INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ACCOUNTING, FINANCE AND MANAGEMENT SCIENCES

HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY
EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2225-8329

Prediction of Oil Price using ARMA Method for Years 2003 to 2011

Abdalali Monsef, Amir Hortmani, Khadijeh Hamzeh

To Link this Article: http://dx.doi.org/10.6007/IJARAFMS/v3-i1/10282 DOI:10.6007/IJARAFMS /v3-i1/10282

Received: *14* January *2013,* ***Revised:*** *19* February 2013*,* ***Accepted:*** *05* June 2013

Published Online: 21 March 2013

In-Text Citation: (Monsef et al., 2013)

To Cite this Article: Monsef, A., Hortmani, A., & Hamzeh, K. (2013). Prediction of Oil Price using ARMA Method for Years 2003 to 2011. *International Journal of Academic Research in Accounting Finance and Management Sciences*, *3*(1), 235–247.

Vol. 3, No. 1, 2013, Pg. 235 - 247

http://hrmars.com/index.php/pages/detail/IJARAFMS	JOURNAL HOMEPAGE

INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN ACCOUNTING, FINANCE AND MANAGEMENT SCIENCES
HRMARS MANAGEMENT ACADEMIC RESEARCH SOCIETY
EXPLORING INTELLECTUAL CAPITAL
www.hrmars.com
ISSN: 2225-8329

Prediction of Oil Price using ARMA Method for Years 2003 to 2011

Abdalali Monsef[1], Amir Hortmani[2], Khadijeh Hamzeh[3]
[1]Department of Economics, Payame Noor University, Iran, [2]Islamic Azad University, Dehaghan Branch, Iran, [3]Isfahan Payame Noor University, Iran
Email: monsefali@yahoo.com, Amir_hortamani@yahoo.com, Pnu_alm@yahoo.com

Abstract
One main purpose of economic analysis is prediction of economic variables. For this reason, various methods have been developed in this context. One important challenge in prediction of time series is precise prediction without any computational complexities. It is usually assumed that autoregressive moving-average (ARMA) models have this ability with a high accuracy. In this study, ARMA method has been used to predict time series of oil price. To determine the order of this process Akaike and Schwartz's Bayesian criteria have been used. The results show that the best answers are obtained by ARMA (1, 0) model for static predictions and ARMA (3, 2) model for dynamic predictions.
Keywords: Prediction, Time Series Models, ARMA, Oil Price

Introduction
Predicting or foreseeing future conditions and accidents as a key element in risk management and control decision makings is very important for many organizations and institutions. In other words, every organization needs to predict its future situation to make conscious decisions because final efficiency of each decision depends on the nature of a sequence of accidents which occurs subsequent to the decision. The ability to guess non-controllable aspects of these accidents before making any decision allows a better selection comparing to the cases in which this ability is absent, so management systems typically have a prediction function to plan and control operation of an organization.

On this basis, it seems that awareness about the returns of the investment is essential for investing on different projects. In today's world in which uncertainty is increasing day by day, stepping in the path of non-returnable and costly investment without economic evaluation and estimation of its benefits and disadvantages is an idiotic act. Today investors often try to stimulate future and take a look at the project outlook aiming at minimizing risks due to their investments and increasing their expected benefits. Lack of awareness about outcome of the project and paying no attention to investments returns can lead to irreparable results.

In this context, oil and gas projects are especially important due to their high costs. On the other hand, uncertainty in this kind of projects is rather more because prices of these energy carriers and also execution cost of this kind of projects have always been fluctuating because of its variable and exclusive technology. Therefore, before taking any action for investment on these projects future should be simulated by looking through the pas to reach maximum social benefits by execution of these projects. In other words, the trend in earlier prices of these products in global markets should be studied and based on these prices the future prices can be simulated using special methodologies. In this way, the risk due to investment on these projects can be reduced.

The most important challenge which governments and investor companies encounter is uncertainty about the price of this valuable natural material. This important is doubled when one country bases the majority of its economic on oil and supplies the main part of its budget by direct sale of crude oil.

On the other hand, refineries, investment companies and other active companies in operational oil projects are very interested to beau are oil prices trend and their values in the future.

In this paper, two hypotheses were initially considered:

1) Short term prediction of oil price is feasible by the best ARMA model and its trend in ascending.

2) Evaluation criteria confirmed the accuracy of prediction by the best ARMA model and will predict the oil price well.

