1. Introduction

Crude oil is considered one of the most important commodities since it constitutes a decisive factor in the configuration of prices of all the other commodities while its price fluctuation is an indication and also a cause of important changes in global economies. The rise, the stability or the decline of crude oil prices have a direct impact in the economies of various states but also in the more general international economy.

Both academics and practitioners try to investigate the factors that affect crude oil prices and to develop predictive models for crude oil pricing. A better forecast of the expected crude oil price helps market participants to improve their plans and decisions. According to Kaboudan (2001) oil prices follow cyclical patterns over time. They tend to escalate for an extended period. Reverse direction then perhaps escalate again. Periodicity is not constant and variations within an escalating or a decreasing period are typical. Also, global demand for petroleum products is highly seasonal and it is higher during the winter months, when countries increase their use of distillate heating oil and residual fuels. Supply of crude oil, including both production and net imports, also shows a similar seasonal variation.

In this paper, we try to investigate the factors that affect crude oil price for the period 1988 – 2008 comparing a linear and a non-linear approach. First we use wavelet analysis to extract the driving forces and dynamics of the crude oil price and returns processes. Wavelets analysis has been proved to be a valuable tool for analyzing a wide range of time-series and they have already been used with success in image processing, signal de-noising, density estimation, signal and image compression and time-scale decomposition (Daubechies, 1992, Mallat, 1999, Wojtaszczyk, 1997). Wavelet techniques are being used in finance, for detecting the properties of quick variation of values, Zapranis & Alexandridis (2008a). Wavelet decomposition is considered a powerful tool for approximation.

Moreover we examine if a wavelet neural network estimator can provide some incremental value in understanding crude oil price process. Wavelet networks proposed by Zhang & Benveniste (1992) as an alternative to feedforward neural networks. Wavelet networks are one hidden layer networks that use a wavelet as an activation function instead of the classic sigmoid function. The activation function can be a wavenet (orthogonal wavelets) or a wave frame (continuous wavelets). Wavelet networks are performing excellent in predicting nonlinear behaviours, Gao & Tsoukalas (2001). Wavelets show local characteristics hence the hidden units of the wavelet network affect the prediction of the network only in a local range, Postalcioglu & Becerikli, (2007).

Wavelet networks have been used in a variety of applications so far. They first have been used in static and dynamic input-output modelling Zhang & Benveniste (1992), Postalcioglu & Becerikli, (2007) and proved that wavelet networks need less training iterations. Szu *et al*. (1992) used for classification of phonemes and speaker recognition. In Gao & Tsoukalas (2001) wavelet networks considered one of the most promising tools to solve electricity load prediction problems. In Subasi *et al.* (2005) wavelet networks used for classification of electroencephalography (EEG) signals while Khayamian *et al*. (2005) used wavelet networks as a multivariate calibration method for simultaneous determination of test samples of copper, iron and aluminum. Finally, Zapranis & Alexandridis (2008b) used wavelet neural network to forecast cash money withdrawals.

The rest of the paper is organized as follows: In the next section we give a review of the relevant literature on crude oil prices behaviour and forecasting techniques. In section 3, we present our methodology. More precisely, in section 3.1 we describe the available data. In section 3.2, we present the results for the linear approach. In section 3.3 wavelet analysis is introduced and applied to crude oil prices and returns. In section 3.4 and 3.5 the linear model is improved using the results from wavelet analysis. In section 3.6 a wavelet neural network is introduced. Moreover, we give our forecasting framework and we explain how a wavelet neural network can be used for predicting future returns. In section 3.7 a combination of wavelet analysis and a wavelet network is used. In section 4 a scenario and stretch analysis is performed to the significant predictors. Finally in section 5 we conclude.

2. Literature Review

In this section a review of the relevant literature is presented with focus on forecasting models for crude oil prices. Forecasting approaches can be separated in single-factor and multi-factor models. In the first group, future predictions of oil prices produced based on the lagged oil prices while in the second one, future predictions of oil prices produced based on correlated to the oil price variables such as consumption, supply, inventories or financial indexes. Early works use different models of the GARCH family. Moosa and Al-Loughani (1994) investigated the relationship between spot and future WTI oil prices using various econometric tests and a GARCH-M(1,1) model. They conclude that future prices are neither unbiased nor efficient forecasters of spot oil prices. Sadorsky (1999) initially investigated the relationship between oil price movements and macroeconomic and financial variables. Using GARCH and vector autoregression models with monthly data concluded that while changes in the oil prices have important influence in economy, economic activity have little impact in oil prices. In addition shocks in oil price volatility have asymmetric effects on the economy. In a latter paper, Sadorsky (2006) compared several univariate and multivariate statistical models (such as GARCH, TGARCH, AR, random walk, historical mean, moving average, VAR, BIGARCH etc.) to forecast daily volatility of oil futures price returns. The out-of-sample performance of the above models was estimated with mean squared error (MSE), mean absolute deviation (MAD) and the Theil U statistic. Sandorsky's results indicate that the majority of the models outperform the simple random walk model with GARCH model to perform satisfactory. Panas and Ninni (2000) examined the existence of chaotic structure for oil products in the Rotterdam and Mediterranean petroleum markets. They found strong evidence of chaos in a number of oil products. Also, Adragni *et al.* (2001) investigated for chaotic structure in oil future prices. They found highly non-linearities which can be explained by ARCH-type models.

Kaboudan (2001) investigated the monthly crude oil price forecasting performance of three methods: genetic programming, neural networks and random walk. Specifically, the explanatory power of the following variables is examined: monthly world crude production, OECD consumption, world crude oil stocks, monthly changes in known US stocks and lagged FOB crude oil prices of US imports. Kaboudan's results suggest that genetic programming outperforms both random walk and neural networks. Morana (2001) following the work of Barone-Adesi *et al.* (1998) used a semiparametric approach to forecast Brent oil prices. The approach was based on the estimation of a GARCH model and a historical simulation of the residuals in order to forecast the variance for period $T+1$. In the same work the reliability of the one-month forward price as a forecast indicator for the oil price was investigated using confidence intervals. The results indicate that the wider the intervals are the least reliable is the one-month forward price. Tang and Hammoudeh (2002) suggest that nonlinear approaches based on the Target Zone Theory can improve oil price forecasts. They developed their model using the average monthly OPEC basket price. Bernabe *et al.* (2004) proposed a stochastic multi-model approach to describe the mechanisms of oil price determination which incorporates multiple equilibria information.

Yousefi *et al.* (2005) investigated the usefulness of wavelets to crude oil price forecasting. They used averaged monthly WTI spot prices and estimated forecast for 1, 2, 3, and 4 months ahead. Also, NYMEX oil future prices were used in order to evaluate the efficiently of future markets. Wang *et al.* (2005) proposed a non-linear method that uses text mining, econometrics and intelligent algorithms (TEI@I) to forecast crude oil prices. This method integrates six modules. According to their results their proposed methodology seems to outperform the individual ANN and ARIMA models in terms of root mean square error (RMSE). Ye *et al.* (2005) proposed a simple regression model to forecast monthly West Texas Intermediate crude oil price using OECD petroleum inventories. They also investigated the effects on monthly crude oil price of various changes to inventories, oil supply and demand.

Rehrl and Friedrich (2006), developed the LOPEX (Long-term Oil Price and EXtraction) model in order to forecast world oil prices and supply up to the year 2100. For this reason they performed oil price scenarios with various resource bases. However, the performance of the model is limited to the underlying assumptions. Postali and Piccheti (2006) investigated the behaviour of oil prices. Specifically, they performed various tests (unit root tests, *J*-statistic, etc.) in order to evaluate the advantages and disadvantages of various stochastic processes using annual oil prices. Postali and Piccheti argue that even a simple Geometric Brownian Motion can be used as an approximation for the movement of international oil prices with hearteningly results. Dees *et al.* (2007) proposed an econometric model for the investigation of the relation between oil prices, oil demand, oil supply (from OPEC and non-OPEC countries) and economic activity. They conclude that OPEC behaviour and the mechanisms of oil price determination can be explained by their model. Amin-Naseri & Gharacheh (2007) proposed an artificial intelligence model using feed-forward neural networks, genetic algorithm and *k*-means clustering. The model was applied for monthly WTI crude oil price forecasting. According to their results, this model outperforms forecasts provided by the econometric model of the STEO but as well as forecasts provided by previous works. Xie *et al.* (2006) used support vector machines for crude oil price forecasting. They compared their proposed model with ARIMA and neural network models and remarked that support vector machines outperform the above models. Shambora and Rossiter (2007) developed an artificial neural network model based on moving average crossover inputs to predict future oil prices. Comparing the performance of the ANN model they conclude that the ANN outperforms the buy-and-hold, the twenty-day average and the random walk models.

Gori *et al.* (2007) studied the relationship between oil prices and consumption under three scenarios of oil price behaviour: parabolic, linear and chaotic. Yu *et al.* (2007) proposed a multiscale neural network learning paradigm based on empirical mode decomposition (EMD). Their model performed well at forecasting WTI oil prices. Askari & Krichene (2008) investigated the oil prices dynamics during 2002–2006.

Specifically, they modelled oil prices as a jump–diffusion process and as a Levy process with a variance-gamma distribution. Fan *et al.* (2008) proposed a generalized pattern matching based on genetic algorithm (GPMGA) for crude oil price forecasting. Their method was evaluated using monthly Brent and WTI oil prices and gave promising results. Zhang *et al.* (2008) used Ensemble Empirical Mode Decomposition (EEMD) to WTI crude oil price analysis and forecasting. They decomposed the crude oil price time-series into independent intrinsic modes and tried to give economical meaning to each mode.

3. Methodology

3.1 Data Description

In this section the available dataset is described and the main statistics of the crude oil prices and returns are presented. In addition we describe the necessary transformations in the explanatory variables. All data for this study were obtained from the US Department of Energy except from the Euro/Dollar exchange rate and the open interest for the crude oil contracts which were obtained from the Π-trading website[4]. The variable to forecast is West Texas Intermediate (WT) crude oil price. More precisely the dependent variable is the monthly crude oil log-returns. The series used start at January 1988 and taken at monthly closing price.
As extensively presented in the previous section various models proposed in literature in order to forecast crude oil prices or returns. Early models were based on explanatory variables such as production and consumption. In the late 90s mean reverting models were proposed were only price lags were considered. Recently practitioners include in their models variables such as the petroleum stocks held by OECD and OPEC as well as the Euro/Dollar exchange rate while others use variables such as the S&P 500 index, the AMEX index or the consumer, producer and the industry price indexes. On the other hand the weak dollar, the limited crude oil supply and mainly the increasing Chinese demand considered as the major factors that can explain the increase in energy prices. Also, it is believed that the increasing number of traded future contracts and the open interest push prices further.
In this study we examine the impact of 17 different explanatory variables as well as price lags up to four months on crude oil prices and returns. Many of the variables were proposed in previous studies while others are used for the first time. In appendix A all explanatory variables as well as the crude oil prices can be found. For simplicity we present the abbreviation of each variable in appendix A. The aim of this study is to find which of the above variables proposed in bibliography are significant in crude oil returns forecasting and present the time evolution of the impact of each variable. One must observe that some variables cannot be used together due to linear combination restrictions e.g. the world production equals the OPEC and the non-OPEC production.
All variables were tested for unit root using the Phillip-Perron test. Table 2 shows that non-stationarity in the mean was a common problem among the original variables. These variables were detrended by taking first differences. In the remaining four variables (*COECD*, *PUSA, SOECD, SUSA)* a linear trend is clear. The detrended series exhibit strong autocorrelation and the Durbin-Watson static is always less than 1 for all four variables indicating strong positive serial correlation in the transformed

[4] Π-trading website: http://www.pitrading.com/free_market_data.htm

variable. To overcome this, the original variables were transformed by taking the log-returns. Appendix B shows the transformed variables. The letter *d* in the new variables names indicates that the original variable transformed using the 1st differences method while the letters *dl* indicate that the original variable transformed by taking the log-returns. In Table 3a and Table 3b the descriptive statistics for the transformed and the original variables can be found.

3.2 The Linear Case

In order to obtain a better insight in crude oil dynamics, in this section we try to fit a linear model to crude oil price returns. Today's available data were used to forecast one month ahead crude oil price returns. Our aim is to find a parsimonious model while minimizing the variables needed and maximizing the explained variability of the dependent variable. To do so the stepwise selection method is used. Initially no predictors were included in the model. The coefficients and the corresponding *p*-values for all predictors are calculated as if it were the last variable to enter the model. Then the variable with the smallest *p*-value is included in the model if it is lower than a threshold equal to 0.1. The new model is re-estimated and new *p*-values are computed. If a new *p*-value of an already included variable is greater than 0.1 then this variable is removed.

