MODELING OIL PRICES IN NAMIBIAN MARKET USING QUANTITATIVE TECHNIQUES

W. NANGOLO*, R.B. GNITCHOGNA[1], F. GIDEON[2], E. LUNGU[3]

May 31, 2017

ABSTRACT

This paper contributes to the debate on oil prices modelling by proposing a quantitative analysis of the dynamics of nominal oil prices in Namibia. The study carries out a post assessment and validation of Random Walk and Mean Reversion processes in modelling and forecasting futures dynamics of nominal oil prices . In this regard, the study used data from the Namibian oil market from April 1989 to April 2014. In essence, despite the recent trends of computational and VAR models for nominal oil prices, modelling and forecasting efforts have circumscribed random walks or mean reversion models as viable options. Popular baseline and benchmark models include Random Walk, Random Walk with drift, AR, MA, ARMA models. The study investigates the suitability of the models in capturing and predicting the data behaviours. Analysis are conducted on chosen sub-period as well as on the overall data. From our results, a major highlight is that there is enough evidence to conclude that oil nominal prices in Namibia are indeed stationary and therefore suitable for analysis. The results also indicated an increasing stochastic trend that confirms that oil prices behave randomly at times. Summary analysis pointing to the fact that oil prices are mean reverting with occurrences of random walk.

KEY WORDS: Oil Prices, Log Returns, Random Walk, Mean Reversion, Augmented Dickey-Fuller test,Variance Ratio Test, Multiple Variance Ratio Test, Ranks-Sign Test, Q-Statistic Test and Johnsen Cointegration Test.

A.M.S. SUBJECT CLASSIFICATION: 60G50, 62J10, 20F11, 15B35

* *Department of Mathematics, University of Namibia, Private Bag 13301 Widhoek, Namibia, e-mail: wnangolo@unam.na. Telephone: +264-61-2063949, Fax: +264612063791*

[1] *Department of Mathematics, University of Namibia, Private Bag 13301 Widhoek, Namibia, e-mail: rgnitchoga@unam.na. Telephone: +264-61-2063962, Fax: +264612063791*

[2] *Department of Mathematics, University of Namibia, Private Bag 13301 Widhoek, Namibia, e-mail: fgideon@unam.na. Telephone: +264-61-2063793, Fax: +264612063791*

[3] *Department of Mathematics, Botswana International University of Science and Technology, P o Box 35, Palapye, Botwana, e-mail:lunguliz@gmail.com, Telephone: +267-71-2216767, Fax:+267-72976077*

1 INTRODUCTION

Many studies have been carried out to test the financial market's efficiency ever since the article of Black and Scholes [2]. Apart from stocks and equities, oil prices are key components of the financial market sector and as such, attract a lot of researches from economists, statisticians, mathematicians, particularly in developed economies. However, the emerging markets in sub-Saharan Africa and in Namibia thereof have received little attention. The oil market is relevant not only to academics, but to consumers and policy makers as well, since a clear understanding of the functioning of the market will translate into better decision making in terms of trade policy. An understanding of oil prices behaviour is important in formulating policies aimed at attaining macroeconomic stability in the country, as oil price fluctuations certainly offset economic targets .

With the increasing globalisation, nations are exposed to growing international community. Trading in both goods and services are affected to a larger extent by the fluctuating of oil prices. For instance, the high increases in oil prices will lead to high prices on other commodities and financial market. On the other hand the decrease in oil prices, has apparently fewer effects on lowering other commodities prices.

For many years, random walk and mean reversion processes have been topical; the theory has challenged academics and finance practitioners and such, has occupied an important place in finance, see [25]. This research is aimed at contributing to the literature by extending the analysis to a small regional oil market (Namibia oil market).

Oil prices have been increasing at an unprecedented pace over the past years and the current oil market show a dramatic decrease in prices. To explain this phenomenon, one needs to identify the tensions on oil prices and their implications to other fossil fuels. This study examine quantitative properties of oil prices in the recent past years, descriptive statistics and some measure of dispersion are presented. We look at the mathematical background underpinning both random walk and mean reversion processes. Independence properties, Markov properties, reflection principle (see definition **1.1.1**), and many others are discussed in this paper. Using the data from the Ministry of Mines and Energy (MME) in Namibia, a procedure including a battery of tests statistics is deigned to investigate whether the change in oil prices in the Namibian oil market follows a random walk or a mean reversion model.

Additional subsidiary research questions are addressed. Is there a trend that governs the movement of oil prices? If there is a trend, it is a deterministic or stochastic trend? Can we predict oil prices accurately?

1.1 Mathematical Background of the models

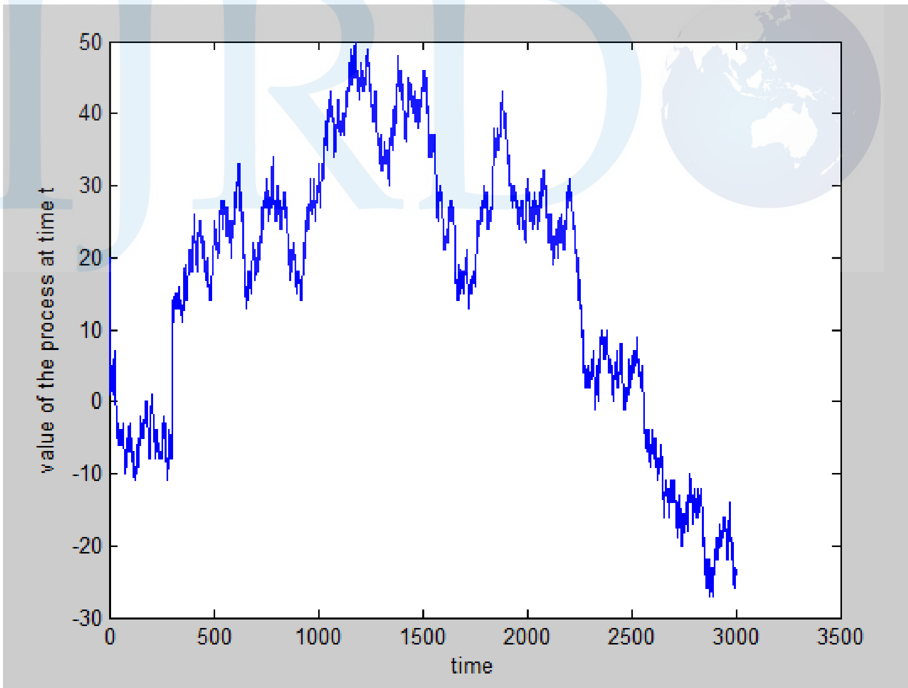
Continuous time Markov processes such as oil prices have been the subject of much attention in recently years. The evolution oil prices is a topic of major concern in Africa and all over the world. We briefly present here random walk and mean reversion models used in modelling time series processes.

1.1.1 *Markov Property and the reflection principle*

A process is Markovian, with respect to a filtration $\mathbb{F}_t$, if for any fixed time t, the future of the process is independent of $\mathbb{F}_t$ given X_t. A stochastic process has the Markov property if the future of the process is conditionally independent of the past given the present statistic. The reflection principle state that Brownian motion reflected at time T is still a Brownian motion, see [20].

1.1.2 *Random walk*

One of the simplest and yet most important models in oil prices change is the random walk model. The term random walk was first introduced in mathematics by Karl Pearson in 1905. Pearson [21] conceptualised a random walk as a mathematical formalisation of a path that consists of a succession of random steps away from previous positions. The steps are independently and identically distributed (iid) in the sense that there are more occurrences of the randomness of their nature than consequences of preceding steps. For example, the price of a fluctuating stock, the financial status of a gambler, or the drunk man finding his way home. **Figure 1.1** below gives an illustrative random walk process.



