Investment Knowledge
Vol.2, Issue.7, Autumn 2013

ARIMA and ARFIMA Prediction of Persian Gulf Gas-Oil F.O.B

H. Amadeh
Assistant Professor of Allame Tabatabaee University

A. Amini
M.A. of Energy Economics, Allame Tabatabaee University (Corresponding Author)
A.amini68@gmail.com

F. Effati
M.A. of Energy Economics, Allame Tabatabaee University

Receipt: 10 , 6 , 2013 Acceptance: 16 , 8 , 2013

Abstract
Gas-oil is one of the most important energy carriers and the changes in its prices could have significant effects in economic decisions. The price of this carrier should not be more than 90 percent of F.O.B price of Persian Gulf, legislated in subsidizes regulation law in Iran. Time series models have been used to forecast various phenomena in many fields. In this paper we fit time series models to forecast the weekly gas-oil prices using ARIMA and ARFIMA models and make predictions of each category. Data used in this paperstarted with the first week of the year 2009 until the first week of 2012 for fitting the model and the second week of 2012 until 13th week of 2012 for predicting the values, are extracted from the OPEC website. Our results indicate that the ARFIMA(0.0.-19,1) model appear to be the better model than ARIMA(1,1,0)and the error criterions RMSE, MSE and MAPE for the forecasted amounts is given after the predictions, respectively

Keywords: Gas-oil, ARIMA, ARFIMA, F.O.B price of Persian Gulf, Prediction

1- Introduction

Energy is a strategic commodity in international levels. Every economic policy in this field could have direct and indirect effects on energy supply and demand. These results affects price and productions of other commodities even social welfare.

Gas-oil is one of the most important energy carriers that its price and consumption has essential effect on the other economics sections such as agriculture, transportation and power generation units which gas-oil is one of the base fuels for their consumption.

Between years 1974-2007 except the war duration, gas-oil consumption in Iran had a high growth trend, 17.5 million liter per day in 1974 consumption increased to 87.3 million liter per day in 2007. These statistics indicates the average growth rate of 2.45 per year in gas-oil consumption. Between 1998 and 2004, domestic production satisfied consumptions, but with increasing level of domestic consumption from 2005, it has begun to be imported from other countries. Growing rate of demand in fuel consumer industries and low energy prices, caused higher level of gas-oil consumption and higher level of importing values.

Growing trend of gas-oil consumption and smuggling that was originally caused by its subsidized low level prices, made governments to improve subsidized regulation law. In this law, considering the increase of gas-oil prices, it is legislated that increasing level of its prices should not be more than 90 percent of Persian Gulf F.O.B, so Persian Gulf F.O.B is a benchmark pricing for domestic production. With respect to necessity of increasing prices in future, knowing the gas-oil upcoming prices has significant importance. Our aim in this paper is to predict future trend of its prices. Short-run prediction would make it easier for the country to decide the efficient prices, as it has important effects on transportation and agriculture.

In this study, the data accumulated and recorded weekly from the site of OPEC from the first week of 2009 up to second week of 2012. We have used the STATA12 software to analyze the data.

Time series forecasting is one of the most important types of quantitative models in which past observations of same variable are collected and analyzed to develop a model describing the underlying relationship (Aryal & Wang, 2003), This modeling approach is particularly useful when little knowledge is available on the underlying data generating process or when there is no satisfactory explanatory model that relates the prediction variable to other explanatory variables (Zhang, 2003). Forecasting procedures include different techniques and models. Moving averages techniques, random walks and trend models, exponential smoothing, state space modeling, multivariate methods, vector autoregressive models, co-integrated and casual models, method based on neural, fuzzy networks or data mining and rule-based techniques are typical models used in time series forecasting (Ragulskis & Lukoseviciute, 2009).Auto-regressive integrated moving average (ARIMA) models are one of the most important and widely used linear time series models. The popularity of ARIMA model is due to its statistical properties as well as the well-known Box-Jenkins methodology (Box & Jenkins, 1976) in the model building process. Although ARIMA models are quite flexible in that they can represent several different types of time series and also have the advantages of accurate forecasting over o short period of time and ease of implementation, their major limitation is the pre-assumed linear form of model. ARIMA models assume that future value of a time series have a linear relationship with current and past values as well as with white noise, so approximations with ARIMA models may not be adequate for complex nonlinear real-world problems. However many researchers have argued the real world systems are often nonlinear (Zhang *et al.*, 1998). These evidences have encouraged academic researchers and business practitioners in order to develop more predictable forecasting models that linear models (Khashei & Bijari, 2011).