First we use all data range from 1/1988 to 1/2008. Although early models propose GARCH models (Sadorky, 1999, Morana, 2001) surprisingly a simple least square method performs better than alternative methods like the ARCH/GARCH family models. This is evidence that the dynamics of the crude oil prices have drastically changed the last years. Following the stepwise selection procedure the reduced model has only six variables: the AMEX, Consumer Price, Producer, Industry and S&P 500 indexes as well as the stocks held by the OECD. The corresponding *p*-values are less than 0.05 suggesting strong significance. The variables selected by the stepwise method are indexes that depend on oil prices and not vice versa, like the AMEX and S&P 500 indexes. One might argue that the indexes incorporate the needed information from other variables however the information they provide is already included in the oil price and therefore cannot be used satisfactory for forecasting. This is shown in Figure 1a where the R^2=18.09% and the $\overline{R}^2$=15.97% indicating that the fit is not good as expected. The Position of Sign (*POS*) is only 64.85% indicating that the linear model cannot predict if the returns are positive or negative. The Percentage of Change in Direction (*POCID*) is 81% which is relative high but the Independent *POCID* (*IPOCID*) is only 62.18% implying that the model cannot learn the change in direction of the crude oil price returns. The maximum absolute error (Max AE) is 0.329 and the mean absolute error (MAE) is 0.0538. Moreover the MSE is 0.004878. The linear regression coefficients as well as their respective *p*-values can be found in Table 4.

In order to improve our findings we check our data for influential values. The leverage value for each observation is calculated and five observations exceeded the cut-off point. Table 5 shows the observations that must be removed and their corresponding leverage values.

The new regression on the reduced sample has a better fit as expected with R^2 =18.49% and $\overline{R}^2$ =16.34%, however the fit is still not good. Next, stepwise selection is used to the reduced sample. This time one additional variable is selected and the overall fit is marginally improved. The new significant predictor is the OECD stocks

(*dlsoecd*) and the new R^2=19.77% and the $\overline{R}^2$=17.28%. In contrast the *IPOCID* and *POS* criteria reduced to 61.80% and 61.97% respectively.
It is clear that the linear model cannot capture the dynamics of the crude oil price returns. Although removing the influential values and re-selecting new predictors seems to work the overall fitting is not good. In all three methods the normalized mean square error (NMSE) is over 0.80. In the next section a non-standard method is used in order to improve our results.

3.3 Wavelet Analysis.

In order to obtain a better insight of the dynamics of the West Texas Intermediate crude oil prices and returns wavelet analysis was used. Wavelet analysis is a mathematical tool used in various areas of research. Especially, during the last years wavelets are frequently used in order to analyse time-series, data and images. Time-series are represented by local information such as frequency, duration, intensity and time-position and by global information such as the mean states over different time periods. Both global and local information is needed for a correct analysis of a signal. The Wavelet transform is a generalization of Fourier and windowed Fourier transforms. A wavelet is a waveform of effectively limited duration that has an average value of zero. A wavelet family is a set of orthogonal basis functions generated by dilation and translation of a compactly supported scaling function, ϕ (or father wavelet), and a wavelet function, ψ (or mother wavelet). The wavelet family consists of wavelet children which are dilated and translated forms of a mother wavelet:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)$$

where, a is the scale or dilation parameter and b is the shift or translation parameter. Works as Daubechies (1992) and Mallat (1999) give concise treatment of wavelet theory. In this study we use the continuous wavelet transform and the Daubechies wavelet family at level 3. The top part of Figure 2 shows the wavelet decomposition of oil prices while the bottom part of Figure 2 shows the wavelet decomposition of the oil price returns.
A closer inspection of the lower part of Figure 2 reveals four interesting time periods. More precisely there are four breaks. The first is at point 33 and corresponds to 10/1990. Also there is a clear bright area around the break, ranging from 31 (01/07/1990) to 36 (01/02/1991). The break is one month after the start of the Gulf War while point 36 corresponds to the end of the war. In other words, it is clear that the Gulf War had a strong impact on the crude oil returns. Also a decreasing impact of the war on both prices and returns is clear until point 42 (7/1991).
A large bright area is around the second break at point 132 (1/1999). The impact of different events starting at point 110 (2/1997) with an increasing power, is reflected there. From point 129 (10/1998) to 136 (5/1999) there is strong impact and then the impact is decreasing until 147 (3/2000). In this time period there are a series of events that directly affected crude oil prices like the Iraq disarmament crisis (1997), Kosovo war (1997), Asian financial crisis (1997). According to Sadorsky (2006) events like

the Brazil, Russia and Long Term Capital Management each close to bankruptcy (1998) and Y2K scares (1999) also had an impact on crude oil prices.

The third break has a center at 167 (12/2001) which corresponds to the lower price of crude oil after the 9/11 attack. This period starts at point 156 (1/2001) with an increasing strength of impact until point 165 (10/2001). This time period reflects the weakened US economy and the terrorists attack. The crude oil prices were affected strongly by these events until 171 (3/2002) and then their influence declined until 176 (9/2002).

The last break is at point 230 (3/2007) and begins at 222 (7/2006) and is not over yet. Examining the wavelet decomposition of the oil prices we see a very bright area at the end of the figure representing a strong change in the driving dynamics of the crude oil prices. However a better examination of the wavelet decomposition of the returns reveals that the pick of this event already passed. Comparing past events, we expect that the volatility of returns will return to normal in the next months and probably the price of oil will fall and settle around a mean.

Finally, there is a weak break at point 73 (1/1994) indicating the end of a period of steady decline in crude oil prices.

Obviously, wavelet analysis captured a variety of events that affected crude oil prices from 1988 to 2008. Most of these events examined also by other studies and their influence in crude oil prices were confirmed. Analyzing the returns of crude oil prices better insight was obtained. Although examining the crude oil price results to almost the same amount of information the analysis in first case is more definite. Both the original times series provided limited information about past events. On the other hand wavelet analysis brought out events and breaks of the time series that were not originally visible. Moreover examining the wavelet transform of past events we can make assumptions about the future behavior of oil prices.

3.4 Dummy variables.

In this section dummy variables included to the initial model in order to improve our fitting. The dummy variables were built according to wavelet analysis. Four dummy variables were introduced corresponding to the four breaks discussed in the previous section. Each dummy variable takes the value one if the event has a strong impact on the crude oil price return, 0.5 if the event has a weak impact on the returns and 0 if the event has no impact at all on the returns.

$$d_i = \begin{cases} 0 & \text{no impact} \\ 0.5 & \text{weak impact} \\ 1 & \text{strong impact} \end{cases} \quad i = 1..8$$

Running the regression again we observe that the $\overline{R}^2$ reduced to 15.24%. Moreover the coefficients of the dummy variables have a *p*-value greater than 0.1 indicating that they must be removed which results to the initial model. Also a model with separate dummy variables for each level of impact was tested but the results were similar.

3.5 Wavelet analysis and different time periods.

It is clear that the dynamics of crude oil prices change over time. However, observing the price and returns time series representation at appendixes A and B one cannot say when these changes happen. On the other hand wavelet analysis was able to capture the changes of the driving forces of the crude oil price process. In this section we use our findings from the wavelet decomposition in order to separate our initial sample in eight different time periods. Next we fit a multilinear model in each period expecting not only to have a better fit but also to point out the variables that affect crude oil prices and returns each period the most.

Table 8 briefly presents the eight regressions for each period. The $\overline{R}^2$ is always above 21% while in period 3 and 5 is over 95%. The *IPOCID* of the overall fit was increased to 73.11% from 62.18% of the original regression and both *MSE* and *NMSE* halved. Figure 1b shows the original and the fitted data. It is clear that the in-sample forecasts were improved. Observing the significant variables of each period we verify again that driving forces of the crude oil price returns change over time. Until 2001 the returns were dependent on the production, demand and in petroleum indexes. After 2001 the price of the natural gas and the stocks of petroleum held by USA have an impact on the price of crude oil and consequently on the returns. Finally in the last period is confirmed that the open interest on future contracts drives the oil prices higher.

However, one must be careful while reading the results from Table 8. In periods 2, 5, 6 and 8 the number of significant variables is large in comparison to the numbers of the corresponding observations. To avoid the problem of over-fitting the following criterion must hold:

$$\frac{n}{p} > 5$$

where n is the number of observations and p is the number of parameters. To do so, we split the original data set in three equal subsets of 80 observations each one containing at least one major event. The Gulf War influence is enclosed in the first subset while the turbulent period from 1997-2000 is enclosed in the second one. The last subset contains the 9/11 and the most recent observations. Estimating the new models we obtain the significant variables of each subset. Figure 1c shows the original and the fitted data while Table 9 shows the significant variables and the fit of the three sub-periods as well as the overall fitting. Again, the price of natural gas and the total stocks of OECD become more significant in the last period. At the second period the linear model regards only the non-OPEC production and the stocks held by the OECD as significant variables. As a result only the 12.92% of the variability of the returns is explained in this period. In the first subset only the lagged returns and the AMEX and S&P 500 indexes are considered and the $\overline{R}^2$ =54.80%. The *IPOCID* and *POS* criteria are both over 70% which is relative high but the *NMSE* is 0.6119.

So far we used five different frameworks. Each time different significant variables proposed from the linear model. From the previous results it is obvious that the dynamics of crude oil prices changed many times the last 20 years. The recent years, as more information is available, practitioners connected new explanatory variables to the crude oil prices. As a result, examining each time period we find the variables that affect the most the crude oil price. However, one must be very careful at selecting the different time periods since the variable selection method is very sensitive to the size of the data set. Adding or removing few points in the time-series may result to a complete different set of predictors. Wavelet analysis proved to be a useful tool. Successfully captured and represented the periods where the dynamics of the crude oil prices changed. Wavelet analysis gave information on both the duration and the level of these changes. Finally, overfitting problems can be avoided by using wavelet analysis as a guide and corresponding one parameter to at least five observations.

3.6. The wavelet neural network case.

Economists often report that changes in OPEC's production have a direct impact on price of West Texas Intermediate crude oil price. On the other hand an increase in price leads to small changes in demand. However, these variables were not selected by the linear model in all previous scenarios. The findings presented in the previous section and the bad fit of the linear model lead us to consider non-linear models. In this section we use a wavelet neural network first presented by Zhang & Benveniste (1992) in order to forecast crude oil price returns. Wavelet networks are one hidden layer networks that use a wavelet as an activation function instead of the classic sigmoid function. Wavelet networks not only allow constructive initialization methods but also wavelets show local characteristics hence the hidden units or wavelons of the wavelet network affect the prediction of the network only in a local range. The structure of a single-hidden-layer feedforward wavelet network is given in Figure 3.

The network output is given by the following expression:

$$\hat{y}(\mathbf{x}) = w_{\lambda+1}^{[2]} + \sum_{j=1}^{\lambda} w_j^{[2]} \cdot \Psi_j(\mathbf{x}) + \sum_{i=1}^{m} w_i^{[0]} \cdot x_i$$

In that expression, $\Psi_j(\mathbf{x})$ is a multidimensional wavelet which is constructed by the product of m scalar wavelets, $\mathbf{x}$ is the input vector, m is the number of network inputs, λ is the number of hidden units and w stands for a network weight. The multidimensional wavelets are computed as follows:

$$\Psi_j(\mathbf{x}) = \prod_{i=1}^{m} \psi(z_{ij})$$

where

$$z_{ij} = \frac{x_i - w_{(\xi)ij}^{[1]}}{w_{(\zeta)ij}^{[1]}}$$

In the above expression, $i = 1, \ldots, m$, $j = 1, \ldots, \lambda+1$ and the weights w correspond to the translation ($w^{[1]}_{(\xi)ij}$) and the dilation ($w^{[1]}_{(\zeta)ij}$) factors. The weights of the network were trained to minimize the quadratic cost function:

$$L_n = \frac{1}{2}\sum_{p=1}^{n}\left(y_p - \hat{y}_p\right)^2$$

First we check if wavelet neural networks have additional predictive power in forecasting crude oil returns in regard to linear models. To do so, first we produce one month ahead forecasts for the whole time period. The wavelet network is trained with the same variables as in the case of the linear model, table 4. Using a network with 3 hidden units slightly better results were obtained. Both R^2 and $\overline{R}^2$ increased by 2% as well as all error criteria slightly decreased. More importantly the *IPOCID* increased to 66.80% and the *POS* to 65.69%. The improvement is more significant when three sub-periods considered and a different network is trained for each sub-period. The corresponding R^2 are 70.18%, 15.19% and 37.28% while in the case of the linear model were 57.12%, 12.92% and 40.58%. The corresponding $\overline{R}^2$ are 60.41, 13.86 and 30.29 while in the linear case were 54.80%, 10.69% and 39.02%. In general, in all cases the wavelet network not only outperformed the linear model but also was able to capture and forecast both the volume and the sign of the returns of crude oil prices.
The previous simulations were an indication that a wavelet network can improve our forecasts however the overall performance is still at an average level. In order to improve further our results we repeat the same procedure but this time the significant variables were selected by the wavelet network. First, a network is trained in all available data. As in linear case, the significant variables were selected through an iterative procedure. The relevance of each variable to the model is determined by the *Sensitivity-Based Pruning* (*SBP*) criterion originally proposed by Moody & Utans (1992). The *SBP* quantifies the effect on the empirical loss of a variables replacement by its mean and it is much better suited for the testing the significance of the explanatory variables (Zapranis & Refenes, 1999).
The correct topology of each network is selected by minimizing the prediction risk which is calculated using the v-fold cross-validation criterion. In order to calculate the v-fold cross-validation criterion we generate 50 new smaller samples. Each sample has a size of 98% of the original data set and is created from the original set by random selection without replacement. A different network is trained for each sample and is validated on the corresponding remaining 2%. The out-of-sample mean square error of all samples is calculated and the average is called the prediction risk. The topology with the minimum prediction risk is selected.
Table 10 shows the significant variables as well as various fitting criteria. Also, Figure 4a shows the real and the fitted crude oil price returns. Comparing Tables 4 and 10 we observe that the fitting using a wavelet neural network is better. The R^2 increased to 23.08% from 18.09%. Also all error criteria decreased in the case of a wavelet neural network. Only the *POCID* and *POS* criterions are slightly worse than the case of a linear model. More importantly, wavelet neural networks propose a more realistic model. In contrast to the linear case, the production of USA and non-OPEC countries as well as the consumptions and stocks held by the OECD and USA are used as explanatory variables. Also wavelet networks imply that the price of natural gas and the Euro/Dollar exchange rate contribute to the fluctuations of the crude oil

price returns. The findings in this section are in line with the economists' speculations. As in the linear case, removing the values with high leverage improves the fitting. The R^2 is 24.43% while in the linear model it was only 18.49%. Also the *IPOCID* and *POS* criteria are 63.09% and 63.24% respectively. In order to improve further our findings we repeat the variable selection procedure to the reduced sample. The new selected variables are *damex*, *dcic*, *dcl*, *dcusa*, *deu*, *dimports*, *dlcoecd*, *dlpusa*, *dlsoecd*, *dlsusa*, *dpnopec*, *dsp*, *dwp*, *dzot*, and WT_{t-1}. In this model the open interest in future contracts is significant while the price of the natural gas it is not significant. However the fitting criteria are almost the same. More precisely both R^2 and $\overline{R}^2$ were decreased about 0.01%.