A random walk process which is regarded by Fuma [9] as a Weak Form Market Efficiency (WFME), assumes that only past prices data are considered when evaluating future oil prices. This rules out any manner of future prices prediction based on anything other than past oil prices data and it assumes that each successive changes have zero correlation. According to this perspective, a look back at historical prices is worthless.

There are two different types of random walks which are of interest for the vast majority of applications. A simple random walk process and a random walk process with drift. Pearson [21] briefly define these two models as follows: a simple random walk is defined as a process where the current value of a variable is composed of the past value plus an error term defined as a white noise.

While a random walk model is said to have ”drift” if the distribution of steps has a non-zero mean. This process shows both a deterministic trend called the "drift" and a stochastic trend called the "white noise".

1.1.3 *Mean reversion*

"What goes up must come down"- This simple wisdom of law of gravity turns out to be a highly non-trivial fact about prices. This was interpreted as evidence of mean reversion, that is a force that drives prices back to a certain medium level after they went above or below it. In other words, mean reversion is a concept of a process that returns to its mean or average.

Mean reversion or Schwartz [26] one-factor model, allows for capturing stochastic behaviour of oil prices. This is based on the assumption that logarithms of oil prices revert to their long term mean. Some researchers refer to mean reversion as long-run due to supply-demand equilibrium. This means that prices tend to revert to its medium trend, (the trend is a product of factors which include but not limited to, marginal cost, production cost, inventory capacity and competitions). Basically the model has two components: deterministic terms such as, level of mean reversion and speed of mean reversion and a stochastic random which is known as distribution properties such as, white noise and volatility which set the prices back to its trend. Figure 1.2 illustrates graphically a mean reversion.

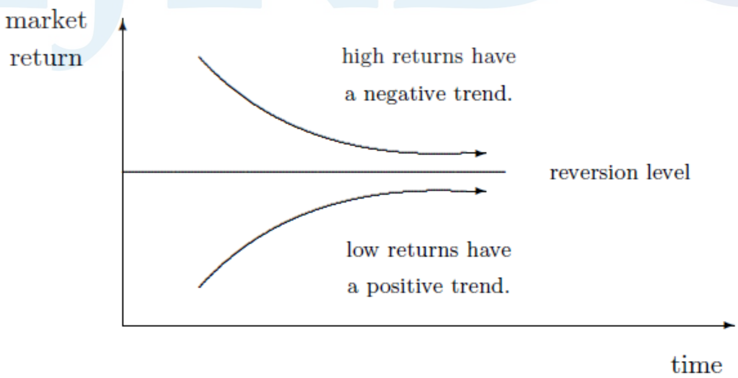


Figure 1.2. The concept of mean reversion [1].

1.2 Models Comparison

Modeling and forecasting efforts have circumscribed random walks or mean-reversion processes as viable options. Let’s look at some possible expectations, **Table 1.1**, gives this summary.

Table 1.1 Comparison of Random walk and Mean Reversion

Random walk	Mean reversion
Changes in price are independent and identical	Prices rise based on previous Prices
Deterministic trend	Volatility determines the price levels
Change in price has no history	Price decision rules are explicit and intuitive
No tractable solution	Lead to tractable solution for the price

This Paper will be organised as follows: The current section provides an introduction to the paper, and gives the basic definitions. Section 2 gives a literature review related to the topic. The third section presents of the modelling techniques that were used . Section 4 and 5 visualises the data and presents the results respectively. Finally the conclusion and remarks and future directions are given.

2 LITERATURE REVIEW

There is nothing strange about massive attention to oil prices, since it is so deeply integrated in the world's economy. Oil is a national strategic resource but also used by roughly each individual household. No surprise that oil prices were never safely under control, and factors behind price changes are still not revealed with absolute confidence. Moreover, with oil significance coming onto new levels, the role it plays in the financial markets and its potential drivers appear to complicate the understanding of their behaviours even more. There are different ways to investigate oil price behaviours. First, it is necessary to understand basic fundamentals of the global petroleum industry and market, and that the oil price is on the background of macro- and microeconomic structures activity.

The first significant attempt to develop a structural model on oil as a scarce resource belongs to Hotelling [11]. Indirectly Hotelling rule is tested by Leinert [14]. His paper finds that crude oil price adjusts or falls to unexpected news about oil field discoveries. One of the attempts to test the Hotelling model belongs to Lin [15], where she used annual data for oil prices and consumption to calibrate the model for different competition settings. She found that prior to 1973 data, the best fitting data model was the one assuming competitive market while afterwards, she indicated that prices in recent years were strongly influenced by OPEC. However the performance of the simple Hotelling model was considered to be poor.

Botha [4] argued that, considerable evidence has demonstrated that in highly competitive and organised markets price changes will be close to random walk. The hypothesis is that in an efficient market characterised by well-informed profit maximising participants, and competing actively with one another, one can not predict future price changes on the basis of the history of prices behaviour only. Successive changes in the variable are done independently from a probability distribution with mean zero. The new price is therefore a realisation of a variable independent of that, that produced the preceding price. He used the idea of Random walk with drift, given by the equation below :

$$Y_t = Y_{t-1} + \mu_t + c, \tag{1}$$

where: Y_t is the spot price log return, Y_{t-1} is the past price log return, μ_t is the white noise and c is the constant called the "drift". This means that on the average the process will tend to move in the direction given by the sign of the constant $'c'$. Botha used data that consist of daily prices on a time period running from April 1968 to September 1975. With the help of empirical results from autocorrelation and running test, he concluded that a random walk model offers a satisfactory explanation of the movement of daily price changes.

Tvedt [28] in her PhD dissertation looks at two alternative ways, Ornstein-Uhlenbeck process and a geometric mean reversion process to model the stochastic nature of the time charter equivalent spot rate in the market for very large crude carries. Earlier, Bjerksund and Steinar [1] had showed that the freight rates follow an Ornstein-Uhlenbeck process. Tvedt followed up this approach, by suggesting a geometric mean reversion process as an alternative to Ornstein-Uhlenbeck process. She argued that Ornstein-Uhlenbeck process does not give a very realistic description of the spot freight rates, because it is not downward restricted, while in reality prices cannot take negative values for example. It is useful to try a process that is downward restricted in order to describe both the spot rate and the time charter equivalent rate. Such a process is given by:

$$dY_t = \alpha(Y^* - \ln Y_t)Y_t dt + \sigma Y_t d\mu_t, \tag{2}$$

where Y_t is the spot price log return, Y^* is the level of mean reversion, α is the speed of mean reversion, σ is the volatility and μ_t is the white noise. A smaller p-value would reject the null hypothesis of random walk. Using the above process, Tvedt found out that the estimated rate of mean reversion α is 0.003289, the volatility σ is 0.1007 and the mean reversion level is 10.58. This statistic rejects the idea of random walk and it supports the existence of mean reversion process in very large crude carries.

Schwartz [26] compared performance of the three models: namely, Mean Reversion, Geometric Brownian Motion (GBM) with convenience yield and the stochastic interest rate. His investigations show evidence of mean reversion in crude oil prices, when he finds that all parameters are statistically significant. He also indicated that oil prices can be modelled as a GBM, but with convenience yield included in the drift.

Blanco and Saronow [3] made theirs, the folksy wisdom that turned out to be a non-trivial factor about stock market, " What goes up must come down". They modified assumptions of a random walk process, to test if energy prices are mean reverting:

$$Y_{t+1} - Y_t = \alpha(Y^* - Y_t) + \sigma\mu_t \tag{3}$$

where Y^* is the level mean reversion, α is the speed of mean reversion, σ is the volatility and μ is the random shock (white noise) to the log return price Y_t. The authors partitioned the periods under consideration into smaller periods of one hour over a single year. The authors, claim that the speed at which prices revert to their long run level may depend on several factors such as nature, magnitude and the direction of the price shock. Nevertheless mean reversion in energy prices is well supported by the empirical studies of Blanco and Saronow in 2001.

