International Journal of Energy Economics and Policy
Vol. 2, No. 1, 2012, pp.41-49
ISSN: 2146-4553
www.econjournals.com

Using SARFIMA Model to Study and Predict the Iran's Oil Supply

Hamidreza Mostafaei
Department of Statistics, North Tehran Branch,
Islamic Azad University, Tehran, Iran. &
Department of Economics Energy, Institute for International Energy Studies,
(Affiliated to Ministry of Petroleum). E-mail: h_mostafaei@iau-tnb.ac.ir

Leila Sakhabakhsh
Department of Statistics, North Tehran Branch, Islamic Azad University,
Tehran, Iran. E-mail: leila.sakhabakhsh@yahoo.com

ABSTRACT: In this paper the specification of long memory has been studied using monthly data in total oil supply in Iran from 1994 to 2009. Because monthly oil supply series in Iran are showing non-stationary and periodic behavior we fit the data with SARIMA and SARFIMA models, and estimate the parameters using conditional sum of squares method. The results indicate the best model is SARFIMA (0, 1, 1) $(0, -0.199, 0)_{12}$ which is used to predict the quantity of oil supply in Iran till the end of 2020. Therefore SARFIMA model can be used as the best model for predicting the amount of oil supply in the future.

Keywords: Long memory; Conditional sum of squares; SARFIMA model; Oil; Iran
JEL Classification: C12; C13; C22; C50

1. Introduction

The recent finance and economic literature has recognized the importance of long memory in analyzing time series data. A long memory can be characterized by its autocorrelation function that decays at a hyperbolic rate. Such a decay rate is much slower than that of the time series, which has short memory. Traditional models describing short memory, such as AR (p), MA (q), ARMA (p, q), and ARIMA (p, d, q) can not describe long memory precisely. A set of models has been established to overcome this difficulty, and the most famous one is the autoregressive fractionally integrated moving average (ARFIMA or ARFIMA (p, d, q)) model. ARFIMA model was established by Granjer and Joyeux (1980). An overall review about long memory and ARFIMA model was model by Baillie (1996). In many practical applications researchers have found time series exhibiting both long memory and cyclical behavior. For instance, this phenomenon occurs in revenues series, inflation rates, monetary aggregates, and gross national product series. Consequently, several statistical methodologies have proposed to model this type of data including the Gegenbauer autoregressive moving average processes (GARMA), k-factor GARMA processes, and seasonal autoregressive fractionally integrated moving average (SARFIMA) models. The GARMA model was first suggested by Hosking (1981) and later studied by Gray et al. (1989) and Chung (1996). Other extension of the GARMA process is the k-factor GARMA models proposed by Giratis and Leipus (1995) and Woodward et al (1998). This paper investigates a special case of the k-factor GARMA model, which is considered by Porter – Hudak (1990) and naturally extends the seasonally integrated autoregressive moving average (SARIMA) model of Box and Jenkins (1976). Katayama (2007) examined the asymptotic properties of the estimators and test statistics in SARFIMA models. There are several methods for estimating the parameters in time series models. In this paper, we estimate the parameters using conditional sum of squares (CSS) method and testing procedures using residual autocorrelations such as the Lagrange multiplier (LM) test are shown.

We intend to forecast the Iran's oil supply in the future. Iran, a member of the Organization of the Petroleum Exporting Countries (OPEC), ranks among the world's top three holders of both proven oil and natural gas reserves. Iran is OPEC's second largest producer and exporter after Saudi Arabia and in 2008 was the fourth-largest exporter of crude oil globally after Saudi Arabia, Russia, and the United Arab Emirates.

This paper is organized of follows: the section 2 gives some definitions and properties for the ARFIMA and SARFIMA processes then we explain using CSS method and LM test. Section 3 illustrates the use of the SARIMA and SARFIMA models and section 4 presents our final conclusions.

2. Materials and Method

2.1 ARFIMA model

Let $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ be a white noise process with zero mean and variance $\sigma_\varepsilon^2 > 0$, and B the backward-shift operator, i.e., $B^k(x_t) = x_{t-k}$. If $\{x_t\}_{t\in\mathbb{Z}}$ is a linear process satisfying

$$\phi(B)(1-B)^d x_t = \theta(B)\varepsilon_t \quad (1)$$

Where $d \in (-0.5, 0.5)$, $\phi(.)$, $\theta(.)$ are polynomials of degree p and q, respectively, given by

$$\phi(B) = 1-\phi_1 B-\phi_2 B^2 - \cdots - \phi_p B^p, \quad \theta(B) = 1-\theta_1 B-\theta_2 B^2 - \cdots - \theta_q B^q$$