3.7. Sub-period modeling using wavelet neural networks

In this section we use both wavelet analysis and wavelet neural networks. As in the case of a linear model, we separate the data in three periods each one consisting of 80 data points. As it was discussed in the previous section the fitting using a linear model is questionable. In the second period the regression suggests that only the non-OPEC production and the stocks held by the OECD must be used as predictors. On the other hand, in the third period variables such as the open interest in future contracts are left out of the model, opposing the general belief in the oil market.

In this section we train a different wavelet network for each sub-period. Also the relevance of each variable to the model is determined by the *SBP*. Table 11 summarizes the selected variables and the fitting criteria for each sub-period as well as the *IPOCID* and *POS* criteria. The *IPOCID* and *POS* criteria increased to 69.74% and 71.12% while the *POCID* is 83.19%. Also, all fitting criteria significantly decreased in comparison to all previous simulations. The Normalized Mean Square Error (NMSE) decreased to 0.499204 while it was 0.611952 in the tree-sub period linear case and 0.772868 in case of wavelet network trained in the full sample. In addition both R^2 and $\overline{R}^2$ were improved. The corresponding R^2 are 63.46%, 32.23% and 50.24% while in the case of the linear model were 57.12%, 12.92% and 40.58%. The corresponding $\overline{R}^2$ are 49.80, 21.35 and 35.79 while in the linear case were 54.80%, 10.69% and 39.02%. The fitting can be found in Figure 4b.

It is clear that both the linear model and the wavelet neural network had an average fitting to the crude oil price returns although wavelet neural networks performed much better. Although wavelet networks weren't able to forecast satisfactory the level of the monthly returns, were able to successfully forecasted both the sign and the change in direction of the monthly returns, two criteria that practitioners are interested. Finally, the model proposed in the linear case is questionable. In most cases the selected explanatory variables were indexes, while variables such as production, consumption and stocks were disregarded. Economist report that practitioners recently take positions on future contracts by looking the volumes of these variables as well as the stocks held or the Euro/Dollar exchange and the natural gas price. These speculations were confirmed by the wavelet networks in contrast to the linear models.

4. Sensitivity and Scenario analysis

In the previous section wavelet analysis was used successfully in order to find structural breaks and changes in the dynamics of the crude oil returns. Next a wavelet neural network was used in order to obtain additional predictive power. In this section we examine the level of the impact of the changes of the significant variables to crude oil returns. In addition we examine the time evolution of the affection of each significant variable to the crude oil returns. Our interest is to examine how future crude oil returns will be affected hence we focus only on the corresponding variables of the last sub-period.

Table 12 shows different sensitivity criteria for each variable. The direct connection from the input variable i to the network output is represented by $w_i^{[0]}$ and can be interpreted similar to the coefficient of a linear regression. *MaxAD* and *MinAD* represent the maximum and minimum absolute derivative respectively while *AvgDM* and *AvgLM* represent the average derivative and elasticity magnitude and *SBP* is the sensitivity-based pruning criterion. Further information about these sensitivity criteria can be found on Zapranis & Refenes (1999). The large value of *AvgLM* of the *dcic,* 5.1698 and the large coefficient indicate that changes in the *dcic* will results to large changes in the crude oil returns. On the other hand the small *SBP* values of *dcl*, *deu*, *dlcoecd*, *dlpusa*, *dlsoecd* and *dwp* indicate that these variables are less significant.

In order to gain a better insight about how sensitive are the returns to each input variable we calculate the corresponding derivatives. In Figure 5 one can observe the time evolution of how sensitive are the oil return to each predictor. In contrast to *SBP* indications, we found that crude oil returns are very sensitive to variables as the OECD's consumption and USA's production. Also, it is clear from Figure 5 that returns become more sensitive to the stocks held by OECD in the recent years. This was expected since the recent years practitioners takes into account the stocks held by both OECD and OPEC. On the other hand Consumer Index sensitivity increases both in level and volatility while USA's production sensitivity increase in level but its volatility had decreased significantly. From Figure 5 we confirm that returns are less sensitive to variables such as the open interest and the Euro/Dollar exchange rate. This is probably due to the fact that the last months the Euro/Dollar exchange rate had an increasing trend with no big fluctuations. Moreover variable selection is very sensitive to the selected time period. Probably the larger selected period (due to model and available data restrictions) than the one suggested by the wavelet analysis (e.g. from 7/2006) resulted to a group of predictors that affect the influence of the euro-dollar exchange rate and the open interest to crude oil returns.

Table 13 shows how returns respond in different scenarios. The top part of the table represent the crude oil return is a particular variable increase or decrease by 5%, 10% and 20%. The level or returns if none variable changes are 0.06196 and is given by the last row of Table 13 It is clear that changes in production results to big changes in oil returns and it is negative correlated (last column) to oil price as expected. Changes in the Euro/Dollar exchange rate result to insignificant changes in oil returns as it was expected from the previous analysis of the sensitivity criteria and the derivatives. However it is clear that an increase to the exchange rate results to an increase to oil returns. Probably this happens because a weakened dollar results to an increase to demand and Table 13 shows that an increase in demand results to an increase of oil returns as it was expected.

Next we examine various scenarios where changes in a category of variables occur. Scenario 1 represents a hypothesis the oil supply is increased. We assume that the USA production increased while the stocks held by OECD decreased by the same percentage. More precisely the first column at the second part of Table 14 shows the corresponding returns when the USA's production increase by 20% while the stocks decrease by 20%. Both these events lead to higher levels of supply. As it was expected returns and prices are negatively affected. Scenario 2 examines the changes in oil returns when both the euro-dollar exchange rate and the trade volume in future contracts increase or decrease at the same time. As we can see, in this case the returns decrease if changes, either negative or positive, in these variables occur.
Observing figure 6 a synchronization of the derivatives of the crude oil returns with respect to *damex*, *deu*, *dlpusa*, *dng* and *dzot* is represented. In scenario 3 we calculate the level of the crude oil returns when *damex*, *deu*, *dlpusa*, *dng* and *dzot* increase or decrease. Scenario 4 results from figure 6 and the pointed event in the middle of 2006. The *damex*, *deu*, *dlpusa*, *dng* variables increase while the *dzot* decrease by the same amount. In scenario 5 *damex*, *deu*, *dlpusa*, *dng* change while *dzot* remains constant. As we can see from Table 14 changes in these variables have an opposite effect on the crude oil returns. Moreover these scenarios lead to large fluctuations of the crude oil returns. In Figure 6 there are periods where changes to *damex*, *deu* and *dzot* occur opposite to the changes of *dlpusa* and *dng.* This is represented in scenario 6. For example when the first three variables increase by 20% and *dlpusa* and *dng* decrease by 20% the crude oil return expected to be 6.591%. Scenario 7 represents the crude oil returns when changes occur only to *damex*, *deu* and *dzot* while scenario 8 represents the crude oil returns when changes occur only to *dlpusa* and *dng.* In both scenarios the returns are negatively correlated. Moreover the level of affection in return is higher under scenario 8.

5. Conclusions

Crude oil plays an important role in national economies and world economic activities while it remains one of the main sources of energy. These reasons have pushed both academics and market participants to the development of predictive models for oil prices. An understanding of the mechanisms of oil price determination and a better forecast of the expected crude oil prices can help market participants to improve their economic activities. In this paper, we tried to investigate the factors that affect crude oil price for the period 1988 – 2008 comparing a linear and a non-linear approach. We studied the dynamics of the West Texas Intermediate crude oil price time series as well as the relation of crude oil price and returns to various explanatory variables. The existence of a unit root on both prices and returns was clear while non-stationarity in the mean was a common problem among the original variables. Next, we used wavelet analysis to thorough examine how the dynamics of the oil price process changed over time. More precisely using a continuous wavelet transform we were able to capture the most important events that had a direct impact on crude oil prices and returns. It must be mentioned that no information were obtained examining the oil prices and returns by alternative methods. Using wavelet analysis we were able to locate structural breaks that originally were not visible. Also, splitting the data to sub-periods according to the continuous wavelet transform better forecasts were obtained.

Next a wavelet neural network was used in order to obtain additional predictive power. Using a linear model resulted to models that contradict economists' analyses and practitioners' beliefs. On the other hand the wavelet network findings are in line with economic studies. Predictors such as the stocks held by the OECD countries and the USA as well as variables as consumption, production, the open interest on future contracts and the euro/dollar exchange rate considered significant. Summarizing, wavelet networks not only have better predictive power in forecasting crude oil returns, but can also model correctly the dynamics of returns. In addition the sensitivity of crude oil prices and returns were examined and the time evolution of the affection of each predictor to the crude oil returns presented.
Examining the wavelet decomposition of the oil prices we see a very bright area at the end of the figure representing a strong change in the driving dynamics of the crude oil prices. However a better examination of the wavelet decomposition of the returns reveals that the pick of this event already passed. Comparing the patterns of past events, we expect that the volatility of returns will return to normal in the next months and probably the price of oil will fall and settle around a mean.
Unfortunately it was not possible to train a wavelet network in the last period, according to wavelet analysis. The last structural break occurred at 7/2006 leaving only 17 available data for study. Examining more data we could have a better insight in how crude oil prices will evolve. Also, further work can be consider in the field of estimation of wavelet neural network prediction intervals while participants on oil market may be more interested in predicting intervals for future oil price movements than simply point estimates.

References

[1] Adragni, B., Chatrath, A., Dhanda, K.K. & Raffie, K. (2001). Chaos in oil prices? Evidence from futures markets. *Energy Economics*, vol. 23, pp. 405-425.
[2] Amin-Naseri, M.R. & Gharacheh, E.A. (2007). A hybrid artificial intelligence approach to monthly forecasting oil price time series. *Proceedings of EANN*, August, Thessaloniki, Greece.
[3] Askari, H. & Krichene, N. (2008). Oil price dynamics (2002–2006). *Energy Economics*, vol. 30, issue 5, pp. 2134-2153.
[4]Bernabe, A., Martina, E., Alvarez-Ramirez, J. & Ibarra-Valdez, C. (2004). A multi-model approach for describing crude oil price dynamics. *Physica A: Statistical Mechanics and its Applications*. vol. 338, issue 3-4, pp. 567-584.
[5] Daubechies, I. (1992). Ten lectures on wavelets. *SIAM*.
[6] Dees, S., Karadeloglou, P., Kaufmann, R.K. & Sanchez, M. (2007). Modeling the world oil market: Assessment of a quarterly econometric model. *Energy Policy*, vol. 35, pp. 178-191.
[7] Fan, Y., Liang, Q. & Wei, Y.M. (2008). A generalized pattern matching approach for multi-step prediction of crude oil price. *Energy Economics*, vol. 30, issue 3, pp. 889-904.
[8] Gao, R. & Tsoukalas, L.H. (2001). Neural-wavelet methodology for load forecasting. *Journal of Intelligent & Robotic Systems*, vol. 31, pp. 149-157.