Rodchenk* [24] extends Pindyck study by applying a shifting trend model. The shifting trend model has an autoregressive process in error terms rather than Pindyck white noise process. The advantage of this model is that, it is not very influenced by the presence of large, long-lived increases and decreases in energy prices and produces of robust long-term forecasts. Rodchenk* used the annual data for 1870-1996 to confirmed Pindyck results. The author states that the shortcoming of the model is the inability to consider the impact of OPEC's decisions.

In 2006, Knetch [13] analysed the concept of convenience yield. The author proved both theoretically and empirically existence of convenience yields and their usefulness for crude oil price forecasting. Geman[10] proposed some mathematical elements towards a definition of mean reversion that would not be reduced to the form of the drift in stochastic differential equations. The author used the West Texas International (WTI) oil prices over the period of January 1994 to October 2004. Using the well known Augmented Dickey-Fuller (ADF) test, he obtained the p-value of 0.651 for spot prices over the period of January 1994 to October 2004. These results reject the mean reversion assumption over the whole period and confirmed that the log crude oil prices follow the random walk during this period.

Chikobvu and Chinhamu [5], investigated whether crude oil prices are mean reversion or follow a random walk process at all time. The authors used the following model to investigate the statistical properties of crude oil price:

$$Y_t = \alpha Y_{t+1} + \mu_t, \tag{4}$$

where, α - the speed of mean reversion and all variables are defined as in equation (1).

Chikobvu and Chinhamu used the Augmented Dickey-Fuller (ADF) test and the Garch model with time-varying properties approach to find whether crude oil prices follow a random walk or a mean reversion process. They used the monthly crude oil prices for the period of January 1980 to September 2010. These data were divided into two segments, namely January 1980 to January 1994 and from February 1994 to September 2010. The ADF test showed that crude oil prices follow a mean reversion process in the first segment of January 1980 to January 1994, as it gave the ADF test statistic for log crude oil price data of -3.599062 with p-value of 0.0829. For the second portion the ADF test statistic for log crude oil prices data is -2.963231 and the p-value is 0.1459 which is more than one percent increase from the previous segment . These results implied that crude oil prices follow a random walk process in the second segment. Results from the Garch model with time-varying parameters also supported that crude oil prices follow a mean reversion on the first period of January (1980 - 1994), random walk on the period of February 1994 to September 2010. The authors suggested additional work to be done as there are still gaps and updates that need to be taken care of, i.e looking at the size of the period under consideration.

3 MODELLING WITH RANDOM WALK AND MEAN REVERSION

3.1 Introduction

In Section 2 we had an overview of the variety of theoretical models and Mathematical tools used to capture oil price movements. The existence of mathematical formulae and concepts depend on the assumptions made about oil prices and how different researchers perceive them. This section is focused on mathematical modelling techniques, their theoretical framework and the implications they have in explaining oil prices fluctuations.

3.2 Data

The data used in this study are from The Ministry of Mines and Energy (MME) in Namibia. The period is from April 1989 to April 2014 (25 years). This is the only period for which data are available in the country. As proposed by [5], huge data set is required to ensure reliable conclusions. The data are divided into five year periods, in a triple bid to neutralise or look into: one possible sectional asymmetric effects of volatility, two impacts of precautionary demand (persistent or transitory effects), and three bigger likelihood of capturing jumps or reverting tendencies of prices, as the later occur mostly for shorter periods of time. This approach has the convenient bonus of facilitating the analysis. All data are transformed into monthly log returns series by taking the first difference in logarithms of the price.

To start with, lets consider a one-period simple return of oil price. Tsay [27] calculate one-period simple returns as follow;

$$1 + R_t = \frac{X_t}{X_{t-1}}, \tag{5}$$

where R_t is the one-period simple return, X_t is the price of the oil at time t. For multi-period, simple return is given by

$$1 + R_t(k) = \frac{X_t}{X_{t-k}}, \tag{6}$$

where k is the number of period under consideration. Equation (6) can also be express as

$$R_t(k) = \frac{X_t - X_{t-k}}{X_{t-k}}. \tag{7}$$

Equation (7) is also know as the Compound return and $R_t(k)$ is called k-period return.

A continuously compounded oil return can be calculated as the natural logarithm of the simple return of the oil prices. From (6), we have

$$r_t = \log(1 + R_t) = \log(\frac{X_t}{X_{t-1}}) = \log(X_t) - \log(X_{t-1}) = Y_t - Y_{t-1}. \tag{8}$$

The continuously compounded multi-periods return is given as the sum of the continuously compounded one period returns and is given by

$$\begin{aligned} r_t(k) &= \log(1+R_t(k)), \\ &= \log[(1+R_t)(1+R_{t-1})(1+R_{t-2})...(1+R_{t-k+1})], \\ &= \log(1+R_t)+\log(1+R_{t-1})+\log(1+R_{t-2})+...+\log(1+R_{t-k+1}), \\ &= r_t + r_{t-1} + r_{t-2} + ... + r_{t-k+1}. \end{aligned} \tag{9}$$

3.3 Random walk model

In Section 1, we looked at the general definition of random walk. A mathematical formulation of a random walk process is given as follows:

$$Y_t = \beta Y_{t-1} + \mu_t, \tag{10}$$

where Y_t is the logarithm of the oil price at time t, β is real constant and μ_t is a sequence of uncorrelated random variables called white noise say, $\{w_i\}$ that are independent and identically distributed (iid). If each w_i is normally distributed, then a white noise is a Gaussian white noise with the following properties:

$$\text{Mean} = E(w_i) = 0, \tag{11}$$

$$\text{Cov}(w_i, wj) = 0. \tag{12}$$

The hypothesis of random walk process is that, oil price returns are independent random variables and if the time intervals are equal, then the returns can be taken to be identically distributed. Let $X(t_i)$ denote the oil price at time t_i, then the simple returns

$$\frac{X(t_1)}{X(t_0)}, \frac{X(t_2)}{X(t_1)}, ..., \frac{X(t_n)}{X(t_{n-1})}, \tag{13}$$

are independent and identically distributed (iid) random variables. To show that oil price follows a geometric random walk, let us consider the following:

$$\begin{aligned} \frac{X(t_n)}{X(t_0)} &= \frac{X(t_n)}{X(t_{n-1})} \cdot \frac{X(t_{n-1})}{X(t_{n-2})} \cdot ... \cdot \frac{X(t_2)}{X(t_1)} \cdot \frac{X(t_1)}{X(t_0)}, \\ &= H(t_n) \cdot H(t_{n-1}) \cdot ... \cdot H(t_2) \cdot H(t_1), \end{aligned} \tag{14}$$

where $H(t_i)$ is a simple return at time t. Therefore $X(t_n)$ can be represented as;

$$X(t_n) = X(t_0) \prod_{i=1}^{n} H(t_i). \tag{15}$$

Taking the natural logarithms of both side we obtain;

$$\begin{aligned} \log[X(t_n)] &= \log[X(t_0) \textstyle\prod_{i=1}^{n} H(t_i)], \\ &= \log[X(t_0)] + \log[\textstyle\prod_{i=1}^{n} H(t_i)], \\ &= \log[X(t_0)] + \sum_{i=1}^{n} \log[H(t_i)]. \end{aligned} \tag{16}$$

Equation (16) is similar to equation (10) if Y_t, Y_{t-1}, μ_t are equated to $\log[X(t_n)]$, $\log[X(t_0)]$, $\sum_{i=1}^{n} \log[H(t_i)]$ respectively and $\beta = 1$, see [23].

