# MODELS FOR OIL PRICE PREDICTION AND FORECASTING

---

A Thesis

Presented to the

Faculty of

San Diego State University

---

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

in

Applied Mathematics

---

by

Fabian Torben Bosler

Fall 2010

# SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Fabian Torben Bosler:

Models for Oil Price Prediction and Forecasting

Barbara Ann Bailey

Barbara Bailey, Chair
Department of Mathematics and Statistics

S Shen

Samuel Shen
Department of Mathematics and Statistics

Marie-Eve Lachance
Department of Finance

July 23, 2010

Approval Date

# DEDICATION

Dedicated to my high school math teacher Mr. Buhman,
who encouraged me to live up to my full potential.
I would not be here today without him.

The night is darkest just before the dawn. And I promise you, the dawn is coming.

– Harvey Dent

# ABSTRACT OF THE THESIS

Models for Oil Price Prediction and Forecasting
by
Fabian Torben Bosler
Master of Science in Applied Mathematics
San Diego State University, 2010

The price of oil affects everyone, everyday. This thesis investigates two different approaches of oil price models for prediction and forecasting. The first is an structural model and is an application of the Stock-Watson method. This model is extended to a structural systematic model, applying the Johansen method. The second approach is a series of univariate time series models starting with ARIMA models, followed by an artificial neural network regression model. All the methods and models are justified with corresponding econometric tests and procedures. The results are very conclusive, as they confirm what was shown in earlier research on different time periods. It is confirmed that in terms of forecasting the nonlinear models perform the best, however the structural models help explaining effects and importance of single variables.

# TABLE OF CONTENTS

# LIST OF TABLES

PAGE

# LIST OF FIGURES

PAGE

# ACKNOWLEDGEMENTS

First and foremost, I want to thank everybody who had a part in the probably best year of my life. I want to thank Barbara Bailey for the excellent supervision; thank you for answering even the dumbest questions. I want to thank Samuel Shen, who as head of the department made this exchange year even possible. I want to thank Marie-Eve Lachance for the interesting class I had with you and the help you gave me with the thesis. I want to thank Carmelo Interlando for the horizon broadening classes I had the pleasure of attending. I want to thank all my friends and especially my girlfriend for their constant support and encouraging words.

# CHAPTER 1
# INTRODUCTION

Oil is a commodity, that unlike any other, affects everyones daily life in a plethora of ways. The modern world is as dependent on oil as the human body is dependent on blood. Oil prices and availability affects transportation be it everyday driving or flights, as well as economical growth, since goods have to be transported and oil is used almost everywhere in the secondary industry. Machines have to be oiled, engines need fuel and some products are entirely oil-based. The insulation on the wires in computers and the circuit boards are made from crude oil. Common products like shampoo, detergent, solvents, paint, ink, tires, lubricants, candle wax and hundreds of thousands of other products are made from oil. Heating and of course the military is heavily dependent on oil prices and availability. The uniqueness of oil and its use in the global economy makes it challenging to find another comparable resource. Furthermore, one must not forget that oil is a non-renewable fossil resource. Presently economic activity and oil consumption are fairly correlated. It is not only for the aforementioned reasons that the interest in oil prices and in particular, in the ability of being able to forecast oil prices is crucial.

## 1.1 MOTIVATION

In essence there are two different types of modelling approaches. The first is a structural model of the price of oil, depending on fundamental data such as demand and supply and is implemented through the use of a linear regression. The second is a time series approach, which includes linear and nonlinear time series analysis. The nonlinear time series analysis includes a neural network autoregressive model, where the complexity of the relationship is modeled with the help of neural network regressions.

Modelling the price of oil is difficult because of the changing variability over time. When the demand for a commodity like oil exceeds supply, the price will rise extremely high, which is due to the fact that demand and supply are quite inelastic in the short run. Even though people will be shocked about higher oil prices, it takes time to adjust habits and consumption. Supply cannot be increased that fast either. If production is on the margin of production then adding more capacity is expensive and takes time. Over time people will adjust their oil consumption which will therefore restore the demand-supply balance. In a period of excess demand, oil producers will use their previously uneconomic wells and rigs, as demand adjusts producers will shut off the most costly wells and rigs first. The wide range

of oil production cost results in increased medium-term volatility. Therefore shifts in demand will cause a much more drastic change in price than before.

A time series plot of the deflated price for crude oil is given in Figure 1.1. Emphasis has to be put on the end of 2007 and the beginning of 2008. A price increase like this has never been seen before and provides quite a challenge to the ambitious statistician. Because such price movements might seem unpredictable, it is a challenge to find a model that performs relatively well, when being confronted with such obstacles.

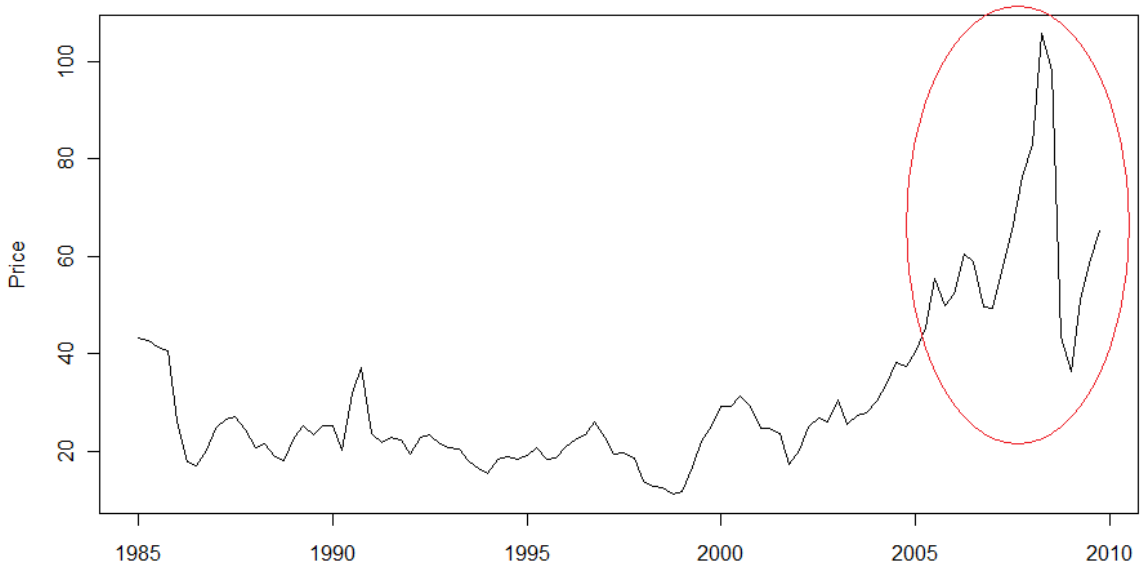


**Figure 1.1. History of crude oil prices.**

Another difficulty in modelling the price of oil arise because of the role of the Organization of Petroleum Exporting Countries (OPEC). OPEC can act as a swing producer and therefore it just covers the remaining demand for oil after non-OPEC supply is used up. In this case OPEC acts as a price taker, however OPEC could also produce at full capacity and take a more competitive role in global oil market, in which case OPEC acts as a price maker. The effect of a dominant producer is easily observed, but harder to model for two reasons: the first being the inability to forecast the behavior of said producer, the second one being the inability to translate a specific behavior into the model.

This thesis is divided into two main modelling approaches. Chapter 2 and Chapter 3 deal with slightly different structural models, and Chapter 4 describes the linear and nonlinear univariate time series models. Chapter 2 deals with a structural approach as proposed by [5]. The model they suggest expands an existing model based on fundamental demand and supply data by refinery utilization rates, a nonlinear effect of OPEC capacity utilization and

conditions in the futures market as explanatory variables. Their reasoning to include such factors is driven by the results of previous research. For example [6], specifies crude oil price as a function of OPEC capacity, OECD crude oil stocks, OPEC quotas and cheating by OPEC on those quotas. Those models perform well in-sample but consistently under-predict real oil prices out-of-sample. They argue in favor of a nonlinear relationship between oil prices and supply. However, my research does not indicate that a nonlinear factor of OPEC capacity utilization is statistical significant, but in fact the inclusion of nonlinearities yields extreme coefficients that are way too sensitive to even slightest changes in capacity utilization. Hence we decided against the inclusion of nonlinearities in chapter 2. The difference in our results from [5] might be caused by the fact that we used an expanded data set compared to theirs. The prevailing relationships between the explanatory variables and oil price is mainly explored with the help of cointegration and the conclusions based on the presence of cointegration amongst multivariate time series. Given that a cointegration relationship is present it is fairly easy to find an equilibrium equation, which describes the long run dynamics in the system. Furthermore an error correction model (henceforth ECM) is derived to represent the short run dynamics, which is also used to conduct forecasting. The second chapter sticks to the methodology and variables proposed in [5] and in essence confirms the results for the expanded time frame up to 2009.

Chapter 3 investigates the implications of multivariate vector autoregressive (VAR) models and the derived structural vector error correction models to determine short-run dynamics and conduct forecasting. This is a new approach to oil price predicting and forecasting and has to my knowledge not been done before. All the variables in a vector autoregressive model are treated symmetrically, that means an equation is estimated for each variable, which tries to explain its movement by its own lags and the lags of all the other variables in the system. However, no external regressors are included in a VAR model. In contrast to that, a structural vector autoregressive model tries to overcome the shortcomings of VAR models by allowing the explicit modelling of simultaneous interdependence between the left-hand side variables. The first step in chapter 3 was to estimate the optimal number of lags to be used for the VAR model, therefore a set of information criteria were applied namely the “Akaike information criterion” as in [2], “HannanQuinn information criterion” as in [10] and the “Schwarz information criterion”, which is closely related to the AIC with the difference that the penalty for excess parameters is greater. After the successful determination of the number of lags needed, a VAR model was estimated for the different possible candidates. The residuals of these VAR models were then tested for normality, serial correlation and the presence of ARCH components. Those tests narrowed the possible candidates further down. Based on the best model a structural vector error correction model (SVEC) was then

estimated by applying some economic restrictions to identify the model. Based on the SVEC model impulse response analysis and forecasting was then conducted.

Chapter 4 investigates a pure univariate time series approach then, in a first step the best fitting linear model is determined. The model selection is done following the Box-Jenkins approach and an ARIMA model is concluded to be the best fitting model. The residuals of the ARIMA model are then checked for any remaining patterns, more precisely nonlinearities. As it turns out, there are indeed patterns remaining, that could not be captured by the linear ARIMA model. The heteroskedasticity in the model is then addressed by fitting a GARCH model. The GARCH model is able to capture the nonlinearity accordingly. The heteroskedasticity in the model is explained by the extreme residuals caused by the immense price swings between 2007 and 2008. In the final analysis an artificial neural network autoregression is fit to the data. An artificial neural network consists of an interconnected group of artificial neurons and computes the output by fitting a linear combination of a nonlinear transformation of the linear combination of the explanatory variables. After fitting the artificial neural network, two different bootstrap approaches for prediction interval determination are investigated, where it turns out that a nonparametric bootstrapping pairs approach achieves better results. An example code for bootstrapping residuals can be found in the Appendix. In the last part of chapter 4 forecasting is conducted based on the fitted neural network.

## 1.2 DATA

We used quarterly data (1986Q1-2009Q4) for average freight on board (F.O.B.) price for all crude oil imported by the US, because it represents the price paid for physical barrels obtained from various sources. OPEC capacity utilization, to represent OPEC's behavior, OPEC's behavior as a dominant producer is hard to predict and therefore to model. Basically OPEC can have two roles, the role of price taker, i.e. just covering the GAP between global demand and non OPEC production, or OPEC can act as a price maker. Being a price maker means producing at maximum capacity to influence the market. OPEC capacity utilization is obtained by dividing the theoretical maximal possible OPEC production by the actual production. It should be kept in mind, that OPEC enacts production quotas on their member countries to control the price. However, not all countries obey their quotas, which means in fact that the countries produce more crude oil then they are entitled to, which then in return lowers the price. Cheating on quotas however is only marginal and therefore captured by an increase in actual production. To implement the effect of conditions in the refining sector on crude oil prices, we collected data on US refinery utilization rates, which vary between 0% and 100%. As stated in [5], global data would be ideal, but US refinery capacity utilization is a satisfactory proxy. Their reasoning is that in a globalized market it is unlikely that utilization

rates in one part of the world will decouple dramatically from other parts, as long as one can ship refined petroleum products. Previous research has shown that the prices of crude oil in different regions of the world are cointegrated, as stated in [3]. If refinery utilization affects crude oil prices, cointegration among prices imposes that refinery utilization rates in different parts of the world also share the same stochastic trend. Assuming refinery utilization rates do not affect crude prices, statistical tests will reject the hypothesis of a relationship between refinery utilization rates and prices, no matter which refinery utilization is used in the model. To include the effect of conditions in the New York Mercantile Exchange, the quarterly average of the near month contract and the quarterly four-month contract for West Texas Intermediate (Cushing, dollars per barrel) of all the days said contracts were traded were included. The difference between the four-month contract and the near month contract is then used to determine whether the market is in contango (the price for four month contracts is greater than for near month contracts) or in backwardation (the opposite). Finally the GDP-deflator is obtained and applied to the F.O.B. price to get real dollar prices.

# CHAPTER 2
# EXPLORING A STRUCTURAL MODELLING APPROACH

A structural modelling approach includes explanatory variables other than just the past data of oil prices into the process. Examples would be demand, consumption, car sales and so on.

## 2.1 METHODOLOGY

To find a meaningful relationship between the explanatory variables and crude oil price we start by examining whether or not a cointegration coherence is present. Cointegration basically means that a vector of variables, which achieve stationarity after differencing, might have linear combinations that already are at or achieve stationary levels. [8] pointed out that integrated variables might share an equilibrium relationship that is either stationary or of a lower degree than the original series. In other words, cointegration is an indicator for co-movements among trending variables that can be used to test for the existence of an equilibrium relationship in the dynamic system. An equilibrium relationship or cointegrating relationship is helpful, because it represents the long-run equation of said system. Based on the equilibrium equation, an error correction model can be derived, which models the short term dynamics, and can therefore be used to forecast the next period given changes in the explanatory variables.

