İSTATİSTİK: JOURNAL OF THE TURKISH STATISTICAL ASSOCIATION

İSTATİSTİK

Vol. 6, No. 1, January 2013, pp. 1–9

ISSN 1300-4077 | 13 | 1 | 1 | 9

RANDOM WALK OR MEAN REVERSION? EMPIRICAL EVIDENCE FROM THE CRUDE OIL MARKET

Delson Chikobvu*

Department of Mathematical Statistics and Actuarial Science

University of Free State, Box 339, Bloemfontein 9300, South Africa

and

Knowledge Chinhamu†

School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal

Private Bag X54001, Durban 4000, South Africa

Abstract : The paper investigates and gives an update on the empirical evidence for or against the random walk theory in the crude oil market. The paper investigates whether there are periods when crude oil prices follow the random walk process and periods when they deviate from the random walk theory (mean reverting). Various studies often give conflicting results for the same study period. Some computations conducted over the period of 2000-2005 lead to inconclusive results (Geman, 2007), suggesting that more work remains to be done in this period and beyond. It is imperative to revisit mean reversion and random walk in the context of crude oil as it has serious implication on modeling crude oil prices. In this paper, a Garch model with time-varying properties is applied to capture periods when the random walk theory may be true and periods when it may be false. This study concludes the existence of mean reversion for crude oil price over the period 1980 to 1994 and a random walk as of February 1994 to the end of the study period in 2010. Prior to February 1994 some models, like arima models, might have been valid. Beyond this period, models need to recognize the random walk in crude oil returns.

Key words: Efficient market, Garch, Crude oil price, Random walk theory, Time varying parameters.

History: Submitted: 15 May 2012; Revised: 4 July 2012; Accepted: 9 August 2012

1. Introduction

The thrust of the study is to investigate whether crude oil prices are mean reverting or follow a random walk process. Crude oil presents an interesting case because it is the fundamental driver of most economic activities in the world. Crude oil is vital in many industries and of great importance to the maintenance of an industrialized modern economy. Higher crude oil prices have a direct impact on macroeconomic variables such as inflation, Gross Domestic Product (GDP), investments, recessions, and other macroeconomic variables (Cheong, 2009). Crude oil returns are related to the global financial markets, including contracts, opinions, risk management and other related financial derivatives.

It is thus important to investigate whether oil price predictions can be done with accuracy or not. Forecasting crude oil future prices remains one of the biggest challenges facing econometricians and statisticians. Some researchers found that crude oil prices follow a random walk, implying that tomorrow's expected crude oil prices should be the same as today's value. If crude oil prices follow a random walk, then prices would be very difficult, if not impossible to predict. It is imperative to revisit mean reversion and random walk in the context of crude oil as it has serious implication on modeling crude oil prices.

* Corresponding author. E-mail address: chikobvu@ufs.ac.za (D. Chikobvu)

† E-mail address: chinhamu@ukzn.ac.za (K. Chinhamu)

This study investigates whether crude oil prices are mean reverting or follow a random walk process at all time periods. The Augmented Dickey-Fuller (ADF) tests and the Garch model with time-varying properties approach are used to investigate mean reversion and random walk process in crude oil prices. Whilst the techniques adopted in this paper may be standard in empirical finance, the approach as presented, with time varying parameters, has not been applied to crude oil to the best of our knowledge. There are still gaps and updating that needs to be done in providing the empirical evidence in the crude oil market.

The rest of this paper is organised as follows. Section 2 gives an overview of the random walk model and related literature. Section 3 describes the empirical methods used in this study. In Section 4, the data and results are reported and discussed. Section 5 provides a conclusion.

2. Literature Review

Geman (2007) used the following model to investigate the statistical properties of crude oil prices:

$$p_t = \varphi p_{t-1} + \epsilon_t$$

to check whether the coefficient φ is significantly different from 1, where p_t is the log of crude oil price. The null hypothesis is the existence of a unit root (i.e. $\varphi = 1$). A p value smaller than 0.05 would reject the null hypothesis with a confidence level higher than 0.95. The bigger the p value, the more the random walk is validated. Geman (2007) used the West Texas International (WTI) oil spot prices over the period January 1994 to 2004. The author got an ADF p value of 0.651 for spot prices of oil prices for the period January 1994 to October 2004. The result rejects the mean reversion assumption over the whole period and confirms that log crude oil price follows a random walk during this period.