The main purpose of this study is presenting the best ARMA model for short-term prediction of oil prices.

Crude Oil Market

As oil market is competitive crude oil prices varies severely, which intensely influences oil transactions patterns. Before oil revolutions in 1970s, oil market had been almost managed by large oil companies, therefore oil contracts mostly were performed under the framework of fixed-term contracts. This pattern had been useful if oil price remained stable because relative stability of oil price allowed producers and consumers to plan better for their production or consumption under fixed-term contracts. In the circumstances that prices change severely; these contracts are not economical because they result in heavy profits or losses (Derakhshan, 17: 2004).

Spot Markets

Firstly, it should be noted that considerable percentage of produced crude oil is directly sold to refineries or other applicants by producers with posted or official prices. This kind of contracts is usually concluded for one month to one year and they can be extendable. Nonetheless, oil companies usually have no accurate estimation of market demand for crude oil, so fixed-term contracts cannot fully fulfill their needs. Using fixed-term contracts causes that oil companies have sometimes oil shortage for sale and sometimes surplus oil. Spot markets which are also called open market can balance surplus and shortage supply of oil companies, i.e. oil companies can sell their surplus supply or buy their oil shortage in spot markets (Derakhshan, 33:2004)

Forward Markets and Futures Markets

Forward markets have been developed for various kinds of crude oil benchmarks such as Brent Blend of North Sea in Europe and WTI in U.S.A and Dubai crude oil in Persian Gulf. To become familiar with forward market, it is enough to note that monthly production of Brent Blend is allocated between the companies to which ownership of North Sea oil belong as cargo lots of 500,000 barrel. However the value 500000 barrels is a computational basis for each cargo lot, this value can changeup to ±10 percent. It is usually planned to load 30 to 40 cargo lots per month. Oil production is monthly allocated to the companies and loading is performed in three days (Derakhshan, 35-36:2004).

In forward transactions of Brent crude oil, the seller gives a buyer a minimum 15 days notice of the intended three dates for loading crude oil cargo during this 15 days. The cargos which are loaded during 15-day period are so called Dated Brent while forward cargos are so called "15-day Brent" (Derakhshan, 36:2004).

Therefore, traders in forward markets for Brent crude oil have a Portefeuille or Portfolio to sell and buy15-day Brent contracts with various prices. These traders can trade on a day basis. Similarly, new buyer can receive the crude oil cargo or transfer its right to third buyer. In this way, buyers can exit the market without receiving crude oil and just by transferring their rights to the others. Traders' profit and loss is determined through "book-outs settlement or briefly settlement (Derakhshan, 36-37:2004).

Crude oil futures market is the completed form of crude oil forward markets because in these markets crude oil is traded in standard units which are so called "contract". The contract trades in units of 1,000 barrels of crude oil whose quality must be standard. Standard crude oil allows traders to be informed about the properties of traded crude oil in Exchange, specially its unspecific gravity (Derakhshan, 37:2004).

Time Series Analysis Methods

Time series patterns which are often used for short-term prediction try to explain the behavior of a variable based on its past values (and possibly other considered variables). These patterns are able to make accurate predictions when the economic pattern has unclear infrastructure. In contrary to econometric patterns which needs statistical data and economic theories, time series patterns just work with statistical data of variables without any need to economic theories. These patterns which relate current values of a variable to its past values are univariate time series models some of which are autoregressive processes, moving average processes, autoregressive moving average processes and autoregressive integrated moving average processes (Nofresti, 7-8:1999).

1. Moving average process: y_t series have moving average process if:

$$y_t = \alpha + \theta(L)\varepsilon_t \tag{1}$$

Where:

$$\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$$
$$\varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \tag{2}$$

Above process is the moving average model of order q which is represented by *MA (q).* It can be easily proved that all moving average processes are stationary.