Various statistical methods are available to test for cointegration (e.g., [8], [11], [18]). We applied the Engle-Granger two step procedure with unsatisfying results, since the error correction term of the short run equation was only statistical significant at the 10% level. Furthermore the two step procedure suffers certain disadvantages for example extremely unreliable standard errors, which makes statistical testing on the long run model impossible. The Engle-Granger two step procedure consists of the following two steps:

- Test the variables to see if they are nonstationary (e.g. use ADF).
- In case of two or more variables being nonstationary, test whether the variables are cointegrated.
    1. Use OLS to regress the dependent variable against the independent variables to obtain the long run equation.

$$x_{1,t} = \beta_0 + \beta_1 x_{2,t}$$

where $x_{1,t}$ is the dependent variable, $x_{2,t}$ is the explanatory variable and $\beta_0, \beta_1$ are the regression coefficients.

2. Test the residuals of the above equation for cointegration, by testing for unit roots. If the residual series is stationary, conitegration is present.

We applied the Stock-Watson Dynamic Ordinary Least Squares (DOLS) method, because the Stock-Watson DOLS method differs from the Engle-Granger two step procedure in estimating the coefficients for the long rung equation. The coefficients are estimated with the dynamic ordinary least squares method. This is done because the Engle-Granger two step procedure yields very inconsistent standard errors, and therefore statistical testing is difficult. The DOLS equation is of the following form

$$y_t = \beta_0 + \sum_{i=1}^{n} \beta_i x_{i,t} + \sum_{i=1}^{n} \sum_{j=-k1}^{+k2} \gamma_{i,j} \Delta x_{i,t-j} + \epsilon_t \tag{2.1}$$

where $y_t$ denotes the dependent variable, $\beta_0$ is the intercept, $\beta_1, \ldots, \beta_n$, are the coefficients used in equation (2.2), $k1$ and $k2$ denote leads(future) and lags(past) of the first order differenced explanatory variables. Usually $k1 = k2$ and the number of lags is determined by evaluating the Akaike Information Criterion, as in [2], which is henceforth refered to as AIC. Then the residuals are tested for cointegration just like in the Engle-Granger two step procedure. In a first step the variables are tested for nonstationarity by conducting a set of tests, mainly the augmented Dicky-Fuller (ADF), which can be found in [7] and the Hyllenberg-Engle-Granger-Yoo test (HEGY), to be found in [11] to test for seasonal unit roots. The univariate HEGY tests include an intercept, time trend and seasonal dummy variables. The HEGY statistic calculated based on the DOLS regression does not include an intercept, time trend or seasonal dummy variables because we expect the error terms to be stationary without a drift and a trend. Otherwise that would have been captured in the DOLS model. The significance levels are from [11]. The univariate ADF tests also include an intercept, time trend and seasonal dummy variables. The ADF tests for the residuals of the long term model do not include an intercept or trend for similar reasons. The respective augmented lag length is chosen by applying AIC. The basic regression for the HEGY statistic, for simplicity with no intercept, time trend or seasonal dummy variables is:

$$\Delta_4 y_{s\tau} = \pi_1 y^{(1)}_{s-1,\tau} + \pi_2 y^{(2)}_{s-1,\tau} + \pi_3 y^{(3)}_{s-2,\tau} + \pi_4 y^{(3)}_{s-1,\tau} + u_{s\tau}$$

where, $\Delta_4 = 1 - L^4$ with $L$ the usual lag operator ($Ly_{s\tau} = y_{s-1,\tau}$). Here $y_{s\tau}$ is $y$ at season $s$ and year $\tau$, that means $s = 1, 2, 3, 4$ and $\tau = 1, 2, \ldots, T$, $u_{s\tau}$ is white noise with constant variance. The auxiliary variables $y^{(1)}_{s\tau}, y^{(2)}_{s\tau}, y^{(3)}_{s\tau}$ are associated with the presence of the factors $(1 - L), (1 + L)$ and $(1 + L^2)$, respectively, of the annual difference $\Delta_4 = 1 - L^4 =$

$(1-L)(1+L)(1+L^2)$.
Specifically,

$$y_{s,\tau}^{(1)} = (1+L)(1+L^2)y_{s\tau}$$
$$y_{s,\tau}^{(2)} = -(1-L)(1+L^2)y_{s\tau}$$
$$y_{s,\tau}^{(3)} = -(1-L)(1+L)y_{s\tau} \quad .$$

Hyllenberg, Engle, Granger and Yoo showed that when $\pi_1 = 0$ the series contains the (nonseasonal or zero frequency) root 1, when $\pi_2 = 0$ the (semiannual) root $-1$ is present. the presence of the (annual) roots $\pm i$ ($i = \sqrt{-1}$) implying $\pi_3 = \pi_4 = 0$ (the stationary alternatives being $\pi_1 < 0, \pi_2 < 0$ and $\pi_3 < 0$ and/or $\pi_4 \neq 0$). Thus, inference on the presence of seasonal unit roots may be carried out through the t-statistics associated to the last three $\pi_i$ coefficients: $t_{\pi_2}$ , $t_{\pi_3}$ and $t_{\pi_4}$. On the other hand, evidence on the presence (absence) of a nonseasonal unit root is given by $t_{\pi_1}$ . However, the analysis of stochastic seasonal nonstationarity becomes simpler if, instead of testing three separate hypotheses, some joint null hypotheses are tested. To that end, one can use the $F$-statistics $F_{34}$, which tests $H_0 : \pi_3 = \pi_4 = 0$.
ADF testing procedure is applied to following model:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p} + \epsilon_t$$

where $\alpha$ is a constant, $\beta$ is the coefficient on a time trend, $p$ is the augmented lag length, $\delta_1, \ldots, \delta_p$ are coefficients of the structural effects and $\epsilon_t$ is white noise. The null hypothesis $H_0 : \gamma = 0$ is tested against $\gamma < 0$. The test statistic is:

$$ADF = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad .$$

The distribution of $ADF$ can be found in the Dicky-Fuller table. The augmented lag length has to be determined when applying the test. We used the Akaike Information Criterion (AIC), as in [2] to determine the best lag length. Following function is minimized:

$$AIC = 2k - 2ln(L)$$

where $k$ is the number of parameters in the statistical model and $L$ is the maximized value of the likelihood function of the estimated model.

## 2.2 TEST RESULTS

The variable $Price$ is again the average freight on board (F.O.B.) price. $Days$ is the number of days of forward consumption, that means OECD stocks divided by OECD demand.

$Caputil$ is the OPEC capacity utilization is obtained by dividing theoretical maximal production by actual production. $Refutil$ is the refinery utilization of American refineries, and reason for the inclusion can be found in [5]. $MM$ is market movement or market conditions and is obtained by subtracting the near month future contract (NYMEX1) from the four month future contract (NYMEX4), which represents general expectation towards the market. As mentioned in section 2.1 ADF and HEGY tests are conducted to determine whether the individual variables are stationary or have annual/seasonal unit roots. Results in (Table 2.1 - panel: Univariate tests) indicate that said variables contain an annual root (indicated by the ADF statistic) and no seasonal roots (indicated by the results that $\pi_2 = 0$ and a joint test $F\pi_3 \cap \pi_4 = 0$ is generally rejected). The ADF statistic rejects the hypothesis of no unit root for all variables, which in turn is confirmed by the $\pi_1$ statistic, which fails to reject the hypothesis of an existing annual unit root, with the exception of $Caputil$.

**Table 2.1. ADF and HEGY Statistics for Annual and Seasonal Unit Roots**

| | ADF | $\pi_1$ | $\pi_2$ | $\pi_3$ | $\pi_4$ | $F\pi_3 \cap \pi_4$ |
|---|---|---|---|---|---|---|
| **Univariate tests** | | | | | | |
| $Price$ | 1.627 | 1.81 | -5.23** | 2.23 | 0.15 | 12.51** |
| $Days$ | 2.227 | 1.64 | -1.34 | -0.56 | -2.09* | 12.20** |
| $Caputil$ | -2.956 | -3.69* | -4.54** | -1.88 | -5.85** | 19.83** |
| $Refutil$ | -0.882 | 0.30 | -2.72+ | -4.71** | -3.44** | 22.36** |
| $MM$ | -3.364+ | -2.04 | -5.82** | -3.95* | -1.12 | 8.09+ |
| **Equilibrium residuals** | | | | | | |
| Equation (2.2) | -5.639** | -6.05** | -4.64** | -4.86** | -2.28* | 12.75** |

Signif. codes: **: P $<$ .01, *: p $<$ .05, +: p $<$ .10

A time series is said to be integrated of order d, in short, I(d), if it achieves stationarity after differencing d times. We can conclude that there are many I(1) trends among the variables, which means the variable is integrated of order one, or in other words is stationary after differencing once, therefore a simple ordinary least squares regression cannot be applied without further adjustments. Due to the presence of I(1) variables, significant relationships among the I(1) variables will be found more often than implied by random chance, as described in [9]. Those relationships are named spurious regressions and are caused by an unseen third factor, which usually is a lurking variable or a confounding variable. To prevent the occurrence of spurious regressions, we tested the variables for cointegration by applying a Phillips-Perron (PP) test, as in [17] and also the aforementioned ADF and HEGY tests on the

residuals $\mu_t$ of the long term equilibrium model given by:

$$\begin{aligned} Price_t =& \beta_0 + \beta_1 Days_t + \beta_2 Caputil_t + \beta_3 Refutil_t \\ &+ \beta_4 (NYMEX4_t - NYMEX1_t) + \mu_t \quad . \end{aligned} \tag{2.2}$$

The coefficients of the equation can be found in Table 2.2.

**Table 2.2. Estimates for Price Equation with Standard Errors in Parenthesis**

| | US F.O.B. Price |
|---|---|
| **Cointegrating Relation equation (2.2)** | |
| Constant | $485.51^{**}$ (64.07) |
| $Days$ | $-2.56^{**}$ (0.39) |
| $Caputil$ | $56.10^{**}$ (10.73) |
| $Refutil$ | $-2.56^{**}$ (0.44) |
| $MM$ | $9.12^{**}$ (1.41) |
| $R^2$ | 0.94 |

Signif. codes: **: P $<$ .01, *: p $<$ .05, +: p $<$ .10

The PP (test-statistic: -6.2755, which is significant at the .01 level), ADF and HEGY (Table 2.1, panel: DOLS residuals) tests clearly show that the residuals are stationary therefore we can conclude that the variables cointegrate. If the residuals were not stationary the nonstationary residuals would indicate a spurious relationship. Since the residuals are stationary we can deduce that the relationship given in equation (2.2) describes the long term equilibrium relationship between price and the right hand side variables.

The coefficients in equation (2.2) are estimated by applying a dynamic least squares model proposed by Stock and Watson, as suggested in [18]. The estimation yields super-consistent estimates for $\hat{\beta}$, in the sense that the OLS estimator $\hat{\beta}$ converges in probability to its true value $\beta$ at a rate proportional to the inverse of the sample size, $T^{-1}$ , rather than at $T^{-1/2}$ as is the standard result in the ordinary case where the variables are I(0), as shown in [18]). The earlier mentioned DOLS equation (equation 2.1) is used to estimate the coefficients in equation(2.2).

Since a cointegrating relationship between oil prices and the explanatory variables could be found, an error correction model can now be estimated to examine the short run

dynamics among the variables. The error correction model (ECM) is obtained by using OLS to estimate the parameters of following equation:

$$\begin{aligned}\Delta Price_t = k + \rho\mu_{t-1} + \sum_{i=1}^{s}\lambda_{1i}\Delta Price_{t-i} + \sum_{i=1}^{s}\lambda_{2i}\Delta Days_{t-i} + \sum_{i=1}^{s}\lambda_{3i}\Delta Captuil_{t-i} \\ + \sum_{i=1}^{s}\lambda_{4i}\Delta Refutil_{t-i} + \sum_{i=1}^{s}\lambda_{5i}\Delta(NYMEX4_{t-i} - NYMEX1_{t-i}) \\ + \delta_1 Q1 + \delta_2 Q2 + \delta_3 Q3 + \delta_4 War + \epsilon_t\end{aligned} \tag{2.3}$$

in which $\Delta$ is the first difference operator, Q1,Q2 and Q3 are dummy variables for the first second, and third quarters respectively. $War$ is a dummy variable for the first Persian Gulf War (1990Q3-1990Q4) and $\mu_t$ are the residuals from the long term equilibrium equation estimated in (2.2). We obtained the number of lags (s) for the explanatory variables by using the Akaike information criterion, as in [2]. Table 2.3 shows the adjustment rate and the $R^2$ of the error correction model. Further attention needs to be paid to the statistical significance of $\rho$ in equation (2.3). The significance of $\rho$ evaluates the hypothesis that the error correction term is meaningful. Furthermore, it tells how much of a disequilibrium will be corrected within a period. A negative value for $\rho$ shows that disequilibrium between oil prices and the right-hand side variables adjusts price towards the equilibrium value, which is given by the cointegration relationship.

**Table 2.3. Estimates for Error Correction Model**

| | US F.O.B. Price |
|---|---|
| **Short run Dynamics equation 2.3)** | |
| Adjustment rate ($\rho$) | $-0.32^{**}$ (0.06) |
| $R^2$ | 0.64 |

Signif. codes: **: P $<$ .01, *: p $<$ .05, +: p $<$ .10

The coefficients for the equilibrium relationship are given in Table 2.2 and are consistent in signs with what we would expect a priori. The effect of $Days$ is negative. An increase in days of forward consumption makes OECD countries less dependent on immediate availability of oil and therefore has a negative effect on price. The sign of capacity utilization by OPEC ($Caputil$) is positive, which is on line with expectations, because if OPEC has to increase their capacity utilization more costly wells have to be used, also an increase is only possible to a certain extent. This in return increases prices as OPEC capacity utilization gets closer to 100%. The effect of refinery utilization ($Refutil$) is negative, a lengthy explanation therefore can be found in [5]. In short the effect is associated with

changes in availability of different crude oil densities. As refinery utilization reaches 100% less valuable densities have to be used for refinerization. Because producers are only set up to use so much of the more expensive and easier to process densities. Also, it is not very surprisingly that the coefficient of $MM$ is positive, which means that contango has a positive effect on prices. Contango means that the four month futures contract is more expensive than the one months future contract, backwardition is the opposite. This is fairly intuitive because it just means, that the market expects oil to become more expensive in the future.

Results for $\rho$ in Table 2.3 show that the cointegrating relationship given by equation (2.2) can be interpreted as a long run equilibrium relationship and equation (2.1) gives the short run dynamics. The parameter $\rho$ is negative, which means a disequilibrium between the right hand variables and price in equation (2.2) will be corrected towards the long run equilibrium. The point estimate for $\rho$ implies that 32% of the difference between the equilibrium and the observed price for crude oil is eliminated within one period.

