MASTER THESIS

A COINTEGRATION ANALYSIS OF NYMEX CRUDE OIL PRICE

Tasiopoulos Konstantinos

October 2019

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
UNIVERSITY OF PIRAEUS

Abstract

The scope of this thesis is to conduct a cointegration analysis on crude oil fundamentals based on (Dagoumas and Perifanis 2018) and examine whether a long run relation exists. In order to achieve this two different techniques were used namely residual based method and system estimation method. In the first case we used Engle – Granger procedure, while in the second case we used Johansen's maximum likelihood estimation. Moreover, we imposed identification restrictions in order to identify purely exogenous shocks to the system and examine how crude oil price will react. Also, we applied Impulse Response Function analysis for presenting the aforementioned reaction graphically for twelve month horizon. The period that we analyzed was from 2008 until 2018 in a monthly-based frequency. The independent variables was from different segments of the oil industry such as purely economic fundamentals in supply and demand framework, from financial sector and macroeconomic variables. Additionally, we evaluated the elasticities in the short and in the long run in order to investigate which variables are responsible for fluctuations for these different periods. Empirical results showed that there is more than one Cointegration relationship and that the crude oil future's price is mainly affected from fundamentals (supply and demand) in the long run but in the short run financial markets could also affect the price.

Key words: Cointegration, Johansen procedure, Engle-Granger procedure, VECM, crude oil price.

Acknowledgements

I would like to thank my supervisor Assistant Professor Dagoumas, for his continuous support and guidance, which helped to the fulfillment of this thesis. I would also like to thank my family and friends, for their moral support and encouragement.

Table of Contents

List of Figures

List of Tables

Chapter 1

LITERATURE REVIEW

Nowadays, there is much debate about the determinants of the crude oil price in a global scale. During the last decade there are several new features in the international crude oil market such as the increasing globalization, the rapid growth of oil demand in non-OECD countries and the increasingly interest in the financial attributes of petroleum. Also, the rapid development of China's economy as the second largest oil consumer and the third highest oil-importing country in the world makes its dependence on imported oil to exceed 65% (Wu and Zhang 2014).

A brief representation of the evolution of crude oil price could be that since 2005 it gradually followed and upward trend and lastly it increased to a record high in July 2008 in $134.56 / brll. Such a peak could be attributed to several global factors like Iraq invasion, strikes around crucial oil exporting nations such as Venezuela and Nigeria, natural disasters like Katrina and Rita hurricanes, but it also to the fact that the global economy has entered into a rapid growth phase. Other dynamics could be the tight relationship between supply and demand, the progressively augmented role of speculators from other finance and commodity trading markets into oil market and as a final point the devaluation of US dollar caused by the reduction of interest rates of Fed pushed up oil prices.

Subsequently, because of the global financial crisis, world oil demand was decreased and oil price dramatically collapsed to $43.05 / brrl in December 2008. In February 2011, oil price exceeded $100 per barrel again in just five months, i.e., $104.03 per barrel. From January 2011 to June 2014, the oil price remained volatile at a high level with a slight upward trend (Wu and Zhang 2014). However, high oil prices led to a significant surge in shale oil production in the US. Hence, in the second half of 2014 oil production in US had been growing rapidly, because of the revolution in shale oil technology and the abolish of the US oil export embargo in late 2015. Furthermore, the substantial growth in OPEC oil supply parallel to the weak oil demand, the stronger US dollar and the substantial growth of speculation, caused a slump in oil prices to $31.93/brrl in the beginning of 2016. Thereafter, oil prices have been slowly rebound to a level above $50/brrl due to the fact that OPEC and non-OPEC producers reached an agreement on production reductions for the first time in late 2016 in the 170th (Extraordinary) Meeting of the OPEC Conference as mentioned in (Perifanis and Dagoumas, 2019) . This is enhanced by the fact that other economies, as in the case of Russia, are similarly oil dependent (Perifanis and Dagoumas, 2017).

In our effort to examine the oil price fundamentals and more specifically whether oil price shocks are caused on the supply or the demand side, we first have to specify if those shocks are exogenous or endogenous in macroeconomic models. Depending on the nature of the shock it would also have different consequences in determining the adequate monetary policy response (Kilian, 2009a). Hence, it is reasonable that oil price determinants have become a widespread field of study.

In our analysis we will highlight the role of the following explanatory factors that could possibly contribute to the settlement of crude oil prices based on the literature. Those are:

- The growing demand due to augmented global economic growth parallel with price/income elasticity of crude oil demand (Demand side)
- Supply shortages whether they are based on concerted practices of oil producing countries(Supply side)

- The behavior of financial market participants or speculation (Financial)
- The increasing role of crude oil inventories
- Turmoil in oil producing countries such as Middle-East (Political)
- The role of the trade weighted US dollar index (Macroeconomic variable)

The aforementioned determinants could coexist together. According to Hamilton (2008) an interpretation or a causal relationship among them could be that increasing demand meets stagnating supply and that prompts speculation about future shortfalls which then leads oil producing countries to accumulate oil reserves. Hotelling (1931) in his seminal paper shows that in a competitive market, the optimum extraction path would be such that the price of the non-renewable resource (in our case the price of oil) will rise over time at a rate equal to the interest rate *r*. Therefore, as the price of oil keeps rising, demand is slowly halted and eventually disappears due to high prices. As mentioned in Hamilton (2009) price should exceed marginal cost even if the oil market were perfectly competitive. Likewise, Hamilton (2009) highlights the role of scarcity rent which is the difference between price of oil and marginal production cost and he also emphasizes the role of income and price elasticity which are both estimated well below unity. In that case, price inelasticity means that if the price of oil goes up, total expenditures on oil go up. Income inelasticity means that as real income (real GDP) goes up, the share of oil expenditures should fall.

The role of demand was examined in Kilian (2008b) where he uses a newly developed measure of global economic activity (Kilian economic index), and proposes a structural decomposition of the real price of crude oil into three main components:

- Supply shocks.
- Shocks to the global demand for all industrial commodities.
- Demand shocks that are specific to the crude oil market.

His analysis showed that oil price increases may have different effects on the real price of oil, depending on the underlying cause of the price increase. Likewise, an increase in precautionary demand for crude oil causes an immediate, persistent, and large increase in the real price of crude oil while an increase in aggregate demand for all industrial commodities causes a slightly delayed but sustained increase in the real price of oil and he concludes that crude oil production disruptions cause a small and transitory increase in a one year horizon. Thus, high oil prices can slow down economic growth, cause inflationary pressures, increase uncertainty and discourage further investment in the oil sector Fattouh (2007).

By examining the supply-side effects in crude oil price, it is a common consensus the fact that if a negative supply shock incurs, aggregate macroeconomic demand would have fallen. This could be translated as an underlying tax on final consumers in favor of oil producers. Furthermore, a supply shock could drive production costs and inflation which in turn prompts central banks to raise their interest rates, thereby further diminishing economic activity. A game changer in the last decade was the rapid expansion of U.S. shale oil production (although capital intensive) which was stimulated by the high price of conventional crude oil, which made this new technology competitive. As a reaction OPEC experienced significant losses in its balance sheets and his long standing dominant position in affecting crude oil price has been tested against shale oil producers.

Increasing speculation could thrive under tight market conditions, geopolitical uncertainties and limited spare capacities. According to Fattouh (2007) despite the fact that inventories have risen, investors believe that in case of a supply shock the current level of inventories would not be enough to absorb the price rise. Practitioners during the last fifteen years have highlighted the role of noisy traders. Black (1986) defines noise traders as agents who sell and buy assets on the basis of irrelevant information rather than on market fundamentals or the arrival of new information. Despite the fact that noise traders may be active in financial markets, the traditional view has been that their role can be ignored because they will continuously lose money and will eventually exit the market. As mentioned in Friedman (1953) "*people who argue that* *speculation is generally destabilizing seldom realize that this is equivalent to saying that speculators lose money since speculation can be destabilizing in general only if speculators on average sell low and buy high*". As mentioned in Alquist and Kilian (2007), Büyükşahin et al. (2009), and Fattouh (2010a), from the starting point of 2003 there was an inflow of investors from outside the oil industry (hedge funds) into oil futures markets. These new financial investors were attracted by high returns. At this point both spot and futures prices of crude oil began to surge and finally reached unprecedented levels. A natural conjecture at the time was that this price surge was caused by the financialization of oil futures markets. On the contrary as mentioned in Fattouh (2013) an alternative explanation could be that financial investors merely responded to the same market forces as other market participants and that both spot and futures prices were driven by the same economic fundamentals. He continued that another possible interpretations could be the low risk-free interest rates parallel to the low returns in other financial markets or that crude oil was seen as a good hedge against inflation risks and a weak U.S. dollar. Finally, Alquist, Kilian and Vigfusson (2012) examine the out-of-sample accuracy of daily and monthly oil futures prices and show that there is no compelling evidence that oil futures prices help forecast the spot price of oil. Thus, providing evidence of a more confined role of speculation in the oil surge of 2008.

Many surveys have underlined that movements in exchange rate of the US dollar have the power to influence commodity prices. Sadorsky (2000) shows that futures prices of crude oil, heating oil and unleaded gasoline are cointegrated with the trade weighted US dollar index. Moreover, Zhang et al. (2008) identify a significant long-term equilibrium and cointegrating relationship between exchange rate of US dollar and crude oil prices. Also, Kilian (2018) mentioned that the real price of oil, through its effects on the terms of trade, could be a primary determinant of long swings in the trade-weighted U.S. real exchange rate. Similarly, Brown and Phillips (1986) and Trehan (1986) suggested that the appreciation of the dollar in the early 1980s lowered the demand for oil outside of the United States and stimulated the supply of oil outside of the United States, contributing to the fall in the real price of oil. What is more, the sustained surge in the real price of oil in the 2000 is often attributed in part to the declining real value of the dollar. In addition, Kilian and Zhou (2018) showed that an exogenous increase in the U.S. real interest rate causes only a modest decline in the real price of oil and this effect tends to be short-lived. The real value of the dollar appreciates strongly and persistently, and the level of global real activity declines. On the same survey he finds that the real depreciation of the U.S. dollar helped reinforce the surge in flow demand caused by the economic boom in emerging economies which is the second most important explanation of this sustained surge in the real price of oil because it accounts for a cumulative increase of 50% in the real price of oil compared with a 65% cumulative increase caused by demand shocks directly associated with the global business cycle.

Another fundamental parameter of crude oil price in the bibliography is the enhancing role of inventories as a hedging factor against fluctuations of oil price. As noted in Fattouh (2007) the current build-up of inventories is a sign of oversupply in the crude oil market while others have argued that this incident was driven by the demand for precautionary inventories. Litzenberger and Rabinowitz (1995) noted that 80-90 percent of the time the oil forward curve is in backwardation meaning that futures prices are lower than spot prices. More generally the fact that oil for future delivery is trading at a large premium over immediate delivery makes the cost of carrying inventories to be covered thus prompting market participants with storage facilities to accumulate inventories and make a profit by selling contracts in the futures market. The relationship between oil prices and oil inventories is negatively correlated. Finally, Fattouh (2007) mentioned the case that the contango market and the associated rise in inventories could occurred together with an upward trend in oil prices. In an effort to explain this absurd relationship, some observers argue that large inventories are no longer seen as a sign of oversupply and hence do not exert downward pressure on prices.

Thus, in an effort to reconcile all the above mentioned parameters we will conduct our analysis in a framework that will included different variables of the abovementioned factors. Hence, from the fundamentals perspective we employed supply and demand variables such as OPEC production, tight oil or shale oil production, real economic activity as a proxy for world demand, trade weighted US dollar index, goldman sachs commodity index in order to examine the 'paper oil' factor and the OECD oil inventories. Finally, the metric of real crude oil nymex price was used as a dependent variable deflated with the CPI 82-84 of US in 2010 as a base price.

Chapter 2

Data Description

Because of the fact that crude oil as a commodity that belongs to a highly liquid market, there are several factors that could determine its price. We know from economic theory that settlement price is determined in the equilibrium point of supply and demand. However, there are a variety of sub categories in the supply side such as OPEC production behavior, US shale oil production but also in the demand side like emerging economies with developing industries and speculators. Hence, for this analysis we will try to develop a model that can be applicable upon those aforementioned factors. The period under study is from 2008 until 2018 on a monthly basis. Thus, 132 observations. The sample size was chosen due to the fact that we want to examine the price fluctuation and the main drivers behind it during the ,approximately, last decade.

The price of crude oil that we have chosen to analyze is the New York Mercantile Exchange (NYMEX) one month forward real settlement price (adjusted for inflation).This time series was taken from EIA's database and the CPI was taken from FED. In order to interpret consumption in our model we used the re-adjusted Killian's global real economic activity index derived from ocean freight rates Killian (2009 and 2018). Also, in the demand side belongs the variable OECD stocks which are the reserves in crude oil that OECD countries have (from EIA). The market supply factors were illustrated by OPEC's crude oil production in thousand barrels per day and by US shale oil (tight oil) production in thousand barrels per day also. This variables was taken from EIA. In order to calculate the overall shale oil production we have to add up each region's production. According to Büyükşahin and Robe (2014a; 2014b) a variable that incorporates the paper oil trading is the S&P GSCI crude oil index. Because of the fact that such a financialization index could play a fundamental role in determining nymex crude oil price, we derived it from investing database. Moreover, macroeconomic factors such as Trade Weighted Dollar Index from FED's database are also included. The rationale behind this selection is based upon the fact that as the Dollar weakens against other currencies it becomes cheaper from oil consuming countries to buy crude oil. The opposite holds true if dollar appreciates against other currencies.

Variables	Description	Data source
N	One month forward crude oil prices in real US dollars	EIA
SHALE	Aggregate tight oil production in thousand barrels per day	EIA
OPEC	OPEC crude oil production in thousand barrels per day	EIA
STOCKS	OECD crude oil reserves in thousand barrels per day	EIA
REA	Killian's readjusted real economic activity	http://www-personal. umich.edu/~lkilian/reaupdate.txt
TWDI	Trade weighted dollar index	FED
GSC	S&P GSCI crude oil index	Investing.com

Figure 2.1: Data sources for all the variables that are used in the analysis (Jan 2008- Dec 2018)

We begin our analysis by transforming all the variables to natural Neperian logarithms in order to examine their respective elasticities. Also, because of their trending nature (stochastic or deterministic) we could use log transformation in order to de-trending them.

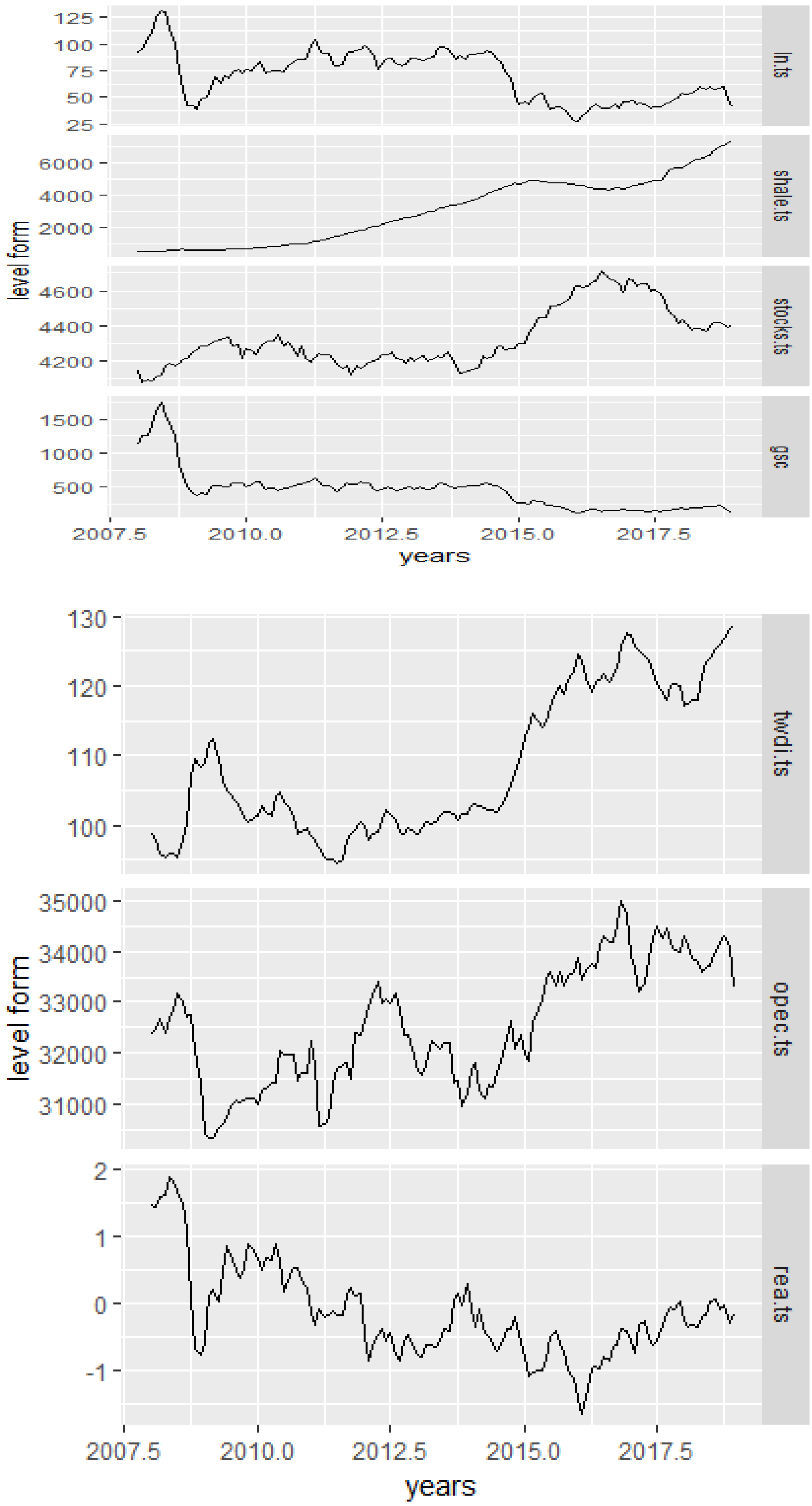


Figure 2.2: Diagram of all variables in log-level form

In addition, because of the fact that our variables are non-stationary in their level form, we took their first differences to make them I(1) stationary. According to Engle and Granger (1987) an integrated series is defined as *a series with no deterministic component that has a stationary, invertible ARMA representation after differencing d times and is said to be integrated of order d, which is denoted as* $x_t \sim I(d)$.