2. Autoregressive moving average process: this process is a combination of autoregressive and moving average processes and is written:

$$\phi(L)y_t = \alpha + \theta(L)\varepsilon_t \tag{3}$$

Equations 1 and 2 are also valid in the above formula. The same constrains as those in autoregressive processes are needed in order that the model remains stationary. These process are called autoregressive moving average models of order p and q and are represented by *ARMA (p,q).*

3. Autoregressive integrated moving average processes: if a time series shows evidence of non-stationary and non-stationary can be removed by differencing it d time, it can be written as:

$$\phi(L)(1-L)^d y_t = \alpha + \theta(L)\varepsilon_t \tag{4}$$

Where p and q represent the order of autoregressive and moving average processes and d denotes the order of differencing required in order that y_t become stationary. This process is referred by *ARMA (p, d, q)*. Here equations 1 and 2 are also valid. In fact, all processes mentioned above are a special (simple) form of this process. For example if $p = q = d = 0$, a random walk process with an acceleration term is obtained and if $p = 1, d = q = 0$, an autoregressive process of order 1 is obtained (Abasi, 148:2007).

In this paper, below criteria were used to determine p and q;

a) Akaike Information Criterion

$$AIC(p) = T\log\hat{\sigma}_p^2 + 2p \tag{5}$$

b) Schwarz Bayesian Information Criterion

$$SBIC = T\log\hat{\sigma}_p^2 + p\log T \tag{6}$$

c) Hannan-Quinn Information Criterion

$$HQIC(p) = T\log\hat{\sigma}_p^2 + 2p\log\log T \tag{7}$$

The following criteria were also employed for comparing predictions of various models:

a) Mean Squared Error(MSE) or Root Mean Squared Error(RMSE)

$$MSE = \frac{\sum_{t=T+1}^{T+n}(\hat{y}_t - y_t)^2}{n} \tag{8}$$

$$RMSE = \sqrt{\frac{\sum_{t=T+1}^{T+n}(\hat{y}_t - y_t)^2}{n}} \tag{9}$$

b) Mean Absolute Deviation(MAD) or Mean Absolute Percentage Error(MAPE)

$$MAD = \frac{\sum_{t=T+1}^{T+n} |\hat{y}_t - y_t|}{n}$$

$$MAPE = \frac{\sum_{t=T+1}^{T+n} \left| \frac{\hat{y}_t - y_t}{y_t} \right|}{n}$$

(10)

c) Their Inequality Coefficient

$$TIC = \frac{\sqrt{\frac{\sum_{t=T+1}^{T+n} (\hat{y}_t - y_t)^2}{n}}}{\sqrt{\frac{\sum_{t=T+1}^{T+n} \hat{y}_t^{\ 2}}{n}} + \sqrt{\frac{\sum_{t=T+1}^{T+n} y_t^{\ 2}}{n}}}$$

(11)

In all above mentioned criteria, t is the sample size and n is the period in which prediction is made (Abrishami, Mehrara, 81-125:2002).

Finally in order to determine whether performance of the models based on above criteria is statistically significant or not, below criteria were used:

1. F test: if square of first model error is greater than second model, this criteria can be written as:

$$F = \frac{MSE_1}{MSE_2} = \frac{\sum_{i=1}^{H} e_{1i}^2}{\sum_{i=1}^{H} e_{2i}^2}$$

(12)

Under the null hypothesis, i.e. equality of both two square values of prediction error, and below conditions this value has an f-distribution with (H, H) degree of freedom. H is the number of prediction periods.

1) Prediction errors have a normal distribution with a zero mean.
2) Prediction errors have no correlation.
3) Prediction errors are contemporaneously uncorrelated.

But taking into account that in most cases these conditions are not met, two following tests are also recommended:

2. Granger-Newbold test: this test has been suggested assuming that two first conditions are met. Assume that:

$$x_t = e_{it} + e_{2t}, z_t = e_{it} - e_{2t}$$
(13)

If first and second model errors are equal, correlation coefficient between x_t and y_t (r_{xz}) is equal to zero. Therefore, under null hypothesis, equality of prediction errors, below value has t distribution with H-1 degree of freedom:

$$GN = \frac{r_{xz}}{\sqrt{\frac{(1-r_{xz}^2)}{H-1}}}$$

(14)

If this correlation coefficient is positive, the first model has a greater error, otherwise second model has a greater error.