[9] Gori, F., Ludovisi, D. & Cerritelli, P.F. (2007). Forecast of oil price and consumption in the short term under three scenarios: Parabolic, linear and chaotic behavior. *Energy*, vol. 32, pp. 1291-1296.
[10] Kaboudan, M.A. (2001). Compumetric forecasting of crude oil prices. *Proceedings of the 2001 Congress on Evolutionary Computation*, vol. 1, pp. 283-287.
[11] Khayamian, T., Ensafi, A.A., Tabaraki, R. & Esteki, M. (2005). Principal componentwavelet networks as a new multivariate calibration model. *Analytical Letters*, vol. 38(9), pp. 1477-1489.
[12] Mallat, S.G. (1999). *A wavelet tour of signal processing*. Academic Press, San Diego.
[13] Moody, J.E. & Utans, J. (1992). Principle Architecture Selection for Neural Networks: Application to Corporate Bond Rating Prediction. *Advances in Neural Information Processing Systems,* vol.4, Moody, Hanson, and Lippmann, eds., Morgan Kaufmann, Palo Alto, San Mateo, CA.
[14] Moosa, I.A. & Al-Loughani, N.E. (1994). Unbiasedness and time varying risk premia in the crude oil future markets. *Energy Economics*, vol. 16(2), pp. 99-105.
[15] Morana, C. (2001). A semiparametric approach to short-term oil price forecasting. *Energy Economics*, vol. 23, pp. 325-338.
[16] E. Panas and V. Ninni (2000). Are oil markets chaotic? A non-linear dynamic analysis. *Energy Economics*, 22, pp. 549-568.
[17] Peters, E. (1994). *Fractal market hypothesis: Applying chaos theory to investment and economics*.Wiley.
[18] Postalcioglu, S. & Becerikli, Y. (2007). Wavelet networks for nonlinear system modeling. *Neural Computing & Applications*, vol. 16, pp. 443-441.
[19] Postali, F.A.S. & Picchetti, P. (2006). Geometric Brownian motion and structure breaks in oil prices: A quantitative analysis. *Energy Economics*, vol. 28, pp. 506-522.
[20] Rehrl, T. & Friedrich, R. (2006). Modeling long-term oil price and extraction with a Hubbert approach: The LOPEX model. *Energy Policy*, vol. 34, pp. 2413-2428.
[21] Sadorsky, P. (1999). Oil price shocks and stock market activity. *Energy Economics*, vol. 21(5), pp. 449-469.
[22] Sadorsky, P. (2006). Modeling and forecasting petroleum futures volatility. *Energy Economics*, vol. 28, pp. 467-488.
[23] Shambora, W.E. & Rossiter, R. (2007). Are there exploitable inefficiencies in the futures market for oil? *Energy Economics*, vol. 29, pp. 18-27.
[24] Subasi, A., Alkan, A., Koklukaya, E. & Kiymik, M.K. (2005). Wavelet neural network classification of EEG signals by using AR model with MLE pre-processing. *Neural Networks*, vol. 18. pp. 985-997.
[25] Szu, H., Telfer, B., Kadambe, S., (1992). Neural Network adaptive wavelets for signal representation and classification. *Opt. Engineering*. 31. 1907—1916.
[26] Tang, L. & Hammoudeh, S. (2002). An empirical exploration of the world oil price under the target zone model. *Energy Economics*, vol. 24, pp. 557-596
[27] Wang, S., Yu, L. & Lai, K.K. (2005). Crude Oil Price Forecasting with TEI@I Methodology. *Journal of Systems Science and Complexity*. vol. 18, issue 2, pp. 145-166.
[28] Ye, M., Zyren, J. & Shore, J. (2005). A monthly crude oil spot price forecasting model using relative inventories. *International Journal of Forecasting*, vol. 21, issue 3, pp. 491-501.
[29]Xie, W., Yu, L., Xu, S. & Wang, S. (2006). A New Method for Crude Oil Price Forecasting Based on Support Vector Machines. *Lecture Notes in Computer Science*, vol. 3994/2006, pp. 444-451.
[30] Yousefi, S., Weinreich, I. & Reinarz, D. (2005). Wavelet-based prediction of oil prices. *Chaos, Solitons and Fractals*, 25, pp. 265-275.
[31]Yu, L., Lai, K.K., Wang, S. & He, K. (2007). Oil Price Forecasting with an EMD-Based Multiscale Neural Network Learning Paradigm. *Lecture Notes in Computer Science*, vol. 4489/2007, pp. 925-932.
[32] Zapranis, A. & Alexandridis, A. (2007) “Weather Derivatives Pricing: Modelling The

Seasonal Residual Variance of an Ornstein-Uhlenbeck Temperature Process With Neural Networks” *Neurocomputing* (accepted, to appear).

[33] Zapranis, A. & Alexandridis, A. (2008a) “Modelling Temperature Time Dependent Speed of Mean Reversion in the Context of Weather Derivative Pricing”. *Applied Mathematical Finance*, vol 15 issue 4, pp.355-386.

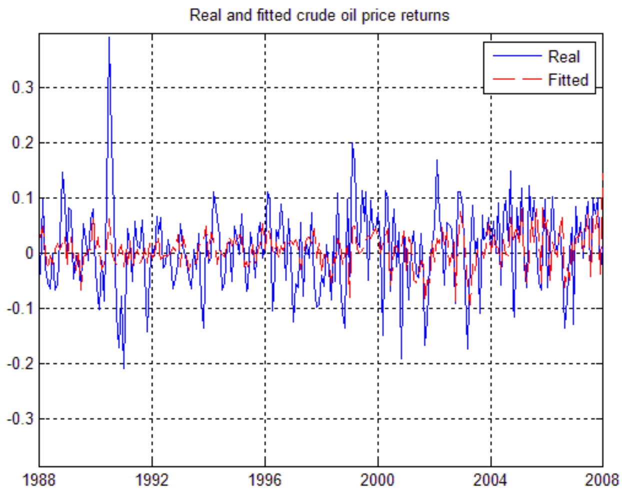
[34] Zapranis, A. & Alexandridis, A. (2008b) “Forecasting Cash Money Withdrawals Using Wavelet Neural Networks”. *Proceedings of AFE*. July, Samos, Greece.

[35] Zapranis, A. & Refenes, A.-P. (1999) *Principles of Neural Model Identification, Selection and Adequacy: With Applications to Financial Econometrics*, Springer-Verlag.

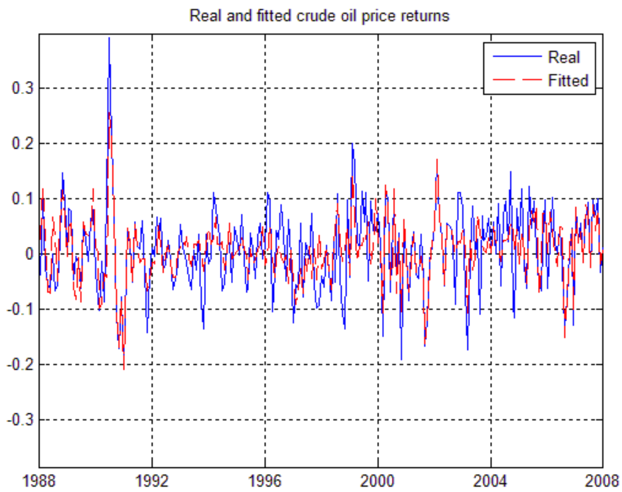
[36] Zhang, Q. & Benveniste, A. (1992). Wavelet networks. *IEEE Trans. Neural Networks*, vol. 3(6), pp 889-898.

[37] Zhang, Q. (1997). Using Wavelet Network in Nonparametric Estimation. *IEEE Trans. Neural Networks*. 8 (2), 227--236

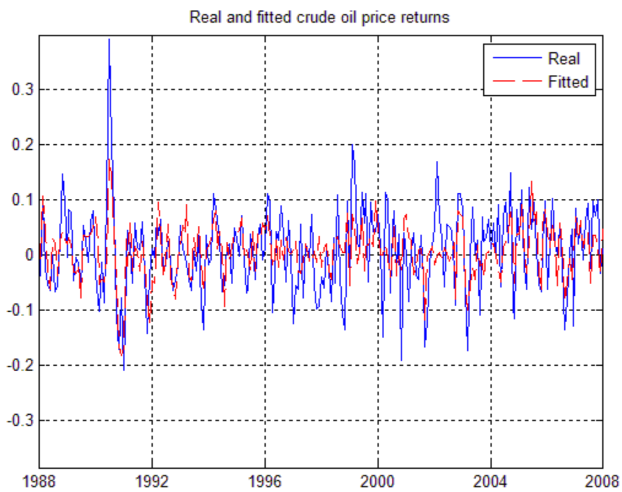
[38] Zhang, X., Lai, K.K. & Wang, S.Y. (2008). A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Economics*, vol. 30, issue 3, pp. 905-918.



(a)



(b)



(c)

Figure 1. Real and fitted crude oil price returns for the (a) whole period (b) for eight sub- periods (c) and for three sub-periods using a linear model.

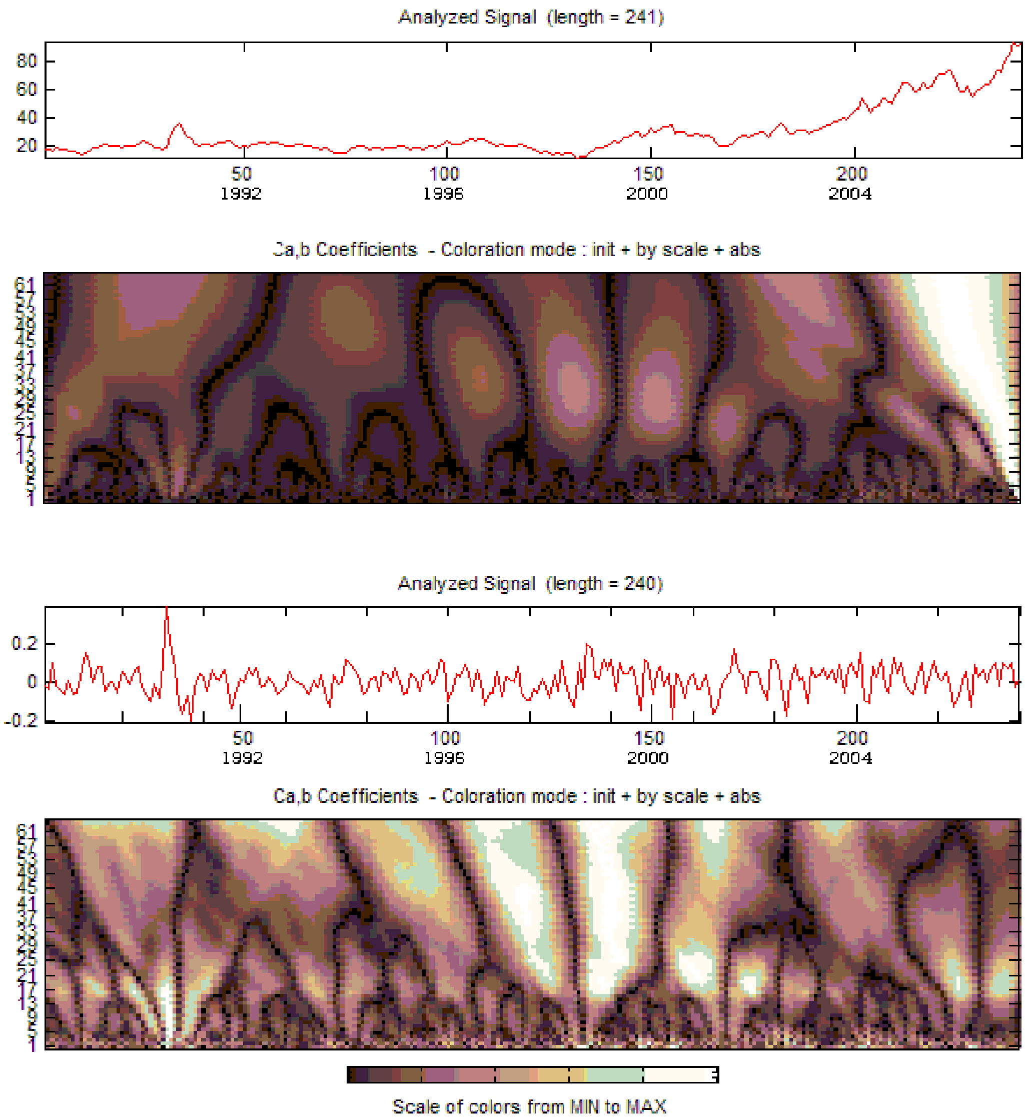


Figure 2. The continuous wavelet transform (db3) of oil Prices (up) and returns (down)