In recent years, studies about long memory have received the attention of statisticians and mathematicians. This phenomenon has grown rapidly and can be found in many fields such as hydrology, chemistry, physics, economic and finance (Boutahar & Khalfaoui, 2011). Models for long memory in mean were first introduced by Granger and Joyeux (1980) and Hosking (1981), following the seminal work of Hurst (1951). The important characteristic of an Autoregressive Fractionally Integrated Moving Average (ARFIMA) process is its autocorrelation function decay rate. In an ARFIMA process, the autocorrelation function exhibits a hyperbolic decay rate, differently from an ARMA model which presents a geometric rate. Long memory in mean has been observed in data from areas such as meteorology, astronomy, hydrology, and economics, as reported in Beran (1994). So in this paper we used both ARIMA and ARFIMA process in order to compare results to understand which models fit the gas-oil market.

2- Literature Review

In Iran, lots of papers published using neural network in order to forecast gas-oil prices but still no ARFIMA process has been made. Furthermore ARIMA models has a few share of the researches for gas-oil market.Al-Fattah in his paper "Time Series Modeling for U.S. Natural Gas Forecasting"presented one methodology for developing forecasting models for predicting U.S. natural gas production, proved reserves, and annual depletion to year 2025 using a stochastic (time series) modeling approach. The methodology is not mechanistic. A mechanistic model would examine individual geologic settings, exploration success, and the physics of gas production and the rate of exploitation for provinces, basins, and reservoirs. However, to do so would result in an extraordinarily massive model that would be difficult, if not impossible, to develop and use. Instead they used a simpler approach which takes advantage of established trends in easily obtained published data. Having adequately validated these time series models using historical data they believed that they can be used to make at least short time forecasts. Comparison of results of this study with other published forecast is also presented (S.M. Al-Fattah, 2005).Babatunde J. Ayeni and Richard Pilat in their paper "Crude oil reserve estimation: An application of the autoregressive integrated moving average (ARIMA) model"explored the possibility of using the Autoregressive Integrated Moving Average (ARIMA) technique in forecasting and estimating crude oil reserves. The authors compare this approach with the traditional decline method using real oil production data from twelve (12) oil wells in South Louisiana. The Box and Jenkins methodology is used to develop forecast functions for the twelve wells under study. These forecast functions are used to predict future oil productions. The forecast values generated, are then used to determine the remaining crude oil reserves for each well (Ayeni & Pilat, 1992).John Elderaand Apostolos Serletis extended the work in Serletis [Serletis, A. (1992). Unit root behavior in energy futures prices. The Energy Journal 13, 119–128] by re-examining the empirical evidence for random walk type behavior in energy futures prices. It tests for fractional integrating dynamics in energy futures markets utilizing more recent data (from January 3, 1994 to June 30, 2005) and a new semi-parametric wavelet-based estimator, which is superior to the more prevalent GPH estimator (on the basis of Monte-Carlo evidence). They found new evidence that energy prices display long memory and that the particular form of long memory is anti-persistence, characterized by the variance of each series being dominated by high frequency (low wavelet scale) components(Elder & Serletis, 2008).Zheng Li and others examined automobile petrol demand in Australia. Their paper is motivated by an ongoing need to review the effectiveness of empirical fuel demand forecasting models, with a focus on theoretical as well as practical considerations in the model-building processes of different model forms. They consider a linear trend model, a quadratic trend model, an exponential trend model, a single exponential smoothing model, Holt's linear model, Holt–Winters' model, a partial adjustment model (PAM), and an autoregressive integrated moving average (ARIMA) model. More importantly, the study identifies the difference between forecasts and actual observations of petrol demand in order to identify forecasting accuracy. Given the identified best-forecasting model, Australia's automobile petrol demand from 2007 through to 2020 is presented under the "business-as-usual" scenario(Li, Rose, & Hensher, 2010).Jose Alvarez-Ramirez and others determined short-term predictability of crude oil markets. They analyzed the auto-correlations of international crude oil prices on the basis of the estimation of the Hurst exponent dynamics for returns over the period from 1987 to 2007. In doing so, a model-free statistical approach—detrended fluctuation analysis—that reduces the effects of non-stationary market trends and focuses on the intrinsic auto-correlation structure of market fluctuations over different time horizons, is used. Tests for time variations of the Hurst exponent indicate that over long horizons the crude oil market is consistent with the efficient market hypothesis. However, meaningful auto-correlations cannot be excluded for time horizons smaller than one month where the Hurst exponent manifests cyclic, non-periodic dynamics. This means that the market exhibits a time-varying short-term inefficient behavior that becomes efficient in the long term(Alvarez-Ramirez, Alvarez, & Rodriguez, 2008).