Where $\phi_i, 1 \le i \le p$, $\theta_j, 1 \le j \le q$ are real constants, than $\{x_t\}_{t\in\mathbb{Z}}$ is called general fractional differentiation ARFIMA (p, d, q) process, where d is the degree or fractional differentiation parameter. If $d \in (-0.5, 0.5)$, then $\{x_t\}_{t\in\mathbb{Z}}$ is a stationary, and an invertible process. The most important characteristic of an ARFIMA (p, d, q) process is the property of long dependence, when $d \in (0.0, 0.5)$, short dependence, when d = 0, and intermediate dependence, when $d \in (-0.5, 0.0)$.

2.2 SARFIMA (p, d, q) (P, D, Q)$_s$ processes

In many practical situation time series exhibit a periodic pattern. We shall consider the SARFIMA (p, d, q) (P, D, Q)$_s$ process, which is an extension of the ARFIMA process (Bisognin and Lopes, 2009).

Definition 1. Let $\{x_t\}_{t\in\mathbb{Z}}$ be a stationary stochastic process with spectral density function $f_x(.)$.suppose there exists a real number $b \in (0,1)$, a constant C_f and one frequency $G \in [0,\pi]$ (or a finite number of frequencies) such that

$$f_x(\omega) \sim C_f |\omega - G|^{-b} \text{ when, } \omega \to G$$

Then, $\{x_t\}_{t\in\mathbb{Z}}$ is a long memory process.

Remark 1. In Definition 1, when $b \in (0,1)$, we say that the process $\{x_t\}_{t\in\mathbb{Z}}$ has the intermediate dependence property (Doukhan et al., 2003).

Definition 2. Let $\{x_t\}_{t\in\mathbb{Z}}$ be a stochastic process given by the expression

$$\phi(B)\Phi(B^S)(1-B)^d(1-B^S)^D(x_t-\mu) = \theta(B)\Theta(B^S)\varepsilon_t \qquad t \in \mathbb{Z} \quad (2)$$

Where μ is the mean of the process, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a white noise process with zero mean and variance $\sigma_\varepsilon^2 = E(\varepsilon_t^2)$, $s \in \mathbb{N}$ is the seasonal period, B is the backward-shift operator, that is $B^{sk}(x_t) = x_{t-sk}$, $\nabla_s^D = (1-B^S)^D$ is the seasonal difference operator, $\phi(.)$, $\theta(.)$, $\Phi(.)$ and $\Theta(.)$ are the polynomials of degrees p, q, P, and Q, respectively, defined by

$$\phi(B)=\sum_{i=0}^{p}(-\phi_i)B^i \qquad \theta(B)=\sum_{j=0}^{q}(-\theta_j)B^j$$

$$\Phi(B)=\sum_{k=0}^{P}(-\Phi_k)B^k \qquad \Theta(B)=\sum_{l=0}^{Q}(-\Theta_l)B^l \tag{3}$$

Where, $\phi_i, 1\le i\le p, \theta_j, 1\le j\le q, \Phi_k, 1\le k\le P$, and $\Theta_l, 1\le l\le Q$ are constants and $\phi_0=\Phi_0=\theta_0=\Theta_0=-1$.

Then, $\{x_t\}_{t\in\mathbb{Z}}$ is a seasonal fractionally integrated ARMA process with period s, denoted by SARFIMA (p, d, q) $(P, D, Q)_s$, where d and D are, respectively, the differencing and the seasonal differencing parameters.

Theorem 1. Let $\{x_t\}_{t\in\mathbb{Z}}$ be a SARFIMA (p, d, q) $(P, D, Q)_s$ process given by the expression (2), with zero mean and seasonal period $s\in\mathbb{N}$. Suppose $\phi(z)\Phi(z^s)=0$ and $\theta(z)\Theta(z^s)=0$ have no common zeroes. Then, the following is true.

(i) The process $\{x_t\}_{t\in\mathbb{Z}}$ is stationary if d + D < 0.5, D < 0.5 and $\phi(z)\Phi(z^s)\neq 0$, for $|z|\le 1$.

(ii) The stationary process $\{x_t\}_{t\in\mathbb{Z}}$ has a long memory property if 0 < d + D < 0.5, 0 < D < 0.5 and $\phi(z)\Phi(z^S)\neq 0$, for $|z|\le 1$.

(iii) The stationary process $\{x_t\}_{t\in\mathbb{Z}}$ has an intermediate memory property if -0.5 < d + D < 0, -0.5 < D < 0 and $\phi(z)\Phi(z^S)\neq 0$, for $|z|\le 1$.