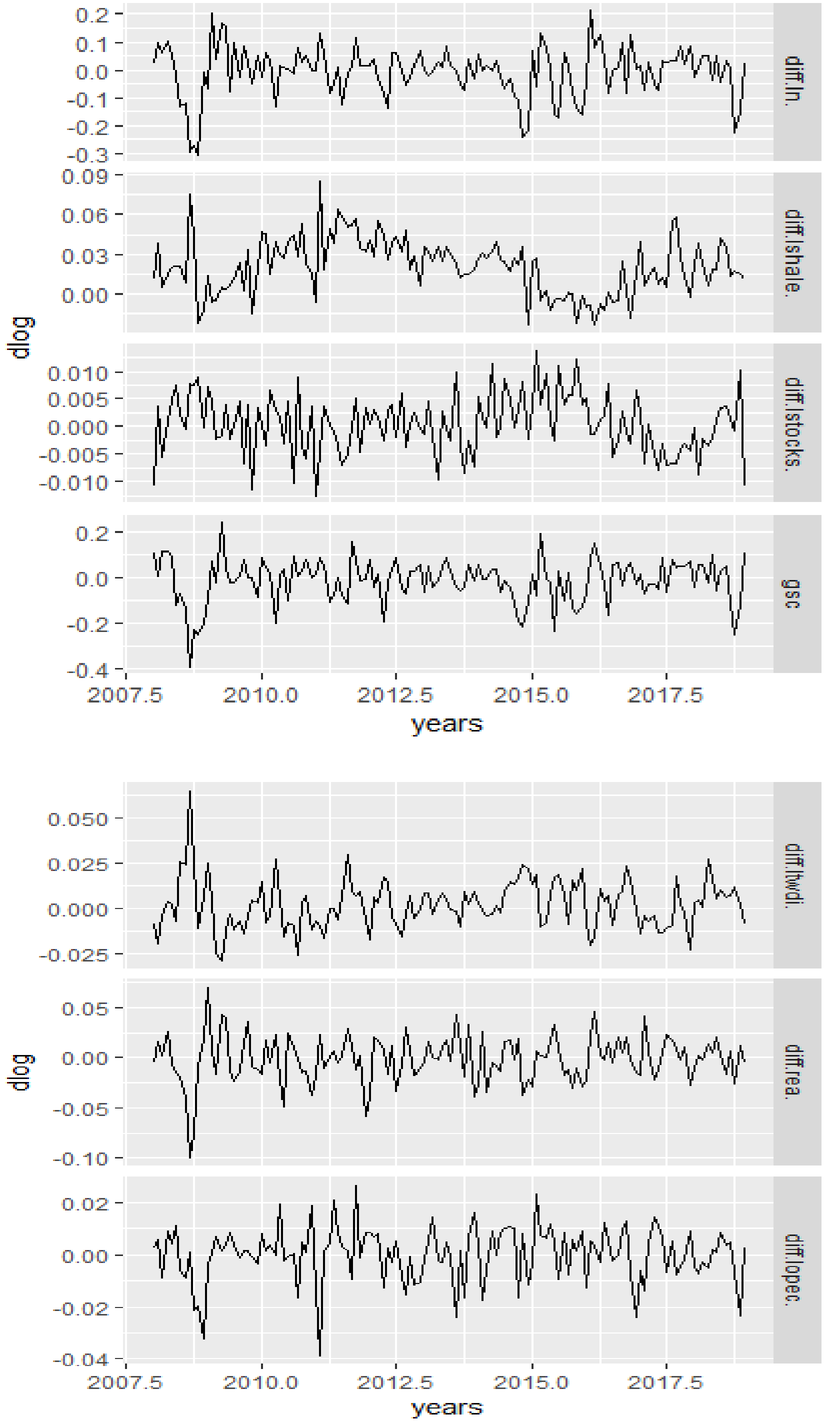


Figure 2.3: Diagram of all variables in log-differenced form

As we notice in the above mentioned plots, it is clear that the log-level form is an explosive process with non-zero mean and a practically infinite variance. On the other hand, the log-differenced form (or the percentage returns) is a mean reverting process with finite variance. Hence, we can proceed with our analysis with our I(1) variables. Finally, it obvious that periods with high volatility (spikes in the plots) are the period of global economic recession (second semester of 2008 until first semester of 2009) and the period of low crude oil prices (last quarter of 2014).

Subsequently, we analyze the descriptive statistics of all our seven variables in their level and differenced form. It is obvious from our data that financial variables such as LN and LGSC that are experiencing stochastic trends are more volatile than the other variables. From excess kurtosis values, we notice that first differences in log variables have a highly non-normal distribution. The negative numbers of skewness show that returns have a left-skewed distribution.

	LN	LSHALE	LSTOCKS	LGSC	LTWDI	REA	LOPEC
Mean	4.183	7.706	8.372	5.930	4.682	-0.015	10.39
Std.Dev.	0.360	0.869	0.038	0.625	0.094	0.068	0.035
Variance	0.129	0.754	0.001	0.391	0.009	0.004	0.001
Min	3.309	6.202	8.316	4.880	4.549	-0.163	10.32
Max	4.873	8.893	8.451	7.469	4.856	0.187	10.46
Median	4.309	8.028	8.362	6.176	4.635	-0.031	10.39
Skewness	-0.33	-0.435	0.758	0.049	0.405	0.916	0.005
Kurtosis	-1.046	-1.388	-0.625	-0.572	-1.378	0.798	-1.103

Figure 2.4: Descriptive statistics of all variables in log-level form

	dLN	dLSHALE	dLSTOCKS	dLGSC	dLTWDI	dREA	dLOPEC
Mean	-0.006	0.021	0.000	-0.016	0.002	-0.001	0.000
Std.Dev.	0.092	0.021	0.005	0.098	0.013	0.024	0.01
Variance	0.008	0.000	0.000	0.009	0.000	0.000	0.000
Min	-0.304	-0.024	-0.012	-0.393	-0.029	-0.100	-0.039
Max	0.212	0.085	0.013	0.242	0.064	0.069	0.027
Median	0.008	0.02	0.000	-0.004	0.002	-0.001	0.001
Skewness	-0.772	0.154	-0.014	-0.75	0.705	-0.569	-0.711
Kurtosis	1.222	-0.108	-0.322	1.229	2.317	2.013	1.289

Figure 2.5: Descriptive statistics of all variables in log-differenced form

Finally, we continue with a representation of a correlation matrix from 1 to -1. As the values are getting closer to 1 we say that they are positively correlated. The opposite holds true if the variables are getting closer to -1 (negatively correlated). For values close to 0 we say that there is no correlation after all. A simple distribution of each time series in log-level form is depicted in the diagram below. We notice that our depended variable LN (NYMEX crude oil price in real dollars) is strongly negative correlated with LSTOCKS and LTWDI while strongly positively correlated with LGSC. In the first case we have that as the depended variable increases the other variables are decreasing while in the second case we have that as the depended variable increases the independent variable increases accordingly. Weakly correlated with LN are the variables LSHALE, REA and LOPEC. However, we notice an interdependence between all variables. That is an issue of multicollinearity that we have to deal with when we proceed with our analysis.

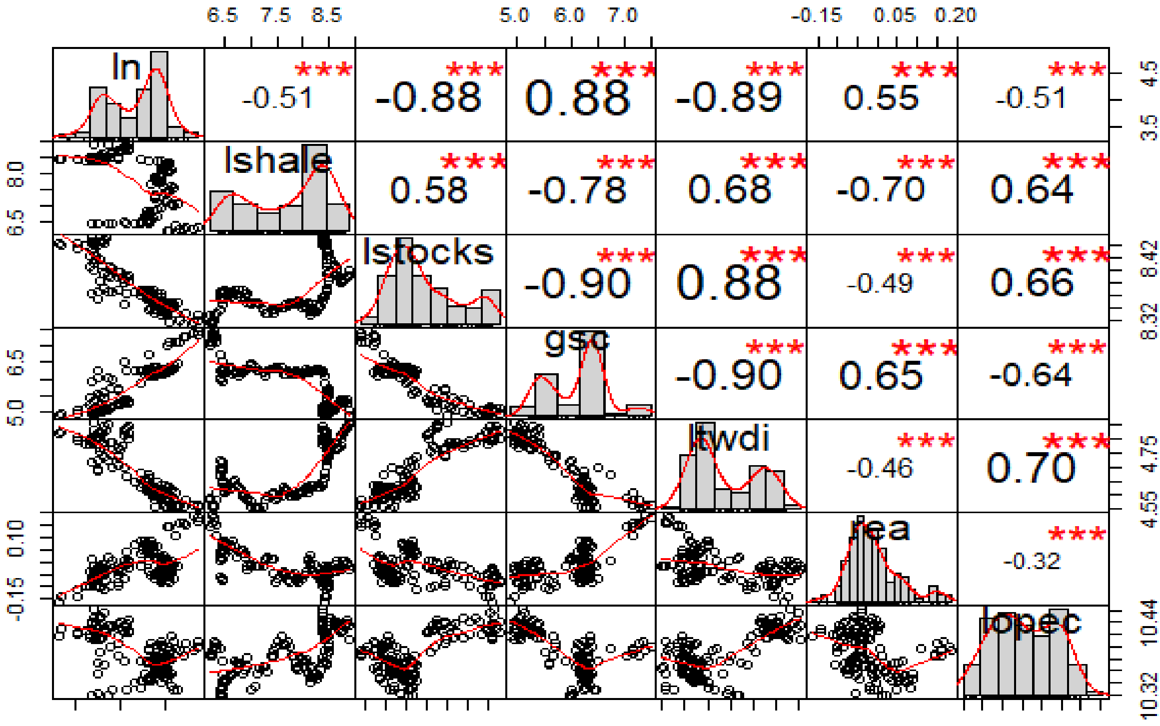


Figure 2.6: Correlation matrix of all variables in log-level form with their respective distributions and scatter plots with regression line.

Chapter 3

3.1 Introduction

Firstly, we will begin our presentation in this chapter by presenting the basic form of a Vector Autoregressive Model (VAR). A VAR consists of a set of K endogenous variables has the following form:

$$y_t = A_1 y_{t-1} + \ldots A_p y_{t-p} + CD_t + u_t \qquad (3.1)$$

Where A_i are $(K \times K)$ coefficient matrices for i= 1, … , p and u_t is a K- dimensional white noise process with time invariant positive definite covariance matrix $E(u_t u_h') = \Sigma u$. The matrix C is the coefficient matrix of deterministic regressors (if present) with (K x M) dimensions and D_t is a column vector with (M x 1) dimensions. As deterministic regressors could perceived constant term, trend and dummy variables. We could also express equation (3.1) as a lag polynomial with the term L denoting the lag operator:

$$A(L)y_t = CD_t + u_t \qquad (3.2)$$

Where: $A(L) = (I_k - A_1 - \ldots - A_p)$.

A fundamental prerequisite of a VAR (p) process is its stability. This could be achieved if the time series included in VAR (p) are stationary with time – invariant means, variances and covariance structure. We are able to examine the stability conditions by evaluating the reverse characteristic polynomial, which can be represented as follows:

$$\det (Ik - A1z - \ldots - Apzp) \neq 0 \text{ for } |z| \leq 1 \qquad (3.3)$$

In case the solution of the above equation (3.3) has a root for z = 1, then either some or all variables in the VAR(p) process are integrated of order one I(1). Also, the stability can be analyzed by considering the companion form and calculating the eigenvalues of the coefficient matrix (see Lütkepohl (2006) for more detailed derivation). As an example we will present a VAR(1) process as:

$$\xi t = A\xi t\text{-}1 + vt \qquad (3.4)$$

$$\xi t = \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \; A = \begin{pmatrix} A_1 & A_2 & & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & & 0 & 0 \\ & \vdots & \ddots & \vdots & \\ 0 & 0 & \cdots & I & 0 \end{pmatrix}, vt = \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

Where the dimension of the vectors ξ_t and v_t is ($Kp\times 1$) and the dimensions of matrix A is ($Kp\times Kp$). In case we have in modulus the eigenvalues of A less than one, then the VAR(p) process is a stable one. For a given sample of the endogenous variables $y_1,...y_T$ the coefficients of a VAR(p) could be estimated efficiently by OLS (Ordinary Least Squares) which can be applied separately to each equation of the whole system. If the error process u_t is normally distributed, then this estimator is equal to the one that has been derived from maximum likelihood.

In general it is possible to interpret a stable AR(p) process as an infinite sum of MA process in the univariate case. This is also possible when we have a stable VAR(p). The representation is given from *Wold theorem* and is the following one:

$$y_t = \Phi_0 u_t + \Phi_1 u_{t-1} + \dots \tag{3.5}$$

where $\Phi_0 = I_\kappa$ and Φ_S are matrices that can be computed as:

$$\Phi_s = \sum_{j=1}^{s} \Phi_{s-j} A_j \tag{3.6}$$

Where $A_j = 0$ for $j > p$

The issue of the appropriate selection of lags in the VAR(p) process can be tackled with the usage of Information Criteria which could select the optimal length without losing degrees of freedom and parallel to that minimize the sum of squared errors. As in the univariate AR(p)-models the information criteria that are widely used in empirical research are the following: Akaike (1981), Hannan and Quinn (1979), Quinn (1980), or Schwarz (1978), or by the *final prediction error* for a detailed exposition of these criteria). These measures are defined as

$$AIC(p) = \log\det(\Sigma_u(p)) + \frac{2}{T} pK^2, \tag{3.7a}$$

$$HQ(p) = \log\det(\Sigma_u(p)) + \frac{2\log(\log(T))}{T} pK^2, \tag{3.7b}$$

$$SC(p) = \log\det(\Sigma_u(p)) + \frac{\log(T)}{T} pK^2, \tag{3.7c}$$

$$FPE(p) = \left(\frac{T+p*}{T-p*}\right)^K \det(\tilde{\Sigma}_u(p)) \tag{3.7d}$$

Where $\Sigma_u(p) = T^{-1}\sum_{t=1}^{T} \hat{u}_t \hat{u}_t'$ and p* is the total number of parameters in each equation and p assigns the lag order. It is shown in Lutkepohl (2006) sample sizes.

3.2 Diagnostic Tests

After estimating our model it is very crucial to test if the residuals are aligned with the model's assumptions (white noises). Hence, we should examine with the usage of the appropriate statistical tests for the absence of serial correlation and heteroscedasticity and see if the error process is normally distributed but for the multivariate case. Also, we should further conduct the structural stability of the model with CUSUM, CUSUM-of-squares, and/or fluctuation tests.

To begin with, the two most commonly applied tests for serial correlation in the residuals of a VAR(*p*) model are: i) Portmanteau test and ii) LM (Langrage Multiplier) test proposed by Breusch (1978) and Godfrey (1978). The Portmanteau statistic according to Lutkepohl (2007) is testing the overall significance of the residual autocorrelations up to lag h and is defined as:

$$Q(h) = T\sum_{j=1}^{h} \mathrm{tr}(C_j^{'} C_0^{-1} C_j^{'}\ C_0^{-1}) \qquad (3.8)$$

Where $\hat{C}_i = \frac{1}{T}\sum_{t=i+1}^{T} \hat{u}_t\, \hat{u}_{t-i}^{'}$.

The test statistic has an approximate $x^2(K^2h\text{-}n^*)$ distribution and n^* is the number of coefficient excluding deterministic terms of a VAR(*p*). The limiting distribution is only valid for *h* tending to infinity at a suitable rate with growing sample size. Hence, the trade-off is between a decent approximation to the x^2 distribution and a loss in power of the test when *h* is chosen too large. By using Monte Carlo techniques, it was found by some researchers that in small samples the nominal size of the portmanteau test tends to be lower than the significance level chosen Ljung& Box (1978), Hosking (1980). Thus, as a consequence the test has low power against many alternatives. For that reason it has been suggested to use the modified test statistic:

$$Q^*(h) = T2\sum_{j=1}^{h} \frac{1}{T-j} tr(C_j^{'} C_0^{-1} C_j^{'}\ C_0^{-1}) \qquad (3.9)$$

The second test is Breush-Godfrey LM-statistic and is based upon the following auxiliary regressions:

$$\hat{u}t = A1yt-1 + \ldots + Apyt-p + CDt + B1\hat{u}tt-1 + \ldots + Bh\hat{u}t-h + \varepsilon t. \qquad (3.10)$$

The null Hypothesis is $H_o : B_1 = \ldots = B_h = 0$, and correspondingly the alternative hypothesis is of the form $H_1 : B_i \neq 0$ for i = 1,2,…h. The test statistic is defined as:

$$LMh = T(K - tr(\tilde{\Sigma}_R^{-1}\tilde{\Sigma}_e)) \qquad (3.11)$$

Where $\tilde{\Sigma}_R^{-1}$ and $\tilde{\Sigma}_e$ assign the residual covariance matrix of the restricted and unrestricted models, respectively.The test statistic LMh is distributed as $\chi^2(hK^2)$.

In the case of heteroscedasticity testing we will use the multivariate ARCH tests Engle (1982), Hamilton (1994) and Lutkepohl (2006). The multivariate ARCH-LM test is based upon the following regression:

$$vech(\hat{u}_t\hat{u}_t^{'}) = \beta 0 + B1vech(\hat{u}_{t-1}\hat{u}_{t-1}^{'}) + \ldots + Bqvech(\hat{u}_{t-q}\hat{u}_{t-q}^{'}) + vt, \qquad (3.12)$$

where v_t assigns a spherical error process and *vech* is the column-stacking operator for symmetric matrices that stacks the columns from the main diagonal on downward. The null hypothesis is $H_0 : = B_1 = B_2 = \ldots B_q = 0$ and the alternative is $H1 : B1 \neq 0 \cap \ldots \cap B_q \neq 0$. The test statistics is defined as:

$$\text{VARCH}_{LM}(q) = \frac{1}{2} TK\,(K+1)\,R_m^2 \tag{3.13}$$

And

$$R_m^2 = 1 - 2/\,K(K+1)\,tr(\hat{\Omega}\hat{\Omega}_0^{-1}) \tag{3.14}$$

Where $\hat{\Omega}$ denotes the covariance matrix of the aforementioned regression model. The test statistic is distributed as $\chi^2(qK^2(K+1)^2/4)$.

In order to test for normality we will apply the multivariate Jarque-Beratest Bera and Jarque (1980), (1981), Jarque and Bera (1987), and Lutkepohl (2006). The univariate versions of the Jarque-Bera test are applied to the residuals of each equation. By using the residuals that are standardized by a Choleski decomposition of the variance-covariance matrix for the centered residuals, a multivariate version of this test can be computed. In this case, please note that the test result is depended upon the ordering of the variables. The test statistics for the multivariate case are defined as:

$$\text{JB}_{mv} = s_3^2 + s_4^2 \tag{3.15}$$

Where

$$s_3^2 = \text{T} b_1' b_1 / 6 \tag{3.16}$$

$$s_4^2 = \text{T}(b_2 - 3_k)'(b_2 - 3_k) / 24 \tag{3.17}$$

Where b_1 and b_2 are the third and fourth non-central moment vectors of the standardized residuals $\hat{u}_t^s = \hat{P}^-(\hat{u}_t - \bar{u}^\wedge{}_t)$ and $\tilde{P}$ is a lower triangular matrix with positive diagonal such that $PP' = \Sigma_u$ which is due to the Choleski decomposition of the residual covariance matrix. The test statistic, JB_{mv} is distributed as $X^2(2K)$ and the multivariate skewness, s_3^2 and kurtosis test s_4^2 are distributed as $X^2(K)$.

3.3 Impulse Response Functions

The main reason that we use impulse response analysis is to investigate the dynamic interactions between the endogenous variables. The whole philosophy of the procedure is based upon the Wold moving average representation of a VAR(p)-process that we have already mentioned. Thus, the $(i,j)^{th}$ coefficients of the matrices Φ_s are interpreted as the expected response of variable $y_{i,t+s}$ to a unit change in variable y_{jt}.