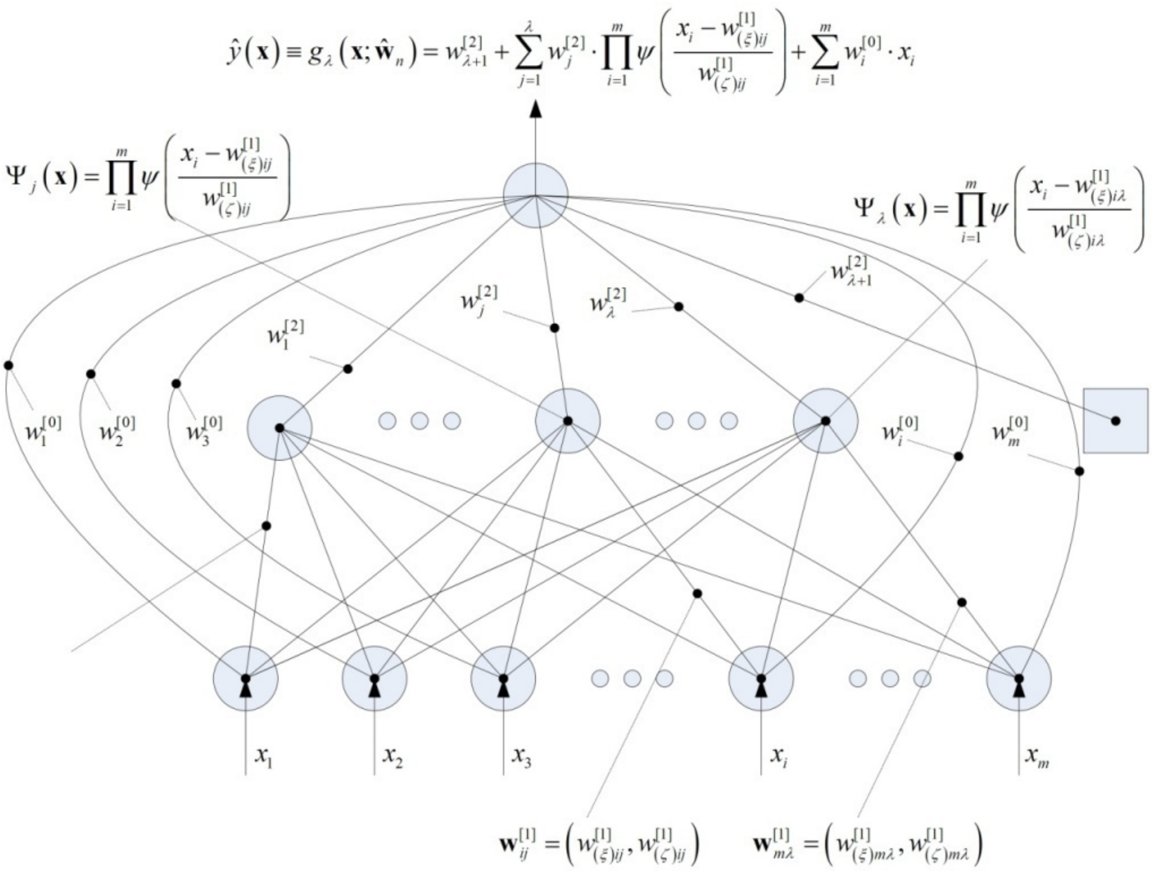
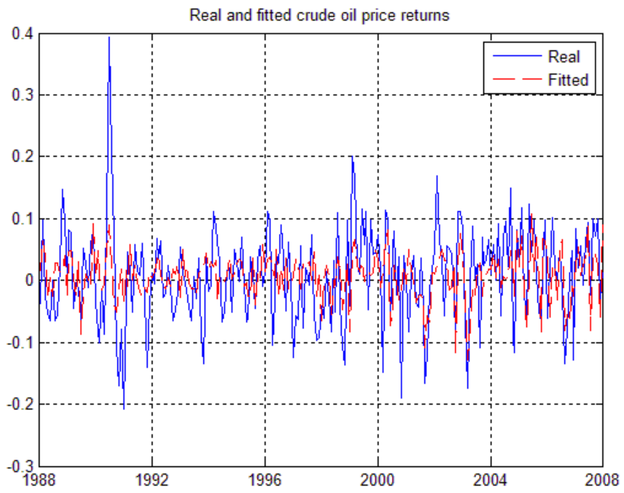
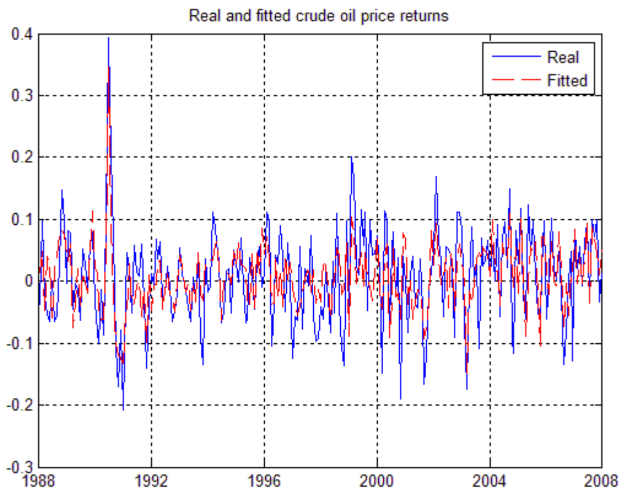


Figure 3. Structure of a wavelet network.



(a)



(b)

Figure 4. Real and fitted crude oil prices for the (a) whole period (b) and for the three sub-periods using a wavelet neural network.

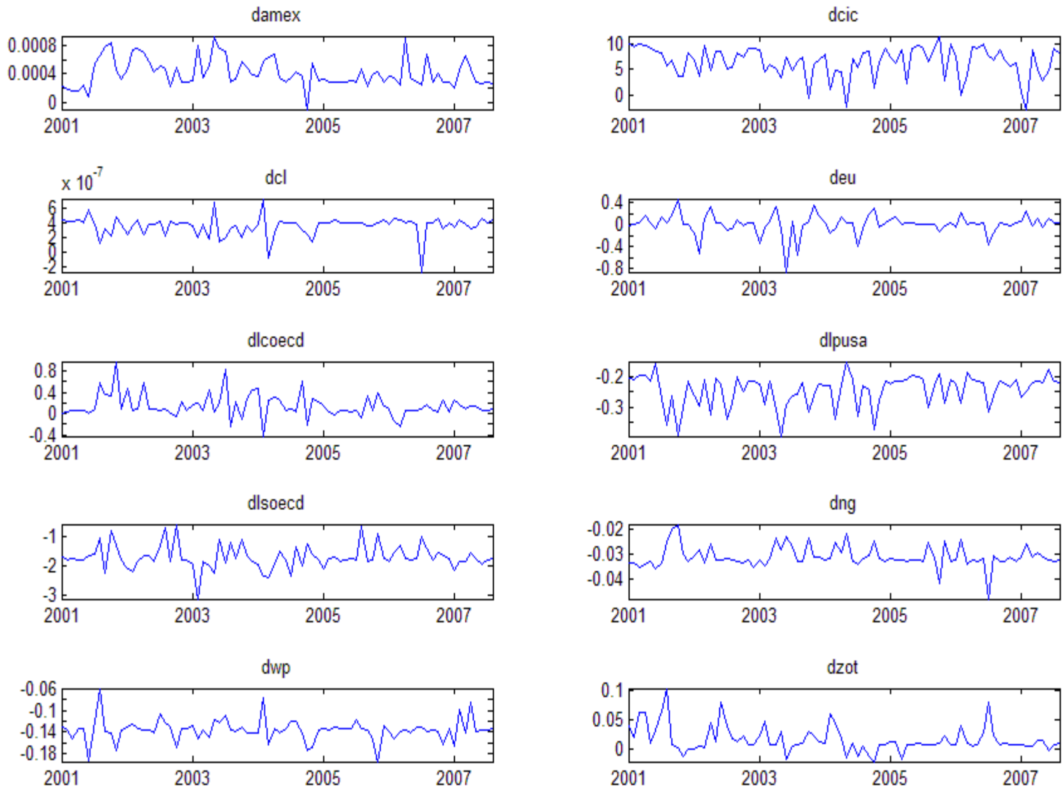


Figure 5. Derivatives of the crude oil returns with respect to the significant variables for the third sub-period.

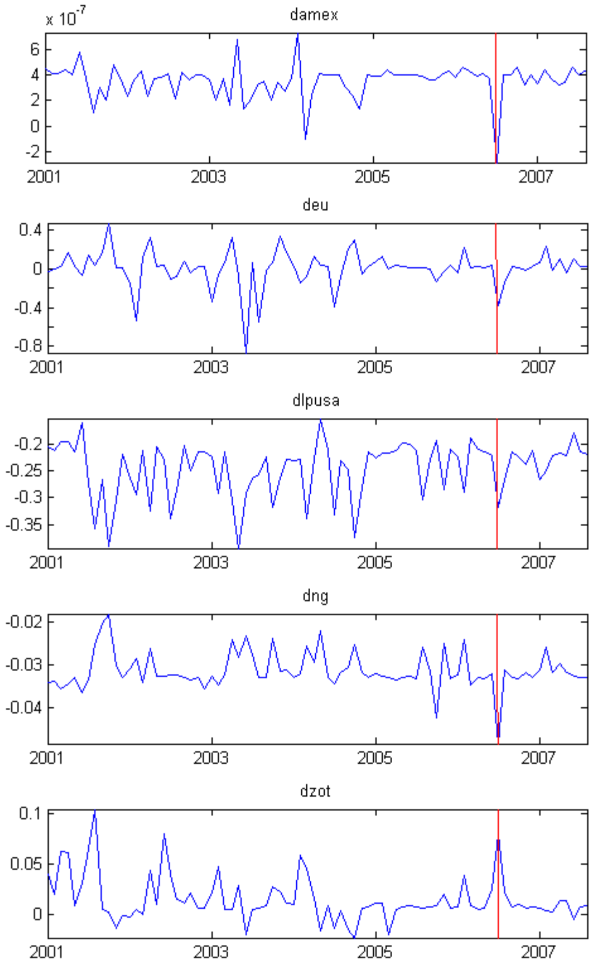


Figure 6. Synchronization of the derivatives of the crude oil returns with respect to the significant variables for the third sub-period.

Table 1: Literature Review.

Author	Year	Title	Method
Zhang *et al.*	2008	A new approach for crude oil price analysis based on Empirical Mode Decomposition	Ensemble Empirical Mode Decomposition (EEMD)
Askari & Krichene	2008	Oil price dynamics (2002–2006)	jump–diffusion process, Levy process of the variance-gamma type
Fan *et al.*	2008	A generalized pattern matching approach for multi-step prediction of crude oil price	Pattern matching, Genetic algorithm
Dees *et al.*	2007	Modelling the world oil market: Assessment of a quarterly econometric model	Econometric model based on demand and supply
Amin-Naseri & Gharacheh	2007	A hybrid artificial intelligence approach to monthly forecasting oil price time series	artificial intelligence
Yu *et al.*	2007	Oil Price Forecasting with an EMD-Based Multiscale Neural Network Learning Paradigm	Neural networks, EMD
Xie *et al.*	2006	A New Method for Crude Oil Price Forecasting Based on Support Vector Machines	Support Vector Machines
Shambora & Rossiter	2007	Are there exploitable inefficiencies in the futures market for oil?	ANN, moving average, random walk
Gori *et al.*	2007	Forecast of oil price and consumption in the short term under three scenarios: Parabolic, linear and chaotic behaviour	Gaussian curve, Chaos, regression
Postali & Piccheti	2006	Geometric Brownian Motion and structural breaks in oil prices: A quantitative analysis	Geometric Brownian Motion, mean reverting process
Sadorsky	2006	Modeling and forecasting petroleum futures volatility	GARCH, TGARCH, AR, random walk, historical mean, moving average, VAR, BIGARCH
Rehrl & Friedrich	2006	Modelling long-term oil price and extraction with a Hubbert approach: The LOPEX model	Hubbert curves, LOPEX model (supply model)
Yousefi *et al.*	2005	Wavelet-based prediction of oil prices	wavelets
Wang *et al.*	2005	Crude Oil Price Forecasting with TEI@I Methodology	ARIMA & ANN
Ye *et al.*	2005	A monthly crude oil spot price forecasting model using relative inventories	Simple regression model

Bernabe *et al.*	2004	A multi-model approach for describing crude oil price dynamics	Gaussian mean-reversion process, nonlinear mean-reversion model
Tang & Hammoudeh	2002	An empirical exploration of the world oil price under the target zone model	non-linear regression, first-generation target zone model
Morana	2001	A semiparametric approach to short-term oil price forecasting	GARCH, historical simulation
Kaboudan	2001	Compumetric forecasting of crude oil prices	genetic programming, neural networks and random walk
Adragni *et al.*	2001	Chaos in oil prices? Evidence from futures markets	Chaos, ARCH-models
Panas & Ninni	2000	Are oil markets chaotic? A non-linear dynamic analysis	Chaos, ARCH-GARCH
Sadorsky	1999	Oil price shocks and stock market activity	GARCH, vector autoregression (VAR)
Moosa & Al-Loughani	1994	Unbiasedness and time varying risk premia in the crude oil futures market	Regression, GARCH-M

Table 2: Phillip-Perron unit root tests.

	Unit Root	P-Value	1st Diff	P. Value		Unit Root	P-Value	1st Diff	P. Value
AMEX	Y	0.9953	N	0.0000	**CIC**	Y	0.9893	N	0.0005
CL	Y	0.5194	N	0.0000	**COECD**	N	0.0105	N	0.0000
CUSA	Y	0.2714	N	0.0001	**EU**	Y	0.9478	N	0.0000
IMPORTS	Y	0.1866	N	0.0000	**NG**	Y	0.0505	N	0.0000
PNOPEC	Y	0.8821	N	0.0000	**POPEC**	Y	0.4210	N	0.0000
PUSA	N	0.0002	N	0.0000	**PWORLD**	Y	0.8707	N	0.0000
SOECD	N	0.0086	N	0.0000	**SUSA**	N	0.0499	N	0.0000
SP	Y	0.8101	N	0.0000	**WP**	Y	0.9918	N	0.0000
ZOT	Y	0.7918	N	0.0001					

Table 3a: Descriptive statistics of the transformed variables.