2.3 CSS method

There are several methods for estimating the parameters in time series models. In this paper, we implement the CSS method to estimate the SARIMA and SARFIMA models of oil supply in Iran. This method is equivalent to the full Maximum Likelihood Estimator (MLE) under quite general conditional homoskedastic distributions. A description of the properties of the CSS estimator and its finite sample performance is presented in Chung and Baillie (1993).

2.4 LM test

This section discusses testing for the integration order, namely, the LM test, which draws on LM tests for the integration order of the ARFIMA model by Robinson (1991), Robinson (1994), Agiakloglou and Newbold (1994), and Tanaka (1999). For the purpose of practical implementation, Godfrey's (1979) LM approach is also used. For the SARFIMA model, we consider the testing problem of the null hypothesis H_0: SARFIMA (p, d, q) $(P, D, Q)_s$ against the alternative:

$H_{A,1}$: SARFIMA (p, d + α_0, q) $(P, D, Q)_s$ Or $H_{A,2}$: SARFIMA (p, d, q) $(P, D + \alpha_s, Q)_s$.

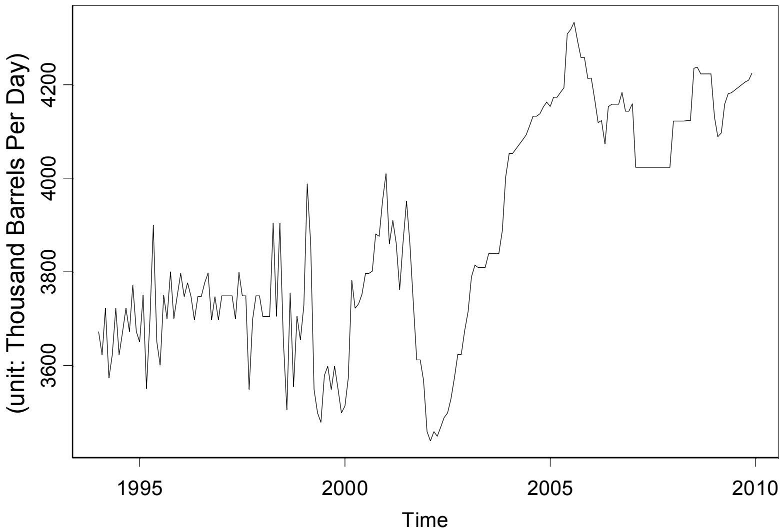
The assumed null model is obtained by imposing the restriction $\alpha_0(\alpha_s) = 0$ and the alternatives are $\alpha_0(\alpha_s) > 0$ or $\alpha_0(\alpha_s) < 0$. We get the p-values for testing the integration order corresponding to tests.

3. Empirical Results

3.1 The data

The data employed in this study are the monthly oil supply in Iran from 1994 to 2009. The data are obtained from the Energy Information Administration of the U.S. Department of Energy. Figure 1 displays the data of oil supply in Iran, $\{x_t\}$.

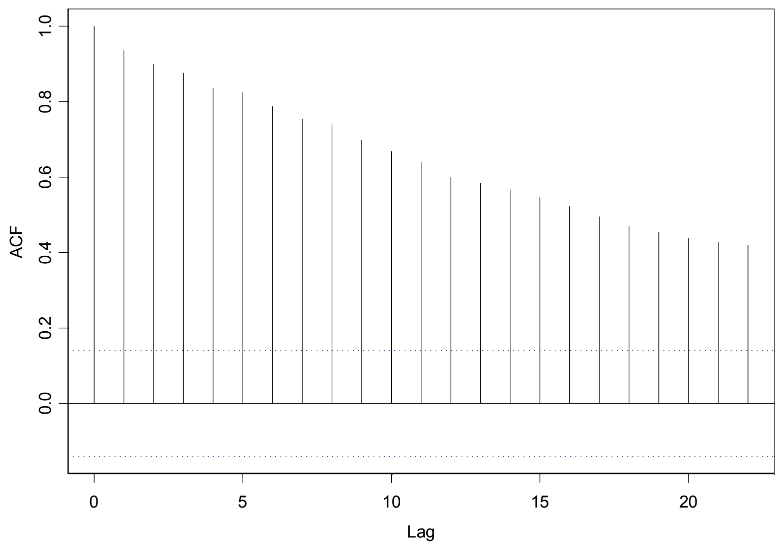
Figure 1. Time plot of oil supply in Iran, 1994-2009