These effects can be cumulated through time $s = 1,2, \ldots$ according to our needs (of course looking beyond some threshold augments the error of the estimation) and hence one would obtain the cumulated impact of a unit change in variable j on the variable i at time s. Moreover, it is often very handy to use *orthogonal impulse responses*. This is plausible if the underlying shocks are less likely to occur in isolation but rather contemporaneous correlation between the components of the error process u_t. This can be the case when the off-diagonal elements of Σ_u are non-zero which is something very common. The orthogonal impulse responses then are derived from a Choleski decomposition of the error variance-covariance matrix $\Sigma_u = PP'$ with P being lower triangular. The moving average representation can then be transformed to

$$y_t = \Psi_0 \varepsilon_t + \Psi_0 \varepsilon_{t-1} + \ldots , \tag{3.18}$$

With $\varepsilon_t = P^{-1}\mathbf{u}_t$ and $\psi i = \Phi_i P$ for $i = 0,1,2\ldots$ and $\Psi_0 = P$. Because of the fact that the matrix P is lower triangular, it follows that only a shock in the first variable of a VAR(p)-process exerts an influence on all the other ones and the other variables cannot have a direct impact on y_{1t}. Hence, an ordinal structure of the error terms is imposed. It is worth mentioning that orthogonal impulse responses were applied in this thesis.

3.4 Forecast Error Variance Decomposition

The forecast error variance decomposition (FEVD) is based upon the orthogonal impulse response coefficient matrices Ψ_n. The main purpose is to analyze the contribution of variable j to the h-step forecast error variance of variable k. If the element-wise squared orthogonal impulse response are divided by the variance of the forecast error variance, $\sum_k^2(h)$we are taking a percentage figure as a result. Thus, we can define the forecast error variance for $y_{k,T+h} - Y_{k,T+h|T}$is defined as:

$$\sigma_k^2(h) = \sum_{n=0}^{h-1}(\psi_{k1,n}^2 + \ldots + \psi_{kK,n}^2) \qquad (3.19)$$

The percentage could be extracted if we divide the term $(\psi_{kj,0}^2 + \cdots + \psi_{kj,n-1}^2)$ with $\sigma_k^2(h)$.The result is:

$$\omega_{kj}(h) = (\psi_{kj,0}^2 + \cdots + \psi_{kj,n-1}^2) \,/\, \sigma_k^2(h) \qquad (3.20)$$

3.5 Unit Root Process

Let us assume that a time series $\{y_t\}$ is an expression of a deterministic trend component and a stochastic one. We can interpret this notation as:

$$y_t = TD_t + u_t\,, \qquad (3.21)$$

$$TD_t = \beta_1 + \beta_2 t \qquad (3.22)$$

$$u_t = \varphi u_{t-1} + \varepsilon_t, \text{where } \varepsilon_t \sim N(0,\sigma^2) \qquad (3.23)$$

Where TD_t assigns for the deterministic linear trend of the above equation and the parameter u_t represents the stochastic part and can be expressed as an AR(1) process. In the case that $|\varphi|< 1$ then y_t is I(0) stationary process around the deterministic trend component TD_t. Otherwise, when $|\varphi|= 1$ then (3) becomes $u_t = u_{t-1} + \varepsilon_t$ or $u_t = u_0 + \sum_{t=1}^{T} \varepsilon_t$ which denotes that y_t is I(1) non-stationary with a stochastic trend. Basically, in order to find if the underlying process is stationary (mean reverting) or non-stationary (unit root process) we have to conduct a *unit root test*. In general, autoregressive unit root tests are based on testing the null hypothesis H_0that $\varphi = 1$ (difference stationary) against the alternative hypothesis H_a that $\varphi< 1$ (trend stationary). The name *unit root* derives because of the fact that under H_0 the characteristic polynomial of u_t, $\varphi(z) = (1 - \varphi z) = 0$,has a root equal to unity.

For illustrative purposes, let us consider a simple AR(1) model like the following one:

$$y_t = \varphi y_{t-1} + u_t, \text{ where } u_t \sim N(0,\sigma^2) \qquad (3.24)$$

as we previously mentioned the null hypothesis that we are going to test is:

H_0: $\varphi = 1$ the process contains a unit root $y_t \sim I(1)$

The test statistic (t – statistic) is given from:

$$t = \frac{\hat{\varphi} - 1}{se(\widehat{\varphi})} \qquad (3.25)$$

where $\hat{\varphi}$ denotes the OLS estimate and *se* the standard error. Also, the AR(1) model could be written as:

$$\Delta y_t = \pi y_{t-1} + u_t \text{ where } \pi = \varphi - 1. \qquad (3.26)$$

Testing $\varphi = 1$ is then equivalent to testing $\pi = 0$. According to Hamilton (1994) if y_t is I(0) stationary process then:

$$\sqrt{T}(\hat{\varphi} - \varphi) \xrightarrow{d} N(0, 1-\varphi^2) \qquad (3.27)$$

where

$$\hat{\varphi} \overset{A}{\sim} N\left(\varphi, \frac{1}{T}(1-\varphi^2)\right) \qquad (3.28)$$

Where A denotes asymptotical *d* denotes convergence in distribution and *A* denotes asymptotically.

Also, the above mentioned t-statistic follows asymptotically N (0,1). The problem that arises is that under the null hypothesis of unit root y_t is not stationary and ergodic. Thus, the sample moments do not converge to fixed constants. According to Phillips (1987) the sample moments of y_t converge to random functions of Brownian motion such as:

$$T^{-3/2}\sum_{t=1}^{T} y_{t-1} \xrightarrow{d} \sigma \int_0^1 W(r)dr \qquad (3.29)$$

$$T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{d} \sigma^2 \int_0^1 W(r)^2 dr \qquad (3.30)$$

$$T^{-1}\sum_{t=1}^{T} y_{t-1} u_t \xrightarrow{d} \sigma \int_0^1 W(r)dW(r) \qquad (3.31)$$

Where the symbol $\xrightarrow{d}$ denotes convergence in distribution and W(*r*) denotes a standard Wiener process (or Brownian motion) on the interval (0,1). Phillips showed that under the unit root null hypothesis (H_0: $\varphi = 1$)

$$T(\hat{\varphi} - 1) \xrightarrow{d} \int_0^1 W(r)\, d\, W(r) \Big/ \int_0^1 W(r)^2 dr \qquad (3.32)$$

$$\text{For } \varphi=1, t \xrightarrow{d} \int_0^1 W(r)\, d\, W(r) \Big/ \left(\int_0^1 W(r)^2 dr\right)^{1/2} \qquad (3.33)$$

Briefly, a Wiener process W (·) is a continuous-time stochastic process, associating each interval $r \in [0, 1]$ a scalar random variable W(r) which satisfies:

- $W_0 = 0$
- for any interval $0 \leq t_1 \leq \dots \leq t_k \leq 1$ the changes $W(t_2) - W(t_1), W(t_3) - W(t_2), \dots, W(t_k) - W(t_{k-1})$ are independent and normally distributed with $W(s) - W(t) \sim N(0, (s - t))$;
- W(s) is continuous in s.

From the above equations we can extract that $\hat{\varphi}$is super consistent which yields that $\hat{\varphi} \xrightarrow{p} \varphi$ (converge in probability) at a faster rate T instead of the usual $T^{1/2}$ in the stationary CLM (central limit theorem) case. Moreover, $\hat{\varphi}$ is not asymptotically normally distributed and its t statistic is not asymptotically standard normal. Thus, in this case the limiting distribution of *t* statistic is called Dickey-Fuller or DF distribution and does not have a closed form representation. Hence, the quantiles of the distribution should be computed through monte carlo simulation Dickey and Fuller (1979). Finally, because of the fact that the normalized bias $T(\hat{\varphi} - 1)$ has a well-defined limiting distribution that does not depend on nuisance parameters it can also be used as a test statistic for the initial unit root null hypothesis ($H_0 : \varphi = 1$).

Furthermore, when we are testing for unit roots, it is crucial to specify the null and alternative hypotheses appropriately to take into consideration the trending behavior of some time series. The type of deterministic terms in the test regression will influence the asymptotic distributions of the unit root test statistics. We will distinct two commonly used cases:

1. **Constant term only (drifting parameter)**

The testing regression is the following one:

$$y_t = c + \varphi y_{t-1} + u_t, \qquad (3.34)$$

The term *c* denotes a non-zero mean under the alternative hypothesis. Thus, we are going to test the following:

$$H_0 : \varphi = 1 \rightarrow y_t \sim I(1) \text{without drifting parameter}$$

$$H_a : |\varphi| < 1 \rightarrow y_t \sim I(0) \text{ with non-zero mean}$$

2. **Constant and Time Trend**

The testing regression is

$$y_t = c + \delta t + \varphi y_{t-1} + u_t,$$

Where *c* is constant and δt denotes the deterministic time trend under the alternative hypotheses. Thus, we are testing:

$$H_0 : \varphi = 1 \Rightarrow y_t \sim I(1) \text{ with drift}$$

$H_a : |\varphi| < 1 \Rightarrow y_t \sim I(0)$ with deterministic time trend

These type of test is used in modeling macroeconomic variables such as GDP

At this point we will present a widely used statistical test for unit roots, the Augmented Dickey Fuller test or ADF. The ADF test tests the null hypothesis that a time series y_t is I(1) (non-stationary) against the alternative that it is I(0) (stationary process). The usage of the word augmented denotes that we will use AR(*p*) parameters to adjust for autocorrelation. Thus, the dynamics in the data have an ARMA structure. The ADF test is based on estimating the following regression:

$$y_t = \beta D_t + \pi y_{t-1} + \sum_{j=1}^{p} \varphi_j \Delta y_{t-j} + u_t, \qquad (3.35)$$

where the parameter Dt is a vector of deterministic terms (constant, trend, seasonal dummies, etc.),$\pi = \varphi - 1$ which under the null hypothesis implies that $\pi = 0$, the sum denotes the p lagged difference terms Δy_{t-j} which are used to eliminate any autocorrelation in the residuals u_t. The error term is also assumed to be white noise. The specification of the deterministic terms depends on the assumed behavior of y_t. Under the null hypothesis, y_t is I(1) which implies that $\varphi = 1$. The ADF t-statistic and normalized bias statistic are based on the least squares estimates of the aforementioned equation and are given by:

$$ADF_t = t = \hat{\varphi} - 1 / \text{se}(\varphi), \text{under the null hypothesis} \qquad (3.36)$$

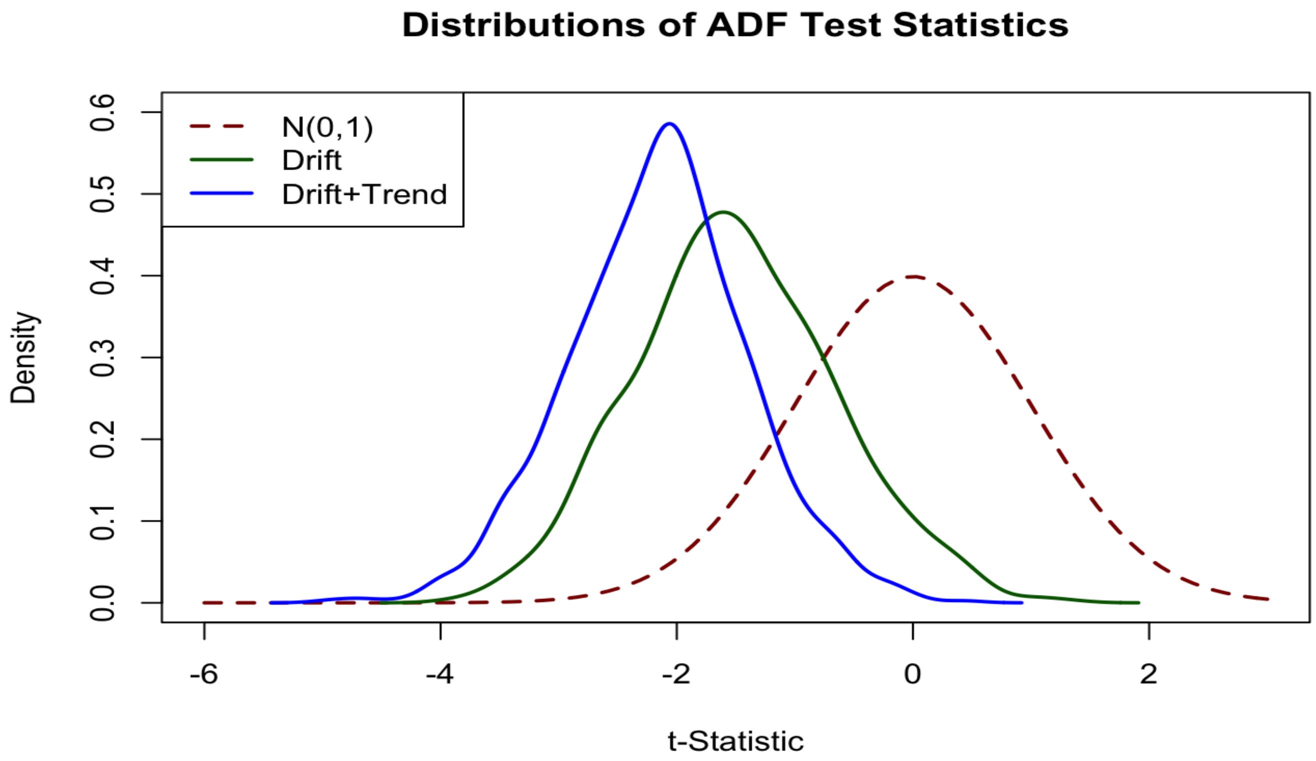


Figure 3.1: ADF distribution under the three different cases that we examined: i) without drift ii) with drift and iii) drift and deterministic linear trend.

In order to select the appropriate lag length *p* we were based on information criteria, such as Akaike (Akaike Information Criteria) (1981) or Schwarz (Bayesian Information Criteria) (1978). Alternatively, the lag order can be determined by testing the residuals for a lack of serial correlation, as can be tested via the Ljung-Box Portmanteau test or a Lagrange multiplier (LM) test.

To conclude, once we correctly identified the lag order *p*, we will use the following procedure. First, we estimate the following ADF-test:

$$\Delta y_t = \beta_1 + \beta_2 t + \pi y_{t-1} + \sum_{j=1}^{p} \varphi_j \Delta y_{t-j} + u_t \qquad (3.37)$$

The following steps are dependent on the behavior of underlying data generating process. This could be according to Pfiaff (2008):

- Stationary around zero mean,
- Stationary around a non-zero mean
- Stationary around a linear trend
- Contains a unit root with zero drift
- Contains a unit root with non-zero drift

So far, the ADF tests have been considered as means to detect the presence of unit roots.

3.6 The Concept of Cointegration and Error- Correction Models

3.6.1 Spurious Regression

In the classical VAR case all variables have to be I(0). In those cases the usual statistical results for linear regression models hold. On the contrary, when we have to deal with variables that are I(1) then the usual statistical results may not hold. This is the case of spurious regression when all the regressors are I(1) and not cointegrated. The following statistical implication will explain in a better way the aforementioned notation.

Let $Y_t=(y_{1t}, ...,y_{nt})'$ denote an $(n\times1)$ vector of I(1) time series that are not cointegrated.

Where $Y_{t} = (y_{1t}, x_{1t})'$ squares regression of y_{1t} on x_{1t} giving the model

$$y_{1t} = \beta_1 X_{1t} + u_t \qquad (3.38)$$

Since y_{1t} is not cointegrated with $X_{1t}$is a spurious regression and the true value of β_1 is zero. The following results about the behavior of β_1 in the spurious regression are due to Phillips (1986):

- β_1 coefficient does not converge in probability to zero but instead converges in distribution to a non-normal random variable not necessarily centered at zero.

- The usual t-statistics for testing that the elements of β_1 are zero diverge to $\pm\infty$ as the sample size goes to infinity $T \to\infty$. Hence, with a large enough sample it will appear that Y_t is cointegrated when it is not if the usual asymptotic normal inference is used.

- The usual R^2 from the regression converges to unity as $T \to\infty$ so that the model will appear to fit well even though it is misspecified.

- In general, regression with I(1) variables only makes sense when the data are cointegrated.

3.6.2 Cointegration and Error - Correction Models

In 1981, Granger (1981) introduced the concept of cointegration into the literature, and the general case was publicized by Engle and Granger (1987) in their seminal paper. Finding a linear combination between two I(d)-variables that yields a variable with a lower order of integration, is the idea behind cointegration. A more formal notion of Cointegration is the following:

The components of vector X_t are said to be cointegrated of order d, b, denoted X_t~CI(d,b), if:

(a) all components of X_t are I(d) and

(b) a vector $\beta \neq 0$ exists so that $Z_t = \beta' X_t$ ~ I(d-b), b>0. The vector β is called the cointegrating vector.

What is really innovative in cointegration theory is the fact that it is now possible to detect stable long-run relationships among non-stationary I(1) variables. Let us consider for illustrative purposes the case of d = 1, b = 1. The variables in the vector x_t are all integrated of order one, but if a linear combination ***a*** exists, then the resultant series u_t is stationary. Individual series are tied to each other by the cointegrating vector, although they are non-stationary. Generally, in economics deviations from a long-run equilibrium between different variables are possible, but these errors are characterized by a mean reversion to its stable long-run equilibrium.

In order to find how to estimate the cointegrating vector ***a*** (and cointegrating systems in general) and how to model the dynamic behavior of *I*(d)-variables, we will examine two widely used methods:

- Single equation methods and
- System methods

In the first one we are interested in estimating a specific cointegrating vector (CI), while in the second one we further determine the number of cointegrating vectors. The most important development is that with the concept of Cointegration is feasible to detect a stable long-run relationships among non-stationary variables. In the case of d = 1, b = 1the components in the vector X_t are all integrated of order one, but if a linear combination β of these exists, then the resultant series u_t is stationary. Individual series are tied to each other by the cointegrating vector, although they are non-stationary.