3- Methodology

*ARIMA Process:*To model a given time series with the ARMA process, the series must be stationary. This means that both the expected values of the series and its auto-covariance function are independent of time. In addition, the series must have stabilized variance and constant mean. Most time series are non-stationary but some can be transformed to a stationary series by differencing. This process is often used to remove the trend, seasonality, and periodic variations of the series, thus rendering the non-stationary time series stationary. The differenced time series, can then be analyzed and modeled like any other stationary time series.

After modeling the differenced time series the output series is transformed back to the original raw data by reversing the order of differencing. An ARIMA model predicts a value in a response time series as a linear combination of its own past values, past errors, and current and past values of other time series. The order of an ARIMA model is usually denoted by the notation ARIMA (*p*, *d*, *q*), where *p* is the order of the autoregressive component, *d* the order of the differencing, and *q* the order of the moving-average process. Mathematically, the ARIMA model is written as**:**

$$\Phi\ (\mathrm{L})(1-\mathrm{L})^{\mathrm{d}}\mathrm{x}_{\mathrm{t}} = \Theta\ (\mathrm{L})\varepsilon_{\mathrm{t}} \quad \textbf{(1)}$$

Where

L*:* the backshift operator (i.e. $L.x_t = x_{t-1}$),

Φ (L): the autoregressive operator, represented as a polynomial in the backshift operator:

$\Phi\ (L) = (1 - \varphi\ ^1L_1 - \varphi\ ^2L_2 - \ldots - \varphi_pL^P)$

Θ (L): the moving-average operator, represented as a polynomial in the backshift operator:

$\Theta\ (L) = (1 + \theta_1L^1 + \theta_2L^2 + \ldots + \theta_pL^P)$

ε_t: the random error

Auto-Correlation Function (ACF): There are two phases to the identification of an appropriate Box - Jenkins model: changing the data if necessary into a stationary time series and determining the tentative model by observing thebehavior of the autocorrelation and partial autocorrelation function. A stationary time series is that it does not contain trend, that is, it fluctuates around a constant mean. Box and Jenkins suggest the number of Lag to be no more than (n⁄4) autocorrelations; the autocorrelation coefficient measures the correlation between a set of observations and a lagged set of observation in a time series. The autocorrelation between x_t and x_{t+k} measures the correlation between pair (x_1, x_{1+k}), (x_2, x_{2+k}),…, (x_n, x_{n+k}). The sample autocorrelation coefficient r_k is an estimate of ρ_k where:

$$r_k = \frac{\sum(x_t-\bar{x})(x_{t+k}-\bar{x})}{\sum(x_t-\bar{x})^2} \qquad \textbf{(2)}$$

With

x_t: The data from the stationary time series

x_{t+k}: The data from k time period ahead of t.

$\bar{x}$: The mean of the stationary time series

Partial Auto-Correlation Function (PACF): The estimated partial autocorrelation function PACF is used as a guide, along with the estimated autocorrelation function ACF, in choosing one or more ARIMA models that might fit the available data. The idea of partial autocorrelation analysis is that we want to measure how $\widehat{x_t}$ and $\widehat{x_{t+k}}$are related. The equation that gives a good estimate of the partial autocorrelation is:

$$\hat{\varphi}_{11} = r_1 \qquad \hat{\varphi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\varphi}_{k-1,j} r_j} \qquad k = 2,3, \ldots \quad \textbf{(3)}$$

Where

$$\hat{\varphi}_{kj} = r_{k-1,j} - \hat{\varphi}_{kk}\hat{\varphi}_{k-1,k-j}, \; k = 3, 4, \ldots ; j = 1, 2, \ldots, k-1 \quad \textbf{(4)}$$

ARFIMA Process (Stationary and invertible ARFIMA process):Let Xt be an ARFIMA (p, d, q) process given by

$$\Phi\,(L)(1-L)^d x_t = \Theta\,(L)\varepsilon_t \quad \textbf{(5)}$$

The process ε_t is white noise with zero mean and finite variance σ_ε^2. The term $(1 - L)^d$ is the binomial power series of L.
The process X_t, given by the expression **(5)**, is called a general fractional differenced zero mean process, where d is the fractional differencing parameter.The process given by the expression **(5)** is both stationary, and invertible if the roots of Φ (L) and Θ (L) are outside the unit circle, and d ∈(−0.5, 0.5). The ARFIMA (p, d, q) process exhibits long memory when d ∈(0.0, 0.5), intermediate memory when d ∈(−0.5, 0.0), and short memory when d = 0 (B. P. Olbermann et al).