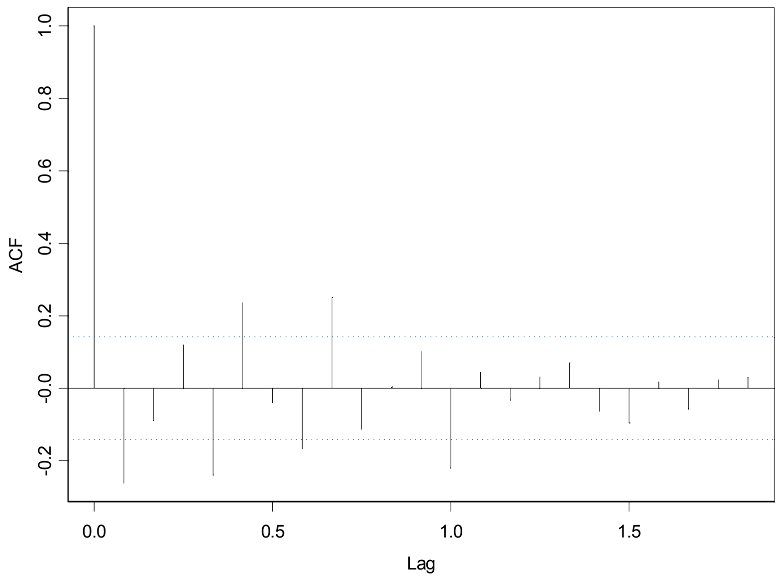
As seen in figure 1, the monthly data are seasonally therefore we consider s=12. Figure 2 displays the autocorrelation function (ACF) of the transformed data. The ACF decays very slowly and exhibits non-stationary.

Figure 2. The sample autocorrelation function (ACF), (a) The ACF of monthly data of Iran's oil supply, (b) The ACF of differenced data

(a)



(b)



As seen in figure 2, there non-stationary in the observed data and this time series doesn't require seasonal differencing. The one approach to trend removal by differencing the series $\{x_t\}$ that the best transform for this data is
$y_t = (1\text{-}B)x_t$.

3.2 Model selection

To search for the best representation of this data, we first fitted differenced data $y_t = (1\text{-}B)x_t$ by the CSS method, where we used a sample mean of $\{y_t\}$, $\bar{y}$ as an estimator of $E(y_t) = \mu$, and set s = 12. AIC and BIC criteria are also used under the assumption of normality [see, e.g., Brockwell and Davis (1991, section 9.3)].

Fitting SARFIMA models or SARIMA models is limited to having SARMA parameters with $0 \le p,q,P,Q \le 3$, and where the total number of estimated SARFIMA parameters (d, D, SARMA parameters, and σ^2) is less than 4. The total number of models is 70. As mentioned earlier, in addition to SARIMA models, SARFIMA models are fitted as well, because we intend to determine if the total oil supply in Iran have long memory. From among these estimation results, we selected models in terms of AIC and BIC that satisfy the following conditions: (i) LM tests are not rejected with the significance level 5% and 10 to 30 degrees of freedom (ii) the SARFIMA parameters all converged. All calculations were made using S-PLUS. Table 1 shows the best five models in terms of AIC model selection with estimators. ID denotes the model identification within 70 models. NE indicates the corresponding parameter is not estimated and is set to be 0. The numbers in parentheses in the column of AIC (BIC) denote the ranking of models in terms of AIC (BIC).

Table 1. Summary of AIC and BIC model selection estimates

σ^2	Θ_1	θ_1	ϕ_2	ϕ_1	**D**	**d**	**BIC**	**AIC**	**ID**
5754.7	NE	0.341	NE	NE	-0.199	NE	(1)2020.4	(1)2010.5	54
5755.1	0.230	NE	NE	NE	NE	-0.264	(2)2020.4	(2)2010.7	46
5696.0	0.243	NE	-0.184	-0.300	NE	NE	(4)2023.7	(3)2010.7	23
5755.9	0.235	0.334	NE	NE	NE	NE	(3)2020.5	(4)2010.8	24
5697.8	NE	NE	-0.186	-0.301	-0.204	NE	(5)2023.8	(5)2010.8	53

SARFIMA (0, 0, 1) (0, -0.199, 0)$_{12}$ model (model ID: 54) is the best model in terms of AIC among the 70 model candidates. From theorem 1, the process $\{y_t\}$ has intermediate memory property.

Table 2 shows the p-values for testing the integration order corresponding to the best five models using the LM test statistics.

