policy* 35, 6549-6553.

Narayan, P., Sharma, S. (2011). New evidence on oil price and firm returns, *Journal of Banking and Finance* 35, 3253–3262.

Nelson, D. B. (1990). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347-370

Nomikos, N. K., & Pouliasis, P. K. (2011). Forecasting petroleum futures markets volatility: The role of regimes and market conditions. *Energy Economics*, 33(2), 321-337.

Park, J., Ratti, R. (2008). Oil price shocks and stock markets in the U.S. and 13 European countries. *Energy Economics* 30, 2587-2608.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return, *Journal of Business* 53, 61-65.

Patton, A.J., Sheppard, K. (2011). Good volatility, bad volatility: signed jumps and the persistence of volatility. Unpublished Manuscript.

Phan, D. H. B., Sharma, S. S., & Narayan, P. K. (2015). Intraday volatility interaction between the crude oil and equity markets. *Journal of International Financial*

Markets, Institutions and Money, forthcoming.

Phan, D.H., Sharma, S.S., Narayan, P.K. (2015). Oil price and stock returns of consumers and producers of crude oil, *Journal of International Financial Markets*, Institutions and Money, 34, 245–262

Phan, D.H., Sharma, S.S., Narayan, P.K. (2015). Stock return forecasting: Some new evidence, *International Review of Financial Analysis*, 40, 38–51

Pindyck, R. S. (1999). Volatility and commodity price dynamics. J. Futur. Mark. 24, 1029-1047.

Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2015). Do Jumps Matter for Volatility Forecasting? Evidence from Energy Markets. *Journal of Futures Markets*, 1-35.

Rogers, L., Satchell, S. (1991). Estimating variance from high, low, and closing prices, *Analysis of Applied Probability*, 1, 500-512

Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics*, 28, 467–488.

Sadorsky, P., & McKenzie, M. D. (2008). Power transformation models and volatility forecasting. *Journal of Forecasting*, 27, 587–606.

Schwert, G. W. (1990). Stock volatility and the crash of '87, *Review of Financial Studies* 3, 77-102.

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, 235(3), 643-659.

Shu, J. H., Zhang, J. E. (2006). Testing range estimators of historical volatility. *Journal of Futures Markets* 26, 297-313.

Wang, G., Yau, Y. (2000). Trading volume, bid–ask spread, and price volatility in futures markets, *Journal of Futures markets*, 20, 10, 943–970.

Wang, T, Wu, J., Yang, J. (2008). Realized volatility and correlation in energy futures markets, *Journal of Futures Markets*, 28, 10, 993–1011.

Wei, C. (2003). Energy, the stock market, and the Putty-Clay investment model. *American Economic Review* 93, 311–23.

Wei, Y., Wang, Y., & Huang, D. (2010). Forecasting crude oil market volatility:

Further evidence using GARCH-class models. *Energy Economics*, 32, 1477– 1484.

White, H. (1980). A Heteroskedasticity Consistent Covariance Matrix and a Direct Test for Heteroskedasticity. *Econometrica*, 48, 817-838

Yang D., and Zhang, Q. (2000). Drift-independent volatility estimation based on high, low, open, and close prices, *Journal of Business* 73, 477-491.

EViews 9 User's Guide II
www.oilprice.com
www.tutor2u.net
www.eviews.com

APPENDICES

Table 1. OLS results for the HAR-RV-1 model

Dependent Variable: LRV_OIL

Method: Least Squares

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.098534	0.057909	1.701546	0.0892
LRV_OIL(-1)	0.369806	0.036815	10.04488	0.0000
@MOVAV(LRV_OIL(-1),5)	0.267783	0.061033	4.387532	0.0000
@MOVAV(LRV_OIL(-1),22)	0.329937	0.052223	6.317809	0.0000

R-squared	0.734805	Mean dependent var	2.991469
Adjusted R-squared	0.734006	S.D. dependent var	0.442497
S.E. of regression	0.228216	Akaike info criterion	-0.113055
Sum squared resid	51.87430	Schwarz criterion	-0.093424
Log likelihood	60.52734	Hannan-Quinn criter.	-0.105594
F-statistic	919.9087	Durbin-Watson stat	2.021286
Prob(F-statistic)	0.000000		

Source: Modified by the author

Table 2. OLS results for the HAR-RV-X-1 model

Dependent Variable: LRV_OIL

Method: Least Squares

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.085610	0.078243	1.094157	0.2742
LRV_OIL(-1)	0.319072	0.037504	8.507637	0.0000
@MOVAV(LRV_OIL(-1),5)	0.257292	0.064871	3.966192	0.0001
@MOVAV(LRV_OIL(-1),22)	0.392880	0.057976	6.776573	0.0000
LVIX(-1)	0.502536	0.103455	4.857518	0.0000
@MOVAV(LVIX(-1),5)	-0.295323	0.137058	-2.154730	0.0314
@MOVAV(LVIX(-1),22)	-0.204176	0.083858	-2.434774	0.0151

R-squared	0.743694	**Mean dependent var**	2.991469
Adjusted R-squared	0.742146	**S.D. dependent var**	0.442497
S.E. of regression	0.224697	**Akaike info criterion**	-0.141150
Sum squared resid	50.13546	**Schwarz criterion**	-0.106795
Log likelihood	77.57482	**Hannan-Quinn criter.**	-0.128093
F-statistic	480.2134	**Durbin-Watson stat**	2.008678
Prob(F-statistic)	0.000000		

Source: Modified by the author

Table 3. OLS results for the HAR-RV-X-2 model

Dependent Variable: LRV_OIL

Method: Least Squares

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.084039	0.070513	1.191815	0.2336
LRV_OIL(-1)	0.363972	0.037144	9.798861	0.0000
@MOVAV(LRV_OIL(-1),5)	0.266390	0.061912	4.302694	0.0000
@MOVAV(LRV_OIL(-1),22)	0.332640	0.054588	6.093672	0.0000
LTYVIX(-1)	0.163309	0.145027	1.126058	0.2604
@MOVAV(LTYVIX(-1),5)	-0.074915	0.185550	-0.403745	0.6865
@MOVAV(LTYVIX(-1),22)	-0.071928	0.105601	-0.681132	0.4959

R-squared	0.735451	Mean dependent var	2.991469
Adjusted R-squared	0.733852	S.D. dependent var	0.442497
S.E. of regression	0.228282	Akaike info criterion	-0.109492
Sum squared resid	51.74802	Schwarz criterion	-0.075138
Log likelihood	61.74602	Hannan-Quinn criter.	-0.096435
F-statistic	460.0919	Durbin-Watson stat	2.020224
Prob(F-statistic)	0.000000		