## 2.3 FORECASTS

A one-step out-of-sample-forecast is conducted based on the short run dynamics equation (2.1). The forecast is graphically displayed in Figure 2.1.

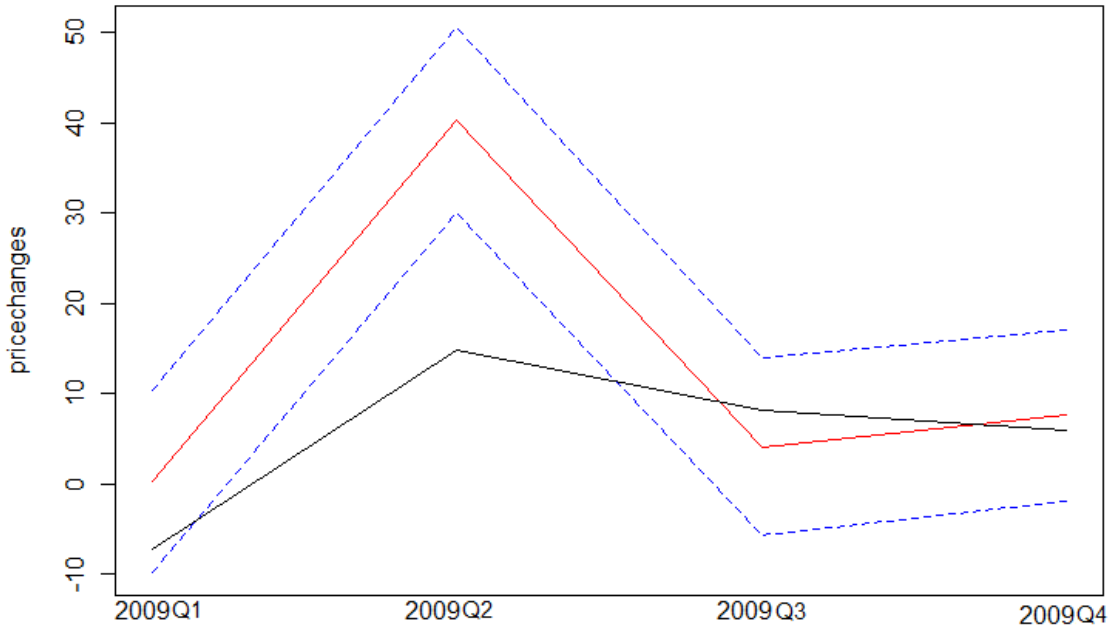


**Figure 2.1. Forecasting 2009 by applying a structural ECM model.**

We will forecast only 2009 because of the extreme price fluctuations between 2007 and 2008. The extreme price increase during 2007 is caused by excess demand and is enhanced by speculations which are in this extent impossible to foresee. The rapid price drop

after 2007 is a correction back to the fair price. Hence 2007 and 2008 are two important years to include in the model to be able to make accurate predictions. The forecast was obtained by using data from the previous periods to estimate price for the actual period by using equation (2.1). The confidence interval is calculated based on following formula:

$$\hat{\beta}^T x_0 - Z_0 < \varphi(x) < \hat{\beta}^T x_0 + Z_0$$

where,

$$Z_0 = t_{n-m,1-\alpha/2} S \sqrt{x_0^T (X^T X)^{-1} x_0}$$

and $\varphi(x)$ is the error correction model, which is estimated in equation (2.3), $n$ is the number of observations, $m$ the number of parameters, $(1 - \alpha/2)$ the confidence level, $S$ the standard error of equation (2.3) and $X$ is the data matrix that was used to estimate equation (2.3). We used a 95% level of confidence to calculate the confidence interval. The dashed lines indicate the confidence interval, the red line is the prediction and the black line are the actual observed values.

# CHAPTER 3
# MULTIVARIATE STRUCTURAL MODELLING APPROACH TO COINTEGRATION

A different approach can be taken by applying the Johansen procedure described in [12]. The main differences in the Johansen procedure is that the tests are maximum likelihood based, that more than one cointegrating relationship can be found and that the multivariate test is based on vector autoregressive models (henceforth VAR) where the EG method, described in [8] is based on a single OLS estimation.

## 3.1 METHODOLOGY

A VAR model tries to explain the endogenous variables based entirely on their own past history and no deterministic regressors. Derived from the VAR model are structural autoregressive models (henceforth SVAR). Those models make short term dynamic modelling possible. The VAR and SVAR models were then unified with the theory of cointegration into vector error correction models (VEC) and structural error correction model (SVEC). An overview for said models can be found in [13] and [14]. We will now provide a brief overview of VAR, SVAR, VEC and SVEC.

### 3.1.1 Vector Autoregressive Model

A basic VAR($p$)-process is of the following form:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{1-p} + u_t \tag{3.1}$$

where $y_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})$ is a vector consisting of K endogenous variables. In our case this means, $y_{1t}$ represents $Price$, $y_{2t}$ represents $Days$, $y_{3t}$ represents $Caputil$ and so forth. The $A_i$ are $K \times K$ coefficient matrices $\forall\ i = 1, \ldots, p$ and $u_t$ is a K-dimensional white noise process (i.e., $E(u_t) = 0$ and $E(u_t u_t^T) = \Sigma_u$).

To explore forecasting and diagnosing the model's short term behavior, i.e., impulse response functions and forecast error variance decomposition one has to understand the Wold moving average decomposition of a VAR($p$) process. The Wold infinite moving average decomposition is defined as:

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \Phi_2 u_{t-2} + \cdots \tag{3.2}$$

where $\Phi_0 = I_K$ and $\Phi_i$ can be calculated recursively by:

$$\Phi_i = \sum_{j=1}^{i} \Phi_{i-j} A_j \ \forall i = 1, 2, \ldots \tag{3.3}$$

where $A_j = 0$ for $j > p$. Forecasts for horizon $h \geq 1$ of a VAR($p$) process can be computed recursively by:

$$y_{T+h|T} = A_1 y_{T+h-1|T} + \cdots + A_1 y_{T+h-p|T}$$

with $y_{T+h|T} = E[Y_{T+h}|Y_T, Y_{T-1}, \ldots]$ and $y_{T+j|T} = y_{T+j}$ for $j \leq 0$. The forecast error covariance matrix is of the following form:

$$Cov\left(\begin{bmatrix} y_{T+1} - y_{T+1|T} \\ \vdots \\ y_{T+h} - y_{T+h|T} \end{bmatrix}\right) = \begin{pmatrix} I & 0 & \cdots & 0 \\ \Phi_1 & I & & 0 \\ \vdots & 0 & \ddots & \vdots \\ \Phi_{h-1} & \Phi_{h-2} & \cdots & I \end{pmatrix} (\Sigma_u \otimes I_h) \begin{pmatrix} I & 0 & \cdots & 0 \\ \Phi_1 & I & & 0 \\ \vdots & 0 & \ddots & \vdots \\ \Phi_{h-1} & \Phi_{h-2} & \cdots & I \end{pmatrix}^T$$

and the matrices $\Phi_i$ are the empirical coefficient matrices of the Wold moving average representation. The $\otimes$ operator is the Kronecker product.

### 3.1.2 Structural Vector Autoregressive Model

A VAR($p$) process can be seen as a reduced form model. A SVAR model is its structural representation and has the following form:

$$Ay_t = A_1^* y_{t-1} + \cdots + A_p^* y_{t-p} + B\epsilon_t \ . \tag{3.4}$$

The structural error terms $\epsilon_t$ are assumed to be white noise with $\sigma^2 = 1$. Also the coefficient matrices $A_i^*$ for $i = 1, \ldots, p$ differ in general from their reduced form counterparts. This becomes obvious by left hand multiplying the above equation by $A^{-1}$, we will then have:

$$\begin{aligned} y_t &= A^{-1}A_1^* y_{t-1} + \cdots + A^{-1}A_p^* y_{t-p} + A^{-1}B\epsilon_t \\ &= A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t \ . \end{aligned}$$

A SVAR model can be used to identify shocks and trace these out by employing impulse response functions through imposing restrictions on the matrices $A$ and/or $B$. It should be noted that the reduced form residuals can be retrieved from a SVAR model by $u_t = A^{-1}B\varepsilon_t$ and its variance-covariance matrix by $\Sigma_u = A^{-1}BB^T A^{-1^T}$. Three kinds of restrictions and their respective SVAR models have to be distinguished.

1. $A$ model: $B$ is set to $I_K$
   (minimum number of restrictions for identification is $K(K-1)/2$).

2. $B$ model: $A$ is set to $I_K$
   (minimum number of restrictions for identification is $K(K-1)/2$).
3. $AB$ model: Restrictions can be placed on both matrices
   (minimum number of restrictions for identification is $K^2 + K(K-1)/2$).

Assuming a Gausian white noise error process the parameter estimation is achieved by minimizing the negative of the concentrated log-likelihood function:

$$lnL_c(A,B) = -\frac{KT}{2}ln(2\pi) + \frac{T}{2}ln|A|^2 - \frac{T}{2}ln|B|^2 - \frac{T}{2}tr(A^T(B^{-1})^T B^{-1} A\tilde{\Sigma}_u)$$

where $\tilde{\Sigma}_u$ is an estimate of the variance-covariance matrix for the error terms of the reduced model and $T$ is as above the number of observations.

### 3.1.3 Vector Error Correction Model

For the VAR equation in section 3.1.1 an error correction model exists, which is of the following form:

$$\Delta y_t = \alpha\beta^T y_{t-p} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + \mu_t \tag{3.5}$$

with

$$\Gamma_i = -(I - A_1 - \cdots - A_i),\ i = 1, \ldots, p-1$$

and

$$\alpha\beta^T = \Pi = -(I - A_1 - \cdots - A_p)\ .$$

The $\Gamma_i$ matrices contain the cumulative long-run impacts. The other specification is of the following form:

$$\Delta y_t = \alpha\beta^T y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + \mu_t \tag{3.6}$$

with

$$\Gamma_i = -(A_1 - \cdots - A_i),\ i = 1, \ldots, p-1$$

and

$$\alpha\beta^T = \Pi = -(I - A_1 - \cdots - A_p)\ .$$

Even though the $\Pi$ matrix is the same in both specifications the $\Gamma_i$ matrices differ, in the second specification the $\Gamma_i$ measure transitory effects where they measure cumulative long-run impacts in the first specification. If cointegration is present the matrix $\Pi = \alpha\beta^T$ is of reduced rank. The dimensions of $\alpha$ and $\beta$ are $K \times r$ where $r$ is the cointegration rank, which indicates how many long-run relationships in between $y_t$ exist. Matrix $\alpha$ is the loading matrix and $\beta$ contains the long-run relationships.

### 3.1.4 Structural Vector Error Correction Model

Applying the same logic that was used to obtain the VEC from the VAR we can transform the SVAR into a SVEC. When the SVEC model is specified and estimated a $B$ model is assumed. That means in other words the matrix $A$ is set to $I_K$. The model is then of the following form:

$$\Delta y_t = \alpha\beta^T y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + B\epsilon_t$$

with $u_t = B\epsilon_t$ and $\epsilon_t \sim N(0, I_K)$. This can be put to use by looking into the Beveridge-Nelson moving average representation of the variables $y_t$ assuming a transitory specification is estimated firstly. The Beveridge-Nelson decomposition is then of the following form:

$$y_t = \Xi \sum_{i=1}^{t} u_i + \sum_{i=j}^{\infty} \Xi_j^* u_{t-j} + y_0^* \ .$$

The Beveridge-Nelson decomposition divides the variables contained in $y_t$ into a part that is integrated of order one and a part that is not integrated. The integrated part is referred to as "common trends" in this case $\Xi \sum_{i=1}^{t} u_i$. The expresion $\sum_{i=j}^{\infty} \Xi_j^* u_{t-j}$ is assumed to be bound, in other words $\Xi_j^*$ converges to zero as $j \to \infty$. In $y_0^*$ the starting conditions are captured. Particular interest lies in long-run effects and shocks, which are incorporated in $\Xi \sum_{i=1}^{t} u_i$. The $\Xi$ matrix is of reduced form and has rank $K - r$, where $r$ is the number of cointegrating relationships within the system, and:

$$\Xi = \beta_\perp \left[ \alpha_\perp^T \left( I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha_\perp^T \ .$$

Since $\Xi$ is of reduced rank only the $K - r$ common trends drive the system. Hence by knowing the rank of $\Pi$ it can be seen that a maximum of $r$ structural errors can have a transitory effect. Therefore at most $r$ columns of $\Xi$ can be set to to zero. After substituting $u_t$ with $B\varepsilon_t$ in the Beveridge-Nelson decomposition the common trends term becomes then $\Xi B \sum_{t=1}^{\infty} \epsilon_t$, the long-run effects of the structural innovations are embodied in the matrix $\Xi B$. The contemporaneous effects of the structural errors can be found in matrix $B$. As in the SVAR models of type $B$ a number of $\frac{1}{2}K(K-r)$ restrictions are needed for the identified SVEC model. The cointegration structure imposes $r(K-r)$ restrictions on the long-run matrix. Of the remaining restrictions at least $r(r-1)/2$ have to be put on the contemporaneous matrix $B$.

## 3.2 TEST RESULTS

We conducted an univariate ADF test to make sure that the variables are integrated of order 1. The ADF test included a time trend and a constant. The results can be found in Table

3.1. It is noteworthy that the results are different than those obtained in Table 2.1. The simple explanation is, two different **R** [1] tests were conducted, the reason therefore is, it is easier to to obtain the deterministic terms with the second one. The inclined reader may look into the package details of both test. The first test conducted is $\mathbf{ADF.test}$ the second one is $\mathbf{ur.df}$. The results are in line with what we would expect from Table 2.1. We clearly fail to reject the null hypothesis of a unit root for $Days$, $Caputil$ and $Refutil$. $FOBdef$ (which is $Price$) and $MM$ are only significant at the 5% level. However it is apparent that we have to reject the null hypothesis of a unit root for all the differenced variables, hence the variables are all $I(1)$.

