Sri Lankan Journal of Applied Statistics, Vol.10, 2009, p.203-216

A Fractional ARIMA Model for the Daily Spot Crude Oil Prices of the Organization of Petroleum Exporting Countries (OPEC)

W.J.R.M. PRIYADARSHANA AND C.D.TILAKARATNE

Department of Statistics, University of Colombo, Sri Lanka

ABSTRACT

This research study uses the fractional integration techniques to model the daily spot crude oil prices of the Organization of Petroleum Exporting Countries (OPEC). Long term dependency or the long term memory describes the higher order correlation structure of a series. If a series exhibits long term memory, then there exists a persistent temporal dependence even between distant observations. In this study the natural logarithm of the daily spot crude oil price were considered for the model building process and the fractional ARIMA (ARFIMA) models were utilized to build adequate model(s) to forecast the indices. The results indicate that during the study period, the daily spot prices of the OPEC series showed a fractional behavior in the series and can be modeled using fractionally differenced parameter (d) of 0.77. Both ARFIMA (2,0.77,3) and ARFIMA (3,0.77,2) models adequately described the process of OPEC daily spot prices during the study period. Thus, the daily OPEC spot price series do exhibit mean reverting and long memory characteristics during this period.

Keywords: Fractional integration, Gaussian semi-parametric method, Long memory, Modelling time series

INTRODUCTION

Oil is an essential commodity in all the economies of the world. It has become one of the triggering indicators to govern the stability of an economy. There is a significant impact on the economy of a country with the fluctuations of the oil prices. In the literature there exists a wide scope of methodologies and techniques that can be used to model oil prices.

In the conventional context, the most popular and the standard practice for building time series models are to utilize autoregressive and mixed models. These models are based on the stationarity and the characteristics of the relevant time series, which are Autoregressive terms (AR) and Moving Average terms (MA). These models are based on integer differencing and it presumes that the order of integration (*I*) to be an integer. Practically, most of the time series have shown that they are either $I(1)$ or $I(0)$. However, numerous studies have suggested that $I(0)$ and $I(1)$ condition is too restrictive and it has been shown that if the differencing operator (d) is between zero and one the process will still be mean reverting. As a result of this, "Fractional differencing" was introduced by Granger and Joyeux (1980) and Hosking (1981) as a model improvement and this has lead to a new era of econometrics.

If a series is fractionally integrated, then it is said to be that there exists a long- term dependency between the observations widely separated in time. Such series exhibit hyperbolically decaying autocorrelations and low-frequency spectral distribution (Barkoulas and Baum, 1997). Furthermore, the presence of long memory represents non linearity in the mean of the process. The conventional integer differenced time series models address only the short-term dynamics. However, the fractional time series approach will capture both the short-term and as well as the long-term dynamical properties of the series.

There is a rich class of publications in the field of fractional integration theory and its applications. Geweke and Porter-Hudak (1983) used three post-war monthly economic time series in order to present a methodology to estimate the long memory parameter d (GPH method) and as well as to formulate a fractional time series model. The formulation of the GPH method is a milestone in the field of long memory time series modelling. Meanwhile, Diebold and Rudebusch (1989) examined persistence in the US aggregate output by estimating fractionally integrated ARIMA models (ARFIMA models). They have considered ten different measures of the US macro economic activities and investigated the post-war real Gross National Product (GNP) per capital in detailed and found evidence of long memory in GNP per capital. The GPH method has been used to estimate the d parameter and ARFIMA method to the model building purposes. Baillie and Bollerslev (1994) examined long memory of the forward premium for the currencies of Canada, Germany and United Kingdom (UK). They have found that there exists long term persistence in the forward premiums considering the fractional behavior of the processes.