Var	Mean	St.Dev.	Max.	Min.	Skewness	Kurtosis
damex	5.819	31.033	156.270	-111.46	0.877	7.441
dcic	0.004	0.003	0.014	-0.01	-0.433	8.588
dcl	873.028	13182.590	56898.730	-76510.20	-0.446	9.379
dcusa	14.502	581.645	1546.675	-1893.90	-0.309	3.599
deu	0.013	0.100	1.086	-0.05	9.950	106.823
dimports	669.393	18795.370	67421	-62595	0.110	3.571
dlcoecd	0.001	0.035	0.076	-0.09	-0.362	2.903
dlpusa	-0.002	0.022	0.076	-0.21	-3.566	38.982
dlsoecd	0	0.010	0.030	-0.04	-0.326	3.485
dlsusa	0	0.013	0.030	-0.05	-0.537	3.866
dng	0.019	0.455	2.470	-2.49	0.221	12.179
dpnopec	11.883	351.531	1106.222	-1168.54	-0.135	3.283
dpopec	59.764	571.275	2212.088	-3522.86	-1.514	12.238
dpworld	71.647	674.348	2548.130	-3547.87	-0.772	7.891
dsp	5.068	37.460	132.160	-163.39	-0.650	5.718
dwp	0.008	0.078	0.337	-0.43	-0.633	9.914
dzot	0.187	0.291	1	-0.80	-0.673	4.269

Table 3b: Descriptive statistics of the original variables.

Var	Mean	St.Dev	Max	Min	Skewness	Kurtosis
amex	488.137	316.262	1559.700	168.950	1.547	4.684
cic	1.626	0.255	2.121	1.159	0.045	2.042
cl	119818.500	57734.520	357066.3	50936.210	1.559	5.172
cusa	18777.460	1471.899	21666.060	16138.770	-0.006	1.712
eu	1.133	0.170	1.570	0.850	0.170	2.321
imports	244541.1	53081.06	327476	134863	-0.220	1.799
coecd	45912.940	3274.550	52036.860	37434.240	-0.428	2.182
pusa	6328.634	883.244	8374.069	4203.964	0.272	2.285
soecd	3865.126	164.455	4257.591	3496.170	0.344	2.452
susa	1613.748	67.058	1784.991	1460.452	-0.061	2.686
ng	3.196	1.952	10.330	1.260	1.198	3.546
pnopec	37695.850	2286.114	41586.160	33402.960	0.137	1.922
popec	27923.460	3338.936	33527.700	19095	-0.379	2.644
pworld	65619.310	5110.198	74431.290	56960.860	0.235	1.799
sp	845.022	410.957	1549.380	257.070	0.036	1.487
wp	0.906	0.483	2.532	0.422	1.726	4.964
zot	89.224	15.413	112.2	67.5	-0.121	1.426

Table 4: Linear regression for the full sample.

Dependent Variable: WTR
Method: Least Squares
Sample (adjusted): 1988M02 2008M01
Included observations: 240 after adjustments

Var.	Coeff.	St. Error	P-Value	
c	-0.0254	0.009436	0.0073	
damex	6.75E-04	0.000163	0.0001	
dcic	5.8607	1.998200	0.0037	
dlsoecd	-1.4647	0.443700	0.0011	
dsp	-3.83E-04	0.000137	0.0057	
dwp	-0.1882	0.066300	0.0049	
dzot	0.05	0.016100	0.0021	
R^2	18.09%		**MAE**	0.0538
$\bar{R}^2$	15.97%		**Max AE**	0.329528
MSE	0.004878		**POCID**	81.09%
RMSE	0.069843		**IPOCID**	62.18%
NMSE	0.819073		**POS**	64.85%

Table 5: Influential Values.

Observations	Leverage
120	0.9467
212	0.5057
225	0.2546
229	0.2652
239	0.3607
Average Leverage	0.0836

Table 6: Linear regression after influential values were removed.

Dependent Variable: WTR
Method: Least Squares
Sample (adjusted): 1988M02 2008M01
Included observations: 235 after adjustments

Var	Coeff.	St. Error	P-Value	
c	-0.024063	0.00954	0.012400	
damex	0.000720	0.00017	5.24E-05	
dcic	6.530437	2.17353	0.002958	
dlsoecd	-1.478698	0.44650	0.001079	
dsp	-0.000417	0.00014	0.003087	
dwp	-0.209573	0.06887	0.002619	
dzot	0.050665	0.01614	0.001923	
$\boldsymbol{R^2}$	18.49%		**MAE**	0.053514
$\boldsymbol{\bar{R}^2}$	16.34%		**Max AE**	0.323574
MSE	0.004871		**POCID**	80.26%
RMSE	0.069792		**IPOCID**	63.52%
NMSE	0.815087		**POS**	63.25%

Table 7: Linear regression after influential values were removed and new variables were selected.

Dependent Variable: WTR
Method: Least Squares
Sample (adjusted): 1988M02 2008M01
Included observations: 235 after adjustments

Var	Coeff.	St. Error	P-Value	
c	-0.023724	0.009515	0.013400	
damex	0.000710	0.000174	6.08E-05	
dcic	6.585519	2.161392	0.002587	
dlsoecd	-1.488437	0.443999	0.000939	
dpnopec	-0.000025	1.33E-05	0.059139	
dsp	-0.000400	0.000139	0.004420	
dwp	-0.213054	0.068505	0.002111	
dzot	0.052935	0.016096	0.001200	
R^2	19.77%		**MAE**	0.053521
$\overline{R}^2$	17.28%		**Max AE**	0.31852
MSE	0.004795		**POCID**	79.40%
RMSE	0.069243		**IPOCID**	61.80%
NMSE	0.802316		**POS**	61.97%

Table 8: Linear regression in eight different sub-periods.

Period	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Range	**1-30**	**30-42**	**43-109**	**110-146**	**147-163**	**164-175**	**176-219**	**220-240**
				Panel A: Selected Variables				
	damex	dcic	damex	dlsoecd	damex	dcl	dcic	damex
	dlcoecd	dcusa	dsp	dsp	dpnopec	dimports	dlcoecd	dcic
	dsp	dimports	WT_{t-1}	WT_{t-2}	dsp	dlsusa	dng	dcl
	WT_{t-1}	dlpusa			dwp	dzot		dcusa
		dpnopec			dzot			dlsusa
		WT_{t-1}						dsp
		WT_{t-2}						dwp
		WT_{t-4}						WT_{t-3}
$\boldsymbol{R^2}$	65.75%	99.97%	27.38%	36.41%	77.77%	97.87%	27.26%	90.91%
$\boldsymbol{\bar{R}^2}$	60.27%	99.90%	23.92%	30.63%	67.67%	96.65%	21.81%	84.30%
	MAE	0.037638	**RMSE**	0.050906	**MSE**	0.002591	**POCID**	84.03%
	Max AE	0.146697	**NMSE**	0.435125	**POS**	74.48%	**IPOCID**	73.11%

Table 9: Linear regression in three different sub-periods.

Period	**1**	**2**	**3**
Range	**1-80**	**81-160**	**161-240**
	damex	dlsoecd	damex
	dsp	dpnopec	dcic
	WT_{t-3}		dlsoecd
	WT_{t-4}		dng
			dwp
			WT_{t-1}
			WT_{t-3}
R^2	57.12%	12.92%	40.58%
$\bar{R}^2$	54.80%	10.69%	39.02%
MAE	0.048276	**Max AE**	0.218018
RMSE	0.060370	**POCID**	82.35%
NMSE	0.611952	**IPOCID**	70.59%
MSE	0.003645	**POS**	71.97%

Table 10: Wavelet networks for the full sample.

		Selected Variables		
damex	dcic	dcusa	dimports	dlcoecd
dlpusa	dlsoecd	dlsusa	dsp	dwp
dng	deu	dpnopec	dzot	WT_{t-1}
WT_{t-2}				
$\boldsymbol{R^2}$	23.08%		**MAE**	0.0533
$\boldsymbol{\bar{R}^2}$	18.25%		**Max AE**	0.3080
MSE	0.004572		**POCID**	77.73%
RMSE	0.067845		**IPOCID**	64.28%
NMSE	0.772868		**POS**	63.59%

Table 11: Wavelet networks in three different sub-periods.

Period	1	2	3
Range	**1-80**	**81-160**	**161-240**
	damex	damex	damex
	dcic	dcic	dcic
	dcl	dcusa	dcl
	dlsusa	dimports	deu
	dng	dlcoecd	dlcoecd
	dpnopec	dlsoecd	dlpusa
	WT_{t-1}	dng	dlsoecd
		dpnopec	dng
		dpopec	dwp
		dsp	dzot
		dwp	
		WT_{t-1}	
R^2	63.46%	32.23%	50.24%
$\bar{R}^2$	49.80%	21.35%	35.79%
MAE	0.043709	**Max AE**	0.155778
RMSE	0.054526	**POCID**	83.19%
NMSE	0.499204	**IPOCID**	69.74%
MSE	0.002973	**POS**	71.12%

Table 12: Variable weights and significance criteria

	damex	dcic	dcl	deu	dlcoecd	dlpusa	dlsoecd	dng	dwp	dzot
$\mathbf{w_i^{[0]}}$	0.2231	0.6321	0.1549	0.0073	0.0207	-0.1830	-0.2157	-0.4772	-0.3044	0.0325
MaxAD	0.0009	11.4217	0.0000	0.8678	0.9849	0.3945	3.1514	0.0483	0.1951	0.1041
MinAD	0.0001	0.1453	0.0000	0.0019	0.0049	0.1517	0.5549	0.0180	0.0588	0.0004
AvgDM	0.0004	6.5593	0.0000	0.1149	0.1841	0.2459	1.7014	0.0313	0.1362	0.0188
AvgLM	0.7568	5.1698	0.5684	0.2008	0.2663	0.3292	1.9489	1.5485	1.8830	0.4186
SBP	0.0221	0.0252	0.0010	0.0016	0.0029	0.0017	0.0047	0.0146	0.0076	0.0143

Table 13: Crude oil returns when significant variables change.

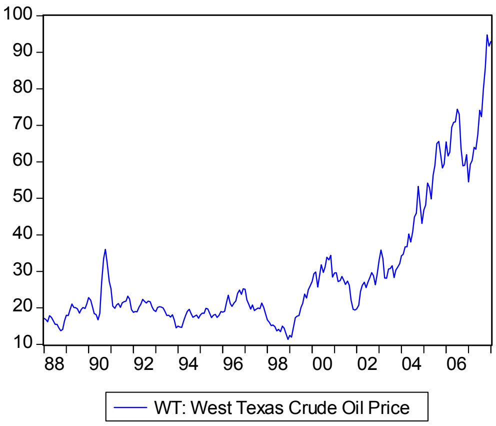
	+20%	+10%	+5%	-5%	-10%	-20%
damex	5.955	6.076	6.137	6.256	6.315	6.433
dcic	5.921	6.061	6.129	6.262	6.326	6.449
dcl	6.184	6.192	6.195	6.197	6.197	6.193
deu	6.198	6.198	6.197	6.196	6.195	6.194
dlcoecd	6.285	6.241	6.219	6.174	6.150	6.103
dlpusa	5.576	5.901	6.052	6.333	6.462	6.696
dlsoecd	6.183	6.190	6.193	6.200	6.203	6.210
dng	6.181	6.189	6.193	6.200	6.204	6.212
dwp	5.991	6.096	6.147	6.245	6.291	6.381
dzot	6.321	6.260	6.229	6.163	6.129	6.059
Mean	6.196					

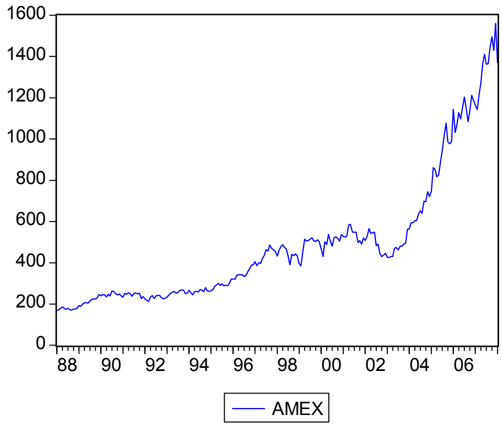
Table 144: Crude oil returns under various scenarios.

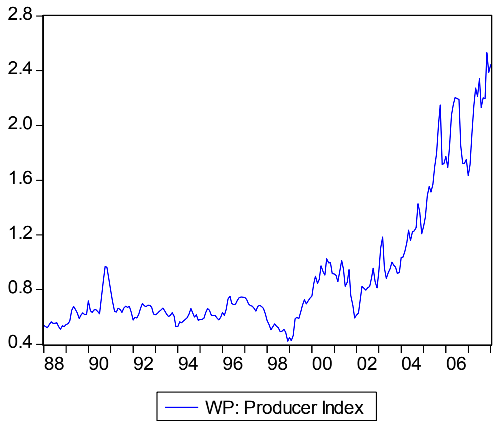
	+20%	+10%	+5%	-5%	-10%	-20%
Scenario 1	5.590	5.908	6.056	6.330	6.455	6.683
Scenario 2	6.186	6.193	6.195	6.197	6.196	6.191
Scenario 3	5.446	5.838	6.021	6.613	6.520	6.807
Scenario 4	5.202	5.711	5.956	6.428	6.654	7.082
Scenario 5	5.329	5.776	5.989	6.396	7.244	6.951
Scenario 6	6.591	6.412	6.392	6.074	5.942	5.653
Scenario 7	6.078	6.940	6.169	6.222	6.246	6.291
Scenario 8	5.560	5.893	6.048	6.336	6.469	6.711
Mean	6.196					

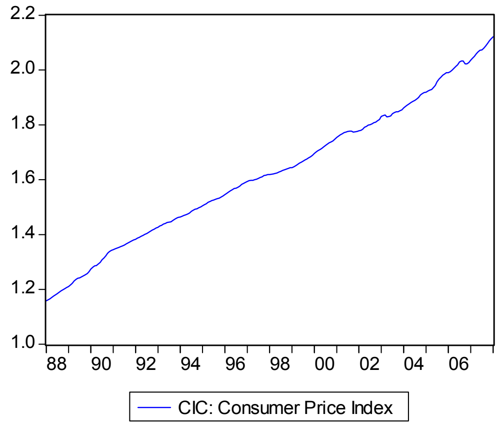
APPENDIX A

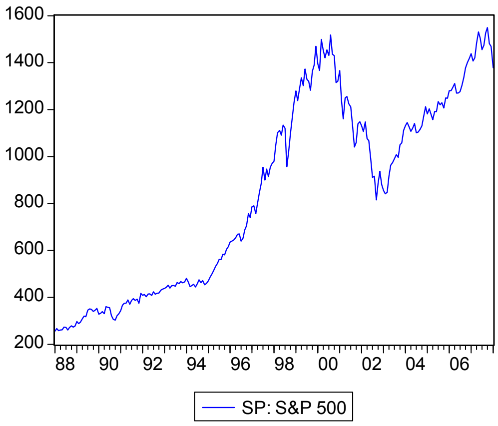
1. Explanatory Variables and Crude Oil Prices.