**Table 3.1. Detailed Results for ADF Test**

| Variable | Deterministic terms | Test statistic | Critical Values | | |
|---|---|---|---|---|---|
| | | | 1% | 5% | 10% |
| $FOBdef$ | constant, trend | -3.80 | -4.04 | -3.45 | -3.15 |
| $\Delta FOBdef$ | constant | -9.46 | -3.51 | -2.89 | -2.58 |
| $Days$ | constant | -3.05 | -4.04 | -3.45 | -3.15 |
| $\Delta Days$ | - | -17.34 | -3.51 | -2.89 | -2.58 |
| $Caputil$ | constant | -3.22 | -4.04 | -3.45 | -3.15 |
| $\Delta Caputil$ | - | -6.87 | -3.51 | -2.89 | -2.58 |
| $Refutil$ | constant | -3.15 | -4.04 | -3.45 | -3.15 |
| $\Delta Refutil$ | - | -13.45 | -3.51 | -2.89 | -2.58 |
| $MM$ | constant, trend | -3.79 | -4.04 | -3.45 | -3.15 |
| $\Delta MM$ | constant | -7.90 | -3.51 | -2.89 | -2.58 |

The next step is to find the appropriate lag length. This was done by employing 4 different information criteria. Following OLS equation was evaluated to determine the optimal lag length: (for simplicity without trend, drift and seasonal dummy)

$$y_t = A_1 + y_{t-1} + \cdots + A_p y_{t-p} + u_t \ .$$

Based on the sample size following criteria are computed:

$$AIC(\rho) = \ln\det(\tilde{\Sigma}_u(\rho)) + \frac{2}{T}\rho K^2$$
$$HQ(\rho) = \ln\det(\tilde{\Sigma}_u(\rho)) + \frac{2\ln(\ln(T))}{T}\rho K^2$$
$$SC(\rho) = \ln\det(\tilde{\Sigma}_u(\rho)) + \frac{\ln(T)}{T}\rho K^2$$

with $\tilde{\Sigma}_u(\rho) = T^{-1}\sum_{t=1}^{T}\hat{u}_t\hat{u}_t^T$ and $\rho$ assigns the lag order. The HannanQuinn information criterion, which can be found in [10] can be used as an alternative to the AIC and the SC. The results can be found in Table 3.2. According to the AIC the optimal lag number should be 4, the HQ criterion suggests 2 and the SC criterion suggests 1 lag. All of these options will be

**Table 3.2. Suggested Number of Lags**

| | AIC(n) | HQ(n) | SC(n) |
|---|---|---|---|
| Proposed number of lags | 4 | 2 | 1 |

considered in the next steps, and based on further diagnostic tests the most appropriate number will then be picked. A first test that we conducted on the different VAR(p) models was a test for serial correlation in the residuals of the respective VAR(p) model, or more precisely the lack of serial correlation. The test conducted is a Portmanteau test with following test statistic:

$$Q_h = T \sum_{j=1}^{h} tr(\hat{C}_j^T \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1})$$

where $\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^{T} \hat{u}_t \hat{u}_{t-i}^T$. The test statistic is approximately distributed as $\chi^2(K^2(h-p))$. With $h$ being chosen to be 16 and $p$ being the number of coefficients excluding the deterministic terms of a VAR(p) model.

The next test we conducted was a test for normality of the residuals of the different VAR(p) models. We chose the Jarque-Bera normality test, and applied the test on the residuals of the respective VAR(p) models. We applied the multivariate version of the Jarque-Bera test which is calculated based on residuals that are standardized with a Choleski decomposition of the variance-covariance matrix for the centered residuals. The multivariate test statistic is given by:

$$JB_{mv} = s_3^2 + s_4^2$$

where

$$s_3^2 = T b_1^T b_1 / 6$$
$$s_4^2 = T(b_2 - 3I_K)^T (b2 - 3I_K)/24$$

in which $b_1$ and $b_2$ are estimates of the third and fourth central non-moment vectors of the standardized residuals $\hat{u}_t^s = \tilde{P}(\hat{u}_t - \overline{\hat{u}}_t)$ respectively with $(\tilde{P})$ being the a lower triangular matrix with a positive diagonal matrix do that $\tilde{P}\tilde{P}^T = \tilde{\Sigma}_u$, which is just the aforementioned Choleski decomposition of the residual covariance matrix. It holds that $JB_{mv} \sim \chi^2(2K)$. The multivariate skewness and kurtosis test, $s_3^2$ and $s_4^2$ respectively are distributed as $\chi^2(K)$.

The last test we conducted was an mulivariate ARCH-LM test, which is based on the following regression equation:

$$vech(\hat{u}_t \hat{u}_t T) = \beta_0 + B_1 vech(\hat{u}_{t-1} \hat{u}_{t-1} T) + \cdots + B_q vech(\hat{u}_{t-q} \hat{u}_{t-q} T) + v_t$$

whereat $v_t$ places a spherical error process and $vech$ is the column stacking operator for symmetric matrices that stacks columns from the main diagonal downward. The vector $\beta_0$ is

of the dimension $\frac{1}{2}K(K+1)$ and the dimensions of the coefficient matrices $B_i$ with $i = 1, \ldots, q$ are $\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1)$. The null hypothesis is: $H_0 : B_1 = \ldots = B_q = 0$ and correspondingly the alternative hypothesis is of the form $H_1 : \exists\ B_i \neq 0$ for $i = 1, 2, \ldots, p$. The test statistic is defined as:

$$VARCH_{LM}(q) = \frac{1}{2}TK(K+1)R_m^2$$

with

$$R_m^2 = 1 - \frac{2}{K(K+1)} tr(\hat{\Omega}\hat{\Omega}_0^{-1}) \ .$$

with $\hat{\Omega}$ assigning the covariance matrix of the ARCH-LM regression it follows that $VARCH_{LM}(q) \sim \chi^2(qK^2(K+1)^2/4)$. The null hypothesis states the absence of ARCH components. The results can be found in Table 3.3.

**Table 3.3. Test Results for Serial Correlation, Normality and ARCH Components**

| Years | Lags | $Q_{16}$ | $p$-value | $JB_5$ | $p$-value | $VARCH_5$ | $p$-value |
|---|---|---|---|---|---|---|---|
| 1986-2009 | 1 | 450.820 | 0.004 | 1775.770 | $< E^{-16}$ | 1207.755 | 0.042 |
| 1986-2009 | 2 | 360.509 | 0.337 | 760.583 | $< E^{-16}$ | 1195.093 | 0.071 |
| 1986-2009 | 3 | 318.865 | 0.585 | 515.813 | $< E^{-16}$ | 1171.912 | 0.161 |
| 1986-2009 | 4 | 300.473 | 0.481 | 140.725 | $< E^{-16}$ | 1138.624 | 0.382 |
| | | | | | | | |
| 1986-2006 | 1 | 467.798 | $< E^{-4}$ | 32.309 | $< E^{-4}$ | 1179.880 | 0.124 |
| 1986-2006 | 2 | 332.118 | 0.746 | 21.098 | 0.020 | 1156.791 | 0.248 |
| 1986-2006 | 3 | 321.009 | 0.552 | 12.842 | 0.232 | 1157.516 | 0.244 |
| 1986-2006 | 4 | 296.841 | 0.540 | 15.392 | 0.118 | 1145.734 | 0.326 |

It can be concluded, that modelling with 1 or 2 lags is too restrictive, the null hypothesis of the absence of ARCH components has to be rejected for the residuals of the $VAR(1)$ and $VAR(2)$ model of the period between 1986 and 2009. For the $VAR(1)$ model there is significant serial correlation. All of the models for the period until 2009 however share the same problem that their residuals are far from being normaly distributed which becomes apparent by having a look at the $p$-value of the $JB_5$ statistic. This is solely caused the extreme price changes during the last three years, as can be clearly seen in comparison with the same models for a period not including the last three years. Figure 3.1 shows the predicted values and the actual price curve in the upper part of the graph and the residuals in the lower part of the graph, the residuals for the last three years are extreme in both up and downswings. For further investigation the nonnormality of the residuals does not pose too big of a problem though, since the confidence intervals will be bootstrapped and therefore do not rely on the assumption of the normal distributed residuals. Furthermore it should be noted that

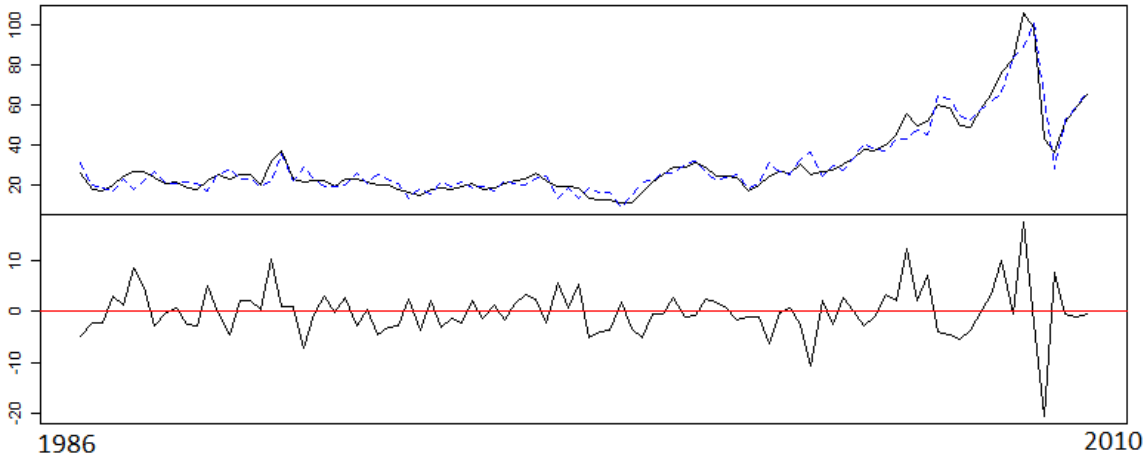


**Figure 3.1. Residuals and predicted deflated freight on board prices.**

not the skewness but the kurtosis part of the Jarque-Bera statistic is responsible for the overall extremely low $p$-values of the test statistics, which is in line with the assumption that the outliers in between 2007 and 2009 are responsible for the low $p$-values.

It remains therefore to look further into a VAR(4) and a VAR(3) model. The next test that is conducted is the actual Johansen test, described in [12] were the matrix $\Pi = \alpha\beta^T$ of the form as in section 3.1.3 is tested for the number cointegration relationships in $\beta$. There are two forms of the Johansen test, the maximal eigenvalue test and the trace test. We will use the trace test to determine the number of cointegrating relationships. The results can be found in Table 3.4.

**Table 3.4. Johansen Test Results**

| | Test Statistic | | Critical Values | | |
|---|---|---|---|---|---|
| $H_0$ | $p = 3$ | $p = 4$ | 90% | 95% | 99% |
| $r = 0$ | 105.74 | 119.03 | 83.20 | 87.31 | 96.58 |
| $r = 1$ | 72.80 | 49.53 | 59.14 | 62.99 | 70.05 |
| $r = 2$ | 44.51 | 27.39 | 39.06 | 42.44 | 48.45 |
| $r = 3$ | 22.18 | 14.26 | 22.76 | 25.32 | 30.45 |
| $r = 4$ | 4.51 | 3.85 | 10.49 | 12.25 | 16.26 |

It can be seen for both, the VAR(3) and the VAR(4) model the null hypothesis of zero cointegrating relationships has to be rejected. The null hypothesis for one cointegrating relationship also has to be rejected for the VAR(3) model but not for the VAR(4) model. Hence it can be concluded that in the VAR(4) model one cointegrating relationship is present, in the VAR(3) model, two are present. But since the AIC suggested the VAR(4) model, further

investigations will only be applied to the VAR(4) model. The VECM model is then reestimated with the restriction of one cointegration relationship and a normalization of the long-run relationship with respect to $FOBdef$. The coefficients for the cointegration relationship and the loading parameters will be provided in Table 3.5.

**Table 3.5. Cointegrating Vector and Loading Parameters**

| Vector | $FOBdef$ | $Days$ | $Caputil$ | $Refutil$ | $MM$ | trend |
|---|---|---|---|---|---|---|
| $\hat{\beta}^T$ | 1.000 | -0.270 | -41.516 | 203.770 | -7.322 | -0.524 |
| $\hat{\alpha}^T$ | | 0.0002 | 0.0014 | 0.0002 | 0.0323 | |

As can be seen, the coefficients of the long-run relationship are quite different from what we would expect from chapter 2 and what seems reasonable from an economic point of view, however we imposed the identification restrictions needed for the estimation of the final SVEC model to be able to forecast, investigate impulse response behavior and conduct a causality test.

For the SVEC model of type B we need $\frac{1}{2}K(K-1) = 10$ linear independent restrictions. From Beveridge-Nelson decomposition it can be concluded that there are $k^* = r(K-r) = 4$ shocks with permanent impacts and only one shock that has a contemporaneous effect, due to r = 1. Since the cointegration relation is interpreted as a stationary price setting relation, the temporary shock is linked with the FOBdef shock variable. Therefore, the five entries in the last column of the long-run impact matrix $\Xi B$ are set to zero. It was found earlier that the matrix $\Xi B$ has reduced rank and therefore setting the first column of $\Xi B$ to zero only imposes $k^* r = 4$ linear independent restrictions. This is why another $\frac{1}{2}k^*(k^*-1)$ entries have to be set to zero. We also assume days of forward consumption is mainly driven by political decisions and therefore shocks in capacity utilization, refinery utilization and market movement do not exert an permanent effect on days of forward consumption, therefore the last three entries of the second row of $\Xi B$ are also set to zero. Furthermore we assumed that shocks in capacity utilization and refinery utilization do not exert a permanent effect on market movement ($MM$), because both backwardation and contago are stable market conditions, which are maintained by self-reinforcing positive feedback loops and extreme shifts between those two states are triggered by exogenous shifts in expectation, as discussed in [5]. Therefore column three and four in the last row are also set to zero. One more restriction has to be imposed, which is done by assuming that shocks in capacity utilization do not exert an immediate effect on refinery utilization, hence $B_{3,4}$ will be set to zero also. The results of the estimated coefficients of the short-run and long-run matrix can be found in Table 3.6 and Table 3.7. With the now identified SVEC model impulse

response forecast can be conducted, in other words to simulate what effect a shock in one of the explanatory variables will have on price in the long run. The long run impulse effects can be seen in Figure 3.2, Figure 3.3, Figure 3.4, Figure 3.5 and Figure 3.6. The confidence bands around the trajectories are bootstrapped. The results are for the most part what could be expected. It appears that after a shock the price overreacts only to be overly corrected in return, therefore it takes a while until price converges towards the permanent impact.