Source: Modified by the author

Table 4. OLS results for the HAR-RV-X-3 model

Dependent Variable: LRV_OIL

Method: Least Squares

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.032057	0.070487	0.454795	0.6494
LRV_OIL(-1)	0.230258	0.038494	5.981710	0.0000
@MOVAV(LRV_OIL(-1),5)	0.194576	0.072208	2.694661	0.0072
@MOVAV(LRV_OIL(-1),22)	0.436039	0.077319	5.639499	0.0000
LOVX(-1)	1.289146	0.142960	9.017509	0.0000
@MOVAV(LOVX(-1),5)	-0.847824	0.191389	-4.429851	0.0000
@MOVAV(LOVX(-1),22)	-0.326692	0.117489	-2.780614	0.0055

R-squared	0.759138	Mean dependent var	2.991469
Adjusted R-squared	0.757683	S.D. dependent var	0.442497
S.E. of regression	0.217822	Akaike info criterion	-0.203298
Sum squared resid	47.11449	Schwarz criterion	-0.168943
Log likelihood	108.6488	Hannan-Quinn criter.	-0.190241
F-statistic	521.6164	Durbin-Watson stat	1.974770
Prob(F-statistic)	0.000000		

Source: Modified by the author

Table 5. OLS results for the HAR-RV-X-4 model

Dependent Variable: LRV_OIL

Method: Least Squares

Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.088447	0.097251	0.909467	0.3633
LRV_OIL(-1)	0.331069	0.037709	8.779511	0.0000
@MOVAV(LRV_OIL(-1),5)	0.266630	0.064280	4.147956	0.0000
@MOVAV(LRV_OIL(-1),22)	0.371476	0.056367	6.590358	0.0000
LGVZ(-1)	0.485157	0.126815	3.825711	0.0001
@MOVAV(LGVZ(-1),5)	-0.304207	0.164648	-1.847618	0.0650
@MOVAV(LGVZ(-1),22)	-0.179023	0.096741	-1.850543	0.0645

R-squared	0.740269	Mean dependent var	2.991469
Adjusted R-squared	0.738700	S.D. dependent var	0.442497
S.E. of regression	0.226194	Akaike info criterion	-0.127874
Sum squared resid	50.80550	Schwarz criterion	-0.093519
Log likelihood	70.93684	Hannan-Quinn criter.	-0.114817
F-statistic	471.6976	Durbin-Watson stat	2.007205
Prob(F-statistic)	0.000000		

Source: Modified by the author

Table 6. OLS results for the HAR-RV-X-5 model

Dependent Variable: LRV_OIL

Method: Least Squares

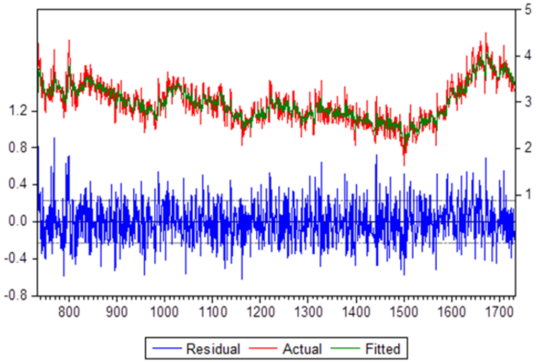
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.105334	0.059834	1.760450	0.0786
LRV_OIL(-1)	0.349188	0.037539	9.301894	0.0000
@MOVAV(LRV_OIL(-1),5)	0.245639	0.063398	3.874589	0.0001
@MOVAV(LRV_OIL(-1),22)	0.349019	0.058774	5.938362	0.0000
LEUVIX(-1)	0.364659	0.162757	2.240510	0.0253
@MOVAV(LEUVIX(-1),5)	-0.085995	0.211514	-0.406571	0.6844
@MOVAV(LEUVIX(-1),22)	-0.250254	0.116074	-2.155998	0.0313

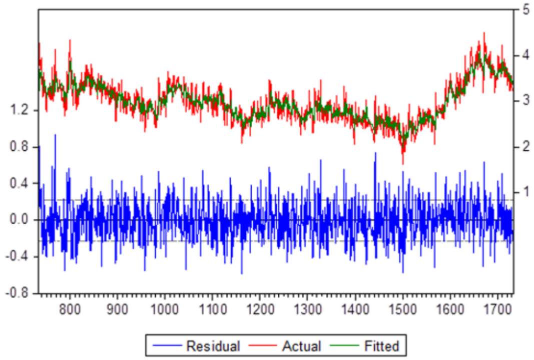
R-squared	0.738149	Mean dependent var	2.991469
Adjusted R-squared	0.736567	S.D. dependent var	0.442497
S.E. of regression	0.227115	Akaike info criterion	-0.119744
Sum squared resid	51.22023	Schwarz criterion	-0.085389
Log likelihood	66.87177	Hannan-Quinn criter.	-0.106687
F-statistic	466.5381	Durbin-Watson stat	2.009644
Prob(F-statistic)	0.000000		

Source: Modified by the author

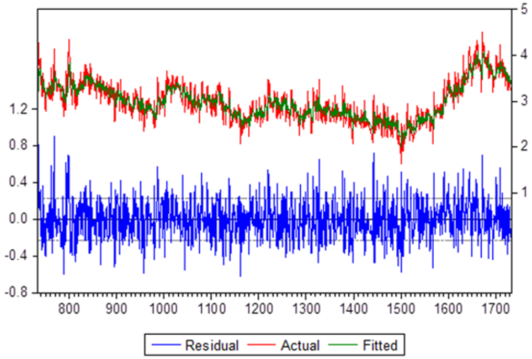
Figure 1: Actual, Fitted, Residual Graph