When considering the models used for forecasting the oil prices apart from the conventional economical methodologies, Gil-Alana (2001), discussed and tested a fractionally integrated model with a mean shift for the US and UK **real oil prices**. Hence, they have utilized the concept of long memory for the univariate time series in the oil industry. However, this study was based on the real oil prices instead of spot oil prices, where, the real oil price is obtained by multiplying the nominal value by an index. They suggested, that the US and UK real oil prices may both be $I(1)$, and by including a mean shift then the series become $I(d)$ with $d < 1$. Thus, these series are mean reverting. It is noticeable that they have used different versions of Robinson tests (Robinson, 1995) for the long- memory parameter identification, not the GPH method mentioned above. Cuaresma *et al.* (2007) examined the role of asymmetric cycles to model and forecast oil prices. They have identified the oil price forecast improved significantly when this asymmetry is explicitly modeled. Even though they have proposed this method and compared with the autoregressive models, they have ignored the process of fractional approach which deals with the non linearity and as well the long term dependency. In the literature, modelling **spot crude oil** prices of the OPEC by employing fractional time series technique was not found. Hence, this research aims to fill the gap and to formulate a fractionally integrated model to forecast the OPEC spot oil prices.

The organization of the paper is as follows. Section 2 presents the materials and methods used for the study. Results and discussions are given in the Section 3 and finally Section 4 presents conclusions and recommendations of the study.

MATERIALS AND METHODS

Daily spot crude oil prices from 4th January 2003 to 4th September 2008 of the OPEC were obtained from the daily basket from the OPEC official web site (www.opec.org). All the spot prices are in the US dollars per crude oil barrel. At first, using the time plots and suitable transformation technique variance was stabilized. Literature has pointed out (Gil-Alana, 2001) that those time series can be modeled with high precision if the logarithm of the data is being considered. Testing for integer integration orders was done by applyong two unit root tests; Phillips-Perron test (PP: Phillips and Perron, 1988) and KPSS test (Kwiatkowski *et al.*, 1992). In PP test the unit root null hypothesis is tested against the alternative of trend stationarity while the KPSS tests the trend stationarity null hypothesis against the unit root alternative hypothesis. The Gaussian semi-parametric method (Robinson, 1995) was employed to estimate the fractional integration parameter d. Since, the Robinson's method is not restricted to using a small fraction of the ordinates of the empirical periodogram of the series than that of the GPH method. Model adequacy and selection will be tested by using the Akaike Information Criteria and Schwarz Information Criteria and by considering the residual structure of the time series. The final conclusion on selecting the best model(s) is based on the composite results of the model adequacy and the robustness of forecasting.

Auto-Regressive Integrated Moving Average (ARIMA)

The ARIMA model is an extension of the mixed models. That is a mixture of Auto-Regressive (AR) and Moving Average (MA) processes. Hence, it is a generalization of the time series modelling. If the time series is having the order of integration *(d)*, then it is called an integrated model. Since, the stationary model which is fitted to the differenced series, data has to be summed or integrated to provide a model for the non-stationary data.

The general integrated moving average process (abbreviated ARIMA process) is of the form:

$$\phi(L)(1-L)^d(X_t-\mu)=\theta(L)\epsilon_t \; ; \quad \epsilon_t \sim \text{i.i.d.}(0,\sigma^2) \tag{1}$$

where, L is the lag operator, μ is the process mean, $\phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$, $\theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$, and all roots of $\phi(L)$ and $\theta(L)$ lie out side the unit circle and d is an integer. However, if d allowed to be real values and if $0 < d < 1$, the standard ARIMA model can be extended to allow fractional differencing. Thus, the process will focus on the fractional integration values. In Equation (1), a binomial expansion on $(1L)^d$ can be introduced as the fractional differencing operator (filter):

$$(1-L)^d=\sum_{k=0}^{\infty}(-1)^k\binom{d}{k}L^k$$

$$=1-dL-\frac{d(d-1)}{2!}L^2-\frac{d(d-1)(d-2)}{3!}L^3+\dots \tag{2}$$

Furthermore, by considering Equation (1) and without mean assumption for simplicity, it can be written as:

$$\phi(L)(1-L)^d X_t=\theta(L)\epsilon_t \tag{3}$$

$$(1-L)\,\phi(L)(1-L)^{d-1}X_t=\theta(L)\epsilon_t \tag{4}$$

Following Diebold and Rudebusch (1989),

$$(1 - L) X_t = \theta(L)\epsilon_t \Big/ \phi(L)(1 - L)^{d-1}$$

$$X_t - X_{t-1} = (1 - L)^{1-d} B(L)\epsilon_t \ ;$$

where B(L) = $\theta(L) \big/ \phi(L)$

$$X_t = X_{t-1} - (1 - L)^{1-d} B(L)\epsilon_t \qquad (5)$$

The stochastic process X_t is both stationary and invertible if all roots of $\phi(L)$ and $\theta(L)$ lie outside the unit circle (Granger and Joyeux ,1980). However, a shock does not have a permanent effect since the process is slowly mean reverting. Assuming that d ε(0, 0.5), Hosking (1981) showed that the correlation function, $\rho(.)$, of an ARFIMA process is proportional to k^{2d-1} as $k \to \infty$. Contrary to the faster, geometric decay of autocorrelations of a stationary ARMA process, those of the ARFIMA process decay hyperbolically to zero as $k \to \infty$. Hence, the ARFIMA process is said to exhibit long memory, or long range positive dependence when d ε(0, 0.5),. When d ε(-0.5, 5),..the process exhibits intermediate memory, or long-range negative dependence. while it exhibits a short memory when d =0, corresponding to a stationary and invertible ARMA model. When d ε(0.5, 1)..the process is mean reverting, even though it is not covariance stationary, as there is no long run impact of an innovation to future values of the process.

Autocorrelation structure reveals information about the serial dependence in the data set. Table 1 reproduced from Diebold and Rudebusch (1989) in order to provide some insight of how the autocorrelations vary between the ARIMA (short term memory) and ARFIMA (long memory) models. ARIMA (1,0,0) and ARFIMA (0,0.3,0) models were considered and it can be identified that the autocorrelations die off fairly quickly (no autocorrelation after lag five) in the ARIMA (1,0,0) process. However, for the ARFIMA (0,0.3,0) there is still autocorrelation even at lag 100. Therefore, it can be seen that the ARFIMA models has the ability to capture the long term dynamism which display a significant dependence even between the observations widely separated in time.

Table 1: Autocorrelation function for ARIMA (1,0,0) and ARFIMA (0,0.3,0)

Model	Lag (k)								
	1	2	3	4	5	10	25	50	100
$(1-0.5L)\ Y_t = \epsilon_t$	0.50	0.25	0.13	0.06	0.03	0.00	0.00	0.00	0.00
$(1-L)^{0.3}\ Y_t = \epsilon_t$	0.50	0.40	0.35	0.32	0.30	0.24	0.18	0.14	0.11

Source: Diebold and Rudebusch (1989)

The Gaussian semi-parametric method

Robinson (1995) developed and proposed a Gaussian semi-parametric approach to estimate the parameter *d*. This estimate is consistent, more efficient and asymptotically normal in the range of $d \in (0.0.5)$. It is assumed that the spectral density of the time series, denoted by $f(.)$, behave as

$$f(\xi) \sim G\xi^{1-2H} \quad as\ \xi \to 0^{+} \tag{6}$$

where, $G \in (0,\infty)$, ξ is the angular frequency and $H \in (0,1)$. The self similarity parameter H relates to the long-memory parameter d by $H = d + ½$. The estimate for H, denoted by $\hat{H}$, is obtained through minimization of the Equation (4) with respect to H:

$$R(H) = \ln \hat{G}(H) - (2H-1)\frac{1}{v}\sum_{\lambda=1}^{v} ln\xi_{\lambda} \tag{7}$$

where, $\xi_{\lambda} = \frac{2\pi\lambda}{T}$ denotes the Fourier frequencies of the sample, $v = g(T) \ll T$ is the number of Fourier frequencies included (that is number of low frequency ordinates) and $\hat{G}(H) = \frac{1}{v}\sum_{\lambda=1}^{v}\xi_{\lambda}^{2H-1} I(\xi_{\lambda})$. Here, T is the number of observations in the study. The discrete averaging is carried out over the neighborhood of zero frequency and, in asymptotic theory, v is assumed to tend to infinity much more slowly than does T. Gaussianity is no where assumed in the asymptotic theory. The

Gaussian semi-parametric estimator is $v^{1/2}$ consistent and the variance of the limiting distribution is free of nuisance parameters and equals $^{1}/_{4v}$.