**Table 3.6. Estimated Coefficients of the Short-Run Impact Matrix with Standard Errors in Parenthesis**

| EQUATION | $\varepsilon_t^{FOBdef}$ | $\varepsilon_t^{Days}$ | $\varepsilon_t^{Caputil}$ | $\varepsilon_t^{Refutil}$ | $\varepsilon_t^{MM}$ |
|---|---|---|---|---|---|
| $FOBdef$ | 2.96 | 0.42 | 3.16 | -1.13 | 1.47 |
| | (0.58) | (0.70) | (0.93) | (0.909) | (0.77) |
| $Days$ | -0.01 | 1.35 | -0.35 | -0.07 | -0.46 |
| | (0.12) | (0.25) | (0.25) | (0.22) | (0.23) |
| $Caputil$ | -0.01 | $< E^{-3}$ | 0.08 | 0.00 | $< E^{-3}$ |
| | $(< E^{-3})$ | $(< E^{-3})$ | $(< E^{-3})$ | 0.00 | $(< E^{-3})$ |
| $Refutil$ | $< -E^{-3}$ | $< E^{-4}$ | $< E^{-4}$ | 0.01 | $< E^{-3}$ |
| | $(< E^{-3})$ | $(< E^{-3})$ | $(< E^{-3})$ | $(< E^{-3})$ | $(< E^{-3})$ |
| $MM$ | -0.33 | 0.26 | -0.32 | -0.21 | 0.45 |
| | (0.06) | (0.15) | (0.12) | (0.12) | (0.11) |

**Table 3.7. Estimated Coefficients of the Long-Run Impact Matrix with Standard Errors in Parenthesis**

| EQUATION | $\varepsilon_t^{FOBdef}$ | $\varepsilon_t^{Days}$ | $\varepsilon_t^{Caputil}$ | $\varepsilon_t^{Refutil}$ | $\varepsilon_t^{MM}$ |
|---|---|---|---|---|---|
| $FOBdef$ | 0.00 | 3.01 | 1.24 | -1.59 | 1.70 |
| | (0.00) | (1.32) | (0.44) | (0.44) | (0.57) |
| $Days$ | 0.00 | 1.20 | 0.00 | 0.00 | 0.00 |
| | (0.00) | (0.31) | (0.00) | (0.00) | (0.00) |
| $Caputil$ | 0.00 | -0.02 | 0.03 | 0.01 | -0.01 |
| | (0.00) | (0.01) | (0.01) | (0.01) | (0.01) |
| $Refutil$ | 0.00 | -0.01 | $< E^{-3}$ | 0.01 | $< -E^{-4}$ |
| | (0.00) | (0.01) | $(< E^{-3})$ | $(< E^{-3})$ | $(< E^{-3})$ |
| $MM$ | 0.00 | 0.20 | 0.00 | 0.00 | 0.30 |
| | (0.00) | (0.11) | (0.00) | (0.00) | (0.05) |

The only result that is surprising is the impulse response function of $Days$ on $FOBdef$, it was to be expected that $Days$ again has a negative effect on price. This could be because this particular model might be misspecified, one assumption for the VAR, SVAR,

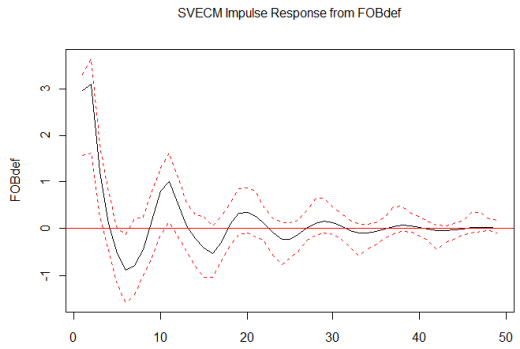


**Figure 3.2.** $FOBdef \rightarrow FOBdef$**.**

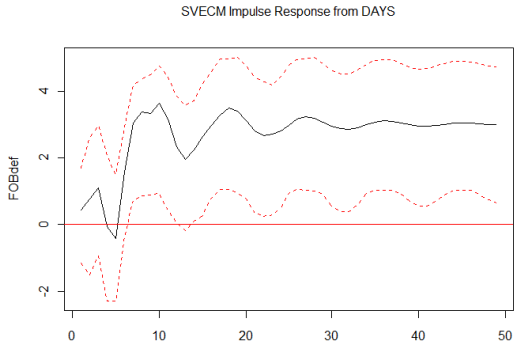


**Figure 3.3.** $Days \rightarrow FOBdef$**.**

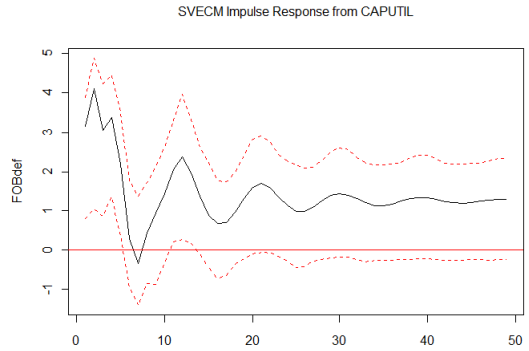


**Figure 3.4.** $Caputil \rightarrow FOBdef$**.**

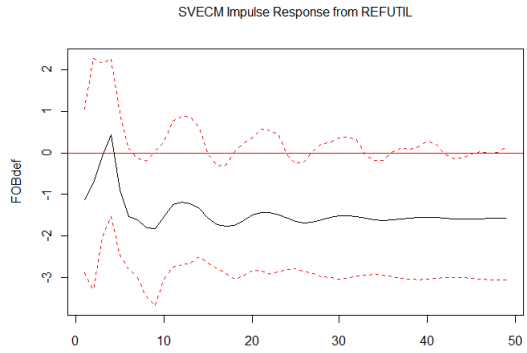


**Figure 3.5.** $Refutil \rightarrow FOBdef$**.**

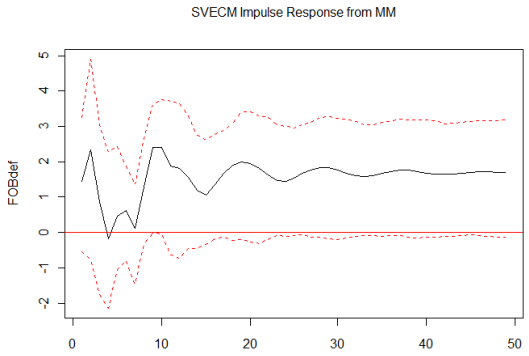


**Figure 3.6.** $MM \rightarrow FOBdef$**.**

VEC and SVEC models are endogenous variables. It is only reasonable to assume that maybe not all of the variables are endogenous. Further possible investigations regarding that topic are subject to further research.

## 3.3 FORECASTS

Considering the nonnormal distribution of the residuals, we applied bootstrapping to obtain a confidence interval for the one-step out-of-sample forecast. The forecast is visually

displayed in Figure 3.7. For reasons of comparability and the difficulty to model 2007-2008 accordingly again 2009 was forecasted. The forecast was obtained by using data from the previous periods to estimate price change for the actual period by using the estimated short term dynamics to see how changes in the individual variables affect the whole system. It can be concluded that the forecast performs worse than the one in section 2.3, which is due to the possible misspecification of the model. The assumptions of endogenous variables should be checked in further research.

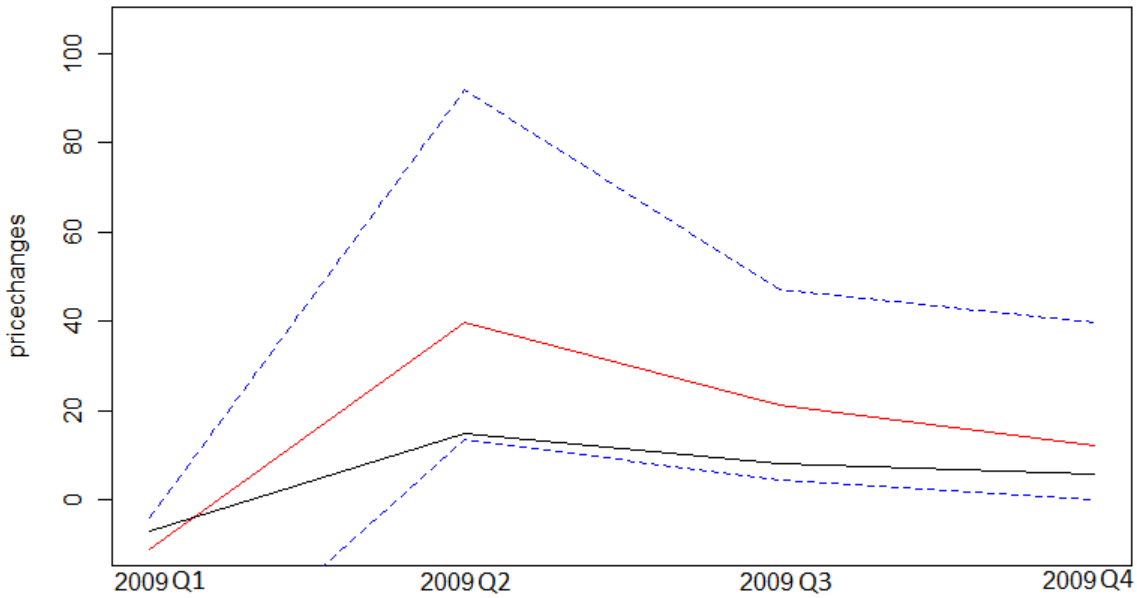


**Figure 3.7. Forecasting 2009 by applying a structural vector error correction model.**

# CHAPTER 4

# UNIVARIATE TIME SERIES APPROACH

This chapter will deal with a strictly univariate approach to the problem. That means we are only looking at the history of price to determine future price movement. We will consider no other explanatory variables in this chapter and apply various methods of time series analysis.

## 4.1 METHODOLOGY

In a first step the stochastic behavior of the time series of price is analyzed for signs of linearity by applying an Autoregressive Integrated Moving Average (ARIMA) model. In a second step the stochastic behavior of the time series is investigated in terms of nonlinearity by applying a GARCH model. In the third part of the chapter a neural network regression is fit to the data to provide an even better nonlinear autoregressive model fit. A forecast is produced for the three different univariate time series models. The ARIMA model is chosen by applying the Box-Jenkins methodology.

## 4.2 AUTO REGRESSIVE INTEGRATED MOVING AVERAGE

An autoregressive integrated moving average (ARIMA) model is a generalization of an autoregressive moving average (ARMA) model. The ARIMA models are generally applied when the data show evidence of nonstationarity. The "I" in ARIMA stands for integrated and denotes the order of integration that is necessary to turn the time series into a stationary time series.

### 4.2.1 ARIMA Methodology

The ARIMA model in general has the parameters: $p$, $d$, and $q$, which are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model respectively, which leads to following notation: ARIMA($p$,$d$,$q$). For reasons of simplicity we will now assume that $d = 0$, which can be easily obtained by just differencing the series $d$ times. The general ARMA(p,q) process is then defined as:

$$X_t = c + \varepsilon_t - \sum_{i=1}^{p} \Phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}$$

The error terms $\varepsilon_t$ are assumed to be (i.i.d) furthermore $\varepsilon_t \sim$ N(0,$\sigma^2$) where $\sigma^2$ is the variance. These assumptions are particularly helpful for further tests to check for nonlinearity not captured by the model. If the error terms are not (i.i.d.) it can be concluded that there is still a nonlinear pattern left in the data, which the ARIMA model was not able to model appropriately. We will later conclude that an ARIMA(2,1,0) and an ARIMA(4,1,0) are the best fitting models. The coefficients for the models are given in Table 4.1.

**Table 4.1. ARIMA Estimates with Standard Errors in Parenthesis**

AR(p): $X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$

| $p$ | $c$ | $\varphi_1$ | $\varphi_2$ | $\varphi_3$ | $\varphi_4$ |
|---|---|---|---|---|---|
| 2 | 0.4490 | 0.4490 | -0.6068 | | |
| | (0.5932) | (0.1382) | (0.1326) | | |
| 4 | 0.2191 | 0.5059 | -0.7120 | 0.2695 | -0.4118 |
| | (0.4923) | (0.1318) | (0.1404) | (0.1532) | (0.1544) |

### 4.2.2 Box-Jenkins Methodology

The Box-Jenkins methodology is applied to an ARMA($p$,$q$) model to determine the most appropriate numbers $p$ and $q$. The first step in the Box-Jenkins methodology is to check for stationarity. This topic however has been covered excessively in the two previous chapters. Thus we already know that the deflated price is integrated of order 1. It is also known that the one time differenced deflated price is stationary. The seasonality was covered in section 2.1 and it was concluded that no significant seasonality is present.

In a second step the order of $p$ and $q$ is determined, this is done by visually inspecting the autocorrelation and partial autocorrelation plot, which can be found in Figure 4.1.

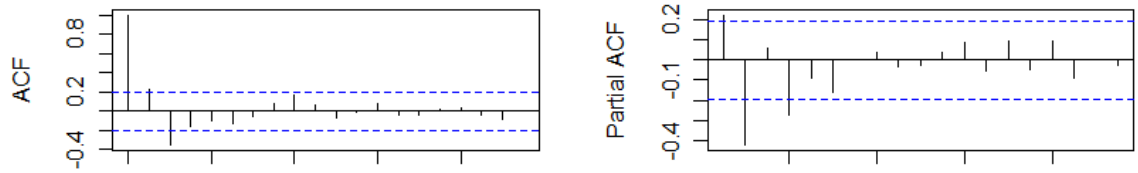


**Figure 4.1. Autocorrelation and partial autocorrelation plots.**

Certain pattern in the ACF plot are associated with a certain order of $p$. In this case we can clearly see an alternating autocorrelation changing from positive to negative and then back again, eventually decaying towards zero. This pattern requires investigation of the partial

autocorrelation plot to finally determine the order of $p$. Also this particular pattern indicates an AR model only, so the MA part is not even needed. The partial autocorrelation of an AR($p$) process becomes zero at lag $p + 1$ and greater. The partial autocorrelation function basically becomes zero at lag 3, then lag 4 is barely outside the confidence interval indicating that the value is essentially zero. For further investigations we will therefore include both models $p = 2$ and $p = 4$. Therefore it can be concluded that an ARIMA(2,1,0) or an ARIMA(4,1,0) is a good fit.

### 4.2.3 Forecasts

Based on the two different ARIMA models a recursively expanding one-step-ahead out-of-sample forecast is created for both models. The forecasts are visually displayed in Figure 4.2 and Figure 4.3. It can be concluded that the ARIMA(4,1,0) model provides a better forecast. However, both of the ARIMA models are doing reasonably well compared to say the SVEC model, which is consistent with [15]. It can be seen that the models are able to capture most of future price movements.