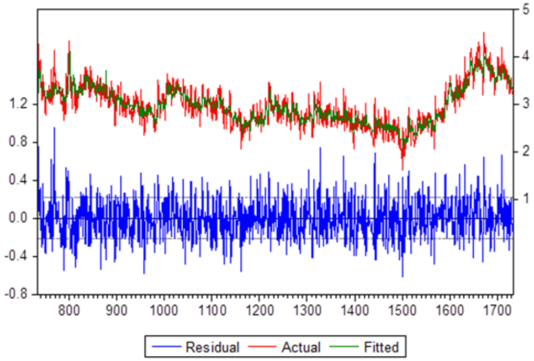
HAR-RV-1 Model



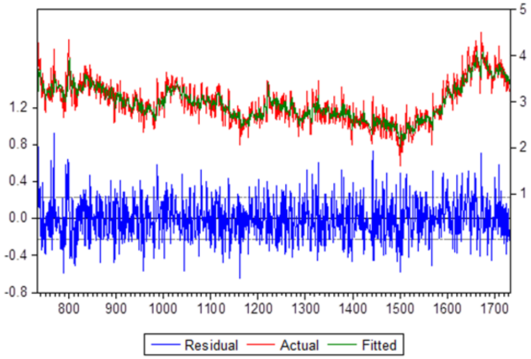
HAR-X-1 Model



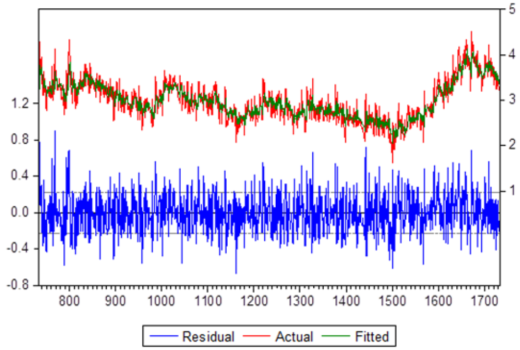
HAR-X-2 Model



HAR-X-3 Model

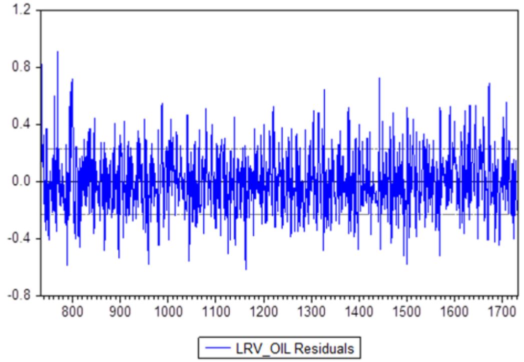


HAR-X-4 Model

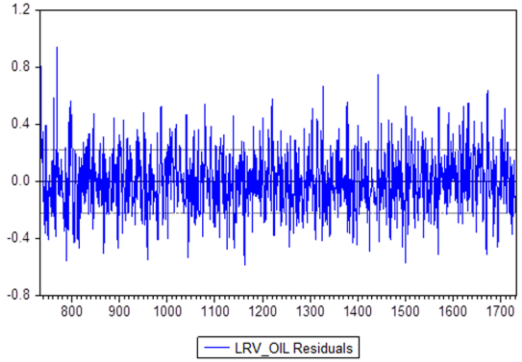


HAR-X-5 Model

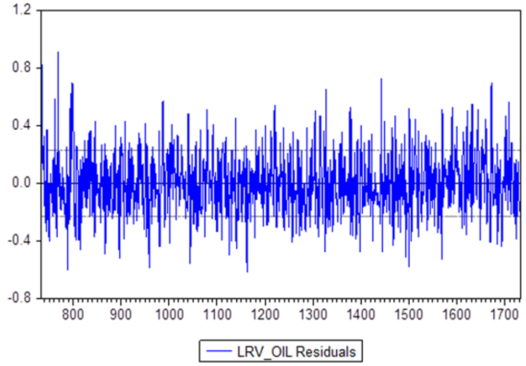
Figure 2: Residual Graph



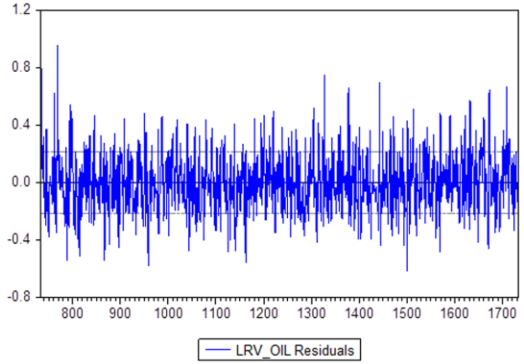
HAR-RV-1 Model



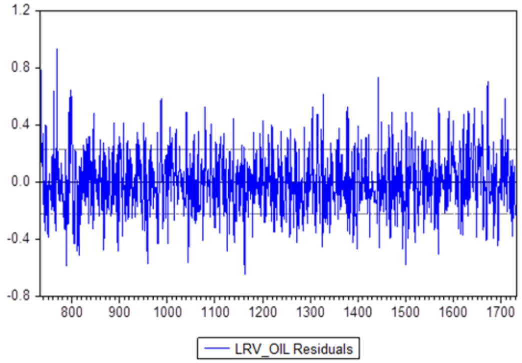
HAR-X-1 Model



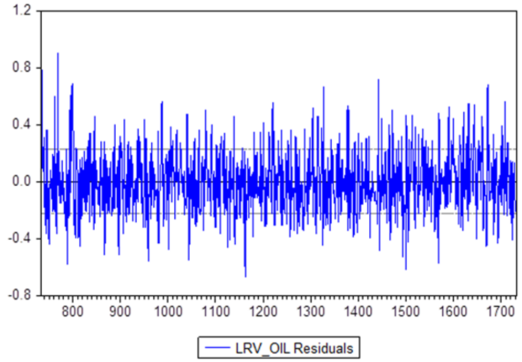
HAR-X-2 Model



HAR-X-3 Model

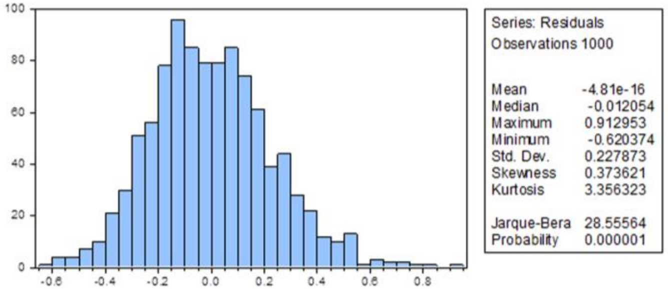


HAR-X-4 Model

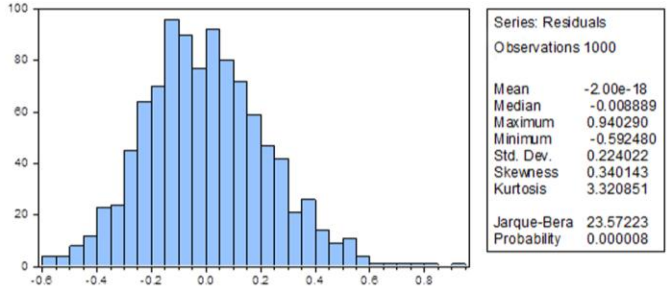


HAR-X-5 Model

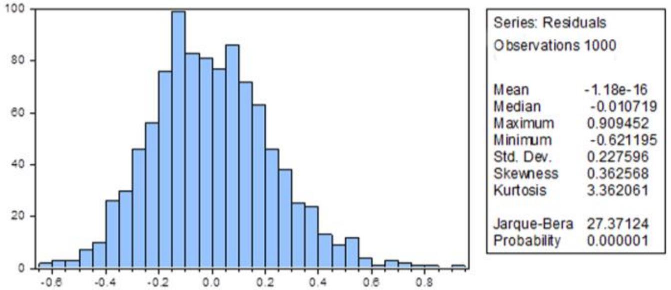
Figure 3: Histograms of Normality Test



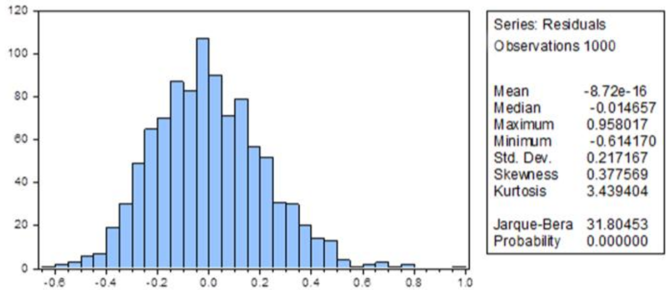
HAR-RV-1 Model



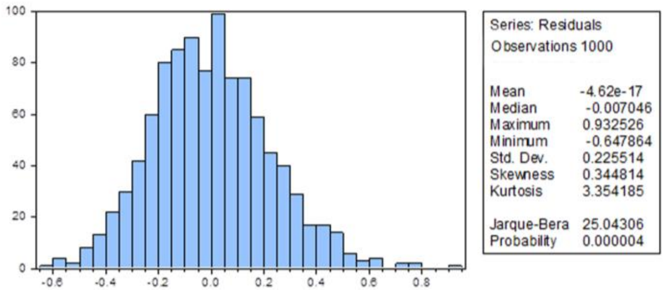
HAR-X-1 Model



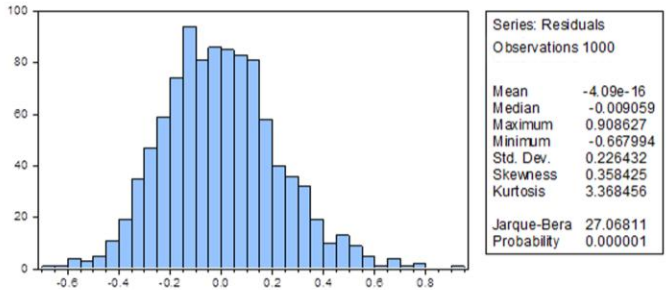
HAR-X-2 Model



HAR-X-3 Model



HAR-X-4 Model



HAR-X-5 Model

Table 7: Results of White Test, ARCH Test and LM Test for HAR-RV-1 Model

Heteroskedasticity Test: White

F-statistic	2.125613	Prob. F(9,990)	0.0251
Obs*R-squared	18.95743	Prob. Chi-Square(9)	0.0256
Scaled explained SS	22.15659	Prob. Chi-Square(9)	0.0084

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.027728	0.142162	0.195045	0.8454
LRV_OIL(-1)^2	0.104238	0.044383	2.348629	0.0190
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),5)	-0.078214	0.121088	-0.645929	0.5185
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),22)	-0.128723	0.086031	-1.496250	0.1349
LRV_OIL(-1)	0.018057	0.102146	0.176779	0.8597
@MOVAV(LRV_OIL(-1),5)^2	0.004815	0.107947	0.044606	0.9644
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LRV_OIL(-1),22)	0.131599	0.164882	0.798142	0.4250
@MOVAV(LRV_OIL(-1),5)	-0.233767	0.172984	-1.351377	0.1769
@MOVAV(LRV_OIL(-1),22)^2	-0.035045	0.084669	-0.413903	0.6790
@MOVAV(LRV_OIL(-1),22)	0.226305	0.165335	1.368771	0.1714

R-squared	0.018957	Mean dependent var	0.051874
Adjusted R-squared	0.010039	S.D. dependent var	0.079669
S.E. of regression	0.079268	Akaike info criterion	-2.222024
Sum squared resid	6.220527	Schwarz criterion	-2.172946
Log likelihood	1121.012	Hannan-Quinn criter.	-2.203371
F-statistic	2.125613	Durbin-Watson stat	1.926307
Prob(F-statistic)	0.025067		

Heteroskedasticity Test: ARCH

F-statistic	14.32151	Prob. F(1,997)	0.0002
Obs*R-squared	14.14702	Prob. Chi-Square(1)	0.0002

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.045284	0.002913	15.54609	0.0000