3.6.2.1 Single equation methods

Engle – Granger residual based method

Engle and Granger (1987) proposed a two-step estimation technique to estimate the cointegrating vector β. Firstly, when the cointegrating vector of a regression bivariate model with I(1) variables is unique ($rank(\Pi) = 1$ as we will see later), then the parameter β can be estimated by OLS (Ordinary Least Squares) and a typical model is the following one:

$$y_t = \alpha_1 + \beta x_t + u_t \qquad (3.39)$$

where u_t ~ I(0) stationary by assumption

$$(\hat{\beta} - \beta) = \sum_{t=1}^{T} x_t u_t \,/\, \sum_{t=1}^{T} x_t^2 \qquad (3.40)$$

Because of the fact that $x_t \sim I(1)$ we have to multiply $(\hat{\beta} - \beta)$ by a scaling factor T (which is the sample size) in order to obtain a non-degenerate asymptotic distribution, so (3.41) amends into:

$$T(\hat{\beta} - \beta) = \frac{1}{T}\sum_{t=1}^{T} x_t\, u_t \;/\; \frac{1}{T^2}\sum_{t=1}^{T} x_t^2 \qquad (3.41)$$

What is worth mentioning here in (3.41) is the fact that $(\hat{\beta} - \beta)$ converges to zero at rate T instead of the typical $\sqrt{T}$ that we have in the stationary case. Thus, what we say in this case is that the estimator $\hat{\beta}$ is super consistent. Also, the estimator $\hat{\beta}$ remains consistent in the case of a simultaneous equation system (bivariate VAR(p) case) because of the uniqueness of the cointegrating vector if it exists (r =1). Although the cointegrating vector can be superconsistently estimated, Stock (1987) has shown that the limiting distribution is non-normal. For that reason, as in the case of spurious regressions the typical t and F statistics are not applicable. Hence, we understand that The Engle – Granger estimator $\hat{\beta}$ has a non-standard distribution which is depended on the data generating process (DGP). Because of the existence of unit roots we need to use Brownian motion and monte-carlo simulations in order to construct a null hypothesis and a distribution. A typical example is the following one:

For DGP:

$$y_t = \beta x_t + u_t,\ \ u_{xt} \sim N(0,\sigma^2) \qquad (3.42)$$

$$\Delta x_t = \varepsilon_{xt},\ \varepsilon_{xt} \sim N(0,\sigma^2) \qquad (3.43)$$

and the covariance matrix of residuals is $E(u_t \varepsilon_{xt}) = \sigma_{u\varepsilon}$

$$T(\hat{\beta} - \beta) \xrightarrow{L} \frac{\sigma_{ue}}{2}(W_\varepsilon(1)^2 + 1) + \sigma_\varepsilon * h * N(0, \int_0^1 [W_\varepsilon(r)]^2\, dr \,/\, \sigma_\varepsilon^2 \int_0^1 [W_\varepsilon(r)]^2\, dr$$

$$(3.44)$$

Where $h = \sqrt{\sigma_u^2 - \sigma_{u\varepsilon}^2/\sigma_e^2}$ (3.45)

So we have that t-values based on the estimator $\hat{\beta}$ are not normally distributed. Thus, inference in the cointegrating regression is impractical due to the fact that standard inference is not valid. Also, even though the asymptotic bias goes to zero as the sample size increases, the estimator may be substantially biased in smaller samples. Except of the aforementioned drawbacks, the OLS residual based method of Engle – Granger procedure has gained much attention because of its applicability in the case of trending variables, and the fact that the residuals from the static regression are in the case of cointegration integrated of order zero I(0). These residuals are the errors from the long-run equilibrium path of the set of I (1)-variables. Whether this series is stationary (i.e., the variables are cointegrated) can be tested for example with the Dickey-Fuller (DF) test or the augmented Dickey-Fuller (ADF) test. It is worth mentioning that critical values for ADF test are provided by MacKinnon (1991) based on critical surface regressions. Once the null hypothesis (H_0) of a unit root in the error series has been rejected, the second step of the two-step procedure follows. In the second step, an error-correction model (ECM) is specified (Engle-Granger representation theorem). For illustrative purposes, we will present a bivariate case with two cointegrated variables Y_t, X_t which are I(1) (as previously). The general specification of an ECM with two variables are the following ones:

$$\Delta Y_t = \mu_0 + \gamma_1 z_{t-1} + \sum_{i=1}^{K} \psi_{1,i}\, \Delta X_{t-i} + \sum_{i=1}^{L} \psi_{2,i}\, \Delta y_{t-i} + \varepsilon_{1,t} \qquad (3.46)$$

$$\Delta X_t = \xi_0 + \gamma_2 z_{t-1} + \sum_{i=1}^{K} \xi_{1,i}\, \Delta y_{t-i} + \sum_{i=1}^{L} \xi_{2,i}\, \Delta X_{t-i} + \varepsilon_{2,t} \qquad (3.47)$$

The error correction model in the first equation, states that changes in Y_t are explained by their own past values, lagged changes of X_t and the error from the long-run equilibrium in the previous period. The speed of adjustment is determined by the value of the coefficient γ, and should be negative in sign. Otherwise, the system would diverge from its long-run equilibrium path. Furthermore, in the case of two cointegrated I(1)-variables, Granger causality must exist in at least one direction, as can be concluded from these equations and the static regression. This means that at least one variable can help forecast the other. To conclude, by using OLS regression in general is that it can identify only one cointegrating vector even when there are many variables in the system (Dolado et al., 1991). On the other hand, the Johansen method (or system methods) makes it possible to detect all cointegrating relationship in a system of variables.

3.6.2.2 System estimation methods

Because of the fact that single equation methods are based on least squares estimation, we have to deal with two basic obstacles:

- Normalization problems
- *A priori* fixed number of cointegrating equations

In contrast, with system estimation methods the problem of normalization does not appear and also we have to estimate the number of cointegrating equation. In other words, we can estimate more than one CI vectors. From all system estimation methods Box-Tiao (1977), Stock and Watson (1988), we shall discuss Johansen's procedure which is commonly used.

<u>Johansen procedure:</u>

Let us assume a VAR model with Gaussian errors like the following one:

$$Y_t = \Phi D_t + A_1 Y_{t-1} + \ldots + A_k Y_{t-p} + u_t, t = 1, \ldots, T \qquad (3.48)$$

Where Y_t is an n-vector of I(1) variables and D_t contains deterministic terms such as constant, trend seasonal dummies etc. The VAR*(p)* process is stable if

$$Det(I_n - \Pi_1 z - \ldots - \Pi_p z^p) = 0 \qquad (3.49)$$

has all roots outside complex unit circle. Otherwise, some or all of the variables in Y_t are I(1) and they may be cointegrated. We should recall at this point that Y_t is cointegrated if there exists some linear combination of the Y_t variables that is I(0).

Then we proceed with the VECM*(p)* form by transforming the initial VAR*(p)* model

$$\Delta Y_t = \Phi D_t + \Pi Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t \qquad (3.50)$$

Where $\Pi = \Pi_1 + \ldots + \Pi_p - I_n$ and $\Gamma_k = -\sum_{j=k+1}^{p} \Pi_j$ for k = 1, ..., p - 1. The matrix Π is called *long–run impact matrix* and Γ_k are the *short-run impact matrices.* We denote at this point that the VAR*(p)* parameters Π_i may be recovered from the VECM*(p)* parameters Πand Γ_k through the following equations:

$$\Pi_1 = \Gamma_1 + \Pi + I_n, \qquad (3.51)$$

$$\Pi_\kappa = \Gamma_k - \Gamma_{k-1}, k = 2, \ldots, p \qquad (3.52)$$

From the above equation we notice that the variables $\Delta Y_{t, \ldots,} \Delta Y_{t-k+1}$ are all I(0) but the variable Y_{t-1} is I(1). In order for the aforementioned equation to be consistent with I(0) variables, $\Pi_1$should not be of full rank. Hence, ΠY_{t-1} should encompass the cointegrating relations if present. In case the VAR(*p*) process contains unit roots then Πmatrix is singular. If this is the case it has reduced rank, that is $rank(\Pi) = r < n$. We denote the rank of a matrix with the symbol r. We can distinct three cases:

I. $rank(\Pi) = n$, which means that all n linearly independent combinations must be stationary. This could happen only if deviations of y_t around the deterministic components are stationary.
II. $rank(\Pi) = 0$.This implies that $\Pi = 0$ and Y_t is I(1) and that cointegration doesn't exist. In such a case the VECM(p) model could be represented by a VAR(p-1) in first differences such as:

$$\Delta Y_t = \Phi D_t + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1}\Delta Y_{t-p+1} + u_t \qquad (3.53)$$

III. $0 < rank(\Pi) = r < n$. Here we have that Y_t is I(1) with r linearly independent cointegrating vectors and n – r common stochastic trends

Because of the fact that Π matrix has a rank r it can be decomposed as the following product:

$$\Pi = \alpha\beta' \qquad (3.54)$$

Where $\alpha$is an (nxr) matrix and $\beta'$is an (r x n)matrix with rank(α) = rank(β) = r. Then the β'matrix designates the span of the cointegrating space such that $\beta' Y_{t-1}$ are the r cointegrated variables, β'contains the coefficients of the cointegrating vectors and $\alpha$is called loading or adjusted matrix and denotes the speed of adjustment to the long run equilibrium of the error correction terms. So, we can re-write the VECM(p) as:

$$\Delta Y_t = \Phi D_t + \alpha\beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1}\Delta Y_{t-p+1} + u_t. \qquad (3.55)$$

Where $\beta' Y_{t-1} \sim I(0)$ since β′ is the matrix of cointegrating vectors.

At this point, we can notice that the factorization of **Π** matrix as a product of $\Pi = \alpha\beta'$ is not uniquely identified. This is obvious because of the fact that for any r x r nonsingular positive definite matrix **H** we have that

$$\alpha\beta' = \alpha H H^{-1}\beta' = (\alpha H)(\beta H^{-1\prime}) = \alpha^{*}\beta^{*\prime} \qquad (3.56)$$

Thus, we understand that the factorization of **Π** helps us to identify the space spanned by the cointegrating relations. Further restrictions should be taken in order to identify the unique values of $\alpha$and β'.

Briefly, the Johanen's methodology consists of the following steps:

- First we specify a VAR(p) model with the variables of interest and then estimate it.
- Then we construct likelihood ratio tests for the rank of **Π** in order to find the exact number of cointegrating equations,
- We could impose normalization and further restrictions on the cointegrating vectors in our model (based on economic theory),
- Finally, we calculate VECM(p) with maximum likelihood.

According to Johansen (1995) we have to specify the deterministic terms in our VECM (p) beforehand. The deterministic term could have the following form:

$$\Phi D_t = \mu_t = \mu_0 + \mu_1 t$$

The deterministic behavior of the model can be distinct in five cases which are the following:

1. $\boldsymbol{\mu_t = 0}$, which means that we do not have a constant term. Then the restricted VECM is in the form of:

$$\Delta Y_t = \alpha\beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1}\Delta Y_{t-p+1} + u_t, \qquad (3.57)$$

All the variables in Y_t are I(1) without drift and the cointegrating relations $\beta' Y_t$ have mean zero. This case is a theoretical approach and it does not appear into empirical research.

2. $\boldsymbol{\mu_t = \mu_0 = \alpha\rho_0}$, which means that the constant is restricted. Thus, the restricted VECM (p) is:

$$\Delta Y_t = \alpha(\beta' Y_{t-1} + \rho_0) + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t , \qquad (3.58)$$

Again the variables in Y_t are I(1) without drift and the cointegrating relations $\beta' Y_t$ have non-zero mean ρ_0.

3. $\boldsymbol{\mu_t = \mu_0}$, which means that the constant is unrestricted. Then the VECM is:

$$\Delta Y_t = \mu_0 + \alpha\beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t, \qquad (3.59)$$

The variables in Y_t are I(1) with a drifting vector μ_0 and the cointegrating relations $\beta' Yt$ may have a non-zero mean. This case is commonly used in applied research.

4. $\boldsymbol{\mu_t = \mu_0 + \alpha\rho_1 t}$, which means that the linear trend is restricted. Then the restricted form of the VECM is:

$$\Delta Y_t = \mu_0 + \alpha(\beta' Y_{t-1} + \rho_1 t) + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t, \qquad (3.60)$$

The variables in Y_t are I(1) with drift vector μ_0 and the cointegrating relations $\beta' Y_t$ have a linear trend term $\rho_1 t$.

5. The last case is that of a quadratic trend: $\boldsymbol{\mu_t = \mu_0 + \mu_1 t}$, which means that we have an unrestricted constant and trend. Thus, the form of the unrestricted VECM is the following one:

$$\Delta Y_t = \mu_0 + \mu_1 t + \alpha\beta' Y_{t-1} + \Gamma_1 \Delta Y_{t-1} + \ldots + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t, \qquad (3.61)$$

Again the variables in Y_t matrixare I(1) with a linear trend (quadratic trend in levels) and the cointegrating relations $\beta' Y_t$ have a linear trend.

In order to estimate the parameters of our VECM(p) (8) we will use the maximum likelihood estimation technique for the cointegration vectors and the other vector autoregressive parameters which is developed by Johansen (1988), Johansen (1991) and Johansen and Juselius (1990). In this section we will describe briefly the procedure for more details a more thorough presentation is given in Johansen (1995). The Johansen procedure uses canonical correlation analysis and assumes that the errors of the VAR(p) model are Gaussian . This analysis is used due to the fact that our purpose is to reduce the information content of our sample of T observations in the K-dimensional space to a reduced dimensional of r cointegrating vectors. Thus, canonical correlation analysis determines the extent to which the multicollinearity in the data will allow such a smaller r-dimensional space.

First of all, we have to estimate 2 auxiliary regressions with OLS (ordinary least squares) method. The first one is that of ΔY_t (first differences) which is regressed on lagged differences of Y_t. Then we keep the residuals of the above regression and we denote them with R_{0t}. In the second set of auxiliary regressions, Y_{t-p} (the level form) is regressed on the same set of regressors. In this case, the residuals are assigned as R_{1t}. So our regression equation is reduced to the following one:

$$R_{0t} = \alpha\beta' R_{1t} + u_t \qquad (3.62)$$

Then we define the product moment matrix of sums of squares and sums of products of R_{ot} and R_{it} (each one of them is n x n dimensional) :

$$\begin{pmatrix} S_{00} & S_{01} \\ S_{10} & S_{11} \end{pmatrix}$$

The above mentioned matrix could be calculated from the following formula:

$$\hat{S}_{ij} = \frac{1}{T}\sum_{t=1}^{T} R_{it} \acute{R}_{jt} \text{ with } i, j = 0, 1 \qquad (3.63)$$

Johansen (1991) shows that the asymptotic variance of $\beta' R_{it}$ is $\beta' \Sigma_{11}\beta$ and the asymptotic variance of R_{ot} is Σ_{00} and their asymptotic covariance matrix is $\beta' \Sigma_{10}$. In this case Σ_{00}, Σ_{10} and Σ_{11} are denoting the population estimates of S_{00}, S_{10} and S_{11}. Subsequently, we could maximize the likelihood with respect to the loading matrix α by holding β constant and then as a second step maximize $\alpha$with respect to β. Then we will end up with the following notation with the loading matrix:

$$\hat{a}' = (\beta' S_{11}\beta)^{-1}\beta' S_{10} \qquad (3.64)$$

The $\hat{a}'$ is an (r x n) matrix so the maximized likelihood function is given by:

$$L(\beta)^{\frac{-2}{T}} = |S_{00} - S_{01}\beta(\beta' S_{11}\beta)^{-1}\beta' S_{10}| \qquad (3.65)$$

The process of maximization of the above likelihood function with respect to βmeans that we have to minimize of the determinant with respect of β. More clearly we can achieve this by solving the below classical eigenvalue problem:

$$|S_{10}S_{00}^{-1}S_{01} - \lambda S_{11}| = 0 \qquad (3.66)$$

Otherwise, we can modify the above equation and find the eigenvalue of:

$$|S_{11}^{-1}S_{10}S_{00}^{-1}S_{01} - \lambda I| = 0 \qquad (3.67)$$

where I assigns the identity matrix

We notice that the roots of the above equation are the r canonical correlations between R_{1t}.and R_{0t}. In other words, we seek those linear combinations of Y_{t-1} that are highly correlated with linear combinations of ΔY_t. If we assume that λ_i are the canonical correlations given by solving equation the above equation, then $(1 - \lambda_i)$ are the eigenvalues of $(I - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01})$. Because of the fact that the value of the determinant of any matrix is equal to the product of its eigenvalues we can denote the following equation:

$$\prod_{i=1}^{n}(1-\lambda_i) = | I - S_{11}^{-1}S_{10}S_{00}^{-1}S_{01}| = |S_{11} - S_{10}S_{00}^{-1}| / |S_{11}| \qquad (3.68)$$

By using the following determinant identity formula:

$$|C - B'A^{-1}B| = |A - BC^{-1}B'| * |C| / |A| \qquad (3.69)$$

in which we assume that C = S_{00}, A = $\beta' S_{11}\beta$ and B = $\beta' S_{10}$ we could transform (*3.69*) and get:

$$|S_{00} - S_{01}S_{11}^{-1}S_{10}| \,/\, |S_{00}| \qquad (3.70)$$

Then we can maximize the likelihood function which is given by:

$$L_{max}^{-2/T} = |S_{00}| * \prod_{i=1}^{n}(1 - \lambda i), \qquad (3.71)$$

Where λ_i = canonical correlations. In order to determine the number of CI vectors Johansen suggested two likelihood ratio tests. The trace test and maximum eigenvalue test. The trace test is a likelihood-ratio test statistic of the null nested hypothesis that there are at most *r* cointegrating vectors:

$$H_0(r): r = r_0 \text{ vs. } H_1(r_0): r > r_0,$$

and the statistic is given from:

$$\mathcal{LR}_{trace}(r_0) = -T \sum_{i=r+1}^{n}(1 - \hat{\lambda}_i), \qquad (3.72)$$

where $\widehat{\lambda_{r+1}}, \dots, \widehat{\lambda_p}$ are the *n* –*r* smallest eigenvalue of the equation (*3.66*). We can decompose the (*n* x *n*) matrix $S_{11}$by using Cholesky decomposition into a product of a non-singular (*n* x*n*) matrix C in the form of S_{11} = CC′. Then (18) can be transformed into:

$$|\lambda I - C^{-1}S_{10}S_{00}S_{01}C^{-1}| = 0, \qquad (3.73)$$

Johansen (1988) has tabulated critical values for the test statistic in Equation (*3.66*) for various quantiles and up to five cointegration relations; i.e., r =1,… , 5. Because of the existence of unit roots the distribution of L_{tr} statistic is not chi-square it is a function of standard Brownian motions.

Then we proceed with Johansen's Maximum Eigenvalue Statistic with a similar likelihood ratio statistic for the hypotheses:

$$H_0(r_0): r = r_0 \text{ vs. } H_1(r_0): r_0 = r_0 + 1$$

The maximum eigenvalue statistic, is given by:

$$\mathcal{LR}_{max}(r_0) = -T \ln(1 - \widehat{\lambda_{r0+1}}) \qquad (3.74)$$

Similar with the trace statistic, the asymptotic null distribution of $\mathcal{LR}_{max}(r_0)$is not chi-square but instead is a complicated function of Brownian motion, which depends on the dimension n − r_0 and the specification of the deterministic terms. Critical values for this distribution are tabulated in Osterwald-Lenum (1992) for the aforementioned five trend cases discussed previously.

Chapter 4

4.1 Stationarity tests

We begin our presentation in this chapter by performing ADF (Augmented Dickey Fuller test) and PP (Phillips – Perron) tests for stationarity. A prerequisite for Cointegration analysis is that all of our variables under study must be in the same order of integration. Thus, our first step is to proceed with our aforementioned statistical tests in order to find if that holds true. We used also PP test because of the fact that ADF tends to identify unit roots in marginal cases such as a coefficient between 0.95 and 0.99. Furthermore, we have to identify the optimal lag length for each variable in order to implement the statistical testing rigorously. T

The information criterion that we used is the Akaike (AIC). The optimal number of lags is the one that minimizes the value of the Akaike information criterion or put it differently minimize the residual sum of squares. As we mentioned in the previous chapter, both ADF and PP are testing the null hypothesis of non-stationarity (presence of a unit root) against the alternative hypothesis of stationarity. Also, we included an intercept and a deterministic trend parameters in the log-level form of all variables. It is worth mentioning that due to the presence of strong seasonality we seasonally adjust the variables LSTOCKS and LOPEC with the X-13ARIMA-SEATS technique.

Level	No. of lags	ADF (drift)	ADF (drift & trend)	PP (drift)	PP (drift & trend)
LN	1	0.2482	0.2051	0.3863	0.3496
LSHALE	4	0.6423	0.8126	0.6226	0.9651
LSTOCKS	7	0.2690	0.0624	0.6036	0.7836
LGSC	1	0.4910	0.2128	0.6139	0.3253
LTWDI	7	0.9086	0.7880	0.8747	0.6557
REA	4	0.0171	0.0607	0.0788	0.2243
LOPEC	0	0.4661	0.0911	0.3940	0.1713

Figure 4.1: Unit root tests for all variables in log form. P-values below 5% rejects the null hypothesis of the existence of unit root.