ARFIMA Process (Non-stationary ARFIMA process): Now, we define the process **(5)** with the parameter $d^* = d + 1$, where d ∈ (0.0, 0.5), and the model **(5)** becomes

$$\Phi\,(L)(1-L)^{d*} x_t = \Theta\,(L)\varepsilon_t \textbf{(6)}$$

The process **(6)** is non-stationary when $d* \geq 0.5$; however, it is still persistent.For $d^* \in [0.5, 1.0)$ it is level-reverting in the sense that there is no longrun impact of an innovation on the value of the process (Velasco, 1999). The level-reversion property no longer holds when$d^{*} \geq 1$.

*Long Memory Analysis (The Modified Rescaled Range (R/S) Analysis):*The first test for long memory was used by the hydrologist Hurst (1951) for the design of an optimal reservoir for the Nile River, of where flow regimes were persistent. Hurst gave the following formula:

$$(R/S)_n = cn^H \quad \textbf{(7)}$$

$(R/S)_n$is the rescaled range statistic measured over a time index n, c is a constant and H the Hurst exponent. This shows the how the R/S statistic is scaling in time. The aim of the R/S statistic is to estimate the Hurst exponent which can characterize a series. Estimation of Hurst exponent can be done by transforming **(7)** to:

$$\log{(R/S)_n} = \log c + H.\log(n) \textbf{(8)}$$

And H can be estimated as the slope of log/log plot of$(R/S)_n$vs. n.For a time series Xt (t = 1… N), the R/S statistic can be defined as the range of cumulative deviations from the mean of the series, rescaled by the standard deviation.Although Mandelbrot (1972) gave a formal justification for the use of this test, Lo (1991) showed that this statistic was not robust to short memory dependence and modified this statistic. Lo defined modified R/S statistic as:

$$(R/S)_n = \frac{\left[\underset{0\le k\le n}{\text{Max}} \sum_{t=1}^{k}(X_t-\bar{X}_n) - \underset{0\le k\le n}{\text{Min}} \sum_{t=1}^{k}(X_t-\bar{X}_n)\right]}{\sigma(n)} \textbf{(9)}$$

Where

$$\sigma_n^2(q) = \sigma_x^2(q) + \frac{2}{n}\sum_{j=1}^{q} W_j(q)\left[\sum_{i=j+1}^{n}(x_i - \bar{x}_n)\left(x_{i-j} - \bar{x}_n\right)\right] \textbf{(10)}$$

If q = 0, Lo’s statistic reduces to Hurst’s R/S statistic. This statistic is highly sensitive to the order of truncation q but there is no a statistical criteria for choosing q in the framework of this statistic. If q is too small, this statistic does not account for the autocorrelation of the process, while if q is too large, it accounts for any form of autocorrelation and the power of this test tends to its size. Given that the power of a useful test should be greater than its size; this statistic is not very helpful. For that reason, Teverovsky et al. (1999) suggest to use this statistic with other tests. Since there is no data driven guidance for the choice of this parameter, the default values for q = 2, 4, 6, 8, 10 are considered. At 5% significance level, the null hypothesis of no long memory process is rejected if the modified R/S statistic does not fall within the confidence interval [0.809, 1.862].

Box-Jenkins Stages

Box and Jenkins propose a practical three-stage procedure for finding a good model. The three- stage Univariate Box-Jenkins (UBJ) procedure is summarized schematically in details are:

Stage 1: Identification: At the identification stage we use two graphical devices to measure the correlation between the observations within a single data series. These devices are called an estimated auto-correlation function (ACF) and an estimated partial auto-correlation function (PACF). The estimated ACF and PACF measure the statistical relationships within a data series in a somewhat crude (statistically inefficient) way. The next step at the identification stage is to summarize the statistical relationships within the data series in a more compact way than is done by the estimated ACF and PACF. We use the estimated ACF and PACF as guides to choosing one or more ARIMA models that seem appropriate.