RESID^2(-1)	0.115937	0.030636	3.784377	0.0002

R-squared	0.014161	Mean dependent var	0.051304
Adjusted R-squared	0.013172	S.D. dependent var	0.077641
S.E. of regression	0.077128	Akaike info criterion	-2.284691
Sum squared resid	5.930943	Schwarz criterion	-2.274867
Log likelihood	1143.203	Hannan-Quinn criter.	-2.280957
F-statistic	14.32151	Durbin-Watson stat	2.066365
Prob(F-statistic)	0.000163		

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	6.593819	Prob. F(2,994)	0.0014
Obs*R-squared	13.09353	Prob. Chi-Square(2)	0.0014

Dependent Variable: RESID
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.138318	0.069642	-1.986119	0.0473
LRV_OIL(-1)	1.203860	0.338610	3.555294	0.0004
@MOVAV(LRV_OIL(-1),5)	-0.630953	0.184610	-3.417752	0.0007
@MOVAV(LRV_OIL(-1),22)	-0.527345	0.159577	-3.304648	0.0010
RESID(-1)	-1.072815	0.299096	-3.586864	0.0004
RESID(-2)	-0.290164	0.092840	-3.125421	0.0018

R-squared	0.013094	Mean dependent var	-4.81E-16
Adjusted R-squared	0.008129	S.D. dependent var	0.227873
S.E. of regression	0.226945	Akaike info criterion	-0.122235
Sum squared resid	51.19508	Schwarz criterion	-0.092788
Log likelihood	67.11734	Hannan-Quinn criter.	-0.111043
F-statistic	2.637528	Durbin-Watson stat	1.991608
Prob(F-statistic)	0.022274		

Table 8: Results of White Test, ARCH Test and LM Test for HAR-X-1

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.268968	Prob. F(2,991)	0.2816
Obs*R-squared	2.554443	Prob. Chi-Square(2)	0.2788

Dependent Variable: RESID
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.035078	0.081589	-0.429937	0.6673
LRV_OIL(-1)	0.347062	0.225931	1.536140	0.1248
@MOVAV(LRV_OIL(-1),5)	-0.158896	0.118994	-1.335324	0.1821
@MOVAV(LRV_OIL(-1),22)	-0.175745	0.131466	-1.336808	0.1816
LVIX(-1)	-0.074171	0.113466	-0.653684	0.5135
@MOVAV(LVIX(-1),5)	-0.043375	0.141577	-0.306373	0.7594
@MOVAV(LVIX(-1),22)	0.116497	0.114240	1.019757	0.3081
RESID(-1)	-0.314043	0.200717	-1.564606	0.1180
RESID(-2)	-0.073909	0.063988	-1.155043	0.2484

R-squared	0.002554	Mean dependent var	-2.00E-18
Adjusted R-squared	-0.005498	S.D. dependent var	0.224022
S.E. of regression	0.224636	Akaike info criterion	-0.139707
Sum squared resid	50.00739	Schwarz criterion	-0.095538
Log likelihood	78.85367	Hannan-Quinn criter.	-0.122920
F-statistic	0.317242	Durbin-Watson stat	1.988302

Heteroskedasticity Test: ARCH

F-statistic	10.63098	Prob. F(1,997)	0.0011
Obs*R-squared	10.53991	Prob. Chi-Square(1)	0.0012