1st diff.	No. of lags	ADF (no drift)	ADF (drift)	PP (no drift)	PP (drift)
ΔLN	0	0.0000	0.0000	0.0000	0.0000
ΔLSHALE	8	0.0000	0.0000	0.0000	0.0000
ΔLSTOCKS	2	0.0003	0.0037	0.0000	0.0000
ΔLGSC	0	0.0000	0.0000	0.0000	0.0000
ΔLTWDI	1	0.0000	0.0000	0.0000	0.0000
ΔREA	3	0.0000	0.0000	0.0000	0.0000
ΔLOPEC	0	0.0000	0.0000	0.0000	0.0000

Figure 4.2: Unit root tests for all variables in Δlog form. P-values below 5% rejects the null hypothesis of the existence of unit root.

Obviously, our variables are not stationary in their level form but they are stationary in first differences. Hence, we conclude that they are I(1) stationary.

4.2 Engle – Granger procedure

Because of the fact that our variables are I(1) stationary we should examine if they are cointegrated. More precisely we are testing if there is a long run equilibrium between those variables. The first method that we will use is the Engle – Granger procedure which is also known as two step method. In the first step we have to calculate with OLS a static regression which is consisted of NYM as the dependent variable. The model is as follows:

$$\text{LnNYM} = \beta_0 + \beta_1\text{LnSHALE} + \beta_2\text{LnOPEC} + \beta_3\text{REA} + \beta_4\text{LnSTOCKS} + \beta_5\text{LnGSCI} + \beta_6\text{LnTWDI} + \beta_7\text{TREND}$$

The Ln operator denotes that our variables are in a logarithmic form except REA which is expressed as a percentage and the deterministic Trend variable. After estimating the static regression which represents the long run relation among the variables we notice that there are some signs of spurious regression. The regression is spurious when we regress one random walk with other independent random walks. The fact that is spurious is because the regression will most likely indicate a non-existing relationship. A convenient way to distinct a spurious regression is if the adjusted R^2 is higher than the Durbin-Watson statistic which is the case in our static model. Hence we have to examine if the residual is nonstationary (cannot reject the null hypothesis of the unit root test). If this is not the case and our residuals are stationary then a long run relationship of lower order exists and we have a cointegrating relationship (or an ECM according to Granger's representation theorem).

Variables	Coefficients	Std. Error	Probability Chi sq.
C	18.04 [a]	4.59	
SHALE	-0.23 [a]	0.05	
OPEC	-0.50	0.35	
REA	0.27	0.18	
STOCKS	-0.52	0.54	
GSCI	0.79 [a]	0.08	
TWDI	-2.09 [a]	0.24	
TREND	0.01[a]	0.001	
Breusch-Godfrey			0.0000
Breusch-Pagan			0.0138
Adj R^2	0.9512		
Durbin Watson	0.6862		
Shapiro-Wilk	0.1633		

[a]Indicates significance at all levels (1%, 5% and 10%).

Table 4.1: *Static model*

The statisticaly significant variables at all levels of significance are SHALE, GSCI, TWDI and the TREND. We use the TREND as a dummy due to the fact that there are some independent variables such as SHALE which have a deterministic trend. The signs of the regression are correct and align with the economic theory as the supply side factors such as SHALE and OPEC have a negative sign while the proxy factor for the demand side REA has a positive sign which denotes that if demand for oil increases so will the price. The financial index GSCI which measures the performance of the commodity market increases as the price of oil increases while OECD STOCKS decreases in an effort to cover OECD consumption as the price of oil increases. Finally, the TWDI variable has a correct sign which comlies with economic theory.

From a statistical testing perspective, we have already mentioned the relationship between adjusted R^2 and the DW statistic. Furthermore, Breusch-Godfrey statistical test for autocorrelation in residuals does not accept the null hypothesis of no autocorrelation in residuals a fact that is obvious from the *figure 2*. The persistent autocorrelation parallel with the fact that the Breusch-Pagan statistic also does not accept the null hypothesis of no heteroscedasticity among the residuals leads as to the assumption that additional unit root testing has to be done in the residuals before reaching any conclusion. Hence, we conducted ADF test to examine the null hypothesis which supports the existence of a unit root which means that the residuals of the static model are non-stationary which finally supports the spurious regression hypothesis. The opposite will denote the existence of a cointegrating relationship.

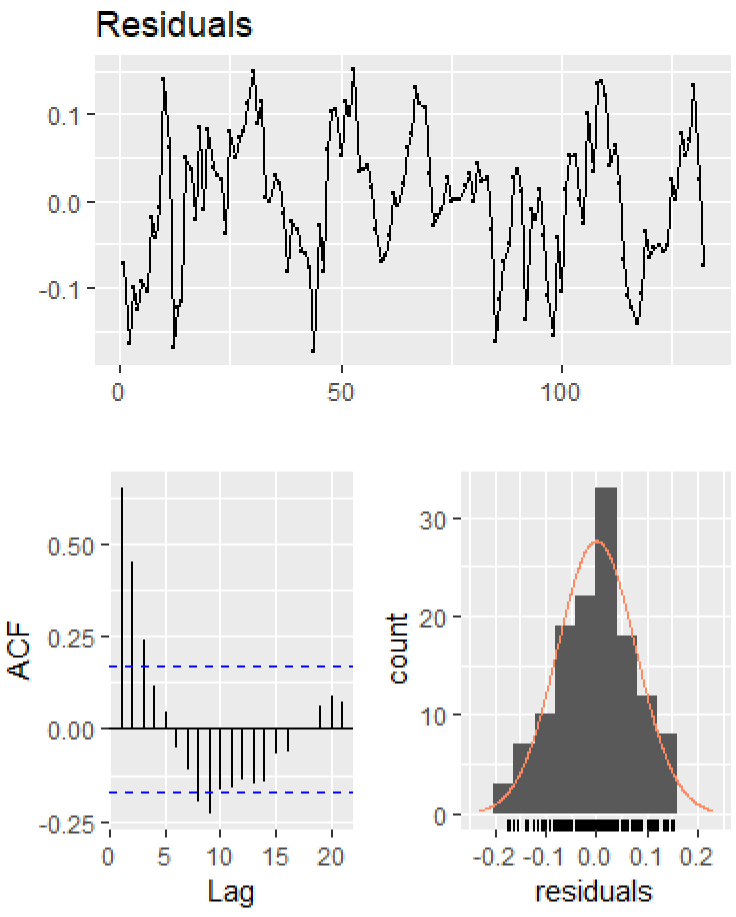


Figure 4.3: *Plot of residuals of the static model.*

The value of ADF test for the residuals without including a constant and trend (because we have already included a constant in the static regression) is -4.6618 which rejects the null hypothesis of unit root (-4.6618<-3.46 for 1% level of significance). Thus, our residuals are stationary and a long run relation exists between our variables. This supposition is in acceptance with the visualization of the residuals behavior in *figure 2*. In this case we proceed with the second step of the Engle-Granger residual based Cointegration method which denotes that we have a unique cointegrating relationship. Moreover, because of the existence of Cointegration the vector OLS estimate $\hat{\beta}$ $(n \times 1)$ is superconsistently which means that it converges in a higher rate T than the usual OLS estimator $T^{1/2}$ in I(0) cases to the true parameter β.

Variables	Coefficients	Std. Error	Probability Chi sq.
C	0.003	0.007	
Δ(SHALE)	0.53[b]	0.25	
Δ(OPEC)	0.45	0.46	
Δ(REA)	0.34	0.20	
Δ(STOCKS)	0.90	0.92	
Δ(GSCI)	0.73 [a]	0.09	
Δ(TWDI)	-1.63 [a]	0.44	
$\Delta(NYM_{t-2})$	0.16 [a]	0.05	
$\Delta(SHALE_{t-1})$	-0.46[c]	0.25	
ECT_{-1}	-0.38 [a]	0.06	
Breusch-Godfrey			0.8084
Breusch-Pagan			0.3436
Adj R^2	0.6610		
Durbin Watson	2.0485		
Shapiro-Wilk	0.9844		

[a]Indicates significance at all levels (1%, 5% and 10%).
[b]Indicates significance at levels 5% and 10%.
[c]Indicates significance at level 10%.

Table 4.2: *Error correction model.*

From *table 4.2* it is evident that our Error Correction Model (ECM) has the following form:

$$\Delta \text{LnNYM} = \beta_0 + \beta 1 \Delta \text{LnSHALE} + \beta 2 \Delta LnOPEC + \beta_3 \Delta \text{REA} + \beta_4 \Delta \text{LnSTOCKS} + \beta_5 \Delta \text{LnGSCI} + \beta_6 \Delta LnTWDI + \beta_7 \Delta \text{LnSHALE}_{t-1} + \beta_8 \Delta \text{LnNYM}_{t-1} + \alpha_1 \text{ECM}_{t-1}$$

The ECM term is a lagged term in order to find the state of error correction of the model. That means that we want to deduce if the coefficient of the error correction mechanism has a negative sign and if it is statistically significant. This means that the model is correcting with a pace that is equal to 0.38% during a month (because we use monthly frequency data) which is the value of the a_1 of the ECM term. From the other independent variables we see that the statistically significant ones are β_5 a positive sign which complies with the theory. That is the case with the TWDI variable which is elastic and with a negative sign and with the lagged term of SHALE and NYM which are inelastic. Those variables have the ability to adjust in the short run while the others don't. The fact that the SHALE variable at time *t* has a positive sign is really odd because as production grows, price tends to fall. An explanation for this might be the fact that it is not an easy case to adjust the scheduled production of crude oil within the same time interval. The fact that the lagged term is negative and significant supports this point of view. The other factors are statistically insignificant in this analysis especially in the case of OPEC and STOCKS. We used lagged variables due to the fact that time series in general are very susceptible to autocorrelation issues.

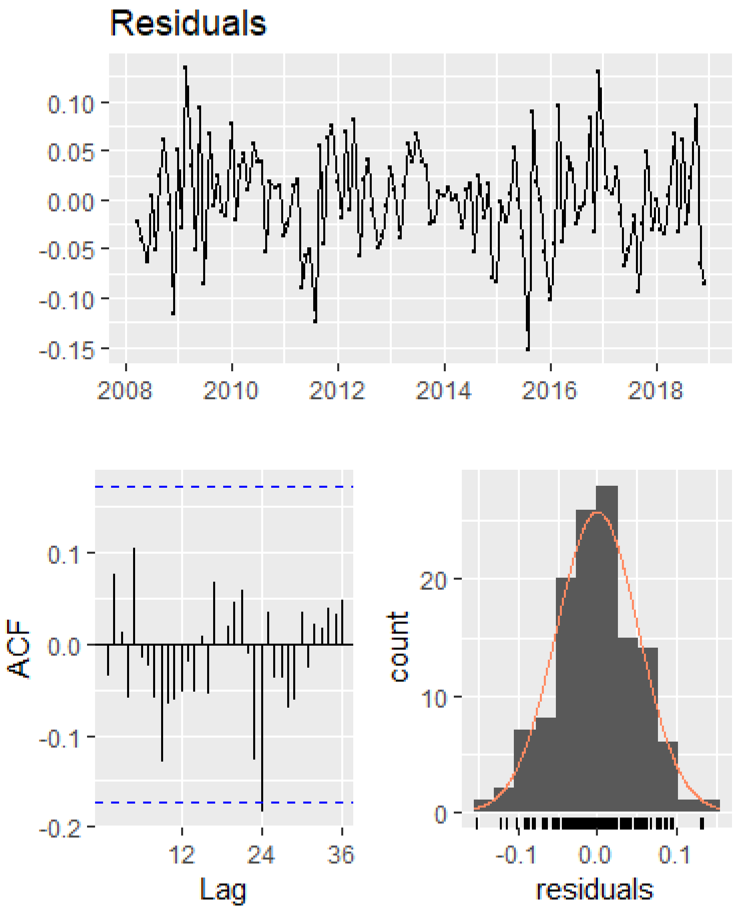


Figure 4.4: *Plot of residuals of ECM.*

The residuals of ECM as its concluded from the statistical testing, are i) homoscedastic due to the fact that Breusch-Pagan statistic supports the null hypothesis of homoscedasticity ii) uncorrelated as it is depicted in *figure 2* and iii) they are normally distributed around a zero mean which is an indispensable factor in hypothesis testing procedure because of the fact that Central Limit Theorem assumes normality. Finally, the adjusted R^2 has an acceptable rate which means that the model has the ability to explain approximately 66% of the variability of our dependent variable NYM while the other percentage remains unexplainable and is represented through the residuals.

As a conclusion, we could derive the fact that there is an existence of a long run relationship between the variables which is expressed through the a_1 coefficient. On the other hand, there are some limitations with Engle Granger procedure in our case. The fact that we used many independent variables made us assume if there are existed more cointegrating relations. This hypothesis cannot be tested with the above method because it is only limited to the estimation of a unique cointegrating relationship. For that reason we will continue our analysis with the Johansen procedure which allows for more cointegrating vectors.

4.3 Johansen procedure

In order to make use of the Johansen procedure properly first we have to estimate a VAR model in levels. Initially, we used the information criteria to find the optimal number of lags to minimize the sum of squared errors. In *table 3* below we notice that the majority of the criteria complying that the number of lags to be used in the model is two.

	1	2	3	4
AIC	-55.630	-56.389	-56.170	-56.052
HQ	-55.048	-55.354	-54.683	-54.112
SC	-54.197	-53.842	-52.508	-51.276
		Number of Lags		
	AIC	HQ	SC	
	2	2	1	

Table 4.3: *Appropriate number of lags for VAR.*

Subsequently, we continue with the estimation of VAR (*2*) with the OLS method. Before continuing with our results, we will conduct a likelihood ratio test that is available in *urca* package in R which is testing the null hypothesis for not including a linear trend in VAR. The test statistic is distributed as χ^2 with (p-r) degrees of freedom and for three cointegrating vector. Thus, 4 degrees of freedom in our case.

	Test statistic	p - value
LR test	9.75	0.04

Table 4.4: *LR test for linear trend in VAR.*

The statistical test cannot accept the null hypothesis for 5% level of significance so we conclude that we have to insert a deterministic trend variable in our model. From the whole system of VAR model we will present the part which is related with our dependent variable LNYM but the residual testing is for the whole VAR system of equations. Also, we have to examine the stability of the model. This could be tested from the roots of characteristic polynomial. As we notice, all roots are inside the unit circle although marginally. The roots are depicted in the table below:

Roots											
0.97	0.97	0.97	0.78	0.78	0.76	0.66	0.66	0.51	0.51	0.47	0.21

Table 4.5: *Roots of the characteristic polynomial.*

Then we proceed with the estimation of VAR (*2*) and present the results in the following table:

.

Variables	Coefficients	Std. Error	t value	p value
C	14.29	4.50	3.18	0.00[b]
Trend	-0.0002	0.002	-0.09	0.93
$LNYM_{t-1}$	0.67	0.11	5.92	0.00[a]
$LSHALE_{t-1}$	-0.22	0.39	-0.55	0.57
$LSTOCKS_{t-1}$	-1.75	1.36	-1.28	0.20

$LGSCI_{t-1}$	0.59	0.10	5.57	0.00[a]
$LTWDI_{t-1}$	-0.63	0.58	-1.09	0.27
REA_{t-1}	0.58	0.27	2.10	0.04[c]
$LOPEC_{t-1}$	0.03	0.62	0.05	0.96
$LNYM_{t-2}$	-0.17	0.11	1.53	0.12
$LSHALE_{t-2}$	0.19	0.38	0.51	0.61
$LSTOCKS_{t-2}$	0.38	1.41	0.27	0.78
$LGSCI_{t-2}$	-0.69	0.10	-6.84	0.00[a]
$LTWDI_{t-2}$	-0.20	0.59	0.34	0.73
REA_{t-2}	-0.48	0.29	-1.65	0.10
$LOPEC_{t-2}$	0.03	0.61	0.06	0.95
Adj. R^2	0.96			
PT.asympt. (mult.)				0.54
ARCH LM (mult.)				0.51
Normality (mult.)				0.34