However, Geman (2007) noted that a mean reversion pattern of crude oil prices prevails for a shorter period from 1994 to 2000, and it changes into random walk as of 2000. The author used three state variable models which incorporate stochastic volatility.

Bessembinder et al. (1995) analysed the relation between crude oil price levels and slope of the futures term structure defined by the difference between a long maturity future contract and its first nearby. Assuming that future prices are unbiased expectations (under the real probability measure) of future spot oil prices, an inverse relation between prices and the slope constitutes evidence that investors expect mean reversion in spot prices, as it implies lower expected future spot prices when prices rise. The authors concluded the existence of mean reversion of oil price over the period 1982-1991, however the same computations conducted over the period 2000-2005 leads to inconclusive results (Geman, 2007). Thus more work remains to be done in this period and beyond.

Bernard et al. (2008) argue that research on crude oil price dynamics for modeling and forecasting has brought out several unsettled issues. Although statistical support is claimed for various models of price paths, many of the competing models differ with respect to their fundamental properties. One such property is mean reversion. Pindyck (1999) says that unit root tests are inconclusive in the analysis of real prices observed on a yearly basis. The author concludes that due to the persistent characteristic of crude oil price, a very long and practically unavailable series is required to perform reliable tests.

3. Methodology

A simple model for log returns

We define the natural logarithmic return (simple log return) of crude oil at time t as:

$$r_t = \log(P_t/P_{t-1}) = \log(P_t) - \log(P_{t-1}) = p_t - p_{t-1}$$

where P_t is the price of crude oil at time t.

The simplest model which can be used to test for the random walk is the simple autoregressive (AR(1)) model, namely:

$$r_t = \beta_0 + \beta_1 r_{t-1} + \epsilon_t \tag{3.1}$$

where $r_t = p_t - p_{t-1}$, is the log return of crude oil price, β_0 and β_1 are the parameters that need to be estimated and $\epsilon_t \sim \text{IID}(0, \sigma^2)$, $p_t = \log(P_t)$ is the natural logarithm of the price of crude oil at time t. If the crude oil price follows a random walk, $\beta_1 = 0$ and so

$$p_t = \beta_0 + p_{t-1} + \epsilon_t \tag{3.2}$$

the random walk with drift parameter β_0.

The natural logarithmic transformation reduces the impact of heteroscedasticity that may be present when you have large data sets with high frequency. The transformation also ensures that predicted crude oil price is positive when anti-logs are taken. The model however does not cater for changing volatility.

Three versions of the random walk are distinguished by Cambell, Lo and MacKinley (1997) and also cited in Jefferis and Smith (2005: p.59) which depend on the assumptions of the error term, namely ϵ_t. Under the first model, the error terms are independently and identically distributed with a zero mean and constant variance, denoted by $\epsilon_t \sim \text{IID}(0, \sigma^2)$. In the second model, the error terms are independent but not identically distributed, which allows for unconditional heteroscedasticity in the ϵ_t or $\epsilon_t \sim \text{NID}(0, \sigma_t^2)$. The problem of heterogeneously distributed processes is relevant, since crude oil prices have been found to display heteroscedasticity. In the third random walk model, the error terms are uncorrelated and neither independent nor identically distributed as mentioned in the research of Jefferis and Smith (2005). This paper will also focus on the third model, with volatilities changing over time.

Equation (3.1) has constant parameters and the error terms are assumed to follow the usual classical assumptions. With financial markets, the assumption of constant variance may be inappropriate as empirical evidence frequently finds that returns have a variance which changes systematically. Equation (3.1) cannot readily capture gradual deviations towards/from the random walk over successive observations.

Garch approach with time varying parameters

Emerson et al. (1997) and Zalewska-Mitura and Hall (1999) have developed, using a Garch approach, a test with time-varying parameters which detects changes towards/from the random walk where the error process does not have a full set of NIID properties. The model checks for changes towards/from the random walk and allows the error process to deviate from the property of being normally independent and identically distributed. The test does 3 things: first, it checks for the random walk; second, it detects changes from/towards the random walk, and third, it will operate with a stochastic series for which the error process might not have a full set of NIID properties.