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.044576	0.002810	15.86358	0.0000
RESID^2(-1)	0.100231	0.030741	3.260518	0.0011

R-squared	0.010550	Mean dependent var	0.049606
Adjusted R-squared	0.009558	S.D. dependent var	0.074593
S.E. of regression	0.074236	Akaike info criterion	-2.361133
Sum squared resid	5.494462	Schwarz criterion	-2.351310
Log likelihood	1181.386	Hannan-Quinn criter.	-2.357399
F-statistic	10.63098	Durbin-Watson stat	2.057274
Prob(F-statistic)	0.001150		

Heteroskedasticity Test: White

F-statistic	1.116608	Prob. F(27,972)	0.3110
Obs*R-squared	30.08380	Prob. Chi-Square(27)	0.3104
Scaled explained SS	34.42297	Prob. Chi-Square(27)	0.1541

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.90E-05	0.335391	-0.000235	0.9998
LRV_OIL(-1)^2	0.069054	0.047191	1.463289	0.1437
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),5)	-0.110131	0.136217	-0.808500	0.4190
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),22)	-0.067429	0.100358	-0.671890	0.5018
LRV_OIL(-1)*LVIX(-1)	0.151688	0.207749	0.730151	0.4655
LRV_OIL(-1)*@MOVAV(LVIX(-1),5)	-0.198881	0.254274	-0.782153	0.4343
LRV_OIL(-1)*@MOVAV(LVIX(-1),22)	0.083107	0.150538	0.552066	0.5810
LRV_OIL(-1)	0.028692	0.160270	0.179023	0.8580
@MOVAV(LRV_OIL(-1),5)^2	-0.005402	0.140022	-0.038581	0.9692
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LRV_OIL(-1),22)	0.174766	0.217430	0.803779	0.4217
@MOVAV(LRV_OIL(-1),5)*LVIX(-1)	0.039880	0.339778	0.117369	0.9066
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LVIX(-1),5)	0.056635	0.473706	0.119556	0.9049
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LVIX(-1),22)	-0.148620	0.295894	-0.502275	0.6156
@MOVAV(LRV_OIL(-1),5)	-0.057614	0.268597	-0.214500	0.8302
@MOVAV(LRV_OIL(-1),22)^2	-0.072452	0.110977	-0.652856	0.5140
@MOVAV(LRV_OIL(-1),22)*LVIX(-1)	0.003174	0.299665	0.010591	0.9916
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LVIX(-1),5)	0.100971	0.440142	0.229405	0.8186
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LVIX(-1),22)	-0.011069	0.310087	-0.035696	0.9715
@MOVAV(LRV_OIL(-1),22)	-0.104410	0.256987	-0.406286	0.6846
LVIX(-1)^2	-0.069042	0.298959	-0.230941	0.8174
LVIX(-1)*@MOVAV(LVIX(-1),5)	0.350341	0.756292	0.463235	0.6433
LVIX(-1)*@MOVAV(LVIX(-1),22)	-0.400955	0.401535	-0.998554	0.3183
LVIX(-1)	-0.050884	0.400860	-0.126936	0.8990
@MOVAV(LVIX(-1),5)^2	-0.231742	0.591932	-0.391500	0.6955
@MOVAV(LVIX(-1),5)*@MOVAV(LVIX(-1),22)	0.363275	0.673207	0.539619	0.5896
@MOVAV(LVIX(-1),5)	-0.572118	0.532062	-1.075284	0.2825
@MOVAV(LVIX(-1),22)^2	-0.082073	0.280188	-0.292921	0.7696
@MOVAV(LVIX(-1),22)	0.786934	0.371150	2.120259	0.0342

R-squared	0.030084	Mean dependent var	0.050135
Adjusted R-squared	0.003142	S.D. dependent var	0.076416
S.E. of regression	0.076296	Akaike info criterion	-2.280788
Sum squared resid	5.658112	Schwarz criterion	-2.143371
Log likelihood	1168.394	Hannan-Quinn criter.	-2.228560
F-statistic	1.116608	Durbin-Watson stat	1.888580
Prob(F-statistic)	0.311007		

Table 9: Results of White Test, ARCH Test and LM Test for for HAR-X-2

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	6.382781	Prob. F(2,991)	0.0018
Obs*R-squared	12.71767	Prob. Chi-Square(2)	0.0017

Dependent Variable: RESID
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.113251	0.077319	-1.464720	0.1433
LRV_OIL(-1)	1.172776	0.334543	3.505604	0.0005
@MOVAV(LRV_OIL(-1),5)	-0.605759	0.180701	-3.352264	0.0008
@MOVAV(LRV_OIL(-1),22)	-0.516057	0.159176	-3.242046	0.0012
LTYVIX(-1)	-0.089238	0.146462	-0.609291	0.5425
@MOVAV(LTYVIX(-1),5)	-0.068195	0.186222	-0.366205	0.7143
@MOVAV(LTYVIX(-1),22)	0.133253	0.111922	1.190592	0.2341
RESID(-1)	-1.044531	0.295369	-3.536353	0.0004
RESID(-2)	-0.281358	0.091211	-3.084711	0.0021

R-squared	0.012718	Mean dependent var	-1.18E-16
Adjusted R-squared	0.004748	S.D. dependent var	0.227596
S.E. of regression	0.227055	Akaike info criterion	-0.118291
Sum squared resid	51.08991	Schwarz criterion	-0.074121
Log likelihood	68.14563	Hannan-Quinn criter.	-0.101504
F-statistic	1.595695	Durbin-Watson stat	1.990933
Prob(F-statistic)	0.121766		

Heteroskedasticity Test: ARCH

F-statistic	14.21581	Prob. F(1,997)	0.0002
Obs*R-squared	14.04408	Prob. Chi-Square(1)	0.0002

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.045195	0.002908	15.54227	0.0000
RESID^2(-1)	0.115490	0.030631	3.770386	0.0002

R-squared	0.014058	Mean dependent var	0.051177
Adjusted R-squared	0.013069	S.D. dependent var	0.077531
S.E. of regression	0.077022	Akaike info criterion	-2.287441
Sum squared resid	5.914651	Schwarz criterion	-2.277618
Log likelihood	1144.577	Hannan-Quinn criter.	-2.283707
F-statistic	14.21581	Durbin-Watson stat	2.066312
Prob(F-statistic)	0.000173		

Heteroskedasticity Test: White

F-statistic	1.428020	Prob. F(27,972)	0.0733
Obs*R-squared	38.15376	Prob. Chi-Square(27)	0.0755
Scaled explained SS	44.43210	Prob. Chi-Square(27)	0.0187