Stage 2: Estimation: At this stage, we get precise estimates of the coefficients of the model chosen at the identification stage. We fit this model to the available data series to get estimates. This stage provides some warning signals about the adequacy of our model. In particular, if the estimated coefficients do not satisfy certain mathematical inequality conditions, that model is rejected.

Stage 3: Diagnostic checking: Box and Jenkins suggest some diagnostic checks to helpdetermine if an estimated model is statistically adequate. The results at this stage may also indicate how a model could be improved. This leads us back to the identification stage. We repeat the cycle of identification, estimation, and diagnostic checking until we find a good final model.

In order to compare obtained ARIMA and ARFIMA models,the criteria chosen to measure the accuracy of the forecast in this study are the mean absolute error (MAE), the root mean square error (RMSE) and the mean squarederror(MSE). These criteria are computed as:

$$MSE = \frac{\sum_{i=1}^{n}(x_i - z_i)^2}{n} \quad, RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - z_i)^2}{n}} \quad, MAE = \frac{\sum_{i=1}^{n}|x_i - z_i|}{n}$$

4- Finding

Data used in this research from the first week of 2009 up to second week of 2012 is shown in figure (1).

Figure 1- Data

As argued above to model a given time series with an ARIMA process, the series must be stationary, so we used unit root test to examine the possibility of non-stationary model.

*Unit Root Test:*A unit root test determine whether a time series variable is non-stationary using an autoregressivemodel. One of the most famous tests is the augmented Dickey- Fuller test. This test used the existenceof a unit root as the null hypothesis. In this study, the Augmented Dickey-Fuller (ADF) test is proposed to examine the stationarity (unit root) of the gas-oil prices of Persian Gulf, Table (1) shows the ADF test for F.O.B of Persian Gulf gas-oil.

Table 1 - Unit Root Test Result

Unit Root Test	Statistic	Result
ADF	-1.155	Significant in 5 and 10 percent critical value

Results strongly confirm at the standard 5% significance level, weekly prices arenot stationary in levels. So we transformed series into its first differenced series and examined it again with unit root test. Table (2) shows the ADF test for first differenced gas-oil prices.

Table 2- Unit Root Test Result

Unit Root Test	Statistic	Result
ADF	-9.871	H_0: series has unit root rejected

Results suggest that the series is stationaryat 5% significance level and the possibility of null hypothesis (H_0: series has a unit root) is rejected.

Autocorrelation and Partial Autocorrelation

Autocorrelation (ACF) is one of the major tools in time series modeling (as guidance in choosing terms to include in an ARIMA model). The partial autocorrelation function (PACF) is also one of the major tools in time series modeling (as guidance in choosing terms to include in an ARIMA model).

ACF and PACF for Persian Gulf Gas-Oil Prices

Figure (2) shows Auto-Correlation Function forgas-oil prices in Persian Gulf. Autocorrelations are computed for 20 lags using Equation (2).

The ACF shows a large positivesignificant spike at lag 1 (this means that the autocorrelation of the successive pairs of observations within 1 time period is not within sampling error of zero). All of the other autocorrelations (for lags 2 to 20) are within the 95% confidence limits. This pattern is typical to autoregressive (AR) process of order one.

Figure (3) shows the partial autocorrelation whichare computed using Equation (3). This figure shows that PACF has a large positive significant spike at lag 1 (this means that the partial autocorrelation of the successive pairs of observations within 1 time period is not within sampling error of zero. All the other partial autocorrelations (for lags 2 to 20) are within the 95% confidence limits.

Considering the ACF and PACF diagrams in addition to AIC and BIC criterion, the best value for the MA and AR are determined. Also the diagram shows only the first data has a root out of the cycle, which would be the degree of 1 for AR process. In order to this criterions, ARIMA (1,1,0) would better explain price changes in gas-oil markets. But as there were this possibility of existing better models to fit the data, other possible models were examined by AIC and BIC criterions. ARIMA (1,1,0) had the lowest criterions, so we used this model to fit the time series. Results are shown in table (3).