The test is based on the following set of equations to constitute the model

$$r_t = \beta_{0t} + \beta_{1t} r_{t-1} + \delta\sigma_t^2 + \mu_t \tag{3.3}$$

$$\mu_t | \psi_{t-1} \sim N(0, \sigma_t^2) \tag{3.4}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \tag{3.5}$$

$$\beta_{0t} = \beta_{0t-1} + v_{0t}\,; \quad v_{0t} \sim N(0, \sigma_{\beta_0}^2) \tag{3.6}$$

$$\beta_{1t} = \beta_{1t-1} + v_{1t}\,; \quad v_{1t} \sim N(0, \sigma_{\beta_1}^2) \tag{3.7}$$

in which σ_t^2 is the conditional variance of the error term ϵ_t, a Garch (1,1) model. ψ_t is the information set available at time t. α_0, α_1 and γ_1 are parameters needed to model the changing volatility. This model has three important characteristics. First, the intercept, β_{0t} and slope coefficient β_{1t} can change through time. However, the special cases where either or both of these are constant are also included. Secondly, this model incorporates an error process in which the variance changes systematically over time. Thirdly, the mean of the log return depends on its conditional variance (level of risk).The basic insight is that risk-averse investors will require compensation for holding a risky asset such as crude oil.

A maximum likelihood search procedure with a standard Kalman filter is used to estimate the model with equation (3.3), the measurement equation, and the set of equations given by (3.5), (3.6) and (3.7), the state equations. The Kalman filter sequentially updates coefficient estimates and generates the set of β_{it}'s, $i = 0,1$ and $t = 1 \cdots T$ and their standard errors. If the crude oil log returns follow a random walk with no drift, then a $100(1-\alpha)\%$ confidence band for each of β_{0t} and β_{1t} should contain zero. The method will be applied to crude oil prices in this paper. The focus of this study is to find out if crude oil prices follow a random walk process or is mean reverting.

ϵ_t and v_{it} $i = 0,1$ are playing the role of both disturbances and state variables. This is somewhat unusual. The Kalman filter is being used in the context of a model with Garch errors. The Kalman filter in its present form is not operable. This is because past values of error terms are unobservable. Nevertheless we may proceed on the basis that the model can be treated as though it were conditionally Gaussian, and we will refer to the Kalman filter as being quasi-optimal(Harvey et al, 1992) (Moonis and Shar, 2002).

Extending the model

Zalewska-Mitura and Hall (1999) have an extension to the model in the previous section.

The test is based on the following set of equations:

$$r_t = \beta_{0t} + \sum_{i=1}^{p} \beta_{it} r_{t-i} + \delta\sigma_t^2 + \epsilon_t \tag{3.8}$$

$$\epsilon_t|\psi_{t-1} \sim N(0, \sigma_t^2) \tag{3.9}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \gamma_1\sigma_{t-1}^2 \tag{3.10}$$

$$\beta_{it} = \beta_{it-1} + v_{it}\ ; \quad v_{it} \sim N(0, \sigma_{\beta_i}^2) \tag{3.11}$$

Such a model can again be modeled using the standard Kalman filter. The parameters required to estimate time-paths of β_{it}'s and δ, $\alpha_0, \alpha_1, \gamma_1$ and all p values can be found by maximizing the Likelihood Function. If the series r_t is a random walk, the $100(1-\alpha)\%$ confidence bands for each of the β_{it}'s must contain zero.

Reasons for modeling Garch effects

Like many econometric time series, crude oil prices exhibits periods of unusually large volatility followed by periods of relative tranquility. In such instances, the assumption of a constant variance (homoscedasticity) is in appropriate. The volatility of crude oil prices is not constant over time, a property called heteroscedasticity. Modelling such varying variances involves Garch modelling. A distinguishing feature of a Garch model is that the error variance may be correlated over time because of phenomenon of volatility clustering. Thus it is appropriate to use the Garch model which incorporates an error process in which the variance is allowed to change systematically over time. Hence, the model can detect gradual departures from the random walk (weak form efficiency) through time.

Building AR (p) models

An important step in the model identification process is to find the order of the autoregressive process for the log returns. There are three basic steps to follow to fit AR(p) models to time series data. These steps involve plotting the data, possibly transforming the data, identifying the dependence orders of the model, parameter estimation, and diagnosis and model choice. The Box Jenkins methodology using auto-correlations is used to identify the order of the model.

Model selection for ADF Tests

The lag order, in addition to a sample size can affect the finite sample behavior of the ADF test. Proper correction for the lag effect in implementing the ADF test is desirable. Because appropriate values for the ADF test can be easily computed with desirable accuracy from response surface equation for any sample size and leg length, the analysis should be useful in practical applications (Cheung and Lai, 1995). The number of the augmenting lags (p) is determined by minimizing the Schwarts Bayesian information Criterion (SBI) or minimizing the Alkaike Information Criterion(AIC). In this study the SBI is used and the software automatically selects the appropriate lag length and hence the model.