In this process the most critical section is the inclusion of the number of low frequency ordinates for the analysis in order to estimate the *d* parameter. That is, $g(T)$ and it is defined as $g(T) = T^{0<r<1}$. Number of studies has shown that there exists a great deal of subjectivity in choosing the correct number of low frequency ordinates in the analysis. Diebold and Rudebusch (1989), Cheung and Lai (1993), Barkoulas and Baum (1997 and 1998) pointed out that, if $g(T)$ is too large then the estimate *d* will be contaminated due to the medium and high frequency components of the spectrum and if too few ordinates are included in the estimation process then it will lead to imprecise estimates of *d*. In this study, we report *d* estimates for $T^{0.5 \leq r \leq 0.8}$ to check the sensitivity of our results to the choice of the sample size.

RESULTS AND DISCUSSION

Initially the time series are tested for integer unit root tests to identify the order of integration. The joint conclusions made by the PP test and KPSS test indicate the nature of the time series. Table 2 gives the summary results of the PP test and KPSS tests on the distinct daily spot oil price series.

Table 2: Summary of the PP test and KPSS tests for the natural logarithm of the OPEC daily spot price series

Series	Series status	PP Test (Z_α)	KPSS Test ($\hat{\eta}_\tau$)
Log_e of OPEC (daily)	Level	-11.600***(0.000)	0.3151***
	First Differenced	No test needed+	0.0716

The superscript *** indicates statistical significance at the 1% level of significance. The parenthesis in the PP test statistic indicates the MacKinnon (1996) one-sided p-values.

PP test rejects the null hypothesis at 1% level significance for the level series. Hence, the level series is trend stationary by means of the PP test. However, the

KPSS test gives different conclusion by rejecting the null hypothesis of trend stationarity and it contradicts with the conclusion from the PP test. Therefore, the combined evidence of the PP test and KPSS test indicates the OPEC daily series is neither an $I(1)$ nor an $I(0)$ process. This suggests a fractionally differenced process may be an appropriate representation for the OPEC daily series during the study period. Further to verify the above results and to estimate fractional differencing parameter (d), Robinson's method is used. The results obtained are summarized in Table 3.

Table 3: Estimates of the fractional differencing parameter (d) from the Robinson's method for natural logarithm of the OPEC daily spot price series.

Series	**Factors**	**Power (τ)**			
		0.5	**0.6**	**0.7**	**0.8**
Log_e of OPEC	Ordinates[$g(T)$]	39	83	173	361
	Estimate (d)	1.061	0.766	0.758	0.776
	Test statistic (T)	20.2951	11.6605	16.1074	20.5597
	P>|T|	0.000***	0.000***	0.000***	0.000***

***indicates statistical significance at 1% level of significance.

The results of Table 3 strongly reject the null hypothesis of $d = 0$ at 1% level of significance and suggest there is a fractional behavior in the series. However, for the OPEC daily data series gives a fractional differenced process with d estimate about 0.77. This indicates long term dependency in the OPEC daily series.

From the preliminary analysis, it was found that the natural logarithm of the daily series of the OPEC series during the study period is a fractionally integrated process, In fact the d estimate is about 0.77. Hence, OPEC series will be modeled by utilizing ARFIMA techniques.

Modelling Daily Spot Oil Prices of the OPEC

The OPEC daily series has a fractional process and the natural log of the daily series of OPEC data are differenced by the fractional differencing value (d=0.77) to obtain the fractionally differenced stationary series in order to estimate the AR and MA parameters. Initially based on the Akaike's Information Criterion (AIC) and Schwarz's Information Criterion (SIC) the best fitting models were identified. Then, these models were assessed with their residual structure in order to come up with a more robust model in the forecasting ability. Table 4 shows the AIC and SIC for different ARFIMA models.