Table 4.6: *VAR model for LNYM as dependent*

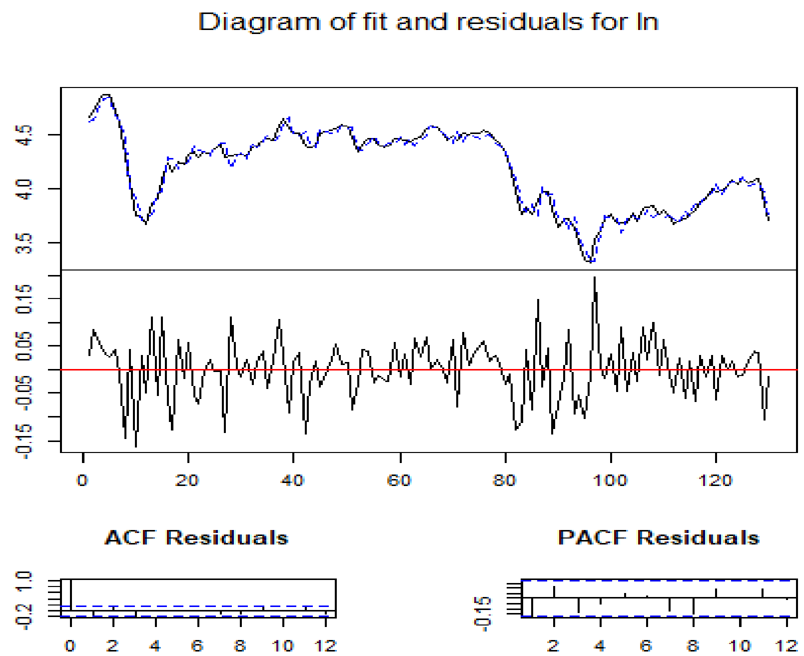


Figure 4.5: *Plot of residuals of VAR (2) as a dependent variable.*

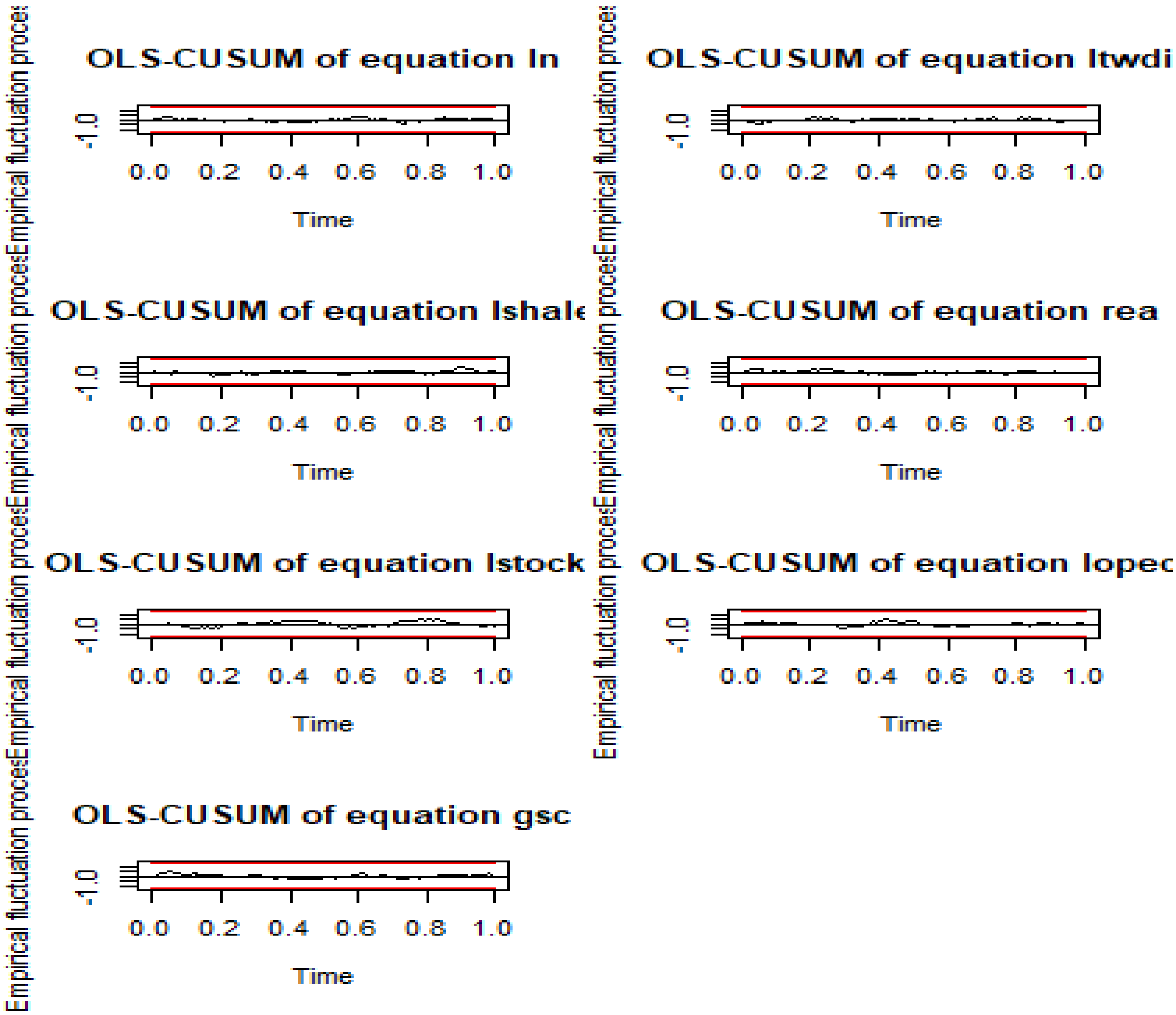


Figure 4.6: *CUSUM test for the stability of VAR(2).*

As we denote from the results, the statistically significant parameters are the lagged variable of LN_{t-1}, $LGSCI_{t-1}$ and REA_{t-1} and the signs comply with economic theory. On the other hand, the independent variables that are *t-2* lags before have absurd signs but are statistically insignificant except $LGSCI_{t-2}$ parameter which has a negative sign. The CUSUM tests confirms the stability of the model and the statistical tests for autocorrelation, heteroscedasticity and normality are in compliance with the Gaussian white noise residuals. Because of the fact that our variables are *I(1)* stationary and the marginal case of the roots of characteristic polynomial we will continue with VECM*(1)* (we lose one lag due to the conversion from VAR in levels to VECM).

Initially, we have to estimate the number of cointegrating relations in the model. This could be possible by sequential likelihood ratio testing from Johansen (1988) by using the trace test and the maximal eigenvalue test. The asymptotic null distribution of those likelihood ratio tests are not chi-square but instead is a multivariate version of the Dickey-Fuller unit root distribution which depends on the dimension (parameters p – cointegrating relations r) and the specification of the deterministic terms. The table below shows the results for both unrestricted constant and restricted trend cases (as mentioned in chapter 3) for both tests.

Rank	Test Statistic	10%	5%	1%
r ≤ 5	10.07	16.85	18.96	23.65
r ≤ 4	20.88	23.11	25.54	30.34
r ≤ 3	25.06	29.12	31.46	36.65
r ≤ 2	34.73	34.75	37.52	42.36
r ≤ 1	36.86	40.91	43.97	49.51
r = 0	75.18	46.32	49.42[b]	54.71[a]

Table 4.7: Cointegration rank: Maximal Eigenvalue Statistic (lambda max), with linear trend in cointegration

Rank	Test Statistic	10%	5%	1%
r ≤ 5	18.59	22.76	25.32	30.45
r ≤ 4	39.48	39.06	42.44	48.45
r ≤ 3	64.53	59.14	62.99	70.05
r ≤ 2	99.26	83.20	87.31	96.58[a]
r ≤ 1	136.12	110.42	114.90	124.75[a]
r = 0	211.31	141.01	146.76	158.49[a]

Table 4.8: Cointegration rank: trace statistic, with linear trend in cointegration

Rank	Test Statistic	10%	5%	1%
r ≤ 5	9.47	13.75	15.67	20.20
r ≤ 4	15.62	19.77	22.00	26.81
r ≤ 3	22.94	25.56	28.14	33.24
r ≤ 2	34.85	31.66	34.40[b]	39.79
r ≤ 1	66.24	37.45	40.30	46.82[a]
r = 0	73.73	43.25	46.45	51.91[a]

Table 4.9: Cointegration rank: maximal eigenvalue statistic (lambda max), without linear trend and with constant in cointegration

Rank	Test Statistic	10%	5%	1%
r ≤ 5	14.09	17.85	19.96	24.60
r ≤ 4	29.71	32.00	34.91	41.07
r ≤ 3	52.64	49.65	53.12	60.16
r ≤ 2	87.49	71.86	76.07	84.45[a]
r ≤ 1	153.73	97.18	102.14	111.01[a]
r = 0	227.47	126.58	131.70	143.09[a]

Table 4.10: Cointegration rank: trace statistic, without linear trend and with constant in cointegration

From the above statistical tests we deduce that the appropriate number of cointegrating relations are 3 except in the case of maximal eigenvalue with linear trend which gave 1 cointegrating relation. After estimating a VECM(*1*) for both one and three cointegrating relations for restricted trend and unrestricted constant we find the most robust one to be trace statistic with trend in cointegrating relationship. The fact that we have three $\beta' y_t$ relations is reinforced from the graphical representation of them.

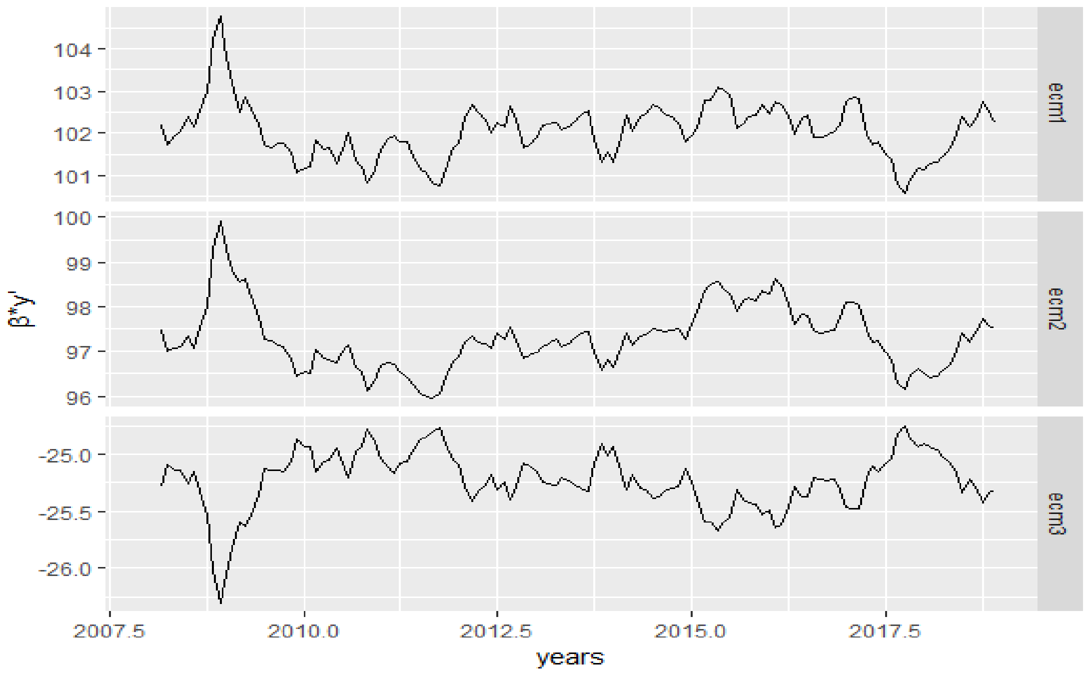


Figure 4.7: *All three cointegrating relations.*

It is worth mentioning that the deterministic trend variable in cointegrating equations is due to the 2008 – 09 global recession. Furthermore, all the above depicted relations are stationary hence, we have a reduced rank regression of matrix Π (as we mentioned in chapter 3) which has a rank of 3. Below we present the error correction parameter for the whole VECM (*1*) system for all variables. All the statistically significant variables for 5% level of significance are correcting in the long run relations. Those parameters are the loading matrix α which is giving the rate of error correction towards long run equilibrium. We obtain the vector α from the factorization of Π matrix into $\alpha\beta'$ vectors. The alpha vector as we aforementioned is the rate of correction while beta vector denotes the space spanned of the long run stationary cointegrating relations.

Variables	ECM_1	ECM_2	ECM_3
LNYM	-0.26[a]	-0.09[b]	-0.29
LSHALE	0.01	-0.05[a]	-0.19[a]
LSTOCKS	0.006[c]	0.004	0.005
LGSCI	-0.20[a]	-0.02	0.06
LTWDI	0.01	-0.01	-0.04
REA	-0.04[a]	-0.04[a]	-0.23[a]
LOPEC	-0.01	0.004	0.03

Table 4.11: Trace with trend model

Finally, for beta matrix we use the Phillips triangular representation and identify the relationships by using two zero restrictions (because we have three cointagrating relations r-1) in each relation and normalize the parameters LN, LSHALE and LSTOCKS of ECM1, ECM2 and ECM3. Below we present the beta matrix:

	ECM1	ECM2	ECM3
LN	1	0	0
LSHALE	0	1	0
LSTOCKS	0	0	1
LGSCI	-0.44	2.49	-0.62
LTWDI	-0.25	2.42	-5.11
REA	0.70	-11.6	3.38
LOPEC	1.94	-3.29	0.46
TREND	-0.001	-0.04	0.004

Table 4.12: *VAR model for LNYM as dependent.*

Consequently, after estimating with maximum likelihood the beta matrixes and identify them, we continue by estimating alpha. It is worth mentioning that if we identify *beta* differently we could end up with different estimate of *alpha*. On the contrary, the other estimates are exactly the same and could be estimated through OLS. Below we present the partial VECM (*1*) with the variable of interest LNYM as the dependent variable.

Variables	Coefficients	Std. Error	t value	p value
ECM_1	-0.26	0.05	-4.85	0.00[a]
ECM_2	-0.09	0.05	-1.96	0.05[b]
ECM_3	-0.29	0.21	-1.44	0.15
C	10.61	2.91	3.65	0.00[a]
$\Delta LNYM_{t-1}$	-0.12	0.11	-1.16	0.25
$\Delta LSHALE_{t-1}$	0.07	0.37	0.21	0.84
$\Delta LSTOCKS_{t-1}$	-1.01	1.37	-0.73	0.46
$\Delta LGSCI_{t-1}$	0.62	0.09	6.56	0.00[a]
$\Delta LTWDI_{t-1}$	-0.01	0.58	-0.02	0.98
ΔREA_{t-1}	0.55	0.29	1.87	0.06[b]
$\Delta LOPEC_{t-1}$	0.10	0.61	0.17	0.86
Adj. R^2	0.48			
PT.asympt. (mult.)				0.82
ARCH LM (mult.)				0.57
Normality (mult.)				0.57

Table 4.13: *VECM (1) of LNYM..*

Again the statistical testing in residuals for the whole system (multivariate testing) are in accordance with uncorrelated, homoscedasticity and normality principles. The ECM parameters are correcting towards long run equilibrium because of negative sign and the first two are statistically significant. From all the other variables only intercept, $\Delta LGSCI_{t-1}$ and ΔREA_{t-1} are statistically significant. Those are the short run adjustments and both are inelastic and with positive sign which complies with theory.

<u>Restrictions on alpha matrix – weak exogeneity tests</u>

In this section we will continue our analysis by imposing some restrictions to identify the *alpha* matrix. Because of the fact that we are considering all variables as endogenous, we need to conduct sequential likelihood ratio tests to examine if some of them are (weakly) exogenous. The hypothesis that we will test is if zero restrictions on *alpha* holds. The restrictions are contained in the ($p \times m$) A matrix such that according to Johansen (1982) $H_4: \alpha = A\psi$, where the elements of matrix ψ contain new unrestricted loadings. In order to test the validity of $H_4: \alpha = A\psi$ hypothesis given the H_1 (r) which is our estimated model

$$-2\ln (Q;H_4|H_1(r)) = T \sum_{i=1}^{r} ln \frac{1- \lambda_{4.i}}{1- \lambda_{\iota}}$$

Which is asymptotically distributed as χ^2 with r(K –m) degrees of freedom. We applied the aforementioned test to all variables and the A matrix was formed as follows:

$$A_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The variables that we tested restrictions upon are with the following order:

LNYM, LSHALE, LSTOCKS, LGSCI, LTWDI, REA, LOPEC, where *m* are the free parameters that are excluded from the exogeneity test and *p* are the number of parameters that we have. In our case the A is (*7 x 6*). The following table presents the results from the likelihood ratio test. The null hypothesis is that the variable is zero thus exogenous against the alternative which is non-zero (endogenous).

Variable	Test stat.	p value	λ_1	λ_1	λ_1	λ_1
$H_{4.1}$ | H_1 (r = 3)	30.23	0.00	0.3375	0.2451	0.1842	0.1540
$H_{4.2}$ | H_1 (r = 3)	30.23	0.00	0.3375	0.2451	0.1842	0.1540
$H_{4.3}$ | H_1 (r = 3)	15.8	0.00	0.3958	0.2346	0.2105	0.1752
$H_{4.4}$ | H_1 (r = 3)	5.17	0.16	0.4391	0.2348	0.2161	0.1584
$H_{4.5}$ | H_1 (r = 3)	7.25	0.06	0.4098	0.2448	0.2329	0.1493
$H_{4.6}$ | H_1 (r = 3)	15.61	0.00	0.4123	0.2468	0.1764	0.1593
$H_{4.7}$ | H_1 (r = 3)	5.14	0.16	0.4376	0.2392	0.2139	0.1740

Table 4.14*: Weak exogeneity test.*

From the above results we conclude that the parameters LGSCI, LTWDI and LOPEC are weakly exogenous. The interpretation of such a result could be that in the causality analysis it is evident that the aforementioned variables cannot cause LNYM because the *alpha loading matrix* does not error correcting towards long run equilibrium. While the other independent variables do cause LNYM.

<u>Restrictions on beta matrix</u>

In the same rationale as in the previous case of *alpha* we will continue with the identification procedure by testing linear restriction on the cointegrating long run vector vector *beta*. Again we will proceed with likelihood ratio tests be testing the null H3 : $\beta = H_3\Phi$ with H_3 ($p \times s$), Φ($s \times r$), and $r \leq s \leq p$, where p are the number of variables, s are the free parameters (those that we are not tested with zero restrictions each time) and r the number of cointegrating relations (in our case 3). Again we will construct the H_3 matrix by imposing zero restrictions sequentially for every variable. An example of H_3 matrix (in our case is an *8x7* matrix because we are also including deterministic trend) for zero restriction in the second variable could be the following:

$$H_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

and is based on the likelihood ratio test statistic:

$$-2\ln(Q;H_3|H_1(r)) = T\sum_{i=1}^{r} ln \quad \frac{1- \quad \lambda_{3.i}}{1- \quad \lambda_\iota}$$

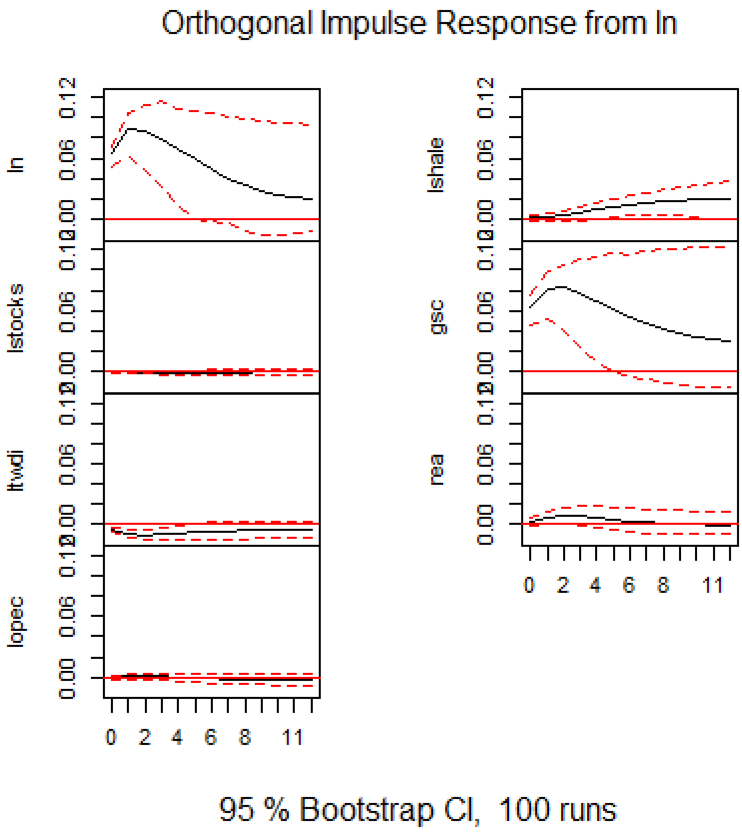
Again $H_1(r)$ denotes our unrestricted model. In the following matrix we present the results for each variable.