3.4 Mean Reversion model

Oil prices paths can also be described by the theoretical stochastic differential equation of the form:

$$dY_t = \alpha(Y^* - Y_t)dt + \sigma\mu_t, \tag{17}$$

where X_t is the oil price at time t, $Y_t = \log(X_t)$ is logarithm of oil price at time t. α and Y^* are constants indicating the speed and the level of mean reversion respectively. σ is a volatility of the log oil prices and is assumed to be constant in this setting. μ_t represents the stochastic behavior and its $\mu_t \sim N(0, dt)$, where dt is the time increment. This process is called "mean reversion".

Knowing that Y_t denotes oil log price, equation (17) can be rewritten as

follows:

$$\begin{aligned} Y_t - Y_{t-1} &= \alpha(Y^* - Y_{t-1})\triangle t + \sigma\mu_t, \\ Y_t &= \alpha Y^*\triangle t + Y_{t-1} - \alpha Y_{t-1}\triangle t + \sigma\mu_t, \\ &= \alpha Y^*\triangle t + (1 - \alpha\triangle t)Y_{t-1} + \sigma\mu_t, \end{aligned}$$

$$Y_t = Q_t + \theta_t Y_{t-1} + \wedge_t, \tag{18}$$

where dt is replaced with a discrete time series $\triangle t$. From equation (18), $\triangle t = \frac{1}{12}$ because data are transformed into monthly log returns. $Q_t = \alpha Y^*\triangle t$ called the regression intercepts, $\theta_t = 1 - \alpha\triangle t$ and the residual $\wedge_t \sim N(0, \sigma^2\triangle t)$. At this point Q and θ are assumed to be constants. Now mean reversion parameters can be evaluated as follows:

$$\alpha = \frac{1-\theta}{\triangle t}, \tag{19}$$

$$Y^* = \frac{Q}{\alpha\triangle t}. \tag{20}$$

Since $\wedge_t \sim N(0, \sigma^2\triangle t)$, variance of the residual can be evaluated as: $var(\wedge_t) = \sigma^2\triangle t$, thus;

$$\sigma = \sqrt{\frac{var(\wedge_t)}{\triangle t}}. \tag{21}$$

3.5 Analysis

We paid attention to our candidates models for our modelling random walks and mean reversion models. We clarified the underlying assumptions, properties and parameters of both models. Let's discuss now the tests statistic that will hint us as whether nominal oil prices are random walk or a mean reverting.

3.5.1 *Testing for Normality*

The main aim of this test is to accept or reject the idea that oil returns are independent and identically distributed (iid), it is aimed at showing that oil returns are normal random variables, with finite mean and variance. To do this, the only two parameters are needed to completely describe the normality of returns: the mean and the variance. Oil return is said to be normally distributed if the log return has fixed mean and variance. Mean and variance are simple return, and the following formulae are used to calculate these two parameters:

$$E(R_t) = e^{\tilde{\mu}+\frac{\tilde{\sigma}^2}{2}} - 1, \tag{22}$$

$$V(R_t) = e^{(2\tilde{\mu}+\tilde{\sigma}^2)}(e^{\tilde{\sigma}^2} - 1), \tag{23}$$

where $\tilde{\mu}$ is the mean and $\tilde{\sigma}$ is the variance of the normally distributed log returns [27]. If E_1 and V_1 are mean and variance of a simple log return R_t respectively, which is lognormally distributed. Then mean and variance of the corresponding log return r_t can be found as in [27] and are given by:

$$E(r_t) = \log\left[\frac{E_1 + 1}{\sqrt{1 + \frac{V_1}{(1+E_1)^2}}}\right], \tag{24}$$

$$V(r_t) = \log\left[1 + \frac{V_1}{(1+E_1)^2}\right]. \tag{25}$$

Since the sum of finite log return of iid normal random variables is normal distributed, then under the same normal assumption of r_t, one can have $r_t(k)$ to be also normally distributed. At this point take note that, this lognormal assumption is not consistent to all the properties of history return since it depends more on the period under consideration k.

3.5.2 *Augmented Dickey-Fuller (ADF) unit root test*

To test whether the log return of oil price Y_t follows a random walk or not, we use the model

$$Y_t = \beta Y_{t-1} + \mu_t, \tag{26}$$

and all the parameters are as defined in equation (10). Considering the null hypothesis $\mathbb{H}_0$: $\beta = 1$. If $|\beta| < 1$, then Y_t is stationary converging to a certain trend and sometimes can be seen as a mean reversion process. If the $|\beta| > 1$, then (26) is not stationary, the mean and variance of Y_t are growing exponentially as time goes on. The maximum estimator of β is the Least Square (LS) estimator;

$$\hat{\beta} = \frac{\sum_{t+1}^{T} Y_{t-1}Y_t}{\sum_{t=1}^{T} Y_{t-1}^2}, \tag{27}$$

for $\mathbb{H}_0$: $\beta = 1$ the appropriate regression residual mean square can be calculated as follows [7];

$$\hat{\sigma}^2_{\mu_t} = \frac{\sum_{t=1}^{T}(Y_t - \hat{\beta}Y_{t-1})^2}{T-1}, \tag{28}$$

where $Y_0 = 0$ and T is the sample size. To calculate the time ratio ($\hat{\tau}$), the following formula is considered;

$$\hat{\tau} = \frac{\hat{\beta} - 1}{\text{std}(\hat{\beta})}, \tag{29}$$

where $\text{std}(\hat{\beta}) = \hat{\sigma}\sqrt{\sum_{t=1}^{T} Y_{t-1}^2}$; see [7].
If we use Monte Carlo method to compare $\hat{\beta}$ and $\hat{\tau}$, the null hypothesis ($\mathbb{H}_0$:$\beta = 1$) can be consider or rejected.

3.5.3 *Variance ratio test*

The variance ratio test was developed by Lo and Mackeinly [17] in 1987. The hypothesis is that; given a stochastic process defined by equation (16) which is iid, then its random walk is linear. This means that the variance of $(Y_t - Y_{t-k})$ is k times the variance of $(Y_t - Y_{t-1})$, where k is the number of periods under consideration. Considering the log return of oil price (Y_t), the variance ratio, $Var(k)$ can be defined as in [17]:

$$Var(k) = \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(1)}, \tag{30}$$

where

$$\hat{\sigma}^2(1) = \frac{1}{T-1}\sum_{t=1}^{T}(Y_t - Y_{t-1} - \hat{\mu})^2, \tag{31}$$

and

$$\hat{\mu} = \frac{1}{T}\sum_{t-1}^{T}(Y_t - Y_{t-1}) = \frac{1}{T}(Y_T - Y_0), \tag{32}$$

while

$$\hat{\sigma}^2(k) = \frac{1}{\eta}\sum_{t=k}^{T}(Y_t - Y_{t-k} - k\hat{\mu})^2, \tag{33}$$

and

$$\eta = k(T-k+1)\left(1 - \frac{k}{T}\right), \tag{34}$$

where Y_T is the last observation of the log return and Y_0 is the initial observation. Under the null hypothesis of random walk, the estimators $\hat{\sigma}^2(1)$ and $\hat{\sigma}^2(k)$ are used to test the presence of random walk in Y_t. The oil prices log return Y_t is said to be a random walk process if $\hat{\sigma}^2(k) - \hat{\sigma}^2(1)$ is approximately zero. Alternatively $\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(1)} - 1$ should converge to zero as well. To calculate the standard normal test statistic used to test the null hypothesis of random walk under homoscedasticity hypothesis $Z(k)$, the following formula is considered as in [17]:

$$Z_1(k) = \frac{(VR(k) - 1)}{\sqrt{\phi(k)}} \sim N(0,1), \tag{35}$$

where

$$\phi(k) = \frac{2(2k-1)(k-1)}{3kT}, \tag{36}$$

and to test the standard normal test statistic for heteroscedasticity-consistent Z_2, we use the following formula:

$$Z_2(k) = \frac{(VR(k) - 1)}{\sqrt{\phi(k)}} \sim N(0,1), \tag{37}$$

where

$$\phi(k) = \sum_{j=1}^{k-1} \left[\frac{2(k-j)}{k}\right]^2 \delta(j), \tag{38}$$

and

$$\delta(j) = \frac{\sum_{t=j+1}^{T} (Y_t - Y_{t-1} - \hat{\mu})^2 (Y_{t-j} - Y_{t-j-1} - \hat{\mu})^2}{\left[\sum_{t=1}^{T} (Y_t - Y_{t-1} - \hat{\mu})^2\right]^2}. \tag{39}$$