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.087277	0.285855	-0.305319	0.7602
LRV_OIL(-1)^2	0.070287	0.045516	1.544243	0.1229
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),5)	-0.047790	0.126533	-0.377688	0.7057
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),22)	-0.120119	0.091293	-1.315749	0.1886
LRV_OIL(-1)*LTYVIX(-1)	0.305644	0.257726	1.185926	0.2359
LRV_OIL(-1)*@MOVAV(LTYVIX(-1),5)	-0.361583	0.335278	-1.078457	0.2811
LRV_OIL(-1)*@MOVAV(LTYVIX(-1),22)	0.180074	0.198726	0.906141	0.3651
LRV_OIL(-1)	-0.108145	0.129670	-0.834001	0.4045
@MOVAV(LRV_OIL(-1),5)^2	-0.015056	0.115262	-0.130624	0.8961
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LRV_OIL(-1),22)	0.112562	0.173829	0.647543	0.5174
@MOVAV(LRV_OIL(-1),5)*LTYVIX(-1)	-0.149456	0.452457	-0.330321	0.7412
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LTYVIX(-1),5)	-0.145186	0.597977	-0.242796	0.8082
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LTYVIX(-1),22)	0.312459	0.351617	0.888633	0.3744
@MOVAV(LRV_OIL(-1),5)	-0.186596	0.212036	-0.880023	0.3791
@MOVAV(LRV_OIL(-1),22)^2	0.009181	0.091571	0.100256	0.9202
@MOVAV(LRV_OIL(-1),22)*LTYVIX(-1)	-0.072171	0.407321	-0.177185	0.8594
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LTYVIX(-1),5)	0.598742	0.549828	1.088962	0.2764
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LTYVIX(-1),22)	-0.718312	0.338650	-2.121105	0.0342
@MOVAV(LRV_OIL(-1),22)	0.320756	0.190344	1.685141	0.0923
LTYVIX(-1)^2	0.254285	0.607492	0.418582	0.6756
LTYVIX(-1)*@MOVAV(LTYVIX(-1),5)	-1.393514	1.569011	-0.888148	0.3747
LTYVIX(-1)*@MOVAV(LTYVIX(-1),22)	0.596331	0.718608	0.829842	0.4068
LTYVIX(-1)	0.302211	0.578377	0.522516	0.6014
@MOVAV(LTYVIX(-1),5)^2	1.434383	1.152219	1.244888	0.2135
@MOVAV(LTYVIX(-1),5)*@MOVAV(LTYVIX(-1),22)	-0.844844	1.053704	-0.801786	0.4229
@MOVAV(LTYVIX(-1),5)	-1.420618	0.774525	-1.834180	0.0669
@MOVAV(LTYVIX(-1),22)^2	-0.020466	0.359203	-0.056977	0.9546
@MOVAV(LTYVIX(-1),22)	1.208195	0.513604	2.352385	0.0189

R-squared	0.038154	Mean dependent var	0.051748
Adjusted R-squared	0.011436	S.D. dependent var	0.079571
S.E. of regression	0.079115	Akaike info criterion	-2.208227
Sum squared resid	6.083930	Schwarz criterion	-2.070810
Log likelihood	1132.114	Hannan-Quinn criter.	-2.155999
F-statistic	1.428020	Durbin-Watson stat	1.911865
Prob(F-statistic)	0.073318		

Table 10: Results of White Test, ARCH Test and LM Test for HAR-X-3

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.191506	Prob. F(2,991)	0.8257
Obs*R-squared	0.386341	Prob. Chi-Square(2)	0.8243

Dependent Variable: RESID
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001011	0.070697	0.014303	0.9886
LRV_OIL(-1)	-0.046513	0.142389	-0.326658	0.7440
@MOVAV(LRV_OIL(-1),5)	0.027766	0.088612	0.313339	0.7541
@MOVAV(LRV_OIL(-1),22)	0.013360	0.104637	0.127677	0.8984
LOVX(-1)	0.038269	0.164856	0.232138	0.8165
@MOVAV(LOVX(-1),5)	-0.020242	0.198991	-0.101721	0.9190
@MOVAV(LOVX(-1),22)	-0.013524	0.142545	-0.094874	0.9244
RESID(-1)	0.044046	0.127314	0.345967	0.7294
RESID(-2)	-0.008469	0.044593	-0.189914	0.8494

R-squared	0.000386	Mean dependent var	-8.72E-16
Adjusted R-squared	-0.007683	S.D. dependent var	0.217167
S.E. of regression	0.218000	Akaike info criterion	-0.199684
Sum squared resid	47.09629	Schwarz criterion	-0.155514
Log likelihood	108.8420	Hannan-Quinn criter.	-0.182896
F-statistic	0.047876	Durbin-Watson stat	1.986868
Prob(F-statistic)	0.999951		

Heteroskedasticity Test: ARCH

F-statistic	3.544348	Prob. F(1,997)	0.0600
Obs*R-squared	3.538877	Prob. Chi-Square(1)	0.0599

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.043813	0.002676	16.37137	0.0000
RESID^2(-1)	0.057634	0.030613	1.882644	0.0600

R-squared	0.003542	Mean dependent var	0.046531
Adjusted R-squared	0.002543	S.D. dependent var	0.071313
S.E. of regression	0.071222	Akaike info criterion	-2.444021
Sum squared resid	5.057399	Schwarz criterion	-2.434198
Log likelihood	1222.789	Hannan-Quinn criter.	-2.440288
F-statistic	3.544348	Durbin-Watson stat	2.055090
Prob(F-statistic)	0.060040		

Heteroskedasticity Test: White

F-statistic	1.172772	Prob. F(27,972)	0.2488
Obs*R-squared	31.54923	Prob. Chi-Square(27)	0.2492
Scaled explained SS	37.94382	Prob. Chi-Square(27)	0.0788