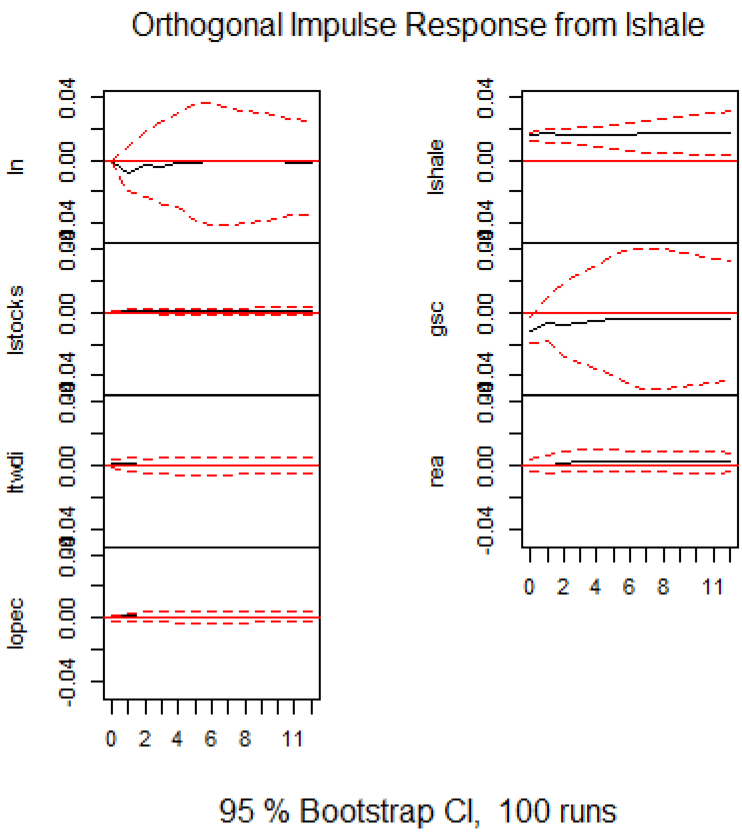
Variable	Test stat.	p value	λ_1	λ_1	λ_1	λ_1
$H_{3.2} \mid H_1 (r = 3)$	10.31	0.02	0.4285	0.2461	0.1876	0.1560
$H_{3.3} \mid H_1 (r = 3)$	2.52	0.47	0.4383	0.2469	0.2206	0.1664
$H_{3.4} \mid H_1 (r = 3)$	4.36	0.23	0.4238	0.2469	0.2295	0.1493
$H_{3.5} \mid H_1 (r = 3)$	20.09	0.00	0.3813	0.2353	0.2024	0.1496
$H_{3.6} \mid H_1 (r = 3)$	13.86	0.00	0.4210	0.2457	0.1764	0.1611
$H_{3.7} \mid H_1 (r = 3)$	3.2	0.36	0.4304	0.2410	0.2335	0.1749
$H_{3.8} \mid H_1 (r = 3)$	7.66	0.05	0.4279	0.2451	0.2059	0.1515

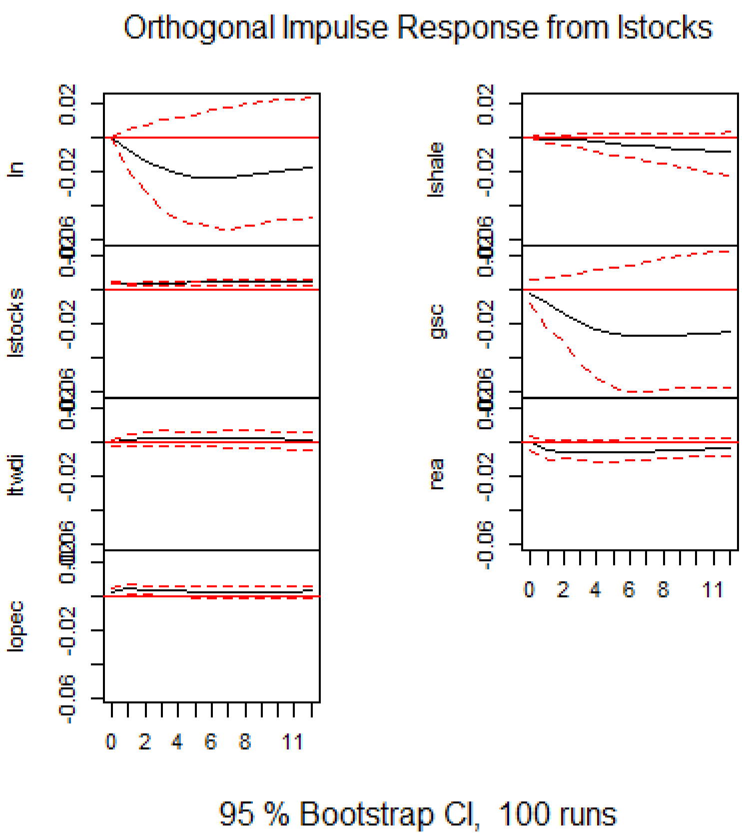
Table 4.15: *Testing restrictions on beta matrix.*

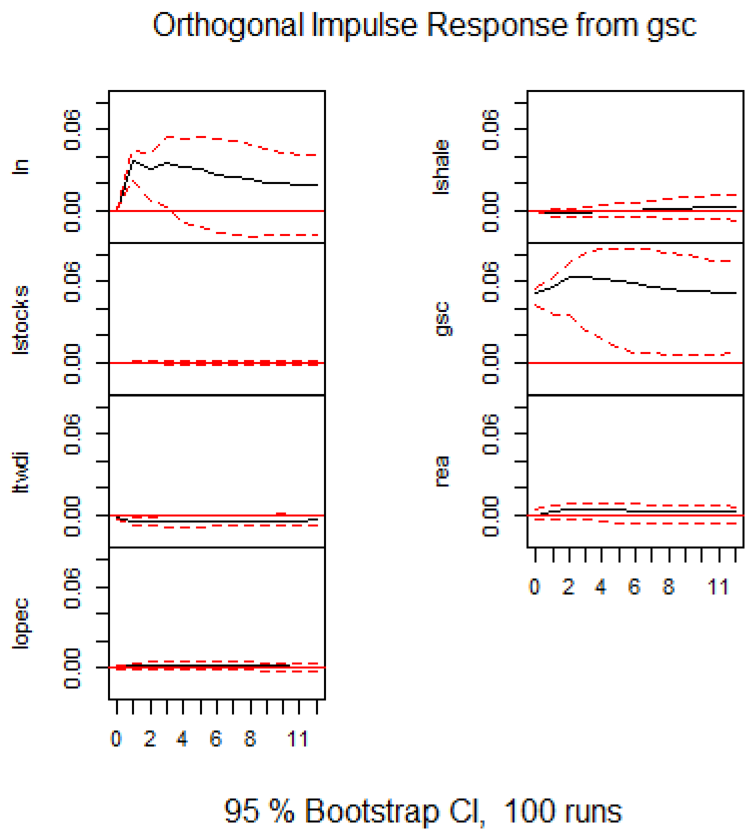
From the above results we conclude that the variables LSTOCKS, LGSCI and LOPEC do not seem to fit in the cointegrating relation because we cannot reject the null hypothesis that they are zero. It is worth mentioning that these tests do not depend on the normalization of the Cointegration relations (Pfaff 2008). The absence of a compact theoretical background (like the ones tested for power purchasing parity from Johansen and Juselius (1992)) for testing restrictions leads us to test a simple hypothesis of zero restrictions sequentially for all the long run parameters in order to eliminate those that could not affect LNYM.

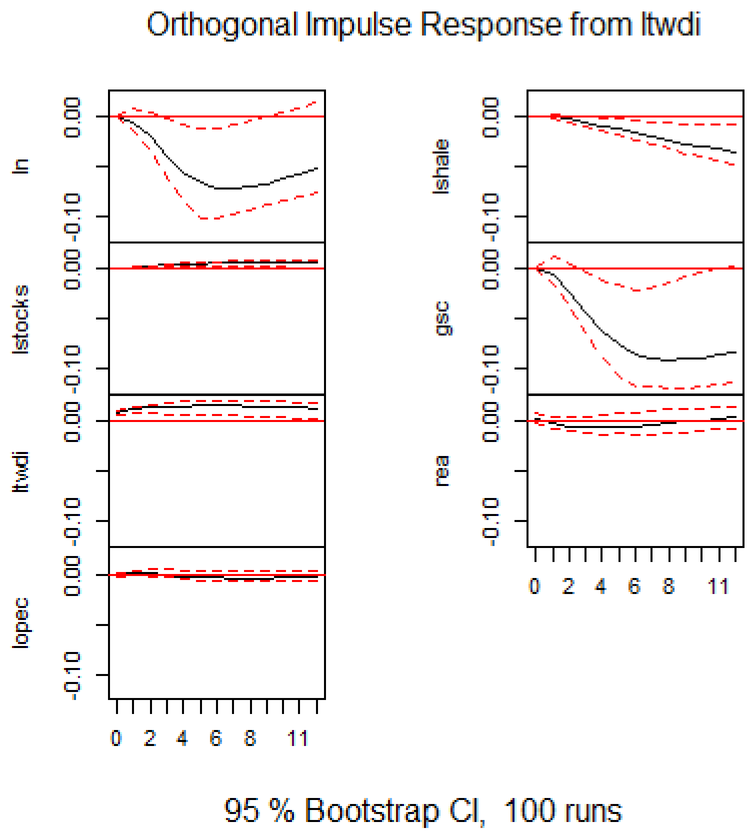
Finally, we transform our VECM (*1*) to VAR(*2*) in order to examine the impulse response functions and the forecast variance decomposition for each variable as we mentioned in Chapter 3. More precisely we want to examine how the system behaves to an exogenous shock in innovations for 95% confidence interval. Since all variables in a VAR model depend on each other, individual coefficient estimates only provide limited information on the reaction of the system to a shock. In order to get a better picture of the model's dynamic behavior, impulse responses (IR) are used. The departure point of every impulse response function for a linear VAR model is its moving average (MA) representation or Wold representation theorem (see Chapter 3). In our analysis in order to identify the shocks of a VAR model we used orthogonal impulse response .The basic idea is to decompose the variance-covariance matrix so that $\Sigma=PP^{-1}$, where P is a lower triangular matrix with positive diagonal elements, which is often obtained by a Choleski decomposition. From this matrix it can be seen that a shock to one variable has a contemporaneous effect on others, but not vice versa. It is worth noting that the output of the Choleski decomposition is a lower triangular matrix so that the variable in the first row will never be sensitive to a contemporaneous shock of any other variable in the system and the last variable will be sensitive to shocks of all other variables. For that reason the outcome of Orthogonal Impulse Responses might be sensitive to the order of the variables. In our case we present the order that we used during the previous restrictions on *alpha* and *beta* matrices. Next, we present the OIRF's for all variables:

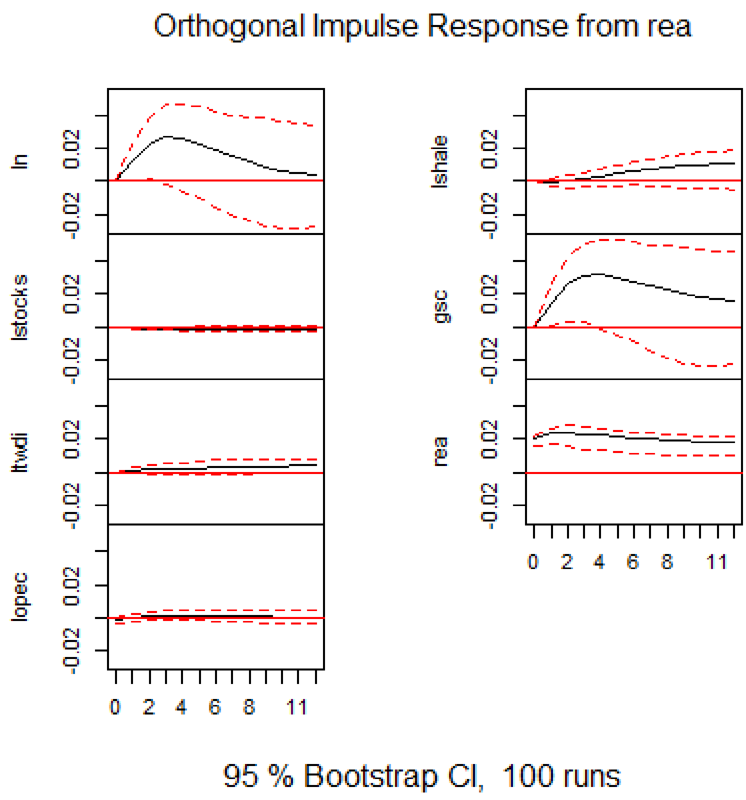



Orthogonal Impulse Response from lshale
ln
lstocks
ltwdi
lopec
lshale
gsc
rea
-0.04
0.00
0.04
0 2 4 6 8 11
95 % Bootstrap CI, 100 runs


Orthogonal Impulse Response from lstocks
ln
lstocks
ltwdi
lopec
lshale
gsc
rea
-0.06
-0.02
0.02
0 2 4 6 8 11
95 % Bootstrap CI, 100 runs


Orthogonal Impulse Response from gsc
ln
lstocks
ltwdi
lopec
lshale
gsc
rea
0.00
0.06
0 2 4 6 8 11
95 % Bootstrap CI, 100 runs


Orthogonal Impulse Response from ltwdi
ln
lstocks
ltwdi
lopec
lshale
gsc
rea
0.00
-0.10
0 2 4 6 8 11
95 % Bootstrap CI, 100 runs



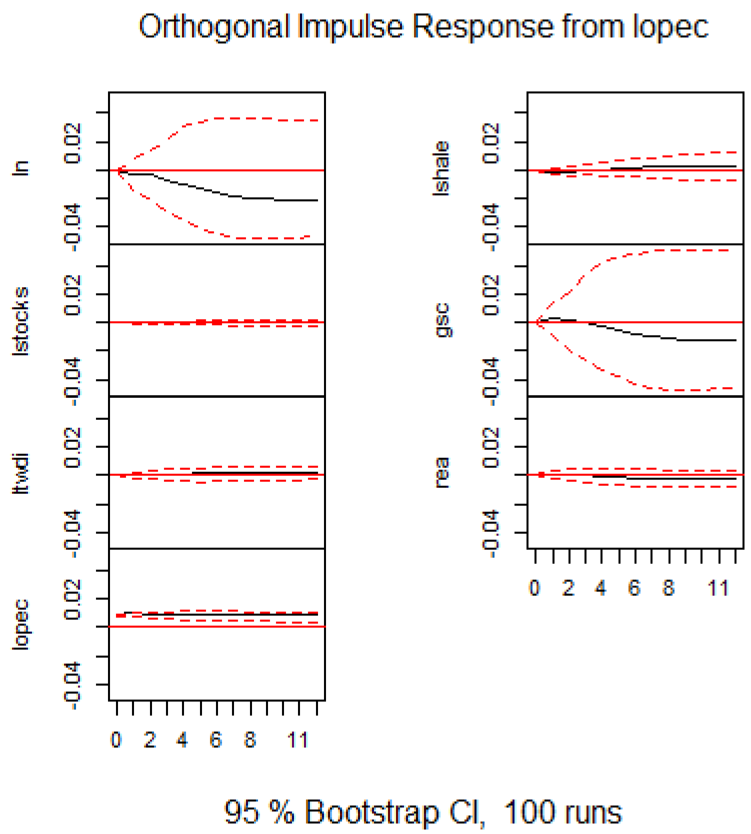


Figure 4.8: *IRF's for all variables.*

It is clear that a shock of all independent variables affect our dependent variable LNYM during the one year horizon that we tested. The results are reasonable due the fact that supply shocks such as LOPEC and LSHALE have a negative effect in LNYM while REA and LGSCI have a positive effect. The variable LSTOCKS also are in accordance with the literature due to the negative effect that it has to LNYM. What is really significant is a standard deviation shock from LTWDI variable which seems to affect the price of crude oil in a more noteworthy way. Furthermore, it is worth mentioning that LGSCI is also affected with the same impact as in the case of LNYM. On the other hand, the other variables do not seem to be affected each other.

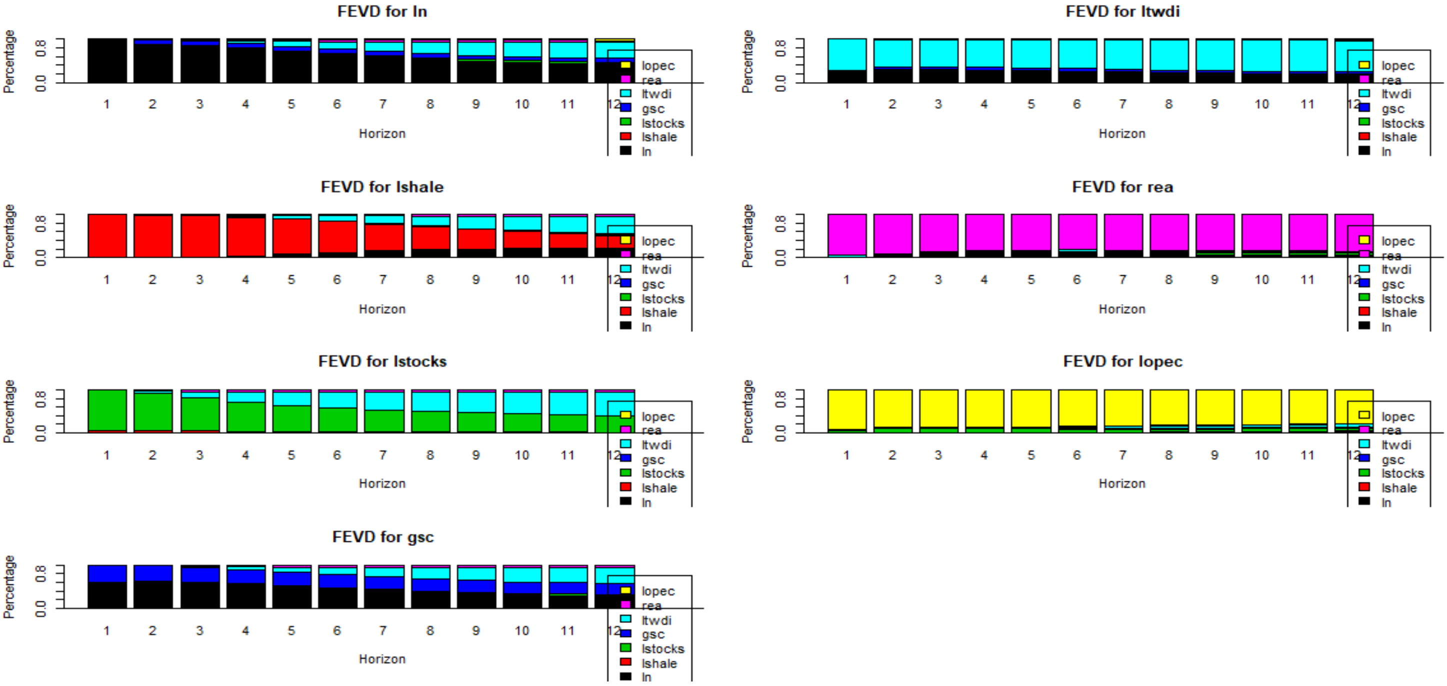


Figure 4.9: *FVD's for all variables.*

A forecast error variance decomposition as we mentioned in chapter 3, is a way to quantify how important each shock is in explaining the variation in each of the variables in the system. Hence, it is equal to the fraction of the forecast error variance of each variable due to each shock at each horizon. For our variable of interest LNYM the results are in compliance with our VECM(*1*) due to the fact that for a one year horizon the variable LTWDI plays a fundamental role for determining the price of LNYM while LGSCI although significant in the first months it decays thereafter. On the other hand, all the other variables appears to exert less influence in the settlement of the final crude oil price. The variables LTWDI,REA and LOPEC behave in a way that they explain they are own variability while LSHALE, LSTOCKS and LGSCI seems to be affected from LTWDI and LNYM. Those results are pretty much the same as in the case of IRF'S.