3. Diebold-Mariano test:

When first and second conditions are not met, this method is highly applicable. In addition, other prediction criteria such as absolute percent prediction error (or other functions such sa*g (e*$_i$*)*) can be used in this method. We have:

$$\bar{d} = \frac{1}{H}\sum_{i=1}^{H}[g(e_{1i}) - g(e_{2i})]$$
(15)

Null hypothesis or equality of predictive power is valid in both two models. Therefore, two situations can be recognized:

a) Terms d_i($d_i = [g(e_{1i}) - g(e_{2i})]$) are uncorrelated: consequently below phrase has a t distribution with H-1 degree of freedom:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\gamma_0}{(H-1)}}}$$

(16)

Where γ_0 is sample variance of d_i series.

b) Terms d_i are correlated. In this circumstances, below static's has a t distribution with H-1 degree of freedom under null hypothesis:

$$DM = \frac{\bar{d}}{\sqrt{\frac{(\gamma_0 + 2\delta_1 + \ldots + 2\gamma_q)}{(H-1)}}}$$

(17)

Where γ_i is the it value of autocorrelation of d_t series. q Is the initial value of γ_i . (Enders, 82-86:2001)

A Review on Previous Researches

Due to importance of the subject, many researchers have been conducted in this area and many methods have been used. There is a wide range of methods from simple linear regression method to advance nonlinear methods. Some of these researches are as follows:

Postali and Picchetti (2006) presented a quantitative analysis for oil price path. They indicted that oil price during 100 years have two endogenous structural breaks. Therefore, null hypothesis of unit root was rejected. Their results showed that oil price can be predicted using geometric Brownian method. In this research, no other prediction method (e.g. ARMA) was used to compare the results.

Malik and Nasereddin (2006) used different prediction methods such as artificial neural network to forecast Gross domestic product of U.S.A with the help of oil prices. They found that neural network method lead to the minimum prediction error. It had been satisfactory if they had compared predictive power of different methods using appropriate criteria.

De Santis (2003) tried to explain why oil prices fluctuate. For this aim, a general equilibrium model was applied to Saudi Arabia which is a big oil producer. The results showed that Saudi Arabia's share in OPEC and some decisions of OECD countries can influence oil prices.

Tang and Hammoudeh (2002) modeled world oil price behavior using first-generation target zone model. They used data during the period of 1988-199 and revealed that OPEC tried to maintain a weak target zone regime for the oil price. In addition they found that based on this model (first-generation target zone) nonlinear models show higher predictive power.

Buchananan, Hodges and Theis (2001) presented a method to predict the direction of natural gas spot price movements. They revealed that traders play an important role in determining the direction of natural gas price movement.

All above mentioned researches presented no long-term or short-term trend for oil price and mostly considered analysis aspects of oil related issues. In other words, theses researches mostly considered descriptive aspects of the research and less used mathematical and computational models and validity tests.

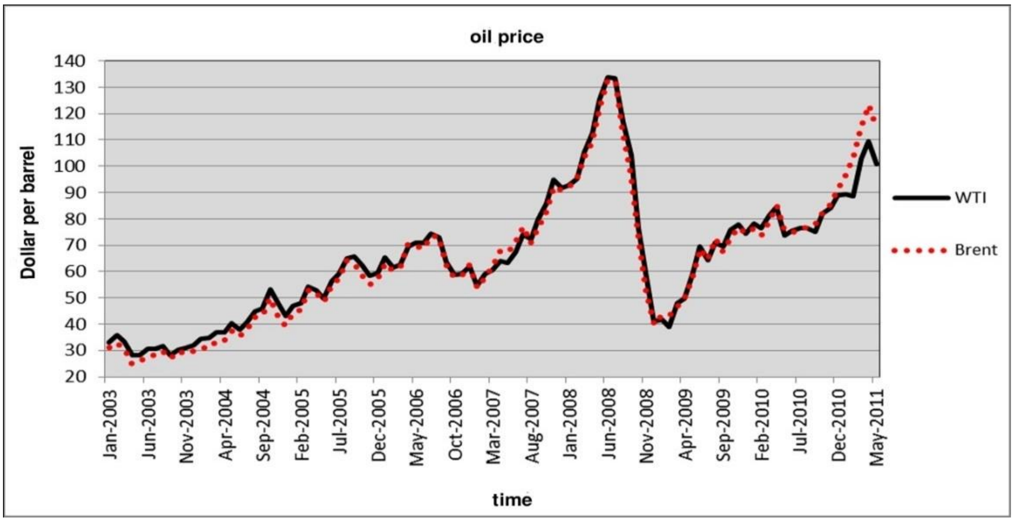
Azar and Rajabzadeh (2003) predicted demand for OPEC oil using Box-Jenkins method. They used annual data during the period of 1960 to 2002. They also employed mean absolute and mean squared error criteria to evaluate the models. However this research is more comprehensive than the other mentioned researches, some deficiencies such as use of too few criteria for model selection, lack of significance testing of difference between validity criteria and use of low orders of ARMA can be observed in this research.