3.5.4 *Multiple Variance ratio test*

Multiple variance ratio test is the extension of Lo and MacKinly [18] conventional variance test as the studentized maximum module (SMM) statistics. Chow and Denning [6] developed a simple multiple variance ratio test to control the test size of variance ratio estimates in a large Type I error, which conventional variance ratio test fails to control. They stress that the two statistics $Z_1(k)$ and $Z_2(k)$ can only test the individual variance ratio for a given k value. Under the null hypothesis any variance ratio should be equal to one, so that it's easy to select all variance ratios with unity. Let us consider the positive integer k_i, then the null hypothesis is $\mathbb{H}_0$: $Var(k_i) = 1$ for $i = 1, 2, 3, 4, ..., n$. The test statistic is defined as follows:

$$Z_1^*(k) = Max_{1 \leqslant i \leqslant n} |Z_1(k_i)|, \tag{40}$$

$$Z_2^*(k) = Max_{1 \leqslant i \leqslant n} |Z_2(k_i)|. \tag{41}$$

The rejection of the null hypothesis depends on the maximum absolute value of the individual variance ratio test statistics. For a large sample size T, the null hypothesis is rejected at the level α significance if Z_1^* is bigger than the $[1 - (\frac{\alpha^*}{2})]$ the percentile of the standard normal distribution, where $\alpha^* = 1 - (1 - \alpha)^{\frac{1}{n}}$. Since the sample size is finite, critical values obtained by [6] will be used.

3.5.5 *Ranks and Signs non-parametric Variance Ratio (VR) test*

The advantages of ranks and signs based tests developed by Wright [30] is, that it can calculate the exact distributions without concerns about the size of distortion. Secondary ranks and signs are regarded as a powerful test if data turn out to be highly non-normal. Wright also proposed non-parametric variance ratio tests using ranks and signs of the returns and shows that they are powerful compare to variance ratio tests.

RANK-BASED VARIANCE RATIO (VR) TEST Suppose $\hat{r}(Y_t)$ is the rank of oil return Y_t among $Y_1, Y_2, Y_3, ..., Y_T$, with a standardised zero mean and a unit variance. Define

$$\hat{r}_{1t} = \frac{\left(\hat{r}(Y_t) - \frac{T+1}{2}\right)}{\sqrt{\frac{(T-1)(T+1)}{12}}}. \tag{42}$$

If we substitute $\hat{r}_{1t}$ with Y_t in the definition of Z_1 test statistic (in equation (33)) we get:

$$R_1(q) = \left(\frac{\sum_{k+1}^{T} (\hat{r}_{1t}^2 + \hat{r}_{1t-1}^2 + ... + \hat{r}_{1t-k+1}^2}{k \sum_1^T \hat{r}_{1t}^2} - 1 \right) \times \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-\frac{1}{2}}, \tag{43}$$

and

$$R_2^*(q) = \left(\frac{\sum_{k+1}^{T} (\hat{r}_{1t}^{*2} + \hat{r}_{1t-1}^{*2} + ... + \hat{r}_{1t-k+1}^{*2}}{k \sum_1^T \hat{r}_{1t}^{*2}} - 1 \right) \times \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-\frac{1}{2}}, \tag{44}$$

where $\hat{r}_{1t}^*$ is a standardised series from any permutation 1,2,3,...,T.

SIGN-BASED VARIANCE RATIO (VR) TEST Let $u(Y_t, k) = 1(Y_t > k) - \frac{1}{2}$. Then $u(Y_t, 0)$ is $\frac{1}{2}$ if $Y_t > 0$ and $-\frac{1}{2}$ otherwise. Let $S_t = 2u(Y_t, 0) = 2u(\mu_t, 0)$. This shows that S_t is iid with mean zero and a unit variance. So each S_t can be equal to 1 with probability $\frac{1}{2}$ and can be -1 otherwise. The sign-based variance ratio test statistic S_1 can be defined as:

$$S_1 = \left(\frac{\sum_{k+1}^{T} (s_t + s_{t-1} + ... + s_{t-k+1}^2}{k \sum_1^T s_t^2} - 1 \right) \times \left(\frac{2(2k-1)(k-1)}{3kT} \right)^{-\frac{1}{2}}. \tag{45}$$

3.5.6 *Q-statistic test*

Q-statistical test well known as "Ljung-Box Q-statistic test" was developed by Ljung and Box [16]. The test is aimed to test the null hypothesis, $\mathbb{H}_0$: Oil price return are autocorrelated. Let consider the stochastic properties of the residuals (white noise) $\mu_t = (\mu_1, \mu_2, \mu_3, ..., \mu_T)$ of the oil price log returns, where μ_t is as defined in equation (1). The residuals can be defined as;

$$\hat{r}(L) = \frac{\sum_{t=L+1}^{T} \mu_t \mu_{(t-L)}}{\sum_{t=1}^{T} \mu_t^2}, \tag{46}$$

where L is a lag ($L = 1, 2, 3......$), T is the sample size, $\hat{r}(L)$ is the L-residual and t is the time. Given the model above, the corresponding Ljung-Box Q-statistic can be calculated as follows;

$$\hat{Q}(\hat{r}) = T(T+2) \sum_{L=1}^{n} \frac{\hat{r}^2(L)}{(T-L)} \tag{47}$$

which is asymptotically distributed as χ^2 with $n - q$ degree of freedom (d.f) where q denotes the number of parameters in the model and n is L_T. If the model is autoregressive-moving average, then q is the sum of the autoregressive order and the moving average order, see [16] for more information.

4 PRESENTATION/EXPLORATION OF DATA

Before proceeding with any in-depth analysis,let us have a look at the raw data in consideration. This will give us a good general impression about the summary statistics, and can also assist detecting possible mistakes at early stage of our analysis. The data cover the period of past 25 years (15 April 1989 - 15 April 2014) with around 9132 observation (records), the data are divided into five equal periods of five years for the analytical reasons presented above, accounting for volatility, impacts of precautionary demand, as well as convenience. All the analysis were conducted on each periods and on the overall data. To start with, we visualise the data. A histogram of the data set, and histograms of data segments. **Figure 4.1** shows the histogram of each period as indicated below.