CHAPTER 5

Conclusions

In the present thesis we conducted a Cointegration analysis on crude oil price determinants for the period 2008 until 2018 in monthly frequency. The goal was to apply two different techniques based on single equation estimating method and system estimating method. In the first case we estimated a static model and then proceed with the ECM as denoted in Engle and Granger (1987) while in the second case we calculated a system of seven variables in a VAR form and then examine if Cointegration among them exists. The results showed that there were three cointegrating relations. We further continued our survey based on identification restriction in cointegrating vectors (*beta* matrix) and the loading vector (*alpha* matrix) in order to determine if there existed any exogenous variable. We find that Goldman Sachs Commodity Index, OPEC and Trade Weighted US dollar Index are exogenous to the system. Hence, they are not correcting to a long run equilibrium making them purely exogenous variables. Also, we find that the role of OPEC doesn't seem to affect NYMEX crude oil price and that could be attributed to the fact that WTI is more like a regional benchmark (due to the fact that supply is land-locked and relatively expensive to ship to certain parts of the globe) rather than BRENT which is a global benchmark (approximately two-thirds of all crude contracts around the world reference Brent Blend). On the other hand, shale oil plays an important role in shaping NYMEX price in the long run. This result comes as no surprise due to the fact that shale oil production is used as a hedging tool for US contrary to production cuts or oversupplies of other producing countries. In the short run only real economic activity (a proxy used for world demand) parallel with Goldman Sachs Commodity Index seems to affect crude oil price. Their coefficients appears to be inelastic and statistically significant. Inelasticity denotes that the coefficients where below unity in responding to a percentage change of the independent variable. The positive signs where in compliance with economic theory as both independent variables are positively correlated with oil price. Finally, we represented our findings graphically by examining the impulse responses which enlighten us about how the system reacts to an exogenous standard deviation shock of one variable and variance decompositions which depicted the variability that could be explained during one year horizon from each variable. Our analysis was implemented using R programming language and with the usage of packages *urca*, *vars*, *dynlm*, *lmtest* and *ggplot2*.

Bibliography

Alquist, Ron, and Lutz Kilian. 2007. "*What Do We Learn from the Price of Crude Oil Futures?*" CEPRWorking Paper No. 6548.

Alquist, Ron, Lutz Kilian, and Robert J. Vigfusson. 2012. "*Forecasting the Price of Oil*." Forthcoming: G. Elliott and A. Timmermann, eds., Handbook of Economic Forecasting 2. Amsterdam: North-Holland.

Black, F. (1986). "*Noise*". Journal of Finance, (41), pp. 529-543.

Box, G.E.P. and G.C. Tiao (1977), "*A Canonical Analysis of Multiple Time Series*," Biometrika, 64, 355-365.

Brown, S.P.A., and K.R. Phillips (1986), "*Exchange Rates and World Oil Prices*," Federal Reserve Bank of Dallas Economic Review, March, 1-10.

Büyükşahin, Bahattin, Michael S. Haigh, Jeffrey H. Harris, James A. Overdahl, and Michel A. Robe.2009. "*Fundamentals, Trader Activity, and Derivative Pricing*." Working paper. Commodity Futures Trading Commission.

Fattouh, B. (2007), "*The drivers of oil prices: The usefulness and limitations of non-structural models, supply-demand frameworks, and informal approaches".* EIB Papers, 12(1), 128-156.

Fattouh, Bassam. (2010a). "Oil Market Dynamics through the Lens of the 2002-2009 Price Cycle". OIES Working Paper 39, Oxford Institute for Energy Studies.

Fattouh, B.A., Kilian, L., & Mahadeva, L. (2013). *"The Role of Speculation in Oil Markets: What Have We Learned So Far?".*

Friedman, M. (1953). "The case for flexible exchange rates: Essays in positive economics", Chicago University Press, Chicago, USA.

Hamilton, J.D. (2008), "*Understanding Crude Oil Prices"*, National Bureau of Economic Research.

Hamilton, J.D. (1994), "*Time Series Analysis ",* Princeton University Press.

He, Yanan & Wang, Shouyang & Lai, Kin Keung. (2010). "*Global economic activity and crude oil prices: A cointegration analysis*." Energy Economics. 32. 868-876. 10.1016/j.eneco.2009.12.005.

Hotelling, H. (1931), "*The economics of exhaustible resources. The Journal of Political Economy"*, 39(2), 137-175.

Johansen, S. (1988). "*Statistical Analysis of Cointegration Vectors*," *Journal of Economic Dynamics and Control*, 12, 231-254.

Johansen, S. [1991], "*Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models*", Econometrica 59(6), 1551–1580.

Johansen, S. and Juselius, K. [1990], "*Maximum likelihood estimation and inference on cointegration — with applications to the demand for money"*, Oxford Bulletin of Economics and Statistics 52(2), 169–210

Johansen, S. (1995). "*Likelihood Based Inference in Cointegrated Vector Error Correction Models"*, Oxford University Press, Oxford.

Kilian, L., (2008a). "*Exogenous oil supply shocks: how big are they and how much do they matter for the U.S. economy?"* Review of Economics and Statistics 90 (2), 216–240.

Kilian, L., (2008b). *Not all oil price shocks are alike: Disentangling demand and supply shock in the crude oil market*. Forthcoming in American Economic Review.

Kilian, L., X, Zhou., (2018). "*Oil Prices, Exchange Rates and Interest Rates."* CESifo Working Paper No. 7484.

Litzenberger, Robert H & Rabinowitz, Nir, (1995). "*Backwardation in Oil Futures Markets: Theory and Empirical Evidence*," Journal of Finance, American Finance Association, vol. 50(5), pages 1517-1545, December.

Lütkepohl, H., (2006).*" New Introduction to Multiple Time Series Analysis"*.

Maddala, G. S., Kim, I. M., (1998). *"Unit Roots, Cointegration, and Structural Change ".*Cambridge University Press.

Osterwald-Lenum, M. (1992). "*A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Statistics*," Oxford Bulletin of Economics and Statistics, 54, 461-472.

Sadorsky, P., (2000).” *The empirical relationship between energy futures prices and exchange rates.”* Energy Economics 22, 253–266.

Perifanis, T. and A. Dagoumas (2017). “*An econometric model for the oil dependence of the Russian Economy*.” International Journal of Energy Economics and Policy 7(4): 7-13.

Perifanis, T. and A. Dagoumas (2019*). “Living in an Era when Market Fundamentals Determine Crude Oil Price.”* The Energy Journal 40 (01), 317-336.

Stock, J.H. and M.W. Watson (1988), "*Testing for Common Trends*," Journal of the American Statistical Association, 83, 1097-1107.

Trehan, B. (1986), “*Oil Prices, Exchange Rates and the Economy: An Empirical Investigation*,” Federal Reserve Bank of San Francisco Economic Review, 4, Fall, 1-40.

Wu G, Zhang YJ. “*Does China factor matter? An econometric analysis of international crude oil prices*. “Energy Policy. 2014;72 (9):78–86.

Zhang, Y., Fan, Y., Tsai, H., Wei, Y., (2008). “*The empirical relationship between energy futures prices and exchange rates”*. Energy Economics 22, 253–266.

Zivot, E., Wang, J., (2006). *“Modeling Financial Time Series with S-PLUS® - 2006”*

Appendix

R code

```
################ time series ##############################
nymex.ts=ts(nymex,start=c(2008,1),end=c(2018,12),frequency = 12)
opec.ts= ts(opec,start=c(2008,1),end=c(2018,12),frequency = 12)
shale.ts= ts(shale,start=c(2008,1),end=c(2018,12),frequency = 12)
stocks.ts= ts(stocks,start=c(2008,1),end=c(2018,12),frequency = 12)
twdi.ts= ts(twdi,start=c(2008,1),end=c(2018,12),frequency = 12)
rea.ts=ts(killian,start=c(2008,1),end=c(2018,12),frequency = 12)
gsci.ts=ts(gsci,start=c(2008,1),end=c(2018,12),frequency = 12)
################ seasonality ##################################
library(seasonal)

sstocks=seas(stocks.ts)
stockssa=final(sstocks) # seasonally adjusted values
sopec=seas(opec.ts)
opecsa=final(sopec) # seasonally adjusted values
################ logs ########################################
lnymex=log(nymex.ts)
lopec=log(opecsa)
lshale=log(shale.ts)
lstocks=log(stockssa)
ltwdi=log(twdi.ts)
ln=log(ln.ts)
lgsci=log(gsci.ts)
################## Engle Granger ##############################
library(vars)
library(urca)

ur.ln=ur.df(ln,type = "trend",selectlags = "AIC")
summary(ur.ln)
```

```
# create trend variable

trend=seq_along(lwti)
length(lwti)
# step1 long run model
engle_granger=lm(ln ~ lshale +  lopec + lgsc + lwcons + lstocks + ltwdi  )
summary(engle_granger)
print(AIC(engle_granger))
#NeweyWest(engle_granger, lag = NULL, order.by = NULL, prewhite = TRUE, adjust = FALSE, diagnostics = FALSE,
#          sandwich = TRUE, ar.method = "ols",  verbose = FALSE)
summary(ur.df(lnymex,type = "none",selectlags = "AIC"))
summary(ur.df(lwcons,type = "trend",selectlags = "AIC"))

dwtest(engle_granger)
bptest(engle_granger)
bgtest(engle_granger,order = 1, order.by = NULL,
    type = c("Chisq"))
ggAcf(resid, lag.max = 12, type = "correlation", plot = TRUE)
ggAcf(resid, lag.max = 12, type = "partial", plot = TRUE)

#residuals
resid=engle_granger$residuals
urresid= ur.df(resid,type = "none",selectlags = "AIC")
summary(urresid) # I(O)
resid.ts=ts(resid,start = c(2008,1), end=c(2018,12),frequency = 12)

# step2 error correction model
library(dynlm)

# with lags
```

```
dtwdi=diff(ltwdi)
dstocks=diff(lstocks)
dshale=diff(lshale)
dopec=diff(lopec)
dwcons=diff(lwcons)
drea=diff(reasa)
dn=diff(ln)
dgsci=diff(lgsci)
mdl1= dynlm(dn ~ dstocks  + dshale  + dtwdi + dwcons   + dgsci + dopec
        + lag(dn,k=-2)   + lag(resid.ts,k=-1))
summary(mdl1)
print(AIC(mdl1))

library(lmtest)
library(tseries)
library(portes)

LiMcLeod(mdl1$residuals,lags=seq(1,12,1),order=0, squared.residuals = T) #,SquaredQ=FALSE
LjungBox(mdl1$residuals,lags=seq(1,12,1),order=0)
bgtest(mdl1, order = 3, order.by = NULL,
    type = c("Chisq"))
bptest(mdl1,studentize = T)
ggAcf(mdl1$residuals, lag.max = 12, type = "correlation", plot = TRUE)
ggAcf(mdl1$residuals, lag.max = 12, type = "partial", plot = TRUE)
dwtest(mdl1)
hist(mdl1$residuals)
shapiro.test(mdl1$residuals)
####################### VAR ################################
library(vars)
library(ggfortify)
library(urca)
```

```
nv=data.frame(ln,lshale,lstocks,lgsc,ltwdi,rea,lopec)
dnv=data.frame(diff(ln),diff(lshale),diff(lstocks),diff(lgsc),diff(ltwdi),diff(rea),diff(lopec))
#descriptive stats
library(e1071)

skewness(diff(ln))
sd(diff(lshale))
var(diff(lstocks))
kurtosis(diff(ln))

# correlation matrix

library("PerformanceAnalytics")
chart.Correlation(nv, histogram=TRUE, pch=19)

### UNIT ROOT TEST ###

urln=ur.df(diff(lshale),type = "drift",selectlags = "AIC")
summary(urln)
ppln=ur.pp(ln,type = "Z-alpha",model = "constant")
summary(ppln)

library(ggplot2)
# level form
lf=data.frame(ln.ts,shale.ts,stocks.ts,gsc)
objectlf=ts(lf,start=c(2008,1),end=c(2018,12),frequency = 12)
# graph
autoplot(objectlf, facets=TRUE) + xlab("years") + ylab("level form")
+ ggtitle("variables")  + xlim("date")

#2
lf2=data.frame(twdi.ts,opec.ts,rea.ts)
```

```
objectlf2=ts(lf2,start=c(2008,1),end=c(2018,12),frequency = 12)
# graph
autoplot(objectlf2, facets=TRUE) + xlab("years") + ylab("level form")
+ ggtitle("variables")  + xlim("date")

object=ts(dnv,start=c(2008,1),end=c(2018,12),frequency = 12)
# graph
autoplot(object, facets=TRUE) + xlab("years") + ylab("dlog")
+ ggtitle("variables")  + xlim("date")

# create date object
date=seq(as.Date("2008-1-1"), as.Date("2018-12-1"), by = "months")

# 2nd group
nv222=data.frame(diff(ltwdi),diff(rea),diff(lopec))
object2=ts(nv222,start=c(2008,1),end=c(2018,12),frequency = 12)
# graph
autoplot(object2, facets=TRUE) + xlab("years") + ylab("dlog")
+ ggtitle("variables")  + xlim("date")
############################### VAR
var=VAR(nv, p = 2,type =c("const"),season = NULL, exogen = NULL)
summary(var)
plot(var)
print(AIC(var))
serial.test(var, lags.pt = 16, type = "PT.asymptotic")
normality.test(var)
arch.test(var,lags.multi = 2,multivariate.only = T)
roots(var)
###############################  VECM ###################################################
# trend in ci

cajo_eigen=ca.jo(nv, type = c("eigen"), ecdet = c("const"),  K = 2,
```

```
        spec=c("transitory"), season = NULL, dumvar = NULL)

cajo_trace=ca.jo(nv, type = c("trace"), ecdet = c("const"),  K = 2,
        spec=c("transitory"), season = NULL, dumvar = NULL)
# constant in ci

cajo_eigen2=ca.jo(nv, type = c("eigen"), ecdet = c("trend"), K = 2,
         spec=c("transitory"), season = NULL, dumvar =NULL)

cajo_trace2=ca.jo(nv, type = c("trace"), ecdet = c("trend"), K = 2,
         spec=c("transitory"), season = NULL, dumvar =NULL)

# with const
summary(cajo_eigen)#3
summary(cajo_trace)#3

# trend
summary(cajo_eigen2)#1
summary(cajo_trace2)#3

# models with trend
cajorls_eigen2=cajorls(cajo_eigen2,r=1,reg.number = NULL) ######## const 3
summary(cajorls_eigen2$rlm)#1

cajorls_trace2=cajorls(cajo_trace2,r=3,reg.number = NULL) ######## const 3
summary(cajorls_trace2$rlm)#3
cajorls_trace2$beta
# models with constant
cajorls_eigen=cajorls(cajo_eigen,r=3,reg.number = NULL) ######## const 3
summary(cajorls_eigen$rlm)#1

cajorls_trace=cajorls(cajo_trace,r=3,reg.number = NULL) ######## const 3
```

```
summary(cajorls_trace$rlm)#3#####
# comparison
print(AIC(cajorls_eigen$rlm))
print(AIC(cajorls_eigen2$rlm))
print(AIC(cajo_trace))
print(AIC(cajorls_trace2$rlm))
######## FINAL ##################################
####################################################################
cajorls_eigen2=cajorls(cajo_eigen2,r=3,reg.number=NULL) ######## trend 3

summary(cajorls_trace2$rlm)
############### restrictions on A (1)
############### restrictions on A (2)
### trace2 with trend
B1= matrix(c( 1, 0, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 0, 0, 1, 0,
       0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0, 1), nrow = 8,ncol = 7,byrow = TRUE)
blr=blrtest(cajo_trace2,B1,r=3)
summary(blr)
library(urca)
### eigen2 with trend
A1= matrix(c( 1, 0, 0, 0, 0, 0,
       0, 1, 0, 0, 0, 0,
       0, 0, 1, 0, 0, 0,
       0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 1, 0,
```

```
    0, 0, 0, 0, 0, 1), nrow = 7,ncol = 6,byrow = TRUE)



alr=alrtest(cajo_trace2,A1,r=3)
summary(alr)
################################## testing residuals
#trace
vvt=vec2var(cajo_trace ,r = 3)    #const
vvt2=vec2var(cajo_trace2 ,r = 2)  #trend

#eigen
vve=vec2var(cajo_eigen ,r = 1)    #trend
vve2=vec2var(cajo_eigen2 ,r = 1)  #constant

library(portes)
#  ARCH EFFECTS
# trace
LiMcLeod(vvt2$resid,lags=seq(1,12,1),order=0,squared.residuals = T)
LjungBox(vve$resid,lags=seq(1,12,1),order=0,squared.residuals = T)

# AUTOCORRELATION

LiMcLeod(vve$resid,lags=seq(1,12,1),order=0) #,SquaredQ=FALSE
LjungBox(vvt2$resid,lags=seq(1,12,1),order=0)

plot(vvt2$resid)
# trace
############## FINAL ###################
############# with trend
######################
normality.test(vvt2,multivariate.only=T)
arch.test(vvt2,lags.multi=2,multivariate.only=T)
```

```
serial.test(vvt2,lags.bg =12,type = "PT.asymptotic")

###############################
############## with constant
normality.test(vvt2,multivariate.only=T)
arch.test(vvt2,lags.multi=2,multivariate.only=T)
serial.test(vvt2,lags.bg =12,type = 'PT.asymptotic')

#eigen
############## with trend
normality.test(vve,multivariate.only=T)
arch.test(vve,lags.multi=5,multivariate.only=T)
serial.test(vve,lags.bg =12,type = "PT.asymptotic")
############## with constant
normality.test(vve2,multivariate.only=T)
arch.test(vve2,lags.multi=12,multivariate.only=T)
serial.test(vve2,lags.bg =12,type = 'PT.asymptotic')
# include trend null=not include

lttest(johansen1_trace, r=2)
lttest(johansen1_eigen, r=1) #do not include trend

print(AIC(vvt))
print(AIC(vve))
#########################################################################
############## VARIANCE DECOMPOSITION
library(vars)
library(tsDyn)
vardec=fevd(vvt2, n.ahead = 12)
plot(vardec, col=c("red","blue","green","grey","black","yellow","purple"))
```

```
############## IMPULSE RESONSE
impvar=irf(vvt2, impulse = NULL, response = NULL, n.ahead = 12,
        ortho = T, cumulative = FALSE, boot = TRUE, ci = 0.95,
        runs = 100)
plot(impvar)
# trace
vvt22=vec2var(cajorls_trace2$rlm,r=3)
############## with trend
normality.test(vve2,multivariate.only=T)
arch.test(vve2,lags.multi=5,multivariate.only=T)
serial.test(vve2,lags.bg =12,type = "PT.asymptotic")
############## with const
normality.test(vve2,multivariate.only=T)
arch.test(vve2,lags.multi=5,multivariate.only=T)
serial.test(vve2,lags.bg =12,type = "PT.asymptotic")
############## VARIANCE DECOMPOSITION
plot(vardec)
vardec=fevd(var, n.ahead = 12)
plot(vardec, col = c("red", "blue", "yellow", "green","purple","black"))
############## IMPULSE RESONSE
impvar=irf(vve2, impulse = NULL, response = NULL, n.ahead = 12,
        ortho = T, cumulative = FALSE, boot = TRUE, ci = 0.95,
        runs = 100)
plot(impvar)
############ STABILITY VAR
stvar=stability(var, type = c("OLS-CUSUM"),
            h = 0.15, dynamic = FALSE, rescale = TRUE,nc=2)
plot(stvar)
#######################################################
```

Το έργο που εκπονήθηκε και παρουσιάζεται στην υποβαλλόμενη διπλωματική εργασία είναι αποκλειστικά ατομικό δικό μου. Όποιες πληροφορίες και υλικό που περιέχονται έχουν αντληθεί από άλλες πηγές, έχουν καταλλήλως αναφερθεί στην παρούσα διπλωματική εργασία. Επιπλέον τελώ εν γνώσει ότι σε περίπτωση διαπίστωσης ότι δεν συντρέχουν όσα βεβαιώνονται από μέρους μου, μου αφαιρείται ανά πάσα στιγμή αμέσως ο τίτλος.

Ημερομηνία: …25…/…10.…/ 2019…

O – H Δηλ.
Τασιόπουλος
Κωνσταντίνος