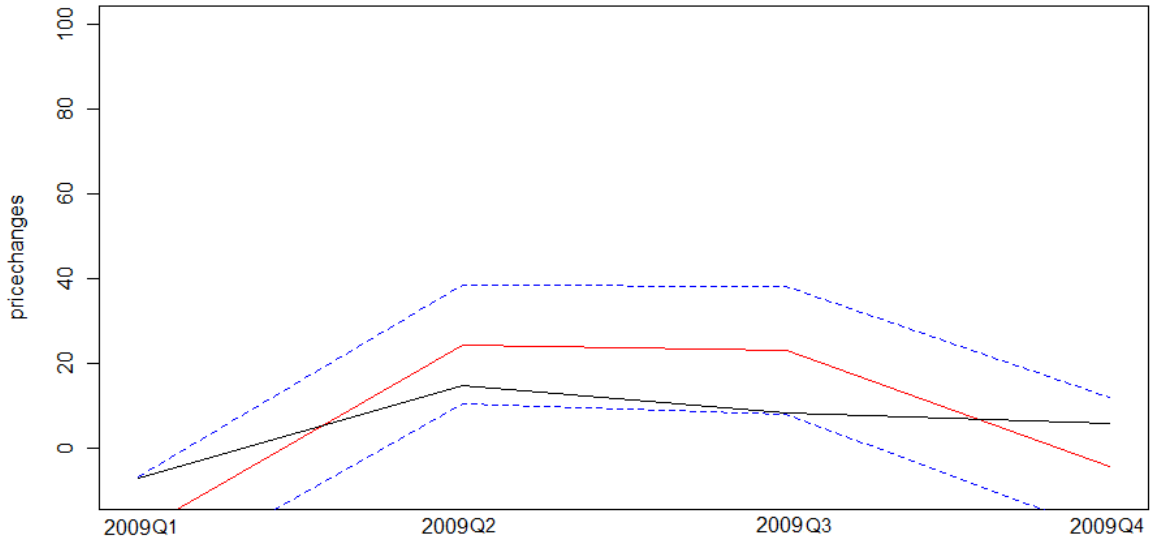


**Figure 4.2. Forecasting 2009 by applying an ARIMA(2,1,0) model.**

Where the black solid line is the actual price change, the red line is the prediction provided the the ARIMA(2,1,0) model. The blue dashed lines are the confidence bands for the ARIMA(2,1,0) model. The second forecast is the forecast for the ARIMA(4,1,0) model, and is also a recursively expanding one-step-ahead out-of-sample forecast.

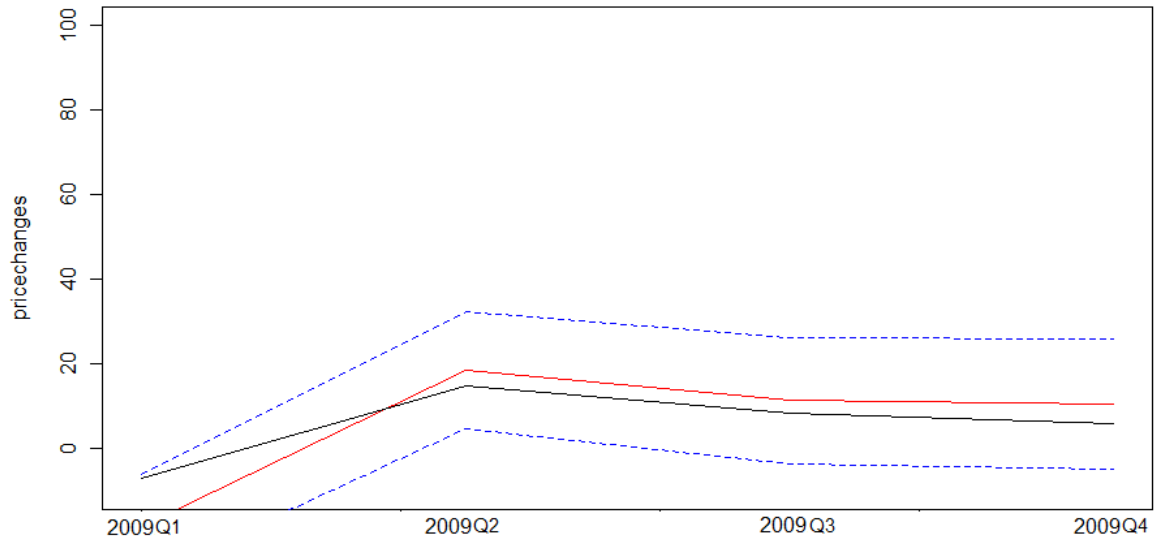


**Figure 4.3. Forecasting 2009 by applying an ARIMA(4,1,0) model.**

Where the black solid line is the actual price change, the red line is the prediction provided the the ARIMA(2,1,0) model. The blue dashed lines are the confidence bands for the ARIMA(4,1,0) model.

## 4.3 General Autoregressive Conditional Heteroskedasticity Model

A test was conducted on the residuals of the aforementioned ARIMA models with the aim to confirm that there are nonlinearities left in the model. The test of choice is the BDS test, described in [4]. The BDS test is a powerful tool, originally developed to test for the null hypothesis of independent and identical distribution. However, when applied to the residuals from a fitted linear time series model, here the ARIMA model, the Brock-Dechert-Scheinkman (BDS) test will be able to spot persisting dependence and the presence of ignored nonlinear structure. The null hypothesis states: that the original linear model captured all dependencies, if the null hypothesis is rejected it can be concluded that there are some nonlinearities remaining and a nonlinear model might be a good idea. The BDS test focuses on the correlation integral, which measures the frequency with which certain patterns are repeated in the data. Given the following time series: $y_t$ for $t = 1, 2, \cdots, T$ the $m$-history is defined as: $x_t^m = (x_t, x_{t-1}, \cdots, x_{t-m+1})$. The correlation integral at embedding dimensions of m is then given by:

$$C_{m,\varepsilon} = \frac{2}{T_m(T_m - 1)} \sum_{m \leq s} \sum_{< t \leq T} I(x_t^m, x_s^m; \varepsilon)$$

where $T_m = T - m + 1$ and $I(x_t^m, x_s^m; \varepsilon)$ is an indicator function which equals one if $|x_{t-i} - x_{s-i}| < \varepsilon$ for $i = 1, 2, \cdots, m - 1$ and else equals zero. The intuition behind this is

that the correlation integral estimates the probability that the distance between any two $m$-dimensional points is less than $\varepsilon$, in other words it estimates the joint probability:

$$Pr(|x_t - x_s| < \varepsilon, |x_{t-1} - x_{s-1}| < \varepsilon, \cdots, |x_{t-m+1} - x_{s-m+1}| < \varepsilon) \ .$$

Now assuming that the $x_t$ are i.i.d said probability should in the limiting case be equal to:

$$C_{1,\varepsilon}^m = Pr(|x_t - x_s| < \varepsilon)^m \ .$$

The BDS statistic is in [4] defined as:

$$V_{m,\varepsilon} = \sqrt{T}\frac{C_{m,\varepsilon} - C_{1,\varepsilon}^m}{s_{m,\varepsilon}}$$

where $s_{m,\varepsilon}$ is the standard deviation of $C_{m,\varepsilon} - C_{1,\varepsilon}^m$ and can be consistently estimated. The BDS statistic converges to $N(0,1)$:

$$V_{m,\varepsilon} \xrightarrow{d} N(0,1) \ .$$

The results of the BDS test can be found in Table 4.2 and Table 4.3.

**Table 4.2. BDS Results for Residuals of ARIMA(2,1,0) with $p$-Values in Parenthesis**

| | $\varepsilon$ | | | |
|---|---|---|---|---|
| Embedding Dimensions | $\varepsilon = 0.5\sigma$ | $\varepsilon = 1\sigma$ | $\varepsilon = 1.5\sigma$ | $\varepsilon = 2\sigma$ |
| | 3.361 | 6.723 | 10.084 | 13.446 |
| 2 | 4.527 | 3.254 | 2.538 | 1.856 |
| | (0.000) | (0.001) | (0.011) | (0.063) |
| 3 | 5.589 | 4.329 | 3.556 | 3.381 |
| | (0.000) | (0.000) | (0.000) | (0.001) |
| 4 | 5.657 | 4.237 | 3.484 | 3.543 |
| | (0.000) | (0.000) | (0.000) | (0.000) |
| 5 | 6.378 | 4.708 | 3.806 | 3.978 |
| | (0.000) | (0.000) | (0.000) | (0.000) |

It becomes obvious that the hypothesis of i.i.d. residuals in the ARIMA models has to be rejected. The $p$-values are for almost every combination of $\varepsilon$ and embedding dimensions of m insignificant. The presence of heteroskedasticity requires the use of a GARCH model to model the changing variance accordingly. The general form of the GARCH(p,q) model is of the following form:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \ .$$

**Table 4.3. BDS Results for Residuals of ARIMA(4,1,0) with $p$-Values in Parenthesis**

| Embedding Dimensions | $\varepsilon$ | | | |
|---|---|---|---|---|
| | $\varepsilon = 0.5\sigma$ | $\varepsilon = 1\sigma$ | $\varepsilon = 1.5\sigma$ | $\varepsilon = 2\sigma$ |
| | 3.2023 | 6.4046 | 9.6069 | 12.8092 |
| 2 | 3.0943 | 2.5422 | 2.4411 | 1.9488 |
| | (0.002) | (0.011) | (0.0146) | (0.0513) |
| 3 | 5.4384 | 4.2698 | 3.5208 | 3.1925 |
| | (0.000) | (0.000) | (0.0004) | (0.0014) |
| 4 | 5.7418 | 4.5284 | 3.5096 | 3.2236 |
| | (0.000) | (0.000) | (0.0004) | (0.0013) |
| 5 | 6.5293 | 5.3067 | 3.9318 | 3.6052 |
| | (0.000) | (0.000) | (0.0001) | (0.0003) |

The GARCH(1,1) model was selected as the best nonlinear model. And a BDS test was conducted on the residuals to check for heteroskedasticity. The results can be found in Table 4.4.

**Table 4.4. BDS Results for GARCH(1,1) with $p$-Values in Parenthesis**

| Embedding Dimensions | $\varepsilon$ | | | |
|---|---|---|---|---|
| | $\varepsilon = 0.5\sigma$ | $\varepsilon = 1\sigma$ | $\varepsilon = 1.5\sigma$ | $\varepsilon = 2\sigma$ |
| | 0.496 | 0.993 | 1.489 | 1.986 |
| 2 | 0.204 | 0.081 | 0.732 | 0.658 |
| | (0.838) | (0.936) | (0.464) | (0.511) |
| 3 | -1.495 | -0.619 | -0.176 | -0.645 |
| | (0.135) | (0.536) | (0.860) | (0.519) |
| 4 | -1.841 | -0.918 | -0.4154 | -1.011 |
| | (0.066) | (0.359) | (0.678) | (0.312) |
| 5 | -1.553 | -1.008 | -0.713 | -1.696 |
| | (0.120) | (0.314) | (0.476) | (0.090) |

The high $p$-values imply we not reject the null hypothesis of i.i.d. residuals anymore, that means the nonlinearities are captured by using a GARCH approach to account for the changing variability over time, hence the nonlinearity is successfully removed from the data.

## 4.4 ARTIFICIAL NEURAL NETWORK REGRESSION MODEL

After finding enough evidence of nonlinearities in the data the next step is the application of an artificial neural network regression. The neural network consists of an interconnected group of artificial neurons and computes the output by fitting a linear combination of a nonlinear transformation of the linear combination of the explanatory variables. It was originally inspired by the nervous system in animals, hence the name. The neural network regression is of the following form, as can be found in [16]:

$$f(x) = \beta_0 + \sum_{k=1}^{M} \beta_k \phi(\mu_k + \gamma_k^T x)$$

where $\phi(u) = (e^u)/(1 + e^u)$ is the logistic distribution function. The parameter vectors are $\beta_i$ for $i = 0, 1, \cdots, M$ and $\mu_k$ and $\gamma_k$, for $1 \leq k \leq M$ adding up to a total number of parameters of $1 + M + M + dM$ with d being the number of explanatory variables or lags that were used. Adding up to a total of $1 + M(d + 2)$ parameters. The parameters are estimated by applying nonlinear least square regression and the number of hidden layers $M$ is determined based on the GCV, the general cross validation: $GCV(p) = RSS(p)/(1 - Cp/n)^2$. The standard procedure is to fit a range of possible hidden layers and then choosing the one model with the lowest GCV, in this paper a $C$-value of 2 is used to protect against over fitting the model. Figure 4.4 shows a graphical display of a one-layer neural network with four input vectors (for example lags of the variable) and five hidden units $M = 5$.

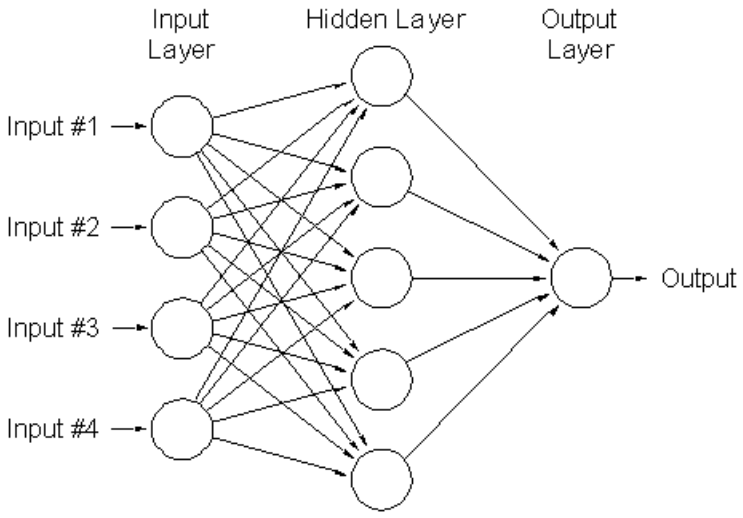


**Figure 4.4. One-layer neural network with four input variables and five hidden units.**

The network thus first forms five so-called hidden variables by linearly combining the input in $M$ different ways and then passing the hidden variables through the logistic distribution function $\phi$. These results are then linearly combined to obtain the final output. The parameter space has a very large number of local minima when nonlinear least squares is used to estimate the neural network parameters. Therefore, a conservative computation is to generate many parameter sets at random and use these as starting values for the minimization procedure. Minimization of the residual sums of squares is done $independently$ for each hidden layer $1, \ldots, M$. A rectangular region of feasible parameters is divided into a series of 250 nested boxes about the origin. For each box, 100 parameter sets are generated at random from a uniform distribution. subsectionTest Results The artificial neural network regression is applied on different lags, overall the model with the lowest GCV is then picked for further applications. The results of the neural network regressions can can be found in Table 4.5.