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.218169	0.267647	-0.815134	0.4152
LRV_OIL(-1)^2	-0.004978	0.049683	-0.100188	0.9202
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),5)	0.013640	0.154436	0.088322	0.9296
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),22)	0.023421	0.149086	0.157096	0.8752
LRV_OIL(-1)*LOVX(-1)	0.156225	0.243518	0.641536	0.5213
LRV_OIL(-1)*@MOVAV(LOVX(-1),5)	-0.101877	0.353022	-0.288586	0.7730
LRV_OIL(-1)*@MOVAV(LOVX(-1),22)	-0.070812	0.224466	-0.315467	0.7525
LRV_OIL(-1)	-0.026399	0.136640	-0.193202	0.8468
@MOVAV(LRV_OIL(-1),5)^2	0.070944	0.176675	0.401549	0.6881
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LRV_OIL(-1),22)	-0.369740	0.332178	-1.113075	0.2660
@MOVAV(LRV_OIL(-1),5)*LOVX(-1)	-0.414350	0.512857	-0.807925	0.4193
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LOVX(-1),5)	0.338215	0.754289	0.448390	0.6540
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LOVX(-1),22)	0.304654	0.485891	0.627001	0.5308
@MOVAV(LRV_OIL(-1),5)	-0.149798	0.278360	-0.538146	0.5906
@MOVAV(LRV_OIL(-1),22)^2	-0.144424	0.252003	-0.573106	0.5667
@MOVAV(LRV_OIL(-1),22)*LOVX(-1)	0.316044	0.564232	0.560131	0.5755
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LOVX(-1),5)	0.475201	0.782794	0.607058	0.5440
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LOVX(-1),22)	-0.127868	0.602039	-0.212392	0.8318
@MOVAV(LRV_OIL(-1),22)	-0.294643	0.328070	-0.898110	0.3693
LOVX(-1)^2	0.946838	0.429885	2.202540	0.0279
LOVX(-1)*@MOVAV(LOVX(-1),5)	-2.057520	1.298411	-1.584644	0.1134
LOVX(-1)*@MOVAV(LOVX(-1),22)	0.093653	0.728931	0.128480	0.8978
LOVX(-1)	0.034494	0.478379	0.072107	0.9425
@MOVAV(LOVX(-1),5)^2	0.559402	1.098516	0.509234	0.6107
@MOVAV(LOVX(-1),5)*@MOVAV(LOVX(-1),22)	0.488669	1.217868	0.401249	0.6883
@MOVAV(LOVX(-1),5)	-0.608074	0.656488	-0.926253	0.3545
@MOVAV(LOVX(-1),22)^2	-0.507735	0.502787	-1.009841	0.3128
@MOVAV(LOVX(-1),22)	1.151651	0.491222	2.344462	0.0193

R-squared	0.031549	Mean dependent var	0.047114
Adjusted R-squared	0.004648	S.D. dependent var	0.073623
S.E. of regression	0.073452	Akaike info criterion	-2.356776
Sum squared resid	5.244092	Schwarz criterion	-2.219359
Log likelihood	1206.388	Hannan-Quinn criter.	-2.304548
F-statistic	1.172772	Durbin-Watson stat	1.891669
Prob(F-statistic)	0.248823		

Table 11: Results of White Test, ARCH Test and LM Test for HAR-X-4

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.178705	Prob. F(2,991)	0.3081
Obs*R-squared	2.373174	Prob. Chi-Square(2)	0.3053

Dependent Variable: RESID
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.038702	0.100857	-0.383735	0.7013
LRV_OIL(-1)	0.373594	0.253153	1.475763	0.1403
@MOVAV(LRV_OIL(-1),5)	-0.181570	0.134956	-1.345408	0.1788
@MOVAV(LRV_OIL(-1),22)	-0.178905	0.137388	-1.302181	0.1932
LGVZ(-1)	-0.080380	0.137175	-0.585969	0.5580
@MOVAV(LGVZ(-1),5)	-0.033134	0.168407	-0.196748	0.8441
@MOVAV(LGVZ(-1),22)	0.113055	0.124802	0.905874	0.3652
RESID(-1)	-0.335053	0.223496	-1.499149	0.1342
RESID(-2)	-0.079630	0.069946	-1.138447	0.2552

R-squared	0.002373	Mean dependent var	-4.62E-17
Adjusted R-squared	-0.005680	S.D. dependent var	0.225514
S.E. of regression	0.226153	Akaike info criterion	-0.126250
Sum squared resid	50.68492	Schwarz criterion	-0.082080
Log likelihood	72.12484	Hannan-Quinn criter.	-0.109462
F-statistic	0.294676	Durbin-Watson stat	1.987913
Prob(F-statistic)	0.967867		

Heteroskedasticity Test: ARCH

F-statistic	8.954405	Prob. F(1,997)	0.0028
Obs*R-squared	8.892502	Prob. Chi-Square(1)	0.0029

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.045588	0.002867	15.90275	0.0000
RESID^2(-1)	0.092146	0.030793	2.992391	0.0028

R-squared	0.008901	Mean dependent var	0.050274
Adjusted R-squared	0.007907	S.D. dependent var	0.076197
S.E. of regression	0.075895	Akaike info criterion	-2.316929
Sum squared resid	5.742790	Schwarz criterion	-2.307105
Log likelihood	1159.306	Hannan-Quinn criter.	-2.313195
F-statistic	8.954405	Durbin-Watson stat	2.056549
Prob(F-statistic)	0.002836		

Heteroskedasticity Test: White

F-statistic	1.657256	Prob. F(27,972)	0.0192
Obs*R-squared	44.00894	Prob. Chi-Square(27)	0.0206
Scaled explained SS	51.07991	Prob. Chi-Square(27)	0.0034

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.452580	0.378979	-1.194207	0.2327
LRV_OIL(-1)^2	0.054761	0.046030	1.189700	0.2345
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),5)	-0.167181	0.132957	-1.257405	0.2089
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),22)	0.053664	0.097794	0.548748	0.5833
LRV_OIL(-1)*LGVZ(-1)	0.334452	0.253335	1.320197	0.1871
LRV_OIL(-1)*@MOVAV(LGVZ(-1),5)	0.073628	0.313606	0.234777	0.8144
LRV_OIL(-1)*@MOVAV(LGVZ(-1),22)	-0.432006	0.184768	-2.338098	0.0196
LRV_OIL(-1)	0.100273	0.196509	0.510271	0.6100
@MOVAV(LRV_OIL(-1),5)^2	0.068203	0.135443	0.503558	0.6147
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LRV_OIL(-1),22)	0.062260	0.207790	0.299630	0.7645
@MOVAV(LRV_OIL(-1),5)*LGVZ(-1)	0.373505	0.398291	0.937769	0.3486
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LGVZ(-1),5)	-0.817808	0.519572	-1.574002	0.1158
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LGVZ(-1),22)	0.520434	0.343952	1.513100	0.1306
@MOVAV(LRV_OIL(-1),5)	-0.369279	0.348993	-1.058126	0.2903
@MOVAV(LRV_OIL(-1),22)^2	-0.065844	0.103792	-0.634386	0.5260
@MOVAV(LRV_OIL(-1),22)*LGVZ(-1)	-0.830711	0.357808	-2.321666	0.0205
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LGVZ(-1),5)	1.161295	0.457651	2.537512	0.0113
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LGVZ(-1),22)	-0.351092	0.305899	-1.147737	0.2514
@MOVAV(LRV_OIL(-1),22)	0.150583	0.316036	0.476476	0.6338
LGVZ(-1)^2	-0.651045	0.352669	-1.846050	0.0652
LGVZ(-1)*@MOVAV(LGVZ(-1),5)	0.963498	1.015040	0.949221	0.3427
LGVZ(-1)*@MOVAV(LGVZ(-1),22)	0.411027	0.527435	0.779295	0.4360
LGVZ(-1)	0.172920	0.608038	0.284391	0.7762
@MOVAV(LGVZ(-1),5)^2	-0.151187	0.722764	-0.209179	0.8344
@MOVAV(LGVZ(-1),5)*@MOVAV(LGVZ(-1),22)	-0.780072	0.695507	-1.121589	0.2623
@MOVAV(LGVZ(-1),5)	-0.857540	0.810404	-1.058163	0.2902
@MOVAV(LGVZ(-1),22)^2	0.114217	0.286685	0.398406	0.6904
@MOVAV(LGVZ(-1),22)	1.139484	0.551347	2.066729	0.0390