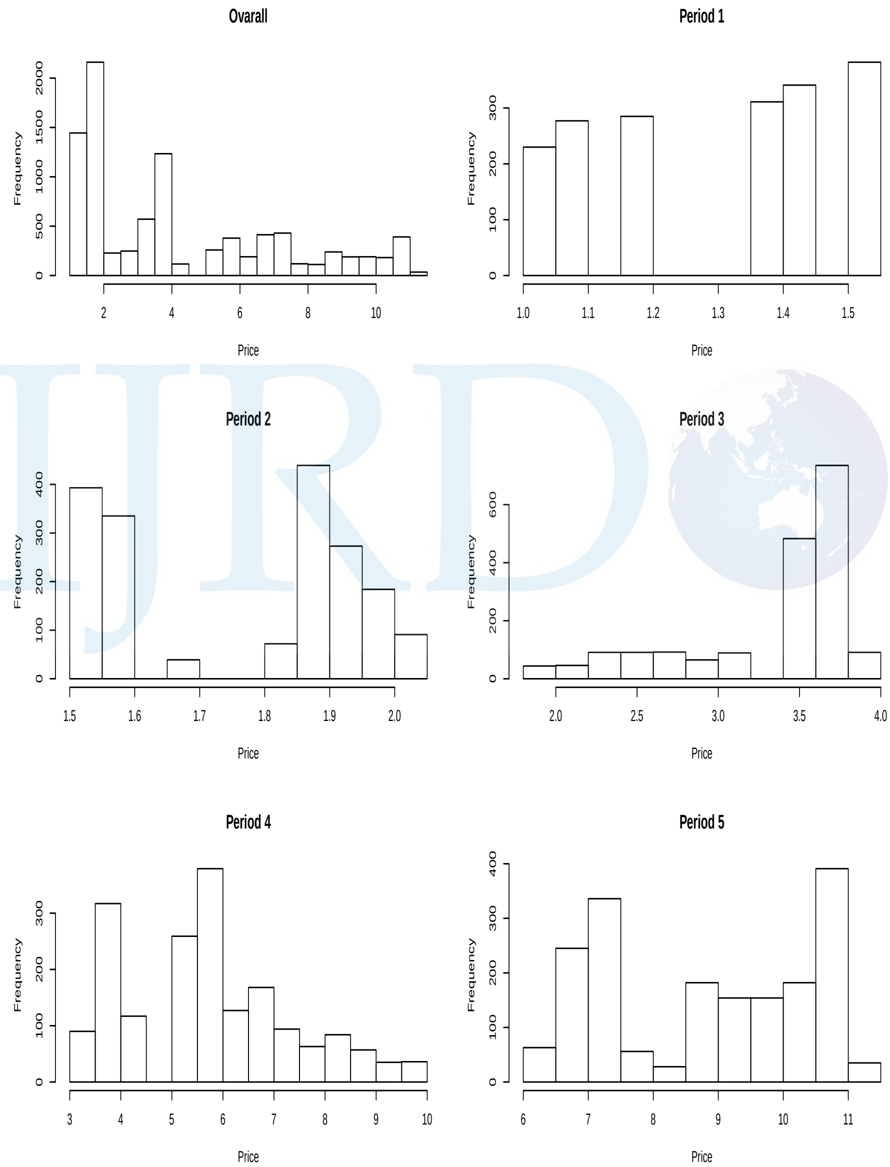


Figure 4.1 Histogram of oil prices by periods.

According to the graphs above, overall the oil prices bellow N$ 2.00 have been mostly observed, while of around N$ 5.0 and more than N$ 12.00 were not really popular. In Period 1, the prices are between N$ 1.00 and N$ 1.50 are almost equally observed. Meanwhile, there was little observations of oil prices between N$1.60 and N$1.80 in Period 2. In Period 3,the price of N$ 3.50 to N$ 4.00 were mostly observed. The oil prices N$ 3.00 to N$ 12.00 have been recorded at the average of 200 observations in Period 4 and Period 5.

The **Table 4.1** below shows the summary of some of the basic statistic of the whole data grouped by periods. **Figure 4.2** gives the plots of nominal prices over each chosen sub-period.

	Mean	**Median**	**Mode**	**Standard Deviation**	**Minimum**	**Maximum**	**Count**
Period 1	1.3027	1.38	1.53	0.18054	1.04	1.53	1826
Period 2	1.7653	1.89	1.51	0.19003	1.51	2.02	1826
Period 3	3.3399	3.54	3.67	0.5449	1.91	3.9	1827
Period 4	5.7999	5.66	3.9	1.6047	3.41	9.73	1826
Period 5	8.8407	9.27	10.74	1.5856	6.03	11.04	1826
Overall	4.2096	3.5	3.9	2.9870	1.04	11.04	9131

Table 4.1 Basic statistic results

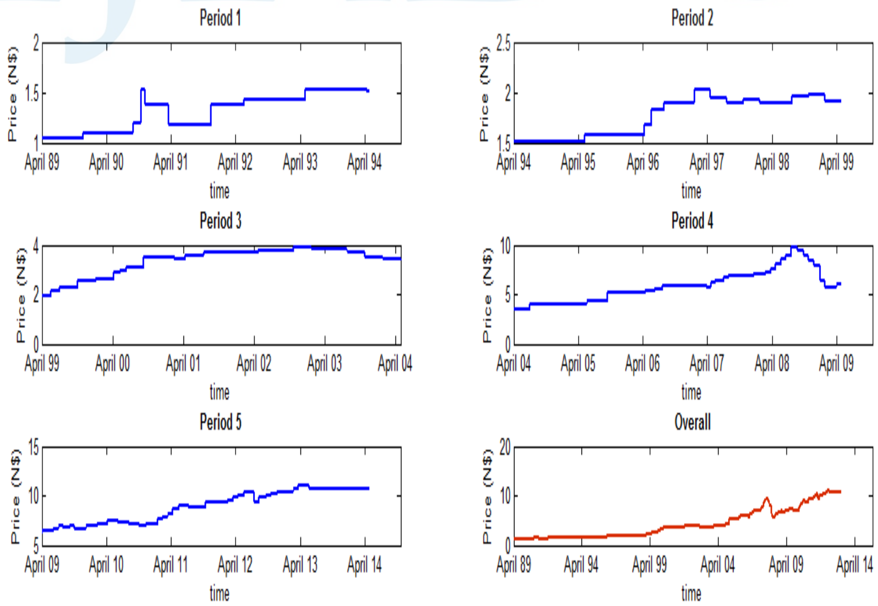


Figure 4.2 Graphs of oil prices by periods.

The basic statistics and the graphs above overall shows that oil prices had been increasing over the periods under consideration. It also indicates that, there are some up and down in the movements of oil prices. The graphs in figure 4.1 shows that in some periods there are slightly smooth in the change, Period 3 and Period 5, while on the other hand, Period 1, period 2 and Period 4 present some dramatic increases/decrease in the oil prices.

From this summary picture the question of whether or not one can find a trend that governs the movement of oil prices is further motivated. The next section will discuss the results and findings.

5 RESULTS AND DISCUSSION

5.1 Introduction

This section gives a report of the results/findings from the analysis techniques presented in Section 3. The results are interpreted and discussed in full details. We try to give a clear meaning, as the main objective of the section is to exhibit clear supportive evidence that elucidates the various questions evoked in the previous section. These results naturally feed the general conclusion on whether oil prices are better described by random walk or mean reversion processes.

All the analysis below are conducted using MATLAB, hValue = 1 indicates the rejection of the null hypothesis the test, hValue = 0 means accepting the test null hypothesis. The pValues indicating the strength at which the null hypothesis of the test are rejected or accepted are also given.

5.2 Testing for Normality

To test for the data normality, Kolmogorov-Smirnov (KS) and Adreason-Darling (AS) tests are performed on the in-sample data. For all sub-periods, the two tests reject the null hypothesis ($\mathbb{H}_0$) at 5% significance, the corresponding p-value of both tests is approximately zero. **Figure 5.1** shows the graph of oil prices compared to the standard normal increment. The graph clearly shows that oil price are not normally distributed. The approach devised for our analysis made provision for that eventuality. Some of the tests work well with non normal data. The rank sign test for example.