Table 4: AIC and SIC values of ARFIMA (p,0.77,q) models for the OPEC daily series.

Number of AR parameters (p)	**Number of MA parameters (q)**			
	0	**1**	**2**	**3**
0	-3389 -3377	-3390 -3379	-3390 -3374	-3398 -3377
1	-3388 -3377	-3390 -3374	-3446 -3425	-3450 -3423
2	-3392 -3376	-3452 -3431	-3456 -3429	-3498[a] -3466[b]
3	-3396 -3375	-3454 -3427	-3496 -3464	-3454 -3416

For each model, AIC and below that, SIC value is reported.

[a] Minimum AIC value

[b] Minimum SIC value

Both AIC and SIC criterions suggest ARFIMA (2,0.77,3) as the best model to fit the series, since it has the minimum value in both criteria. However, there are few other models which have closure values for the minimum of the both criterions and it is more prudent to select those models for further analysis together. Therefore, those models were also be considered together with ARFIMA (2,0.77,3) to test for model adequacy. Therefore, the models considered to further analysis are, ARFIMA (2,0.77,1), (2,0.77,2), (2,0.77,3) and (3,0.77,2). These models were separately fitted and assessed the plausibility and adequacy by considering the parameter estimates and also from the residual analysis. Table 5 gives the parameter estimation of the models together with their significance level. The constant term is also included in this table, since it is significant at most of the models considered.

Table 5: Parameter estimates of the identified ARFIMA models of the OPEC series.

ARFIMA (*p*,*d*,*q*)	**Parameter**	**Estimate**	**Test statistic**	***p* – value**
(2,0.77,1)	Constant	0.00284	0.7300	0.463
	AR 1	-0.7413	-4.59	0.000***
	AR 2	0.0906	2.52	0.012**
	MA 1	0.8575	-5.40	0.000***
(2,0.77,2)	Constant	0.00041	2.6300	0.009***
	AR 1	1.1575	33.74	0.000***
	AR 2	-0.3381	-6.94	0.000***
	MA 1	-1.0821	49.29	0.000***
	MA 2	0.1595	-3.81	0.000***
(2,0.77,3)	Constant	0.00069	3.4600	0.001***
	AR 1	0.8198	5.00	0.000***
	AR 2	-0.0666	-0.42	0.677
	MA 1	-0.7538	4.76	0.000***
	MA 2	-0.0910	0.61	0.545
	MA 3	-0.0574	1.43	0.152
(3,0.77,2)	Constant	0.00044	2.4800	0.013**
	AR 1	0.9882	12.72	0.000***
	AR 2	-0.1472	-1.42	0.155
	AR 3	-0.0419	-1.49	0.136
	MA 1	-0.9105	13.04	0.000***
	MA 2	-0.0015	0.02	0.987

** and *** indicates statistical significance at 5% and 1% level of significance.

Table 5 indicates that the parameters of the ARFIMA (2,0.77,1) are highly significant (at 1% level of significance) except for the constant term. Moreover, all the parameters of the ARFIMA (2,0.77,2) are significant. However, the model suggested by both AIC and SIC criterion do have some insignificant parameters. The higher order parameters are not significant. Furthermore, when considering ARFIMA (3,0.77,2) model, a similar situation can be identified as the ARFIMA (2,0.77,3) model. Hence, the higher order terms are not significant even though the model ARFIMA(3,0.77,2) has the second lowest values in the AIC and SIC criterions. However, by considering the residual analysis of the above models it was noted that the models ARFIMA (3,0.77,2) and ARFIMA(2,0.77,3) have similar structure of residuals. Figure 1 illustrates the summary residuals plots for the ARFIMA(2,0.77,3). The ordered standardized residuals plot is used in order to have an initial idea about the randomness of the residual series and subsequently correlogram and Ljung Box Q statistics (based on *p*-values and 5% significance level) plots were used to further testing down the randomness of the residuals. Furthermore, Ljung Box Q statistics plot indicates that approximately for three weeks time serial independence holds for the residuals structure.