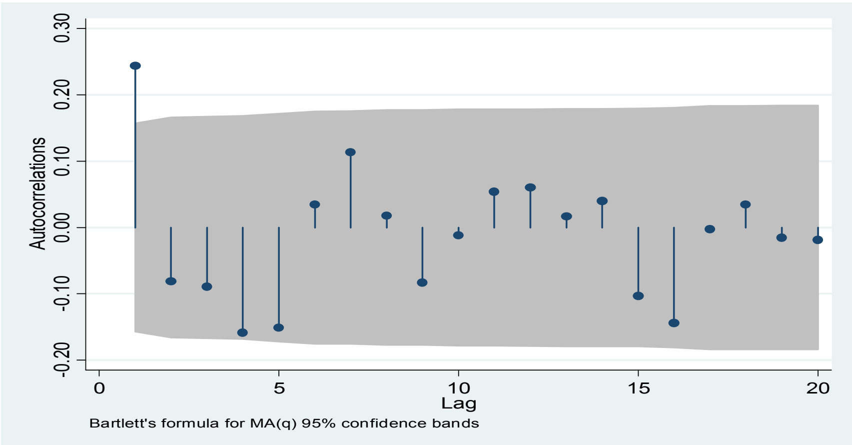

Figure 2- ACF

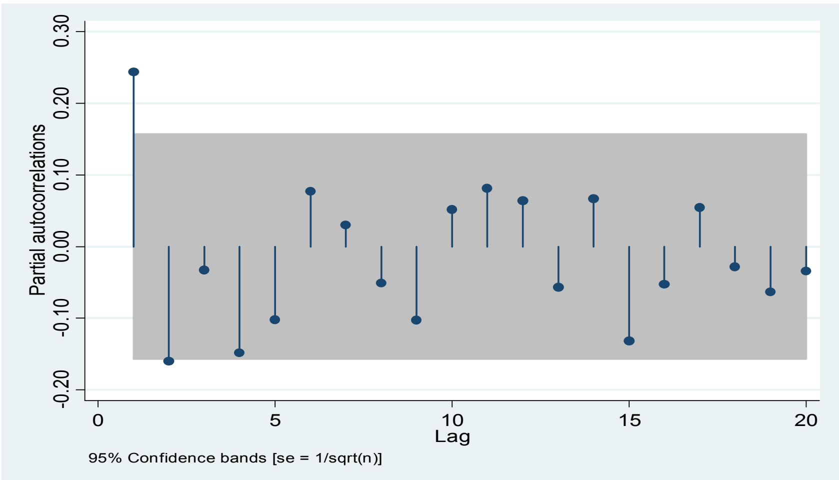

Figure 3-PACF

Table 3-ARIMA Estimation

	Coefficient	Z statistic	Probability
Constant	0.48	1.53	0.1
AR(1)	0.25	3.28	0
TESTS			
Portmanteau (Q)		AIC	BIC
33.75	Rejected with 0.75 prob.	797	806

After estimating AR and MA degree of the model, we checked the white noise process of the residuals. We used Portmanteau test. The results indicated that our ARIMA process has no auto-correlation between its residuals.

The modified rescaled range analysis has been made, and the result indicated that our series has long memory. As we checked above for stationary situation of the model, and we found the model non-stationary, we know d is more than 0.5, so we differed the series and used the modified rescaled range analysis again. The result showed that our differed time series has long memory in order to its statistic (1.76) that is significant in 10 percent criterion values with maximum of 45 lags. Now as we know our series has long memory, we can use ARFIMA process for fitting model. For this purpose we used Stata12 to estimate each parameters of the process.

Table 4- ARFIMA Estimation

	Coefficient	Z statistic	Probability
Constant	0.48	3.31	0
MA(1)	0.44	5.18	0
d	-0.19	-2.18	0
TESTS			
Portmanteau (Q)		AIC	BIC
27.83	Rejected with 0.92 prob.	792	804

Results implicates that ARFIMA (0.-0.19,1) is a better model with lower AIC and BIC respect to other possible ARFIMA processes. Q test was made in order to check correlation between residuals, and rejected the possibility of correlated residuals with 92 percent of probability.

Forecasting

As we estimated all the parameters we need, now we can predict future prices via these two model and compare the results. Predicted values in these models and actual prices for 9 period from 45th week of 2011 until first week of 2012, which is given in the table (5), are in-sample prediction and from second week of 2012 until 13th week of 2012 are used for out-of-sample prediction. The result of comparison between these two models are shown in the diagram (4).