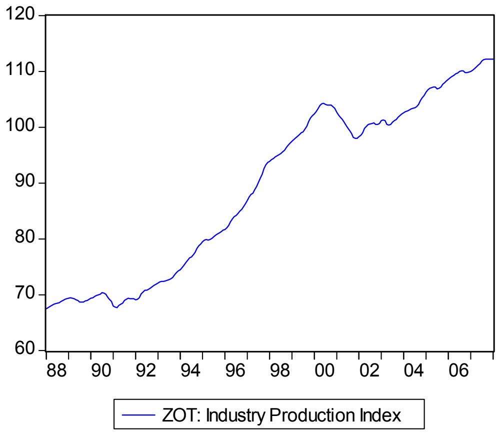


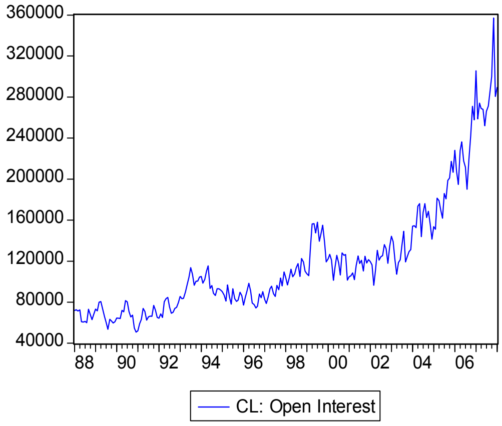


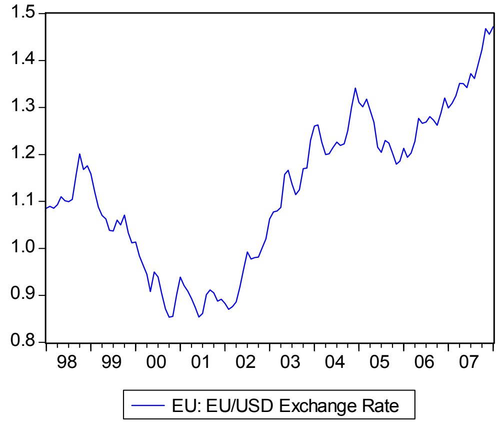


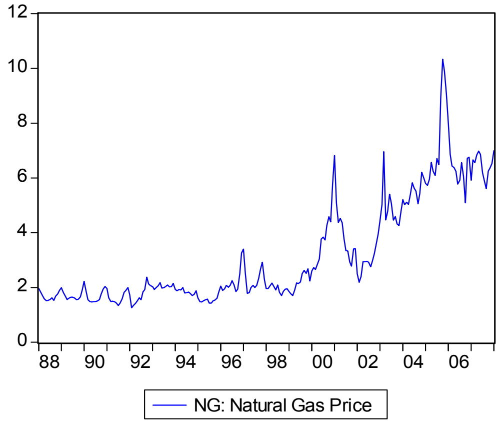


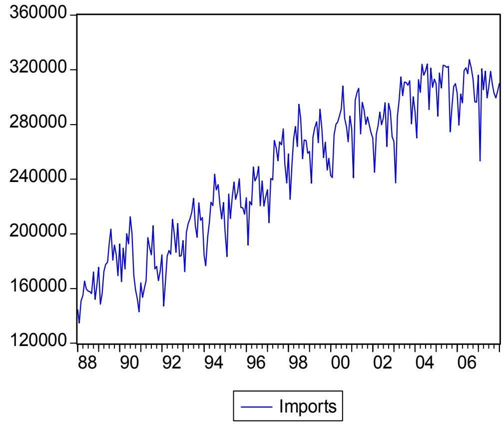


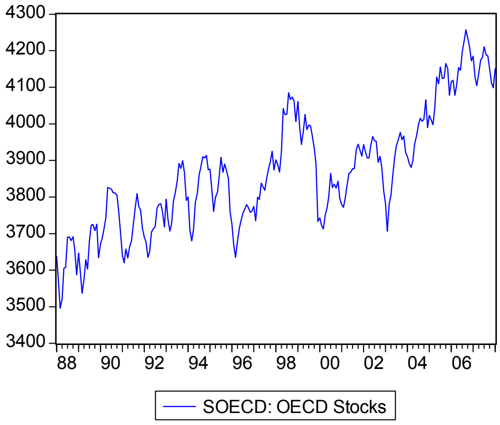


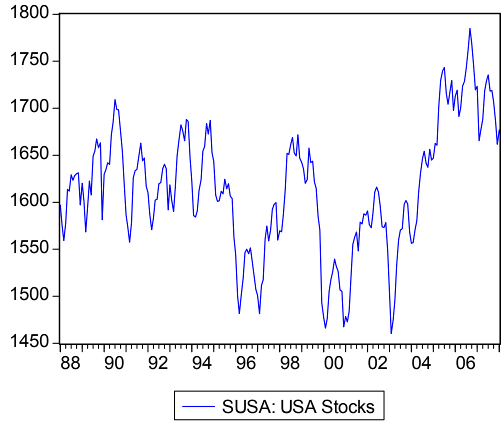

360000
320000
280000
240000
200000
160000
120000
80000
40000
88 90 92 94 96 98 00 02 04 06
CL: Open Interest

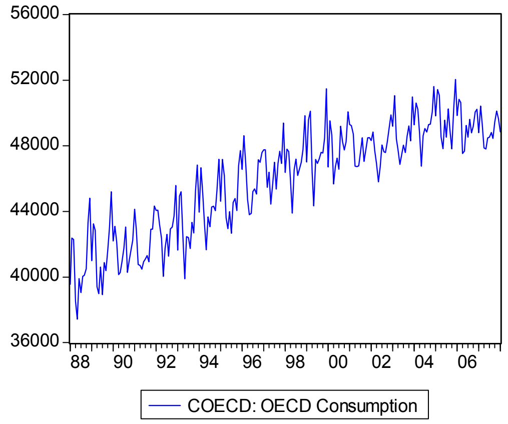

1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
98 99 00 01 02 03 04 05 06 07
EU: EU/USD Exchange Rate

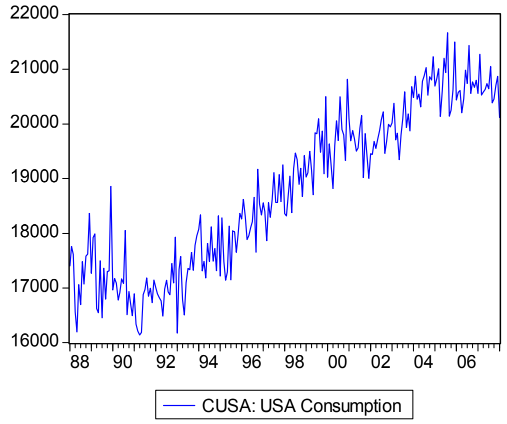

12
10
8
6
4
2
0
88 90 92 94 96 98 00 02 04 06
NG: Natural Gas Price

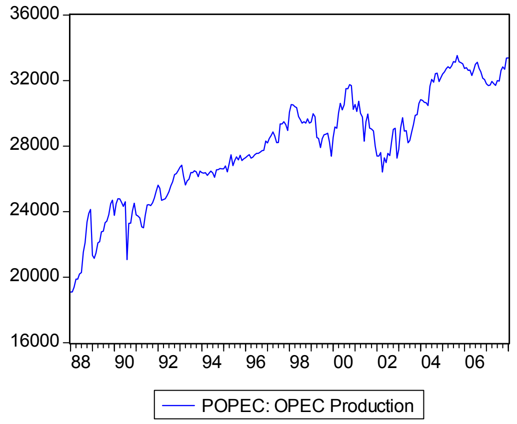

360000
320000
280000
240000
200000
160000
120000
88 90 92 94 96 98 00 02 04 06
Imports

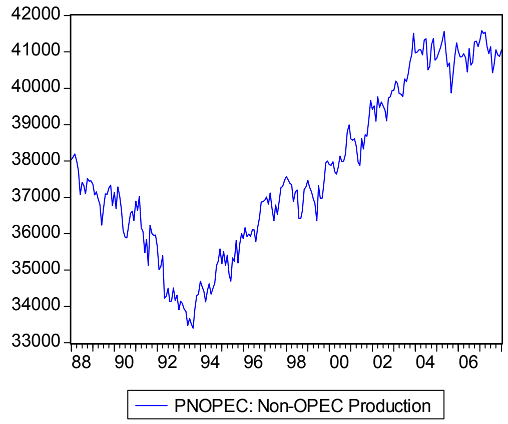

4300
4200
4100
4000
3900
3800
3700
3600
3500
3400
88 90 92 94 96 98 00 02 04 06
SOECD: OECD Stocks

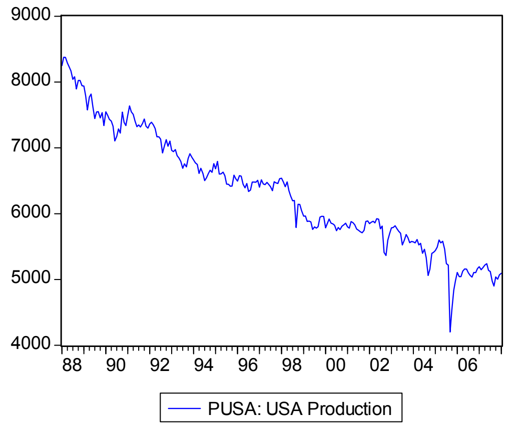

1800
1750
1700
1650
1600
1550
1500
1450
88 90 92 94 96 98 00 02 04 06
SUSA: USA Stocks

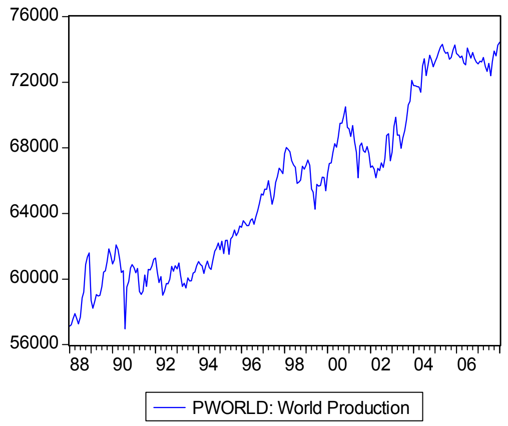

56000
52000
48000
44000
40000
36000
88 90 92 94 96 98 00 02 04 06
COECD: OECD Consumption


22000
21000
20000
19000
18000
17000
16000
88 90 92 94 96 98 00 02 04 06
CUSA: USA Consumption


36000
32000
28000
24000
20000
16000
88 90 92 94 96 98 00 02 04 06
POPEC: OPEC Production