Therefore, both models were used to test for the robustness of forecasting and then the most appropriate and suitable model was selected.

Robustness of Forecasting

Five out of sample forecasts (five operating days starting from the 5th Sep, 2008) were made based on the two models identified above. The criteria that have been used are Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD) and Mean Squared Error (MSE). The minimum of all theses criteria were considered as the most adequate model in terms of robustness of forecasting. Table 6 summarizes the measurements on the forecasting ability of the two ARFIMA models of (2,0.77,3) and (3,0.77,2).

Table 6: Statistical measures of the residuals of forecast from the ARFIMA (2,0.77,3) and ARFIMA (3,0.77,2) models.

Measure	**ARFIMA (2,0.77,3)**	**ARIMA (3,0.77,2)**
MAPE	4.1755%	4.1755%
MAD	5.9477	5.9475
MSE	73.053	73.059

The robustness of forecasts made by the two models is nearly equal in all measurements. Hence, both models equally qualify for the forecasting purposes and a clear separation of adequacy between the two models could not be obtained. Hence, the both models can be used for the forecasting purposes.

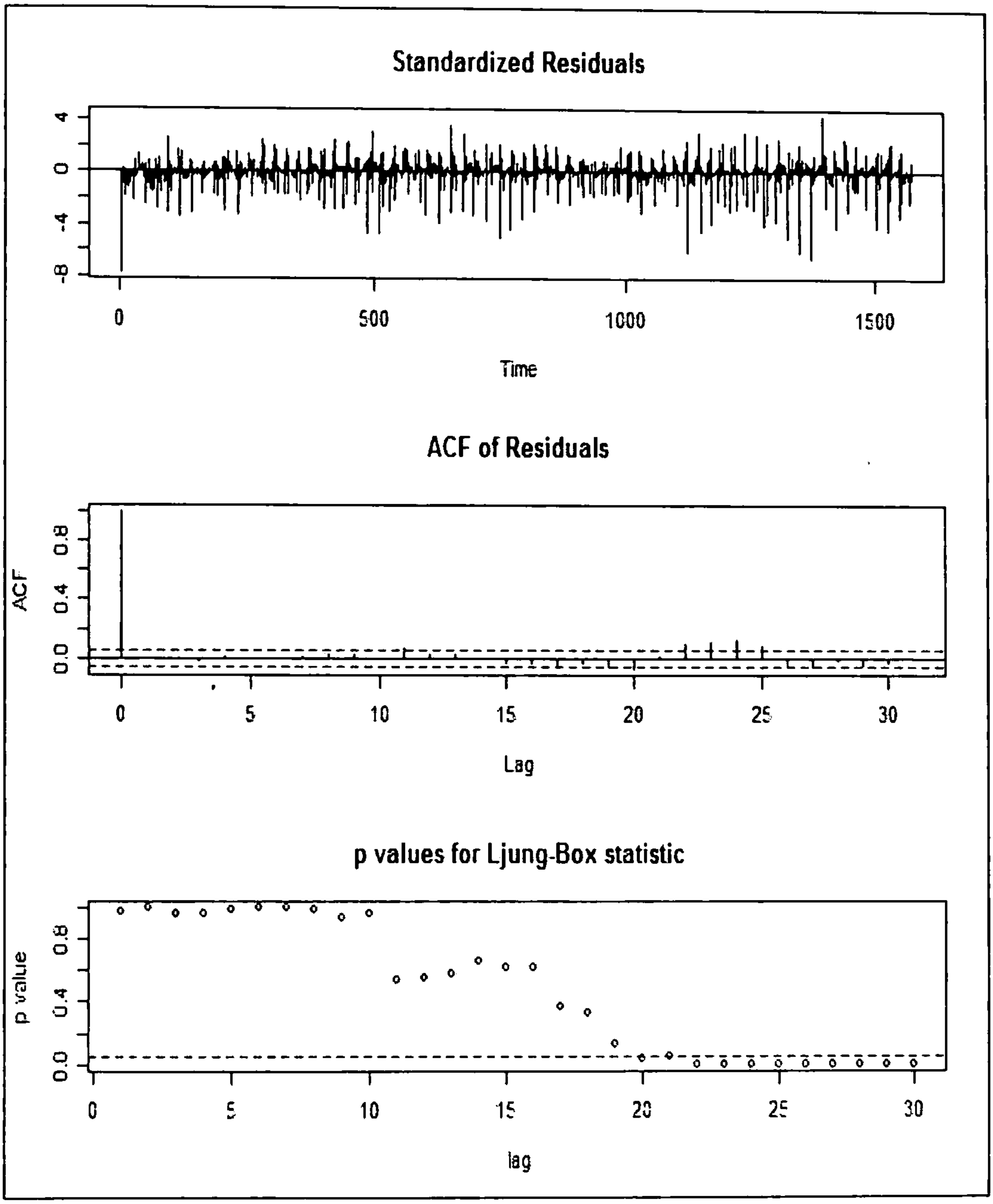

Figure 1: Summary residuals plots of the ARFIMA(2,0.77,3) model

CONCLUSIONS AND RECOMMENDATIONS

This study revealed that the OPEC daily spot price series exhibits a fractionally integrated process (d=0.77) during the study period. Hence, there is a temporal dependence between the spot oil prices of the OPEC even at distance observations and it is mean reverting even though it is not covariance stationary. Therefore, for the OPEC daily spot series ARFIMA (2,0.77,3) and ARFIMA (3,0.77,2) were equally identified as adequate fractional integrated processes as the forecasting models based on the composite results of the model adequacy and the robustness of forecasting.

REFERENCES

Abosedra, S. (2005). Futures versus univariate forecast of crude oil prices. *OPEC Review*, 29(4): 231-241.