4. Empirical results

This section discusses data source and data analysis. The data is on crude oil prices. The section also discusses the results of the random walk process and the results of the ADF test. Lastly, results from the Garch model with time-varying parameters approach are also discussed.

The data used in this study is monthly crude oil price from January 1980 to September 2010 with a total of 369 observations and is quoted in US dollars. The data is a monthly crude spot price of European Brent. This data is from the World Bank and the International Monetary Fund (IMF) and is available from the following website: http://www.mongabay.com/commodities/price-charts/price-of-uk-brent-oil.html.

The data is used to form two sub-segments of data namely, January 1980 to January 1994 and February 1994 to September 2010 segments. Two data segments are used for ADF tests for reasons which will become obvious after observing the results from Garch modelling. The data series from 1980 to 2010 when analysed as a whole with ADF test, the results are somewhat different, for instance, from 1980 to 1994, crude oil price does not follow a random walk model, i.e. it is mean reverting. The data series is also transformed into monthly log returns series by taking the first difference in the logarithm of the prices to give the log returns.

The ADF for the data in the period 1980 to 2010 (Full Data Set)

The ADF test is used to test stationarity for the data set from 1980 to 2010. Conclusions are made in line with Geman's (2007) paper. The ADF test statistic for untransformed crude oil price is -2.946064 with p value of 0.1493. At 10% significance level, the null hypothesis of non-stationary (unit root) is not rejected implying that crude oil prices are non stationary. Non stationarity implies the random walk (German, 2007). Similar results are obtained for log crude oil price data from 1980 to 2010, the ADF test statistic is -2.414055 (p value 0.3716). The p value is greater than 10%, the null hypothesis of unit root is not rejected. This result implies that the log crude oil price is non stationary and thus a random walk.

The ADF test for the data in the period January 1980 to January 1994 (First Segment)

The ADF test statistic on untransformed crude oil price data for the period 1980 to 1994 is -3.599062 with p value of 0.0829. Since p value is less than 10%, the null hypothesis of non-stationary (unit root) is rejected implying that crude oil price is stationary over the period of the first segment. This result is rather surprising and is not consistent with the results of the whole data set ranging from 1980 to 2010. This result suggests that crude oil price is mean reverting over the period 1980 to 1994. The ADF test statistic for log crude oil price data (period 1980 to 1994) is -2.963231. The p value $= 0.1459$ which is more than 10% , meaning that the null hypothesis of unit root is not rejected and hence implying that the log crude oil price are thus a random walk process. To summarize the results thus far, the data series from 1980 to 2010 when analysed as a whole, the conclusion is that the crude oil price follows a random walk model. However a shorter period, called the first segment, the conclusion is that that crude oil price is mean reverting over the period 1980-1994. However if the data is log transformed over the same period, 1980-1994, the conclusion is that the log crude oil price follows a random walk.

The ADF test for the data in the period February 1994 to September 2010 (Second Segment)

The value of the ADF test statistic is -1.703747 (p value $= 0.4278$), that is the hypothesis of a stationarity is not rejected and the monthly crude oil price follows a simple random walk for the period February 1994 to September 2010.The ADF test statistic of -2.816595 (p value of 0.1931) for log crude oil is less negative than the critical value at 10% significant level, the null hypothesis of unit root is not rejected. Thus the log crude oil price is non stationary hence implying a random walk for log crude oil price. The conclusion is crude oil and log crude oil price is a random walk over the period 1994-2010. These results of the random walk or mean reversion seem to depend on the period under consideration and whether the data is log transformed or not. The behavior of a random walk is more pronounced in log crude oil price.The conclusion is log crude oil price is a random walk over the period 1980-1994 but for the untransformed data the conclusion is that crude oil price is mean reverting for the same period (1980-1994).

Results from the Garch model with time varying parameters.