Estimation of the Models and Their Results

Data during the period of early 2003 to May 2011 was used to estimate time series models. Taking into consideration the modeling principles of time series, which were described previously, stationary of respective time series was examined. If the series is either level stationary or trend stationary, it can be easily modeled; otherwise it must be differenced to become stationary. For this aim, ADF test was used. Prices of two well-known benchmark types of oil, i.e. Brent and WTI, always closely track each other and in other words each one is a multiple of the other. Therefore, we used prices of Brent oil which is more well-known throughout the world and based on which most of crude oils are classified.

Curve 1.

Stationary test was performed for considered time series using augmented Dickey-Fuller test. The results in Table 1 show that this series is stationery at the 5% level.



Ref: the research results

Table 1. Dickey-Fuller test results

test statistics Value	Critical values	Significance level
-3/66868	-4/05439	%1
	-3/45632	%5
	-3/15399	%10

Ref: the research results

Dynamic Predictions

Modeling was performed after this test without any need to differencing. Table 2 lists the results of this modeling. As it is observed in this Table, *ARMA* (3, 2) and *ARMA* (3, 3) models are largely different from the other models.

Table 2. Statistics of root mean squared error mean absolute error and Theil inequality coefficient in dynamic situation

model	RootMean Squared Error(RMSE)	Mean Absolute Error(MAE)	Theil Inequality Coefficient (TIC)
ARMA(1,0)	28.48008	24.5612	0.15863
ARMA(1,1)	29.56541	25.51603	0.165551
ARMA(0,1)	42.13073	39.35685	0.255698
ARMA(2,0)	31.42901	27.07276	0.177513
ARMA(2,1)	32.45911	27.90513	0.184176
ARMA(2,2)	32.18736	27.61085	0.182331
ARMA(1,2)	31.10919	26.94821	0.175593
ARMA(0,2)	41.4737	37.64902	0.248952
ARMA(3,0)	32.37511	27.80237	0.183592
ARMA(3,1)	32.43238	27.85843	0.183975
ARMA(3,2)	17.7295	13.09163	0.095154
ARMA(0,3)	39.95993	35.85924	0.244745
ARMA(1,3)	31.21197	26.94054	0.176161
ARMA(2,3)	32.00665	27.3927	0.181084
ARMA(3,3)	21.66974	18.1479	0.116515

Ref: the research results

It should be noted that time series models with autoregressive and moving average of lag greater than 3 were also tested, but since no improvement was observed in prediction statistics, according to parsimony principle they were not used. Based on these facts, either *ARMA* (3, 2) or *ARMA* (3, 3) model must be selected for dynamic predictions. Table 2 shows the results of comparison between two mentioned models. As it is obvious, only Granger-Newbold criterion differentiates between two models and considers *ARMA* (3, 2) prediction as a better prediction than A*RMA* (3, 3), but no significant differences can be distinguished by two other criteria.

Table 3. Prediction evaluation tests: comparison of *ARMA* (3, 2) and A*RMA* (3, 3) models

Test	Statistics	P-Value
1)The F Test	1.352957	0.263943
2)The Granger-Newbold Test	3.500308	0.002742
3)The Diebold-Mariano Test		
3-1) $g(e)=e^2$	1.187595	0.251322
3-2) $g(e)=e^4$	0.711124	0.486651

Ref: the research results

In other words, however absolute values of the error calculated by various criteria are different, these differences is not statistically significant. Based on this fact and parsimony principle, the simplest model, i.e. *ARMA* (3, 2) was selected. In addition, the absolute values of prediction error in this model are smaller than those in other models. Therefore, the best model for dynamic predictions is *ARMA* (3, 2) model and the best order for MA and AR is 2 and 3, respectively.

Static Predictions

At this stage, static predictions were made using the estimated models and desirable statistics were calculated. These statistics are listed in table 4. A problem which arises in this condition is that the results of the majority of models are very close to each other and this makes the selection of two models and their comparison difficult. One solution to this problem is that models with inordinate differences are taken aside (models with no AR process) and among other remained models, the test is performed on two models with maximum differences. If no significant differences are observed between the two models, it can be concluded that the models whose prediction statistics are located in this interval, show no significant differences and again according to parsimony principle the simplest model is selected.