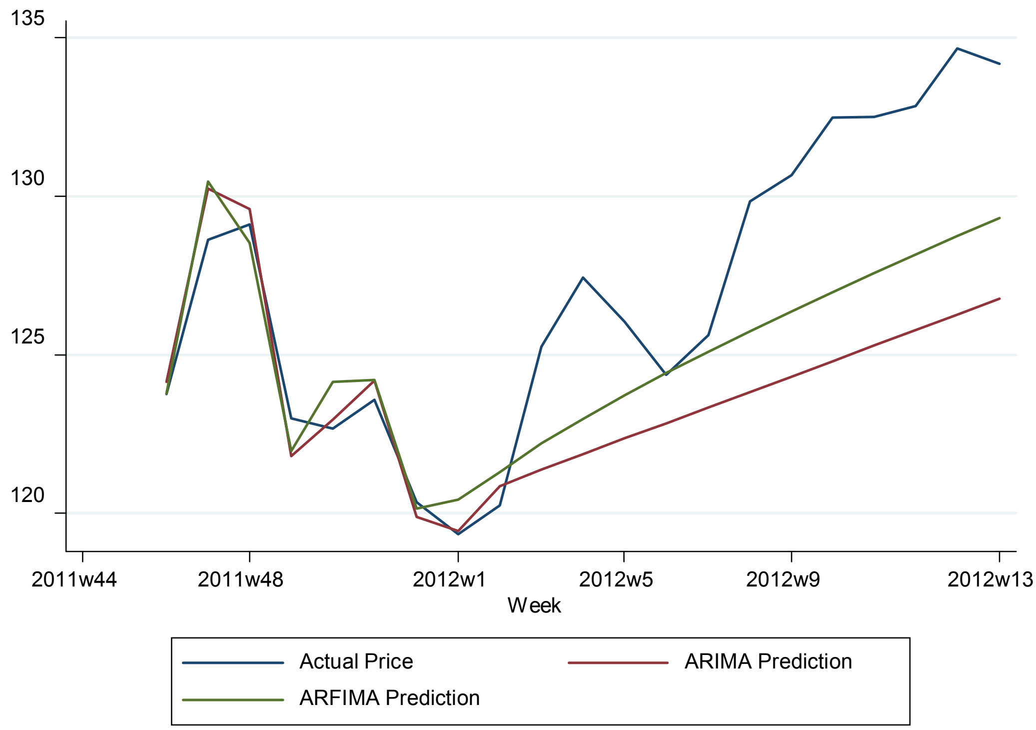

Figure 4- Predicted Values

Table 5- Predicted Values

Sample	Date	ARIMA Prediction	Actual Prices	ARFIMA Prediction
In-Sample	Week 45 year 2011	124.15	123.77	123.78
	Week 46 year 2011	130.25	128.64	130.47
	Week 47 year 2011	129.6	129.11	128.54
	Week 48 year 2011	121.79	122.99	121.96
	Week 49 year 2011	122.95	122.67	124.15
	Week 50 year 2011	124.19	123.59	124.21
	Week 51 year 2011	119.87	120.34	120.13
	Week 52 year 2011	119.42	119.32	120.42
	Week 1 year 2012	120.84	120.24	121.27
Out-of-Sample	Week 2 year 2012	121.36	125.26	122.19
	Week 3 year 2012	121.86	127.43	122.99
	Week 4 year 2012	122.35	126.05	123.73
	Week 5 year 2012	122.84	124.37	124.42
	Week 6 year 2012	123.33	125.62	125.09
	Week 7 year 2012	123.82	129.84	125.74
	Week 8 year 2012	124.31	130.66	126.36
	Week 9 year 2012	124.8	132.48	126.97
	Week 10 year 2012	125.29	132.51	127.57
	Week 11 year 2012	125.79	132.85	128.16
	Week 12 year 2012	126.28	132.67	128.74
	Week 13 year 2012	126.77	134.17	129.32

RMSE, MSE and MAE criteria has been computed and is shown in table(6).

Table 6- RMSE, MSE and MAE Criterions

	RMSE	MSE	MAE
ARIMA	1.82	3.33	1.078
ARFIMA	1.54	2.37	1.054

All these criteria indicate that ARFIMA has better capability for predicting future prices of gas-oil market.

5- Conclusion

In this study, we tried to identify the best ARIMA and ARFIMAmodels for forecasting gas-oil prices of Persian Gulf. For this, we calculated the in-sample and out-of-sample forecasts of the prices and evaluated the performance of the ARFIMA and ARIMA models in terms of their ability to capture best fitted value of prediction.