Baillie, R.T. and T. Bollerslev (1994). The long memory of the forward premium. *Journal of International Money and Finance*, 13: 565–571.

Barkoulas, J.T. and C.F. Baum (1997). Long memory and forecasting in Euroyen deposit rates. *Financial Engineering and the Japanese Markets*, 4: 189-201.

Barkoulas, J.T. and C.F. Baum (1998). Fractional dynamics in Japanese financial time series. *Pacific-Basin Finance Journal*, 6: 115-124.

Box, G.E., G.M. Jenkins and G.C. Reinsel (1994). *Time Series Analysis Forecasting and Control* (3rd ed.). Dorling Kindersley, New Delhi, India.

Caporale, G.M. and L.A. Gil-Alana (2004). Fractional cointegration and real exchange rates. *Review of Financial Economics*, 13: 327-340.

Cheung, Y-W. and K.S. Lai (1993). A fractional co-integration analysis of purchasing power parity. *Journal of Business & Economic Statistics*, 11: 103-112.

Cuaresma, J.C., A. Jumah, and S. Karbuz (2007). Modelling and forecasting oil prices: The role of Asymmetric Cycles. Working Paper, Faculty of Economics and Statistics, University oflnnsbruck; <http://www.uibk.ac.at/fakultaeten/volkswirtschaftundstatistk/forschung/wopec/repec/inn/wpaper/2007-22.pdf>

Diebold, F.X. and G.D. Rudebusch (1989). Long memory and persistence in aggregate output. *Journal of Monetary Economics*, 24: 189-209.

Geweke, J. and S. Porter-Hudak (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis*, 4: 221–238.

Gil-Alana, L.A. (2001). A fractionally integrated model with a mean shift for the US and the UK real oil prices. *Economic Modelling,* 18: 643–658.

Granger, C.W.J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, 14: 227–238.

Granger, C.W.J. and R. Joyeux (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1: 15–39.

Hosking, J.R.M. (1981). Fractional differencing. *Biometrika*, 68: 165–176.

Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54: 159–178.

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

Robinson, P. (1995). Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 13: 1048–1072.