The results of using the Garch model with time varying parameter are presented in this section. Figures 1, 2 and 3 present the results of the changes towards/from the random walk. The figures show the paths of the estimated β_{it}'s, $i = 0, 1, 2$ coefficient (see equation (3.8)) with their respective 95 per cent confidence bands. For the period 1980 to 2010, the best model using Box Jenkins methodology is an AR (2) model:

$$r_t = \beta_{0t} + \beta_{1t} r_{t-1} + \beta_{2t} r_{t-2} + \mu_t$$

Garch model with time varying parameters (period from 1980 to 2010)

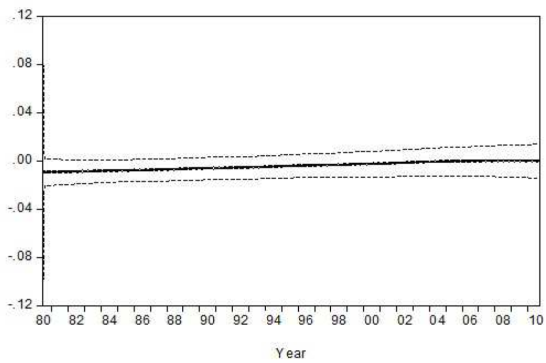

Figure 1. $\hat{\beta}_{0t}$, Drift parameter for crude oil price from 1980 to 2010. The estimates of $\hat{\beta}_{0t}$ are shown by a solid bold line and its confidence limits by dotted lines.

Consider Figure 1, which represents the results of the estimated drift parameter $\hat{\beta}_{0t}$ for the period 1980 to 2010. The estimate, $\hat{\beta}_{0t}$, has constant value of -0.0072 and is insignificantly different from zero considering its 95 percent confidence limit.

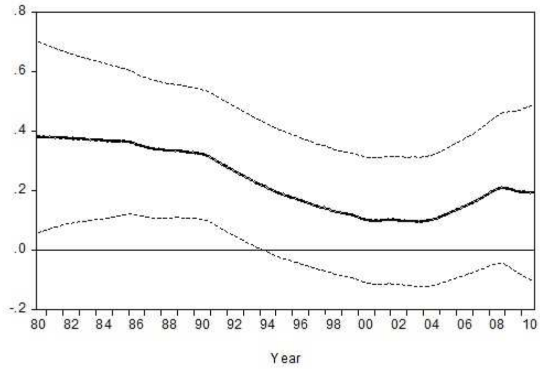

Figure 2. $\hat{\beta}_{1t}$ estimates for crude oil price from 1980 to 2010. The estimates of $\hat{\beta}_{1t}$ are shown by a solid bold line and its confidence limits by dotted lines.

Figure 2 shows the results of the parameter $\hat{\beta}_{1t}$ for the period 1980 to 2010. The estimate $\hat{\beta}_{1t}$ has an initial value of 0.39 and is significantly different from zero at 0.05 level. The magnitude of the estimated parameter gradually declines and first becomes insignificantlly different from zero in February 1994. The parameter remains insignificant for the rest of the period to end at a level of 0.19 in September 2010.

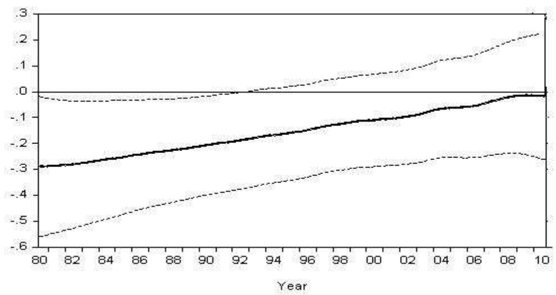

Figure 3. $\hat{\beta}_{2t}$ for crude oil price 1980 to 2010. The estimates of $\hat{\beta}_{2t}$ are shown by a solid bold line and its confidence limits by dotted lines.

Figure 3 shows the results of the parameter $\hat{\beta}_{2t}$ for the period 1980 to 2010. The estimate $\hat{\beta}_{2t}$ has an initial value of -0.29 and is significantly different from zero at 0.05 level. The magnitude of the estimated parameter gradually increase and first becomes insignificantly different from zero in March 1993. The parameter remains insignificant for the rest of the period to end at -0.02 in September 2010. Crude oil price follow the random walk from February 1994. Prior to the year 1994, the finding is that crude oil prices did not follow the random walk process i.e. crude oil price were mean reverting.

5. Conclusion

In this study, an attempt was made to determine whether crude oil price is mean reverting or a random walk process. Two approaches namely the Augmented Dickey-Fuller (ADF) test and the Garch model with time-varying properties are used. Before carrying out formal Augmented Dickey-Fuller (ADF) tests, the autocorrelation function (ACF) correlogram of crude oil price and log crude oil price are examined to investigate stationarity.