The estimation results suggest that the ARFIMA model can better predict the market trend than ARIMA process, indicating that gas-oil market has long memory property. The presence of long-memory properties casts doubt on the weak efficiency of ARIMA prediction of gas-oil markets.

The analyses implicate that ARIMA(1,1,0) and ARFIMA(0,-.19,1) are the best fitted models which ARFIMA has better capability of forecasting with respect to MSE, RMSE and MAPE criterions. This suggests that government should be careful when measuring prices in gas-oil markets. The findings of this study should be useful in facilitating accurate price management, developing pricing models, and determining bestdecision with respect to gas-oil market of Persian Gulf.

References

- Mishra, P. (2012). "Forecasting Natural Gas Price-Time Series and Nonparametric Approach." Lecture Notes in Engineering and Computer Science 2197".
- Saud M. Al-Fattah (2006), Time Series Modeling for U. S. Natural Gas Forecasting, E-Journal of Petroleum Management and Economics, ISSN: 1718-7559.
- Box, G. E. P.., and G. M. Jenkins. 1970. Time series analysis: forecasting and control. Holden Day, San Francisco, CA.
- Ayeni, B. J. and Pilat, R. (1992), Crude Oil Reserve Estimation: An Application of the Autoregressive Integrated Moving Average (ARIMA) Model, Journal of Petroleum Science and Engineering, Vol 8 (1), pp 13-28.
- Olanrewaju I. Shittu, Olaoluwa Simon Yaya (2009). Measuring Forecast Performance of ARMA and ARFIMA Models: An Application to US Dollar/UK Pound Foreign Exchange Rate. European Journal of Scientific Research.

- V. Reisen, B. Abraham (1998). Prediction of Long Memory Time Series Models: A Simulation Study and an Application. Staff Paper. Unknown Publisher.
- Nesrin Alptekin. Long Memory Analysis of USD/TRL Exchange Rate, International Journal of Social and Human Sciences 1 2007.
- Jurgen A. Doornik, Marius Ooms (2001). Computational Aspects of Maximum Likelihood Estimation of Autoregressive Fractionally Integrated Moving Average models, November 29, 2001.
- Lopes, Olbermann and Reisen (2002). Non-stationary Gaussian ARFIMA Processes: Estimation and Application, September 30, 2002.
- James H. Stock and Mark W. Watson (2003). STATA Tutorial to accompany Stock/Watson Introduction to Econometrics.
- Mahendran Shitan, Pauline Mah Jin Wee, Lim Ying Chin, Lim Ying Siew, (2008). Arima and Integrated Arfima Models for Forecasting Annual Demersal and Pelagic Marine Fish Production in Malaysia, Malaysian Journal of Mathematical Sciences 2(2): 41-54 (2008).

▪ Alvarez-Ramirez, Jose, Alvarez, Jesus, & Rodriguez, Eduardo. (2008). Short-term predictability of crude oil markets: A detrended fluctuation analysis approach. Energy Economics, 30(5), 2645-2656.

▪ Ayeni, Babatunde J, & Pilat, Richard. (1992). Crude oil reserve estimation: an application of the autoregressive integrated moving average (ARIMA) model. Journal of Petroleum Science and Engineering, 8(1), 13-28.

▪ Elder, John, & Serletis, Apostolos. (2008). Long memory in energy futures prices. Review of Financial Economics, 17(2), 146-155.

▪ Li, Zheng, Rose, John M, & Hensher, David A. (2010). Forecasting automobile petrol demand in Australia: an evaluation of empirical models. Transportation Research Part A: Policy and Practice, 44(1), 16-38.

▪ S.M. Al-Fattah, SPE, Saudi Aramco. (2005). Time Series Modeling for U.S. Natural Gas Forecasting. International Petroleum Technology Conference.

▪ Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. Journal of Econometrics. 74, 3–30.

▪ Dickey, D. and Fuller, W. (1979). Distribution of the estimators for autoregressive time series with unit root. Journal of the American Statistical Association. 74, 427-431.

- Geweke, J. and Porter-Hudak, S. (1983). The Estimation and Application of Long Memory Time Series Models. Journal of Time Series Analysis. 4, 221-238.
- Granger, C. W. J. and Joyeux. R. (1980). An introduction to long-memory time series models and fractional differencing. Journal of Time Series Analysis. 1, 15-29.