R-squared	0.044009	Mean dependent var	0.050805
Adjusted R-squared	0.017454	S.D. dependent var	0.077992
S.E. of regression	0.077308	Akaike info criterion	-2.254436
Sum squared resid	5.809194	Schwarz criterion	-2.117019
Log likelihood	1155.218	Hannan-Quinn criter.	-2.202208
F-statistic	1.657256	Durbin-Watson stat	1.902721
Prob(F-statistic)	0.019224		

Table 12: Results of White Test, ARCH Test and LM Test for HAR-X-5

Breusch-Godfrey Serial Correlation LM Test results:

F-statistic	2.219272	Prob. F(2,991)	0.1092
Obs*R-squared	4.458884	Prob. Chi-Square(2)	0.1076

Dependent Variable: RESID
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.080229	0.071672	-1.119394	0.2632
LRV_OIL(-1)	0.660269	0.322300	2.048615	0.0408
@MOVAV(LRV_OIL(-1),5)	-0.318816	0.164584	-1.937108	0.0530
@MOVAV(LRV_OIL(-1),22)	-0.297587	0.159769	-1.862608	0.0628
LEUVIX(-1)	-0.095717	0.169057	-0.566185	0.5714
@MOVAV(LEUVIX(-1),5)	-0.162056	0.229784	-0.705254	0.4808
@MOVAV(LEUVIX(-1),22)	0.234673	0.164550	1.426156	0.1541
RESID(-1)	-0.590705	0.285494	-2.069065	0.0388
RESID(-2)	-0.152251	0.087418	-1.741658	0.0819

R-squared	0.004459	Mean dependent var	-4.09E-16
Adjusted R-squared	-0.003578	S.D. dependent var	0.226432
S.E. of regression	0.226837	Akaike info criterion	-0.120212
Sum squared resid	50.99185	Schwarz criterion	-0.076043
Log likelihood	69.10620	Hannan-Quinn criter.	-0.103425
F-statistic	0.554818	Durbin-Watson stat	1.989572
Prob(F-statistic)	0.815205		

Heteroskedasticity Test: ARCH

F-statistic	11.08215	Prob. F(1,997)	0.0009
Obs*R-squared	10.98231	Prob. Chi-Square(1)	0.0009

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 999 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.045437	0.002895	15.69343	0.0000
RESID^2(-1)	0.102477	0.030783	3.328986	0.0009

R-squared	0.010993	Mean dependent var	0.050691
Adjusted R-squared	0.010001	S.D. dependent var	0.077106
S.E. of regression	0.076719	Akaike info criterion	-2.295322
Sum squared resid	5.868222	Schwarz criterion	-2.285499
Log likelihood	1148.513	Hannan-Quinn criter.	-2.291588
F-statistic	11.08215	Durbin-Watson stat	2.057589
Prob(F-statistic)	0.000904		

Heteroskedasticity Test: White

F-statistic	1.473458	Prob. F(27,972)	0.0571
Obs*R-squared	39.32005	Prob. Chi-Square(27)	0.0592
Scaled explained SS	45.91430	Prob. Chi-Square(27)	0.0130

Dependent Variable: RESID^2
Method: Least Squares
Included observations: 1000

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.082406	0.190090	-0.433508	0.6647
LRV_OIL(-1)^2	0.044687	0.046537	0.960259	0.3372
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),5)	-0.027352	0.127504	-0.214519	0.8302
LRV_OIL(-1)*@MOVAV(LRV_OIL(-1),22)	-0.111518	0.098536	-1.131748	0.2580
LRV_OIL(-1)*LEUVIX(-1)	0.560157	0.321166	1.744136	0.0815
LRV_OIL(-1)*@MOVAV(LEUVIX(-1),5)	-0.474154	0.406417	-1.166667	0.2436
LRV_OIL(-1)*@MOVAV(LEUVIX(-1),22)	-0.013878	0.209485	-0.066246	0.9472
LRV_OIL(-1)	0.008930	0.112822	0.079156	0.9369
@MOVAV(LRV_OIL(-1),5)^2	0.102853	0.121036	0.849769	0.3957
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LRV_OIL(-1),22)	-0.127788	0.191113	-0.668654	0.5039
@MOVAV(LRV_OIL(-1),5)*LEUVIX(-1)	-0.299596	0.541666	-0.553101	0.5803
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LEUVIX(-1),5)	-0.748645	0.719035	-1.041180	0.2981
@MOVAV(LRV_OIL(-1),5)*@MOVAV(LEUVIX(-1),22)	1.112175	0.407591	2.728657	0.0065
@MOVAV(LRV_OIL(-1),5)	-0.352576	0.197709	-1.783305	0.0748
@MOVAV(LRV_OIL(-1),22)^2	0.129441	0.118591	1.091492	0.2753
@MOVAV(LRV_OIL(-1),22)*LEUVIX(-1)	-0.364563	0.465910	-0.782474	0.4341
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LEUVIX(-1),5)	1.496444	0.591777	2.528728	0.0116
@MOVAV(LRV_OIL(-1),22)*@MOVAV(LEUVIX(-1),22)	-1.317513	0.363946	-3.620077	0.0003
@MOVAV(LRV_OIL(-1),22)	0.396137	0.203541	1.946221	0.0519
LEUVIX(-1)^2	-0.496870	0.644623	-0.770791	0.4410
LEUVIX(-1)*@MOVAV(LEUVIX(-1),5)	0.761055	1.897364	0.401112	0.6884
LEUVIX(-1)*@MOVAV(LEUVIX(-1),22)	0.295018	1.019761	0.289301	0.7724
LEUVIX(-1)	0.153373	0.455360	0.336817	0.7363
@MOVAV(LEUVIX(-1),5)^2	-0.021102	1.559967	-0.013527	0.9892
@MOVAV(LEUVIX(-1),5)*@MOVAV(LEUVIX(-1),22)	-0.849462	1.524648	-0.557152	0.5776
@MOVAV(LEUVIX(-1),5)	-0.450738	0.588382	-0.766063	0.4438
@MOVAV(LEUVIX(-1),22)^2	0.334949	0.456108	0.734362	0.4629
@MOVAV(LEUVIX(-1),22)	0.340337	0.315489	1.078763	0.2810

R-squared	0.039320	Mean dependent var	0.051220
Adjusted R-squared	0.012634	S.D. dependent var	0.078866
S.E. of regression	0.078367	Akaike info criterion	-2.227240
Sum squared resid	5.969352	Schwarz criterion	-2.089823
Log likelihood	1141.620	Hannan-Quinn criter.	-2.175012
F-statistic	1.473458	Durbin-Watson stat	1.932752
Prob(F-statistic)	0.057150		