**Table 4.5. Artificial Neural Network Regression Results**

| Lags | Hidden Layer | GCV | $R^2$ |
|---|---|---|---|
| 1 | 2 | 25.89 | 93.6% |
| 2 | 3 | 24.66 | 95.54% |
| 3 | 3 | 26.74 | 95.98% |

It is evident that the lowest GCV is obtained by using a model that utilizes two lags. Said model uses three hidden layers. For the forecasting two methods of bootstrapping were applied to obtain confidence intervals for the forecast. The first method is a parametric residual-based bootstrap approach, the second one a nonparametric approach. It follows a short description of the residual-based approach. Let $x_t = f(x_{t-1}, \ldots, x_{t-p}) + \varepsilon_t$. Now a first estimate of $f(\cdot)$, denoted by $\hat{f}(\cdot)$, is obtained by applying the neural network regression, with two lags and three hidden layers. The residuals $\hat{\varepsilon}_t$ are given by $\hat{\varepsilon}_t = x_t - \hat{f}(x_{t-1}, \ldots, x_{t-p})$. Now assuming that the $\hat{\varepsilon}_t$ are approximately independent and identically distributed, bootstrap samples of $x_t$ were obtained by resampling with replacement from $\hat{\varepsilon}_t$ and setting $x_{t,b} = \hat{f}(x_{t-1}, \ldots, x_{t-p}) + \varepsilon_{t,b}$ with $t = p+1, \ldots, n$, where $\varepsilon_{t,b}$ denotes the $b$-th bootstap sample, $b = 1, \ldots, B$ and $n$ denotes the length of the time series. Each of the bootstrapped series is then used to reestimate $f(\cdot)$, and yields another $\hat{f}_b(\cdot)$, which is then used to conduct a one-step ahead out-of-sample forecast. The 90% quantile is then obtained from the $B$ forecasts.

The nonparametric regression-based approach utilizes the following methodology. Let $x_t = f(x_{t-1}, \ldots, x_{t-p}) + \varepsilon_t$. Now the different estimates of $f(\cdot)$ are obtained by bootstrapping the data pairs by sampling with replacement. Then neural network regression, with two lags and three hidden layers is applied, i.e., $x_{t,b} = \hat{f}(x_{t-1,b}, \ldots, x_{t-p})$. Each of the $\hat{f}_b(\cdot)$ is then used to conduct a one-step ahead out-of-sample forecast for 2009. Afterwards the 90% quantile is obtained from the $B$ forecasts.

The results show that the second method is more consistent with what would be expected. A problem with the first method is the assumption of independent and identically distributed residuals. The nature of the data however leads to fairly unequally distributed residuals, an additional hurdle is that the residual-based samples are highly dependent on the initial estimate of $f(\cdot)$, therefore the nonparametric regression-based approach is used to obtain the final confidence band.

### 4.4.1 Forecasts

A one-step ahead out-of-sample forecast was conducted based on the data from 1986 to 2008. The forecast is visually displayed in Figure 4.5. The four quarters of 2009 where then forecasted based on data from the preceding time intervals. The forecast was obtained by using data from the previous periods to estimate price change for the actual period by applying the artificial neural network regression. It can be seen in Figure 4.5 that the forecast is quite accurate, and hereby confirms what was expected. It seems that the neural network does a very good job in capturing the nonlinearities in the data and forecasting.

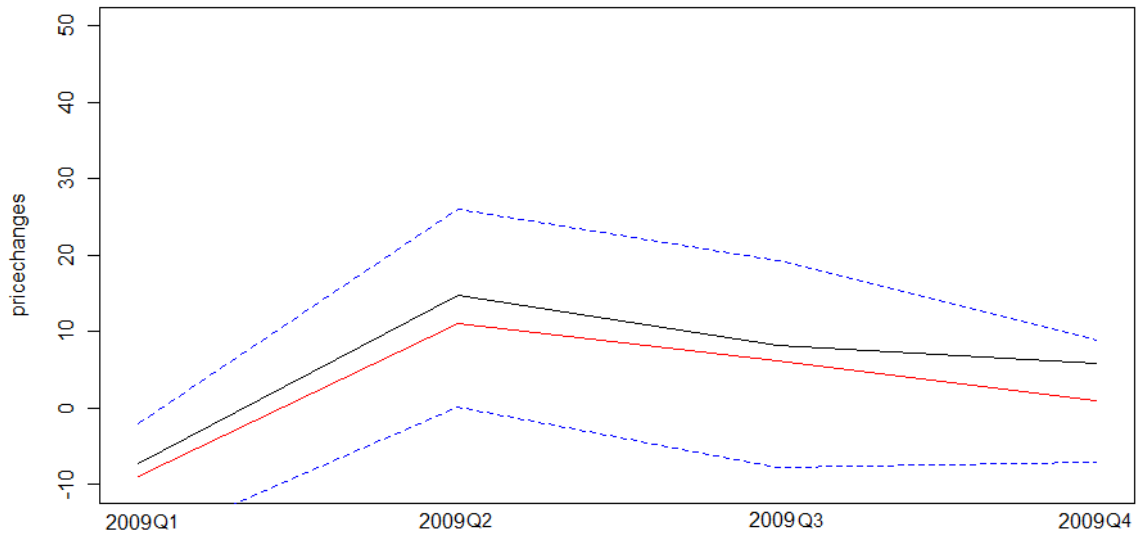


**Figure 4.5. Forecasting 2009 by applying an artificial neural network regression.**

The black line represents the actual price changes, the red line are the predictions provided by the model. The blue dashed lines are the boundaries of the confidence interval.

# CHAPTER 5
# CONCLUSION

After investigating all three approaches for predicting and forecasting the price of oil certain conclusions can be drawn. Most importantly, the statement by [15] that univariate nonlinear time series models (artificial neural network regression) perform better than their linear counterparts, which in turn perform better than structural models in forecasting, is confirmed. Table 5.1 provides a summary of the forecasts given by the different models, and the actual price. It can be clearly seen that the structural models perform the worst, the ARIMA models do quite a good job, but by far the best results are achieved by the artificial neural network model. Even when faced with the challenge of modelling data like the data at hand, the neural network regression is able to model the nonlinear pattern in the data with astonishing accuracy. In my opinion, the neural network regression should be adapted to be the model of choice for forecasting. Clearly a structural approach appears intuitively to be more promising, but the variable selection proves to be quite difficult. Therefore the main conclusion of this thesis is that neural network regression is an excellent statistical modelling tool for cumbersome data. It should however be kept in mind, that structural models provide an excellent tool to investigate the significance and meaning of certain variables. Aalso the structural model can be used to determine, how much change in the dependent variable comes with a unit change in one of the explanatory variables. So in conclusion, the structural models are good for understanding the prediction but lack the power of accurate forecasts, where clearly the time series models perform better.

**Table 5.1. Comparison of Forecasts for 2009**

| Quarter | SW-Method | SVEC | ARIMA(2,1,0) | ARIMA(4,1,0) | NNREG | Actual Change |
|---|---|---|---|---|---|---|
| Q1 | -7.71 | -11.20 | -19.76 | -18.57 | -8.96 | -7.19 |
| Q2 | 40.18 | 39.85 | 24.41 | 18.41 | 11.06 | 14.77 |
| Q3 | -1.04 | 21.23 | 23.11 | 11.24 | 6.25 | 8.24 |
| Q4 | 2.44 | 12.28 | -4.29 | 10.35 | 0.92 | 5.89 |

# BIBLIOGRAPHY

[1] R: A language and environment for statistical computing. `http://www.R-project.org`, 2010.

[2] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Automat. Control*, 19(6):716–723, 1974.

[3] L.J. Bachmeier and J.M. Griffin. Testing for market integration crude oil, coal, and natural gas. *The Energy Journal*, 27(2):55–71, 2006.

[4] W.A. Brock, W.D. Dechert, J.A. Scheinkman, and B. LeBaron. A test for independence based on the correlation dimension. *Econometric Rev.*, 15:197–235, 1996.

[5] S. Dées, A. Gasteuil, R. Kaufmann, and M. Mann. Assessing the factors behind oil price changes. *European Central Bank Working Paper Series*, 855, 2008.

[6] S. Dées, P. Karadelogou, R. Kaufmann, and M. Sánchez. Modelling the world oil market assessment of a quarterly econometric model. *Energy Policy*, 35:178–191, 2007.

[7] D.A. Dickey and W.A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.*, 74:427–431, 1979.

[8] R.E. Engle and C.W.J. Granger. Cointergration and error-correction: representation, estimation, and testing. *Econometrica*, 55:251–276, 1987.

[9] C.W.J. Granger and P. Newbold. Spurious regressions in econometrics. *J. Econometrics*, 2:111–120, 1974.

[10] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 41:190–195, 1979.

[11] S. Hyllenberg, R.E. Engle, C.W.J. Granger, and B.S. Yoo. Seasonal integration and cointegration. *J. Econometrics*, 44:215–238, 1990.

[12] S. Johansen. Cointegration and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.

[13] S. Johansen. *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford, 1995.

[14] H. Lutkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York, 2006.

[15] S. Moshiri and F. Foroutan. Forecasting nonlinear crude oil futures prices. *The Energy Journal*, 27(4):81–96, 2006.

[16] D. Nychka, B. Bailey, and S. Ellner. Funfits: Data analysis and statistical tools for estimating functions. Unpublished report, 1996.

[17] P.C.B Phillips and P. Perron. Testing for a unit root in time series regression. *Biometrika*, 75:335–346, 1988.

[18] J.H. Stock and M.W. Watson. A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica*, 61:783–820, 1993.