Table 4. Statistics of root mean squared error mean absolute error and Thiele inequality coefficient in static situation

model	Root Mean Squared Error(RMSE)	Mean Absolute Error(MAE)	Theil Inequality Coefficient(TIC)
ARMA(1,0)	7.713342	7.194527	0.038664
ARMA(1,1)	6.864355	6.391594	0.034195
ARMA(0,1)	23.06035	21.55685	0.125978
ARMA(2,0)	6.523549	6.015664	0.032375
ARMA(2,1)	6.599234	6.112711	0.032748
ARMA(2,2)	6.578778	6.069127	0.032622
ARMA(1,2)	6.649142	6.14747	0.033006
ARMA(0,2)	15.83849	14.41708	0.083184
ARMA(3,0)	6.60498	6.1111	0.032768
ARMA(3,1)	6.617713	6.132185	0.032841
ARMA(3,2)	17.02588	16.29957	0.090106
ARMA(0,3)	11.36557	9.696469	0.058145
ARMA(1,3)	6.623101	6.14058	0.03287
ARMA(2,3)	6.561896	6.042003	0.032533
ARMA(3,3)	6.66845	5.671258	0.033296

Ref: the research results

Based on Table 4, among models which have AR process *ARMA* (1, 0) and *ARMA* (2, 0) models have maximum differences in prediction statistics, therefore the test is performed on these two models. Table 5 lists the results of this test.

Table 5. Prediction evaluation tests: comparison of *ARMA* (1, 0) and A*RMA* (2, 0) models

Forecast Evaluation		
Test	Statistics	P-Value
1)The F Test	1.398032	0.242158
2)The Granger-Newbold Test	0.260697	0.797458
3)The Diebold-Mariano Test		
3-1) $g(e)=e^2$	-0.508616	0.617559
3-2) $g(e)=e^4$	-0.478079	0.638684

Ref: the research results

Based on the obtained results, none of the tests confirms the difference between the two models. In other words, no statistically significant difference is observed between the two models which had maximum difference in prediction statistics. Therefore, all models whose prediction statistics are located in this interval, have the same performance. Then, according to parsimony principle the simplest model, i.e. *ARMA* (1, 0), is selected and the best order of AR and MA processes is 1 and 0, respectively.

Conclusions and Recommendations

The tests results revealed that among the models, *ARMA* (3, 2) is the best model for dynamic prediction of crude oil price series taking into consideration all aspects of modeling. For static predictions, it should be noted that the majority of the models lead to very close results and comparative test showed that these models have no statistically significant differences. Therefore, the simplest model, i.e. *ARMA* (1, 0), was selected for prediction.

Also it was found that static predictions are always more accurate than dynamic predictions. This is attributes to how data are used in these two kinds of prediction. In real world, dynamic predictions are more important and static predictions are also possible for just one step forward.

Finally the best selected models for static and dynamic predictions were used to make a futuristic prediction for future time periods. It is obvious that in static condition only one step forward prediction was possible and the predicted value was obtained to be 144.5 dollars per oil barrel. In dynamic prediction, two steps forward predictions were obtained to be equal to 83.8 and 83.9 dollars per barrel (for next two month). As discussed, in one step forward predictions, the priority is given to static predictions.

References

Abrishami, H., & Mehrara, M. (2002). *Applied econometrics* (new approaches), first edition, University of Tehran press.

Azar, A., & Rajabzade, A. (2003). Evaluating compound prediction in economics area through classic neural network approaches, *economic journal of University of Tehran*, fall and winter, No. 63.

Masoud, D. (2004). Risk management and derivatives in oil markets, international institute of energy studies.

Hossein, A. (2007). Advance econometrics, first edition, Baradarn press.

Mohammad, N. (1999). *Unit root and Co integration econometrics*, first edition, Rasa cultural services institute.

Vol. 3, No. 1, 2013, E-ISSN: 2225-8329 © 2013 HRMARS

Buchananan, W. K., Hodges, P., Theis, J. (2001). Which way the natural gas price: an attempt to predict the direction of natural gas spot price movements using trader positions. Journal of Energy Economics 23, 279-293.

De Santis, R. (2003). Crude oil price fluctuations and Saudi Arabia's behavior. Journal of Energy Economics 25, 155–173.