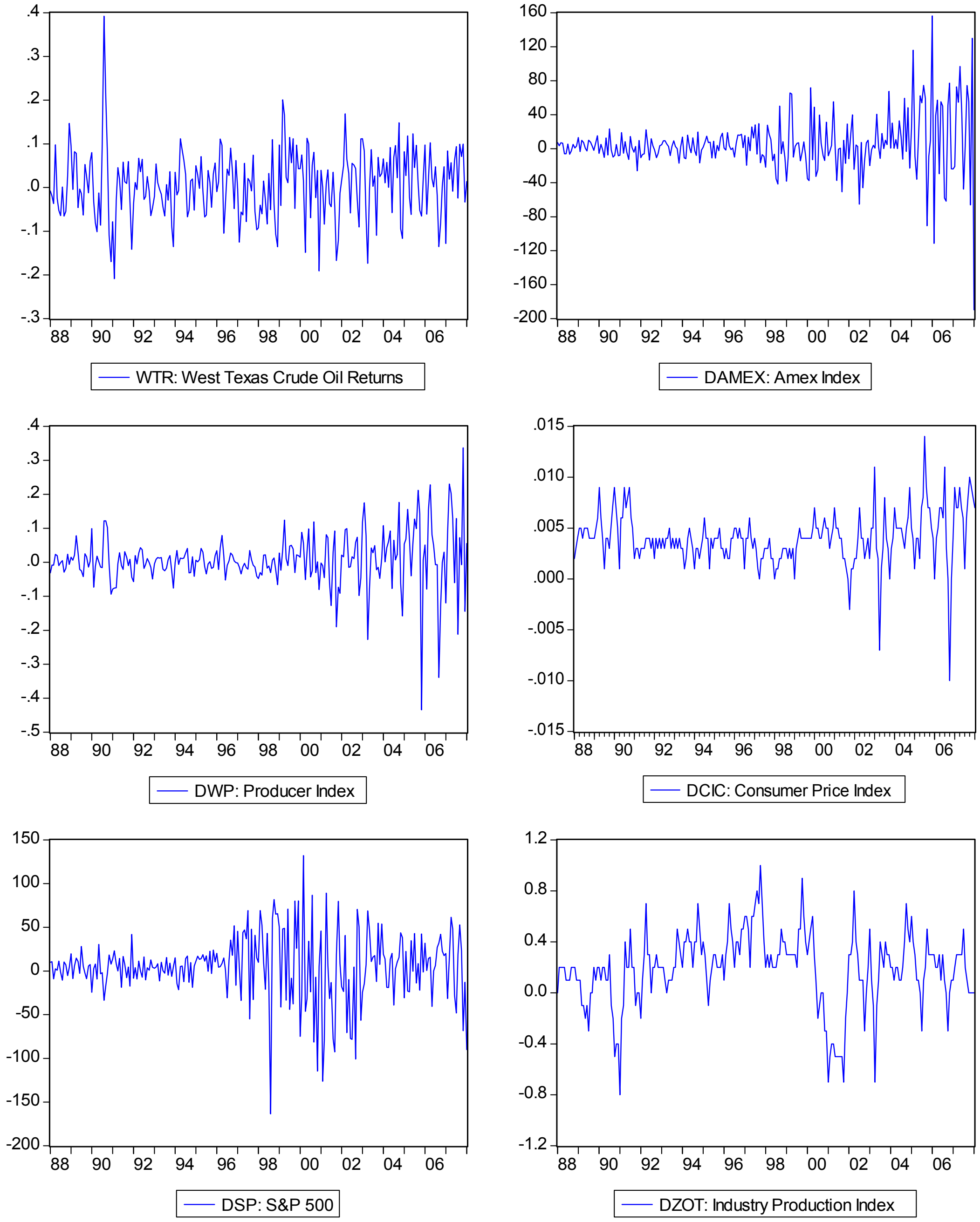

42000
41000
40000
39000
38000
37000
36000
35000
34000
33000
88 90 92 94 96 98 00 02 04 06
PNOPEC: Non-OPEC Production

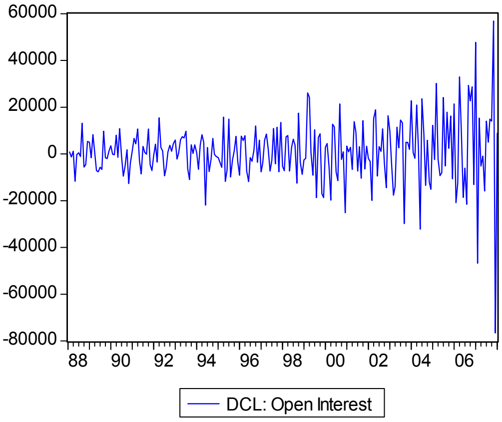

9000
8000
7000
6000
5000
4000
88 90 92 94 96 98 00 02 04 06
PUSA: USA Production

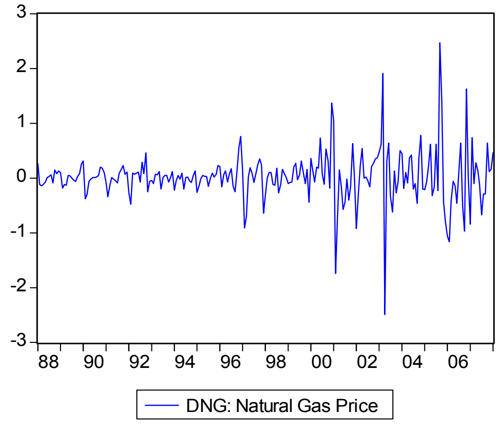

76000
72000
68000
64000
60000
56000
88 90 92 94 96 98 00 02 04 06
PWORLD: World Production

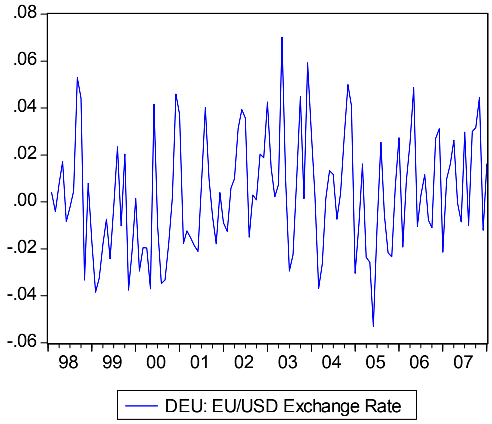
APPENDIX B

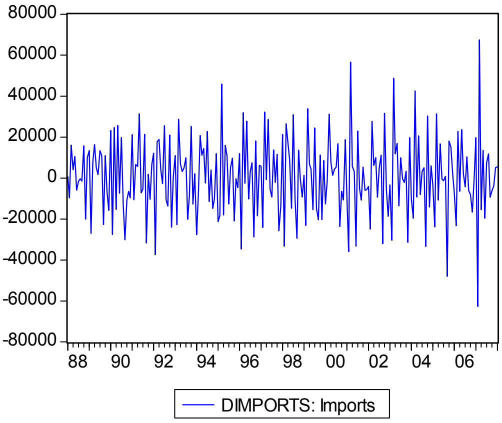
1. Transformed Explanatory Variables and Crude Oil Returns.

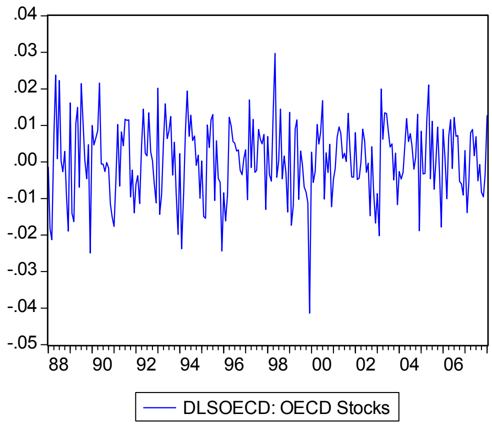


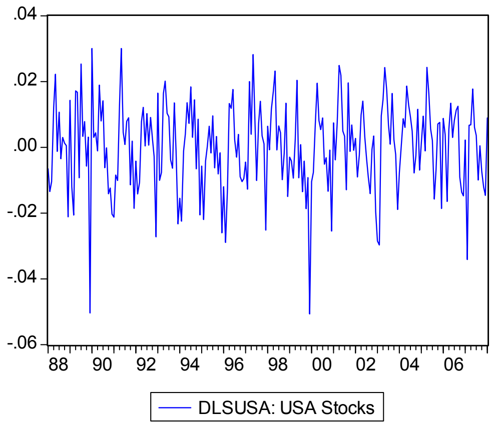

60000
40000
20000
0
-20000
-40000
-60000
-80000
88 90 92 94 96 98 00 02 04 06
DCL: Open Interest

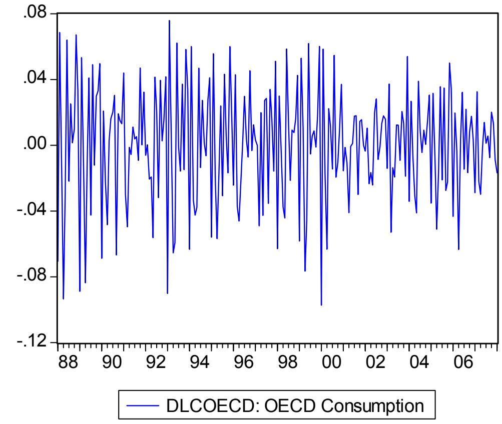

3
2
1
0
-1
-2
-3
88 90 92 94 96 98 00 02 04 06
DNG: Natural Gas Price

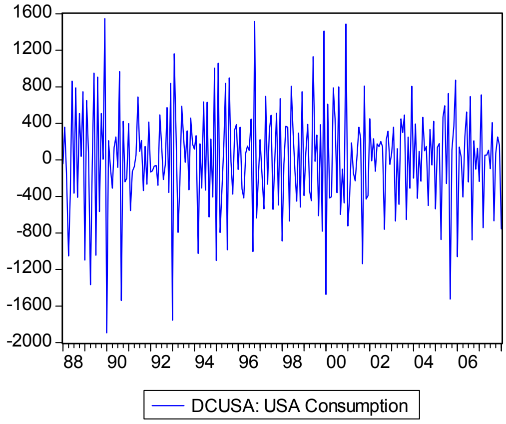

.08
.06
.04
.02
.00
-.02
-.04
-.06
98 99 00 01 02 03 04 05 06 07
DEU: EU/USD Exchange Rate

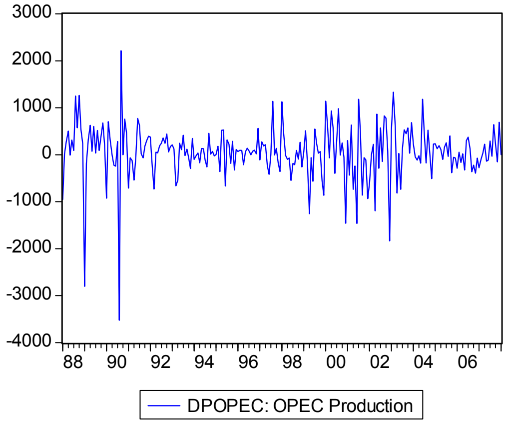

80000
60000
40000
20000
0
-20000
-40000
-60000
-80000
88 90 92 94 96 98 00 02 04 06
DIMPORTS: Imports

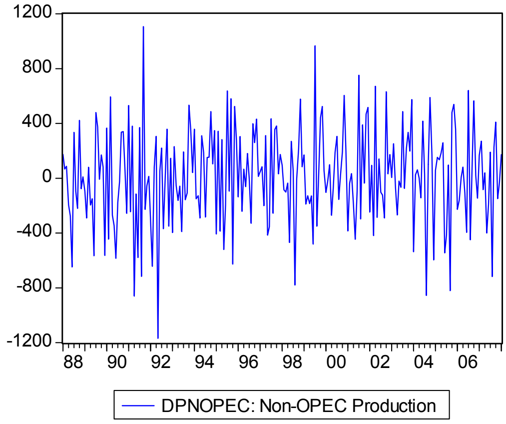

.04
.03
.02
.01
.00
-.01
-.02
-.03
-.04
-.05
88 90 92 94 96 98 00 02 04 06
DLSOECD: OECD Stocks

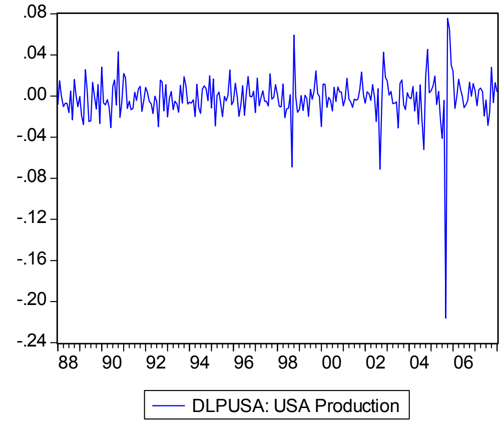

.04
.02
.00
-.02
-.04
-.06
88 90 92 94 96 98 00 02 04 06
DLSUSA: USA Stocks

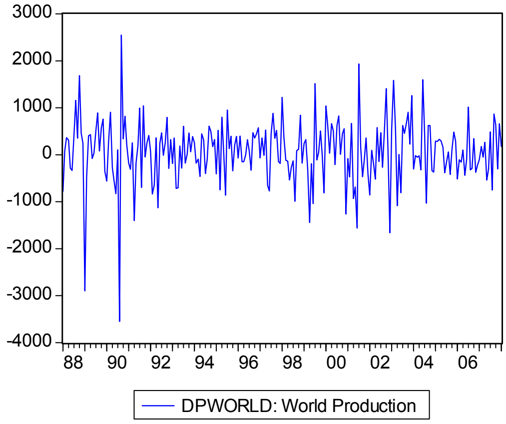

.08
.04
.00
-.04
-.08
-.12
88 90 92 94 96 98 00 02 04 06
DLCOECD: OECD Consumption


1600
1200
800
400
0
-400
-800
-1200
-1600
-2000
88 90 92 94 96 98 00 02 04 06
DCUSA: USA Consumption


3000
2000
1000
0
-1000
-2000
-3000
-4000
88 90 92 94 96 98 00 02 04 06
DPOPEC: OPEC Production


1200
800
400
0
-400
-800
-1200
88 90 92 94 96 98 00 02 04 06
DPNOPEC: Non-OPEC Production


.08
.04
.00
-.04
-.08
-.12
-.16
-.20
-.24
88 90 92 94 96 98 00 02 04 06
DLPUSA: USA Production


3000
2000
1000
0
-1000
-2000
-3000
-4000
88 90 92 94 96 98 00 02 04 06
DPWORLD: World Production