# APPENDIX

# R-CODE

# R-CODE

```
SW-Method:
library(dynlm)

mtsData.DOLS1 <- dynlm(mtsData[,1] ~ mtsData[,-1] + L(d(mtsData[,-1]),1)
+ L(d(mtsData[,-1]),-1) + d(mtsData[,-1]) , start = c(1986,1), end = c(2009,4))
mtsData.DOLS2 <- dynlm(mtsData[,1] ~ mtsData[,-1] + L(d(mtsData[,-1]),1)
+ L(d(mtsData[,-1]),-1) + L(d(mtsData[,-1]),2) + L(d(mtsData[,-1]),-2)
+ d(mtsData[,-1]) , start = c(1986,1), end = c(2009,4))
mtsData.DOLS3 <- dynlm(mtsData[,1] ~ mtsData[,-1] + L(d(mtsData[,-1]),1)
+ L(d(mtsData[,-1]),-1) + L(d(mtsData[,-1]),2) + L(d(mtsData[,-1]),-2)
+ L(d(mtsData[,-1]),3) + L(d(mtsData[,-1]),-3) + d(mtsData[,-1])
, start = c(1986,1), end = c(2009,4))
mtsData.DOLS4 <- dynlm(mtsData[,1] ~ mtsData[,-1] + L(d(mtsData[,-1]),1)
+ L(d(mtsData[,-1]),-1) + L(d(mtsData[,-1]),2) + L(d(mtsData[,-1]),-2)
+ L(d(mtsData[,-1]),3) + L(d(mtsData[,-1]),-3) + L(d(mtsData[,-1]),4)
+ L(d(mtsData[,-1]),-4) + d(mtsData[,-1]) , start = c(1986,1), end = c(2009,4))

residuals.mtsData.DOLS4 = mtsData[,1]
- coefficients(mtsData.DOLS4)[1]
- (coefficients(mtsData.DOLS4)[2]*mtsData[,2]
+coefficients(mtsData.DOLS4)[3]*mtsData[,3]
+coefficients(mtsData.DOLS4)[4]*mtsData[,4]
+coefficients(mtsData.DOLS4)[5]*mtsData[,5])

mtsData.equ.DOLS1 <- dynlm(mtsData.equ[,1] ~ mtsData.equ[,-1]
+ L(d(mtsData.equ[,-1]),1) + L(d(mtsData.equ[,-1]),-1)
+ d(mtsData.equ[,-1]) , start = c(1986,1), end = c(2009,4))
mtsData.equ.DOLS2 <- dynlm(mtsData.equ[,1] ~ mtsData.equ[,-1]
+ L(d(mtsData.equ[,-1]),1) + L(d(mtsData.equ[,-1]),-1)
+ L(d(mtsData.equ[,-1]),2) + L(d(mtsData.equ[,-1]),-2)
+ d(mtsData.equ[,-1]) , start = c(1986,1), end = c(2009,4))
mtsData.equ.DOLS3 <- dynlm(mtsData.equ[,1] ~ mtsData.equ[,-1]
```

```
+ L(d(mtsData.equ[,-1]),1) + L(d(mtsData.equ[,-1]),-1)
+ L(d(mtsData.equ[,-1]),2) + L(d(mtsData.equ[,-1]),-2)
+ L(d(mtsData.equ[,-1]),3) + L(d(mtsData.equ[,-1]),-3)
+ d(mtsData.equ[,-1]) , start = c(1986,1), end = c(2009,4))
mtsData.equ.DOLS4 <- dynlm(mtsData.equ[,1] ~ mtsData.equ[,-1]
+ L(d(mtsData.equ[,-1]),1) + L(d(mtsData.equ[,-1]),-1)
+ L(d(mtsData.equ[,-1]),2) + L(d(mtsData.equ[,-1]),-2)
+ L(d(mtsData.equ[,-1]),3) + L(d(mtsData.equ[,-1]),-3)
+ L(d(mtsData.equ[,-1]),4) + L(d(mtsData.equ[,-1]),-4)
+ d(mtsData.equ[,-1]) , start = c(1986,1), end = c(2009,4))

results.aic <- matrix(0, ncol = 2, nrow = 4)
colnames(results.aic) <- cbind("AIC-simple", "AIC+squared+cubic")

results.aic[1,1] <- AIC(mtsData.DOLS1)
results.aic[2,1] <- AIC(mtsData.DOLS2)
results.aic[3,1] <- AIC(mtsData.DOLS3)
results.aic[4,1] <- AIC(mtsData.DOLS4)
results.aic[1,2] <- AIC(mtsData.equ.DOLS1)
results.aic[2,2] <- AIC(mtsData.equ.DOLS2)
results.aic[3,2] <- AIC(mtsData.equ.DOLS3)
results.aic[4,2] <- AIC(mtsData.equ.DOLS4)

ECM.estimate.DOLS <- function (y, z = c(1986,1), w = c(2008,4)) {
source("C:\\Users\\Nightmare\\Documents\\makeSeason.r")
source("C:\\Users\\Nightmare\\Documents\\makeWar.r")
temp <- dynlm(d(mtsData[,1]) ~ L(residuals.mtsData.DOLS4)
+ L(d(mtsData[,1]),c(1:y)) + L(d(mtsData[,2]),c(1:y))
+ L(d(mtsData[,3]),c(1:y)) + L(d(mtsData[,4]),c(1:y))
+ L(d(mtsData[,5]),c(1:y)) + season1[,1] + season1[,2]
+ season1[,3] + war, start = z, end = w)
return(temp)
}

findbest.ECM.estimate.DOLS <- function (x) {
output <- matrix(0, ncol = 3, nrow =x)
```

```
colnames(output) <- cbind("lags","AIC","Rank")
for (we in 1:x) {
output[we,1] <- we
output[we,2] <- AIC(ECM.estimate.DOLS(we))
}
output[,3] <- rank(output[,2])
return(output)
}

checkmatrix <- findbest.ECM.estimate.DOLS(5)

for (we in 1:5) {
if(checkmatrix[we,3] == 1){check <- we}
}
rm(checkmatrix,we)

mtsData.DOLS.shortrun <- ECM.estimate.DOLS(check)
rm(check)
summary(mtsData.DOLS.shortrun)

###findbest.ECM.estimate###
findbest.ECM.estimate <- function (x) {

source("C:\\Users\\Nightmare\\Documents\\makeEquilibriumOLS.r")
source("C:\\Users\\Nightmare\\Documents\\Function.ECM.estimate.r")

output <- matrix(0, ncol = 3, nrow =x)
colnames(output) <- cbind("lags","AIC","Rank")
for (we in 1:x) {
output[we,1] <- we
output[we,2] <- AIC(ECM.estimate(we))
}

output[,3] <- rank(output[,2])

return(output)
```

```
}

##ECMestimate##
ECM.estimate <- function (y, z = c(1986,1), w = c(2006,4)) {
source("C:\\Users\\Nightmare\\Documents\\makeSeason.r")
source("C:\\Users\\Nightmare\\Documents\\makeWar.r")
temp <- dynlm(d(mtsData[,1]) ~ L(residuals.mtsData.DOLS4))
+ L(d(mtsData[,1]),c(1:y)) + L(d(mtsData[,2]),c(1:y))
+ L(d(mtsData[,3]),c(1:y)) + L(d(mtsData[,4]),c(1:y))
+ L(d(mtsData[,5]),c(1:y)) + season1[,1] + season1[,2]
+ season1[,3] + war, start = z, end = w)
return(temp)
}

##printSW##
source("C:\\Users\\Nightmare\\Documents\\Function.predict.SW.r")

pricechanges <- ts(prediction[,2], start = c(2009,1)
, end = c(2009,4), frequency = 4)
U <- ts(prediction[,3], start = c(2009,1)
, end = c(2009,4), frequency = 4)
L <- ts(prediction[,1], start = c(2009,1)
, end = c(2009,4), frequency = 4)
minx = min(L,ts(tail(diff(mtsData[,1]), n = 4)
, start = c(2009,1), end = c(2009,4), frequency = 4))
maxx = max(U,ts(tail(diff(mtsData[,1]), n = 4)
, start = c(2009,1), end = c(2009,4), frequency = 4))

plot.ts(pricechanges, col="red", ylim=c(minx,maxx))
lines(U, col="blue", lty="dashed")
lines(L, col="blue", lty="dashed")
lines(ts(tail(diff(mtsData[,1]),n = 4), start = c(2009,1)
, end = c(2009,4), frequency = 4))

rm(minx,maxx,U,L)
```

```
##SW-Method##
library(vars)
source("C:\\Users\\Nightmare\\Documents\\loadDATA.r")
source("C:\\Users\\Nightmare\\Documents\\makeMTSdata_woCHEAT.r")
source("C:\\Users\\Nightmare\\Documents\\createVARmodels.r")

vecm <- ca.jo(mtsData, type = "trace",ecdet = "trend"
, K = 4, spec = "transitory")
vecm1 <- ca.jo(mtsData[1:88,], type = "trace",ecdet = "trend"
, K = 4, spec = "transitory")

vecm.r1 <- cajorls(vecm, r = 1)
vecm1.r1 <- cajorls(vecm, r = 1)

LR <- matrix(NA, nrow = 5, ncol = 5)
LR[,1] <- 0
LR[2,3:5] <- 0
LR[5,3:4] <- 0

SR <- matrix(NA, nrow = 5, ncol = 5)
SR[3,4] <- 0

svec <- SVEC(vecm, LR = LR, SR = SR, r = 1
, lrtest = FALSE, boot = TRUE, runs = 100)
svec1 <- SVEC(vecm1, LR = LR, SR = SR, r = 1
, lrtest = FALSE, boot = TRUE, runs = 100)

realization <- matrix(0, ncol = 4, nrow = 100)
dimnames(realization) <- list(format(1:100)
, c("2009Q1","2009Q2","2009Q3","2009Q4"))

###bootstrapping###
for (j in 1:100) {
set.seed(j)
tempv <- sample(96,replace=T)
temp <- ca.jo(diff(mtsData)[tempv,]
```

```
, type = "trace",ecdet = "trend", K = 4, spec = "transitory")
svec2 <- try(SVEC(temp, LR = LR, SR = SR, r = 1
, lrtest = FALSE, boot = TRUE, runs = 100), TRUE)
if(class(svec2) != "try-error"){
for (we in 1:4) {
realization[j,we] <- svec2\$SR[1,]%*%diff(mtsData)[95+we,]
}
}
rm(svec2)
}

for (k in 1:100) {
if(realization[k,1] != 0){
x <- try(class(realization1), TRUE)
if(class(x) == "try-error"){realization1 <- realization[k,]}
else{realization1 <- rbind(realization1,realization[k,])}
}
}
z <- length(realization1)/4
dimnames(realization1) <- list(format(1:x)
, c("2009Q1","2009Q2","2009Q3","2009Q4"))

rm(we,j,k,x,z,realization,temp,tempv,svec2)

##findbest comb##
findcomb <- function (exp, dep, max=F)
{
    numb <- if (max == T){6}else{2^(length(exp)/length(dep)-1)}

    temp3 <- matrix(0, ncol = length(exp)/length(dep)+2, nrow = numb)
    dimnames(temp3) <- list(format(1:numb),
    c("FOBdef2","FOBdef3","DAY2","DAY3","CAPUTIL2",
    "CAPUTIL3","REFUTIL2","REFUTIL3","MM2","MM3","GCV(2)","Rank"))

    for (j in 1:numb) {
    tempv <- digitsBase(j, base=2, 10)
```

```
    temp <- nnreg(cbind(exp[,c(if(tempv[10,] == 1){1}else{0}
    ,if(tempv[9,] == 1){2}else{0},if(tempv[8,] == 1){3}else{0}
    ,if(tempv[7,] == 1){4}else{0},if(tempv[6,] == 1){5}else{0}
    ,if(tempv[5,] == 1){6}else{0},if(tempv[4,] == 1){7}else{0}
    ,if(tempv[3,] == 1){8}else{0},if(tempv[2,] == 1){9}else{0}
    ,if(tempv[1,] == 1){10}else{0})]),dep,1,4)

    temp2 <- summary(temp, noprint = TRUE)

        temp3[j, 1] <- if(tempv[10,] == 1){1}else{0}
        temp3[j, 2] <- if(tempv[9,] == 1){1}else{0}
        temp3[j, 3] <- if(tempv[8,] == 1){1}else{0}
        temp3[j, 4] <- if(tempv[7,] == 1){1}else{0}
        temp3[j, 5] <- if(tempv[6,] == 1){1}else{0}
        temp3[j, 6] <- if(tempv[5,] == 1){1}else{0}
        temp3[j, 7] <- if(tempv[4,] == 1){1}else{0}
        temp3[j, 8] <- if(tempv[3,] == 1){1}else{0}
        temp3[j, 9] <- if(tempv[2,] == 1){1}else{0}
        temp3[j, 10] <- if(tempv[1,] == 1){1}else{0}
        temp3[j, 11] <- temp2[temp\$best.model,6]
}
temp3[,12] <- rank(temp3[,11])

    temp4 <- matrix(0, ncol = 12, nrow = 5)
    dimnames(temp4) <- list(format(1:5),  c("FOBdef2"
    ,"FOBdef3","DAY2","DAY3","CAPUTIL2","CAPUTIL3"
    ,"REFUTIL2","REFUTIL3","MM2","MM3","GCV(2)","Rank"))
    for (we in 1:numb) {
if(temp3[we,12]==1){temp4[1,] <- temp3[we,]}
if(temp3[we,12]==2){temp4[2,] <- temp3[we,]}
if(temp3[we,12]==3){temp4[3,] <- temp3[we,]}
if(temp3[we,12]==4){temp4[4,] <- temp3[we,]}
if(temp3[we,12]==5){temp4[5,] <- temp3[we,]}
}
    print(temp4)
}
```

```
##create VAR models##
var1 <- VAR(mtsData, p = 1, type = "both", season = 4)
var2 <- VAR(mtsData, p = 2, type = "both", season = 4)
var3 <- VAR(mtsData, p = 3, type = "both", season = 4)
var4 <- VAR(mtsData, p = 4, type = "both", season = 4)

var1.1 <- VAR(mtsData[1:88,], p = 1, type = "both", season = 4)
var2.1 <- VAR(mtsData[1:88,], p = 2, type = "both", season = 4)
var3.1 <- VAR(mtsData[1:88,], p = 3, type = "both", season = 4)
var4.1 <- VAR(mtsData[1:88,], p = 4, type = "both", season = 4)

var_wo0708 <- VAR(ts(rbind(as.matrix(mtsData[1:88,])
,as.matrix(mtsData[97:100,])), frequency = 4), p = 4, type = "both", season = 4)

testresults <- matrix(0, ncol = 6, nrow = 8)

for (we in 1:4) {
temp <- VAR(mtsData, p = we, type = "both", season = 4)
temp1 <- VAR(mtsData[1:88,], p = we, type = "both", season = 4)

if(i == 1){x=0}
if(i == 2){x=6}
if(i == 3){x=12}
if(i == 4){x=18}

st <- serial.test(temp, lags.pt = 16, type = "PT.asymptotic")
st1 <- serial.test(temp1, lags.pt = 16, type = "PT.asymptotic")
norm <- normality.test(temp)
norm1 <- normality.test(temp1)
arch <- arch.test(temp, lags.multi = 5)
arch1 <- arch.test(temp1, lags.multi = 5)
testresults[1+x] <- st$serial[1]
testresults[2+x] <- st$serial[3]
testresults[3+x] <- norm$jb.mul$JB[1]
testresults[4+x] <- norm$jb.mul$JB[3]
testresults[5+x] <- arch$arch.mul[1]
```

```
testresults[6+x] <- arch$arch.mul[3]

testresults[25+x] <- st1$serial[1]
testresults[26+x] <- st1$serial[3]
testresults[27+x] <- norm1$jb.mul$JB[1]
testresults[28+x] <- norm1$jb.mul$JB[3]
testresults[29+x] <- arch1$arch.mul[1]
testresults[30+x] <- arch1$arch.mul[3]
}

##ADF-Results##
adfsummary <- matrix(0, nrow = 5, ncol = 4)
dimnames(adfsummary) <- list(c("FOBdef","$Days$"
,"$Caputil$","$Refutil$","MM"),c("teststat","1pct","5pct","10pct"))

adfsummary2 <- matrix(0, nrow = 5, ncol = 4)
dimnames(adfsummary2) <- list(c("FOBdef","$Days$"
,"$Caputil$","REFUTIL","MM"),c("teststat","1pct","5pct","10pct"))

for (we in 1:5) {
adf1 <- summary(ur.df(mtsData[,we], type = "trend"
, selectlags = "AIC"))
adf2 <- summary(ur.df(diff(mtsData[,we])
, type = "drift", selectlags = "AIC"))

adfsummary[we,1] <- adf1@teststat[1]
adfsummary[we,2] <- adf1@cval[1,1]
adfsummary[we,3] <- adf1@cval[1,2]
adfsummary[we,4] <- adf1@cval[1,3]

adfsummary2[we,1] <- adf2@teststat[1]
adfsummary2[we,2] <- adf2@cval[1,1]
adfsummary2[we,3] <- adf2@cval[1,2]
adfsummary2[we,4] <- adf2@cval[1,3]
}
```

```
##bootstraipairs.nnreg##
bstrap.pairs.nnreg <- function (exp,dep,x)
{
    realization <- matrix(0, ncol = 4, nrow = x)
    dimnames(realization) <- list(format(1:x)
    , c("2009Q1","2009Q2","2009Q3","2009Q4"))
    z <- length(dep)

    for (j in 1:x) {
set.seed(j)
tempv <- sample(z-4,replace=T)
set.seed(j)
outnnreg <- nnreg(exp[tempv,],dep[tempv],1,4)
prediction <- predict.nnreg(outnnreg, exp[z-3:z,])

        realization[j, 1] <- prediction[1,1]
        realization[j, 2] <- prediction[2,1]
        realization[j, 3] <- prediction[3,1]
        realization[j, 4] <- prediction[4,1]
}
return(realization)
}

##bootstrapresiduals.nnreg##
bstrp.residuals.nnreg <- function (expo,dep,x)
{
    z <- length(dep)
    set.seed(1)
    temp <- nnreg(expo[1:(z-4),], dep[1:(z-4)], 1,4)
    realization <- matrix(0, ncol = 4, nrow = x)
    dimnames(realization) <- list(format(1:x)
    , c("2009Q1","2009Q2","2009Q3","2009Q4"))

    for (j in 1:x) {
set.seed(j)
tempv <- sample(z-4,replace=T)
```

```
newdep <- temp$fitted.values[,temp$best.model]
+ temp$residuals[,temp$best.model][tempv]
set.seed(j)
outnnreg <- nnreg(expo[1:(z-4),],newdep,1,4)
prediction <- predict.nnreg(outnnreg, expo[z-3:z,],)

        realization[j, 1] <- prediction[1,1]
        realization[j, 2] <- prediction[2,1]
        realization[j, 3] <- prediction[3,1]
        realization[j, 4] <- prediction[4,1]
}
return(realization)
```

Other programs upon request.