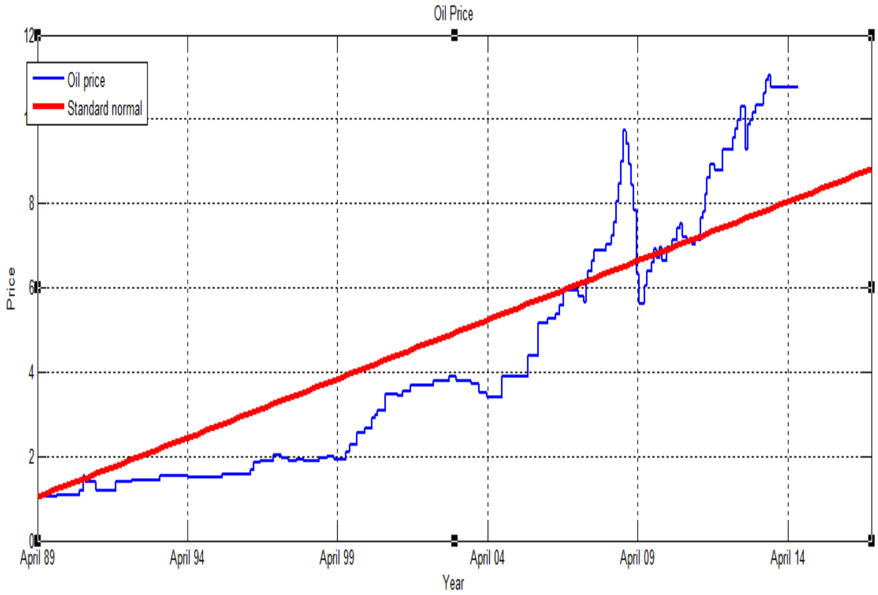


Figure 5.1 Graph of Kolmogorov-Smirnov (KS) test

5.3 Stationarity Test Results

Using the Augmented Dickey-Fuller (ADF) test, the test results show that there is no unit root test at 10%, 5%, or 1% significance level. With the p-value of 0.0010, there is enough evidence to conclude that the oil prices are stationary. All the ADF statistics support the idea that oil prices are stationary under the considered period, see table 5.1 below.

	Period 1	**Period 2**	**Period 3**	**Period 4**	**Period 5**	**Overall Period**
h-Value	1	1	1	1	1	1
p-Value	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
cValue	-1.9416	-1.9416	-1.9416	-1.9416	-1.9416	-1.9416
stat	-42.6966	-42.6966	-42.6966	-42.7083	-42.6966	-95.5353

Table 5.1 ADF test results

5.4 Variance ratio (VR) test

Variance ratio test is used for testing for the presence of randomness in the time series. To test whether oil prices are random walk, each period is divided into six segments to thoroughly investigate whether oil price variations are independently and identically distributed (iid). The results of VR test indicates the presence of random walk in period 4. There is also some signs of random walk in the first five years as well as in the last five years. However the large part of the data in consideration rejects the hypothesis of random walk in oil prices. **Table 5.2** show the outcome results of Variance ratio test. **Figure 5.2** shows stationary differenced returns.

	Segment	1	2	3	4	5	6
Period 1	h-Value	0	0	0	0	1	0
	p-Value	0.94	0.949	0.8797	0.1609	0.0477	0.5345
Period 2	h-Value	0	0	0	0	0	0
	p-Value	0.9976	0.9945	0.9914	0.9485	0.8841	0.8199
Period 3	h-Value	0	0	0	0	0	0
	p-Value	0.9976	0.9945	0.9914	0.9485	0.8841	0.8199
Period 4	h-Value	0	0	0	1	1	1
	p-Value	0.4958	0.5704	0.5674	0	0	0
Period 5	h-Value	0	0	0	0	1	1
	p-Value	0.9652	0.9371	0.8938	0.1835	0.0163	0.0085

Table 5.2 Variance ratio test results

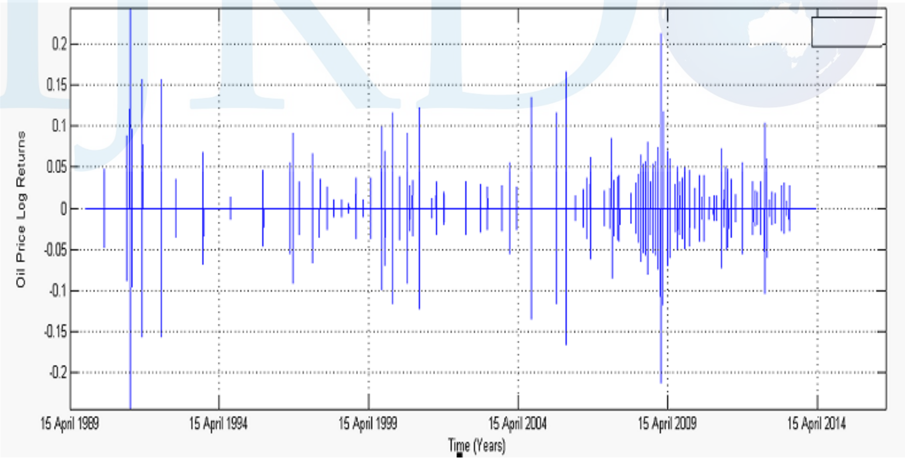


Figure 5.2 Differenced returns.

5.5 Multiple Variance ratio test

This test is aimed at the variance across all the five periods if they are equal. The multiple variance test is conducted at all the five periods. **Table 5.3** below has the outcome results.

Group Summary Table

Group	Count	Mean	Std Dev
1	1826	1.30266	0.18054
2	1826	1.76532	0.19003
3	1826	1.76532	0.19003
4	1826	5.79994	1.60466
5	1826	8.84074	1.58564
Pooled	9130	3.89479	1.01922

Bartlett's statistic	15305.5
Degrees of freedom	4
p-value	0

Table 5.3. Results from Multiple Variance ratio test

The results in **Table 5.3** reject the null hypothesis at 5% significant level. This means that the means are different across the five periods under consideration. Nevertheless, the mean of Period 1 and Period 2 are significantly equal as it was a case with variance ratio VR. **Figure 5.3** seems also to confirm this result.

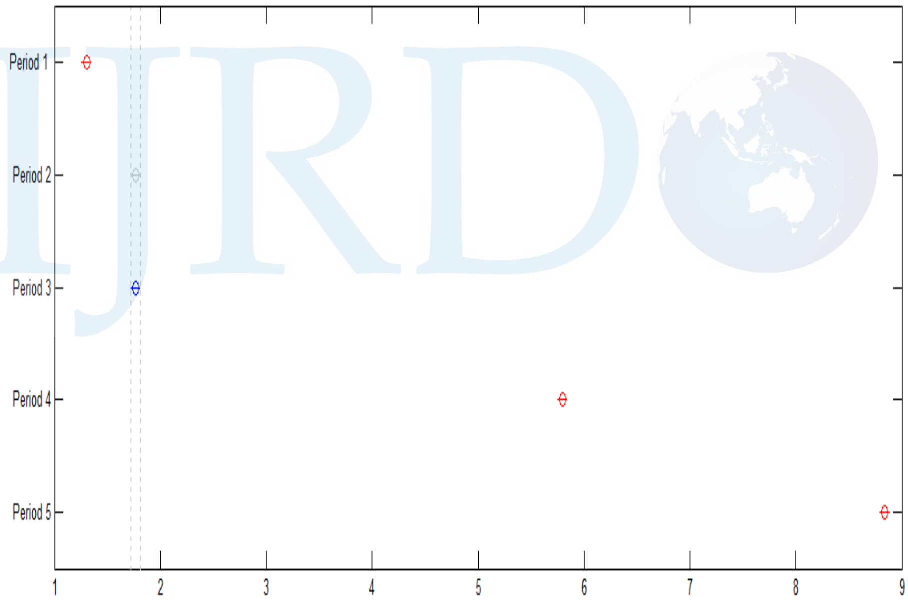


Figure 5.3. Mean level of the five periods.