The untransformed data series from 1980 to 2010 shows evidence of a random walk process when using the ADF test, yet a shorter period (first segment) shows mean reversion for the period January 1980 to January 1994 according to the same test. The test also shows that crude oil price follows a random walk over the period February 1994 to 2010. Thus the results seem to depend on the period under consideration and this is rather puzzling. These results of the random walk or mean reversion also seem to depend on whether the data is log transformed or not. The behaviour of a random walk is more pronounced in log crude oil price. These results show that the ADF test approach has a limitation of depending on the period under consideration.

The Garch model with time-varying parameters approach shows the presence of mean reversion in log crude oil prices over the period January 1980 to January 1994. It shows a random walk as of February 1994. The conclusion of the ADF test seem to depend on the period under consideration. The cut off period of January 1994 for the first segment is suggested by the results of the Garch modelling approach. This approach does not depend on the period under consideration and can be deemed to be better than the ADF test in that sense. The Garch modelling approach confirms the ADF approach when the latter approach uses segmented data with log transformed data.

The results obtained in this paper are consistent with the results by Geman (2007) who concluded that the crude oil price follow a random walk for the period January 1999 to October 2004, using the ADF test. The result of the study also shows some similarity with Bessembinder et al. (1995), who confirm the existence of mean reversion over the period 1982 to 1991. The authors obtained inconclusive results over the period 2000 to 2005 and in this paper, a random walk prevails over that period.

This paper uses more current monthly data on crude oil prices up to September 2010. The Garch time-varying property approach used in the study produces results that are somewhat different to the ADF test. The results are similar only when the data is segmented after observing information from the Garch model. This study concludes the existence of mean reversion for crude oil prices over the period 1980 to 1994 and a random walk as of February 1994. The results also confirm a finding by Geman (2007) that the behaviour of a random walk is more pronounced when using log crude oil price. The ADF test using untransformed data shows that the crude oil price is mean reverting for the period January 1980 to February 1994 yet the log transformed data shows a random walk over the same period.

The paper recommends further study in using time-varying parameters approach to investigate mean reversion and random walk for asset prices of precious metals such as gold and platinum. Statisticians and econometricians should at least use the Garch approach before using the ADF tests to investigate whether prices are mean reverting or a random walk for a period under consideration, and avoid conflicting results which depend on size of the sample (period) under consideration.

References

[1] Bessembinder, H., Coughenour, J., Seguin, P. & Smoller, M. (1995) Mean reversion in equilibrium asset prices: Evidence from the futures term structure. Journal of Finance, 50: 361-375.

[2] Bernard, S.T., Khalaf, L., Kichian, M. & McMahon, S. (2008) Oil prices: Heavy Tails, Mean Reversion and the convenience Yield.
Available on line: http://ideas.repec.org/p/lvl/lagcr/0801.html

[3] Cheong, C.W. (2009) Modeling and forecasting crude markets using ARCH-type models. Energy policy, 2346-2355.

[4] Cheung, Y.W. and Lai, K.S. (1995) Log Order and critical values of the Augmented Dickey-Fuller Test. Journal of Business and Economic Statistics, 13(3): 277-280.

[5] Campbell, J.Y., Lo, A.W. & Mackinlay, A.C. (1997) The Econometrics of Financial Markets, Princeton: Princeton University Press.

[6] Emerson, R., Hall, S.G. & Zalaweska-Mitura, A. (1997) Evolving Market Efficiency with an application to some Bulgarian shares. Economics of planning, 30(1): 75-90.

[7] Geman, H. (2007) Mean Reversion versus Random Walk in Oil and Natural Gas Prices. United Kingdom: University of London.

[8] Harvey, H., Ruiz, E & Sentona, E. (1992) Unobserved component time series models with ARCH disturbances. Journal of Econometrics 52: 129-157.

[9] Jefferies, K. & Smith, G. (2005) The changing efficiency of African stock markets. South African Journal of Economics, 73: 54-67.

[10] Lo, A.W. & Mackinlay, A.C. (1999) A Non-Random Walk down Wall Street. Princeton: Princeton University Press.

[11] Moonis, S.A. & Shar, A. (2008) Teseting the time variation in beta in India. Journal of Emerging Markets Finance, 2(2): 163-180.

[12] Pindyck, R.S. (1999) The long run evolution of energy prices. The Energy Journal 20: 1-27.

[13] Zalewska-Mitura, A. & Hall, S.G.] (1999) Examining the first stages of market performance: a test for evolving market efficiency. Economics Letters, 64: 1-12.

[14] The mongabay website
http://www.mongabay.com/commodities/price-charts/price-of-uk-brent-oil.html [Accessed 22July 2012]

1 306010 010091 >