5.6 Ranks and Signs non-parametric VR test

The Ranks and Signs test is regarded as a stronger test compare to Variance Ratio (VR) test, when it comes to the test for the present of random walk in time series data. As it is the case with VR test, ranks and signs test also confirmed the present of the random walk in both first and four period. Equally the test fails to provide enough evidence to reject presence of random walk in Period 5 as it only produce the p-value slight bigger than the 5% significant level. These results show the presence of random walk in Periods 1, 4 and 5, but reject the null hypothesis of random walk for both Period 2 and Period 3. However additional investigations are needed as far as period 5 is concerned. **Table 5.4** presents the numerical results.

	Period 1	Period 2	Period 3	Period 4	Period 5	Overall Period
h-Value	0	0	1	0	1	1
p-Value	0.375	0.1516	0.0139	0	0.0053	0.0000198
Stas=Zval			2.4606	1.7538	2.7868	4.2667
Signedrank	30.5	66.5	142	262	460.5	3835

Table 5.4 Ranks and Signs test results

5.7 Q-statistic test

According to Lung-Box Q-statistic test, oil prices are autocorrelated. In all periods, the test strongly rejects the null hypothesis at 5%. These results indicate that today prices have a statistically significant influence in determining future prices. Nonetheless historical prices cannot be considered as major indicators of present prices. The best interpretation of this result will say that oil prices only depend on the immediate past prices plus some random adjustments.

5.8 Johansen Cointegration test

The ADF test shows that data are stationary. The best way to test for the cointegration is to use the Johansen cointegration (JC) test. The JC test implies that there is a linear trend among the data. **Table 5.5** Shows the results of the Johansen cointegration test.

****

Results Summary (Test 1)

Data: pat

Effective sample size: 1825

Model: H1

Lags: 0

Statistic: trace

Significance level: 0.05

Period	h	stat	cValue	pValue	eigVal
1	0	31.4692	69.8187	0.9979	0.0083
2	0	16.1946	47.8564	0.9990	0.0045
3	0	7.9487	29.7976	0.9965	0.0017
4	0	4.8815	15.4948	0.8212	0.0014
5	0	2.2422	3.8415	0.1349	0.0012

Table 5.5 Johansen cointegration test results

The above results favour the null hypothesis of the Johansen cointegration test. While there was many developments that can cause permanent changes in the oil prices, there is also a long-run equilibrium relation

6 CONCLUSION AND RECOMMENDATIONS

Our results indicated that nominal oil price returns on the sub-periods of our data set are stationary, according to the ADF test. This implies stability of the models on sub-periods data and that analysis conclusions data can be validated. However the overall data set is non stationary, and not normally distributed, which makes it difficult to get very accurate overall results. Nevertheless, our analysis happens to give some good outcomes, the tests indicate that the prices only depend on the few past prices, but not on the entire history. The presence of random walks in Period 1, Period 4, slightly less in Period 5 asserts that while there are some up and down movements, oil prices are increasing with a positive stochastic trend . The results are more in support of the conclusion that oil prices follow a mean reversion process, with untimely occurrences of random walk. The causes of randomness need to be further investigated looking at the global macroeconomic stability, political factors, OPEP decisions, and level of inventories.

CONFLICTS OF INTEREST DISCLOSURE

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] P. Bjerksund and E. Steinar. Contingent claims evaluation for mean-reverting cash flows in shipping- in real options in capital investment, models, strategies, and applications. *Lenos Trigeorgis*, 3:65–83., 1994.

[2] F. Black and M. Scholes. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659, 1973.

[3] C. Blanco and D. Saronow. Mean reverting process - energy price process used for derivatives pricing and risk management. *Commodities Now*, pages 68–72, 2001.

[4] J.P. Botha. The random walk model and the behaviour of gold price. *Investment Analysts Journal*, 9:19–23, 1980.

[5] D. Chikobvu and K. Chinhamu. Random walk or mean reversion? emprical evindence from the crude oil market. *Journal of The Turkish Statistical Association: iSTATISTICAL*, 6:1–9, 2013.

[6] K.V. Chow and K.C. Denning. A simple multiple variance ratio test. *Journal for Econometrics*, 58:385–401, 1993.

[7] D.A. Dickey and W.A Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, 1979.

[8] C. Dincerler, Z. Khokher, and T. Simin. An empirical analysis of commodity convenience yield. *Social Science Research Journal*, 12:352–467, 2005.

[9] E.F. Fuma. Efficient capital mmarket: A review of theory and emprical work. *The Journal of Finance*, 25(2):383–417, 1970.

[10] J. Geman. Mean reversion versus random walk in oil and natural gas prices. *Journal of Alternative Investments*, 8:219–222, 2007.

[11] H. Hotelling. The economics of exhaustible resource. *The Journal of Political Economy*, 39:137–175, 1931.

[12] L. Kilian. Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Econimic Review*, 99(3):1053–1069, 2009.

[13] T.A. Knetch. Forecast the price of crude oil via convenience yield pprediction. *Deutsche Bundesbank Discussion Paper, Series 1*, 12, 2006.

[14] L. Leirnert. Does the oil price adjust optimally to oil field discoveries. *ETH Swiss Federal Institute of Technology*, 12:60–92, 2012.

[15] C. Lin. Insights from a simple hotelling model of the world oil market. *Natural Resources Research*, 18(1), 2009.

[16] G.M. Ljung and G.E.P. Box. On a measure of lack of fit in time series models. *Bilmetrika*, 65(2):297–303, 1978.

[17] A.W. Lo and A. C. MacKeinlay. Stock market prices do not follow random walk: Evidence from a simple specification test. *Review of Financial Studies*, 1987.

[18] A.W. Lo and C.A. MacKeinlay. Stock prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1:773–785, 1988.

[19] N. Meade. Oil prices-brownian motion or mean reversion? a study using a one year ahead density forecast criterion. *Elselvier, Energy Econimics*, 32:1485–1498, 2010.

[20] P. Morters and Y. Peres. Browian motion. http://www.stat.bekeley.edu/peres/bmbook.pdf, August 2010.

[21] K. Pearson. The problem of the random walk. *Nature*, 72:342, 1905.

[22] R.S. Pindyek. The long run evolution of energy prices. *The Energy Journal*, 20:1–27, 1999.

[23] L. Quigley. Statistical analysis of log returns of financial assets. Msc in financial mathematics, Eagna Chun Ghimh, 2008.

[24] S. Radchenko*. The long-run forecasting of energy prices using the mode of shifting trend. *The Energy Journal*, 21:1–21, 2005.

[25] A. Rashid. Do exchange rate follow random walks? an application of variance ration test. *Pakistan Economic and Social Review*, 44(1):57–79, 2006.

[26] E.S. Schwartz. The stochastic behavior of commodity price: Implications for valuation and hedging. *TheJournal of Finance*, LII No.3:923–973, 1997.

[27] R.S. Tsay. *Analysis of Financial Time Series*. JOHN WILEY & SONS, INC, 2rd edition, 2002.

[28] J. Tvedt. *Market Structure, Freight Rate and Assets in Bulk Shipping*. PhD thesis, Norwegain school of economics and Business Administration, 1995.

[29] Y. Wang and C. Wu. Forecast energy market volatility using garch models: Can multivariate models beat univariated models? *Elselvier, Energy Econimics*, 34:2167–2181, 2012.

[30] J. Wright. An assessment of the investment climate in Zambia. *Journal - American Water Works Association*, 17:45–66, 2000. Washington, USA.