Table 2. P-values for testing the integration order corresponding to the best five models

Alternative hypotheses			Model
$0 \neq \alpha_s$, $0 \neq \alpha_0$	$\alpha_0 = 0$, $\alpha_s < 0$	$\alpha_0 < 0$, $\alpha_s = 0$	
0.0060	5.5×10^{-10}	0.3206	SARFIMA (0, α_0, 1) (0, α_s , 0)
0.00009	0.2471	0.00002	SARFIMA (0, α_0, 0) (0, α_s , 1)
0.5629	0.2858	0.3173	SARFIMA (2, α_0, 0) (0, α_s , 1)
0.4722	0.3430	0.2811	SARFIMA (0, α_0, 1) (0, α_s , 1)
0.0051	0.0002	0.2989	SARFIMA (2, α_0, 0) (0, α_s , 0)

In this table, models ID 54, ID 46 and ID 53 correspond to some models in alternative hypotheses of the first, second and fifth rows of SARFIMA models, and models ID 23 and ID 24 correspond to null hypotheses of the third and forth rows of SARFIMA models.
Our findings as follows: (i) results for SARFIMA (0, α_0, 1) (0, α_s, 0), SARFIMA (0, α_0, 0) (0, α_s , 1), and SARFIMA (2, α_0, 0) (0, α_s , 0) support the estimation of d or D for models ID 54, ID 46, and ID 53. (ii) Except for SARFIMA (0, α_0, 0) (0, α_s, 1), results for SARFIMA models show large p-values for the alternative $\alpha_0 < 0$, $\alpha_s = 0$. (iii) Results for some SARFIMA models show relatively small p-values for the alternative $\alpha_0 = 0$, $\alpha_s < 0$ and $\alpha_0 \neq 0$, $\alpha_s \neq 0$.

The best model for y_t is SARFIMA (0, 0, 1) (0, -0.199, 0)$_{12}$ model therefore the best model for x_t is SARFIMA (0, 1, 1) (0, -0.199, 0)$_{12}$ model.

3.3 Forecasting

Upon determination of appropriate model, it can be used for forecasting. The best model is SARFIMA (0, 1, 1) (0, -0.199, 0)$_{12}$ model which is used to predict the total oil supply in Iran till the end of 2012 and 2020, as shown in figures 6 and 7. Tables 3 and 4 show the results of the In-sample and out-sample forecasts for the SARFIMA model. As seen in figures 6 and 7, total oil supply in Iran has increasing trend for the future.

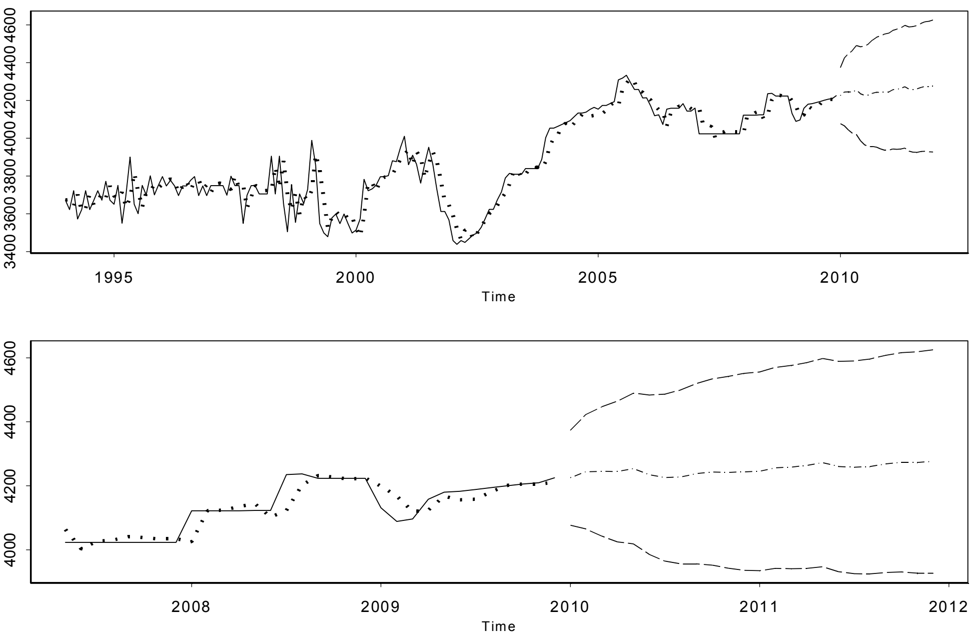


Figure 6. Prediction plot of oil supply in Iran (2010-2012)

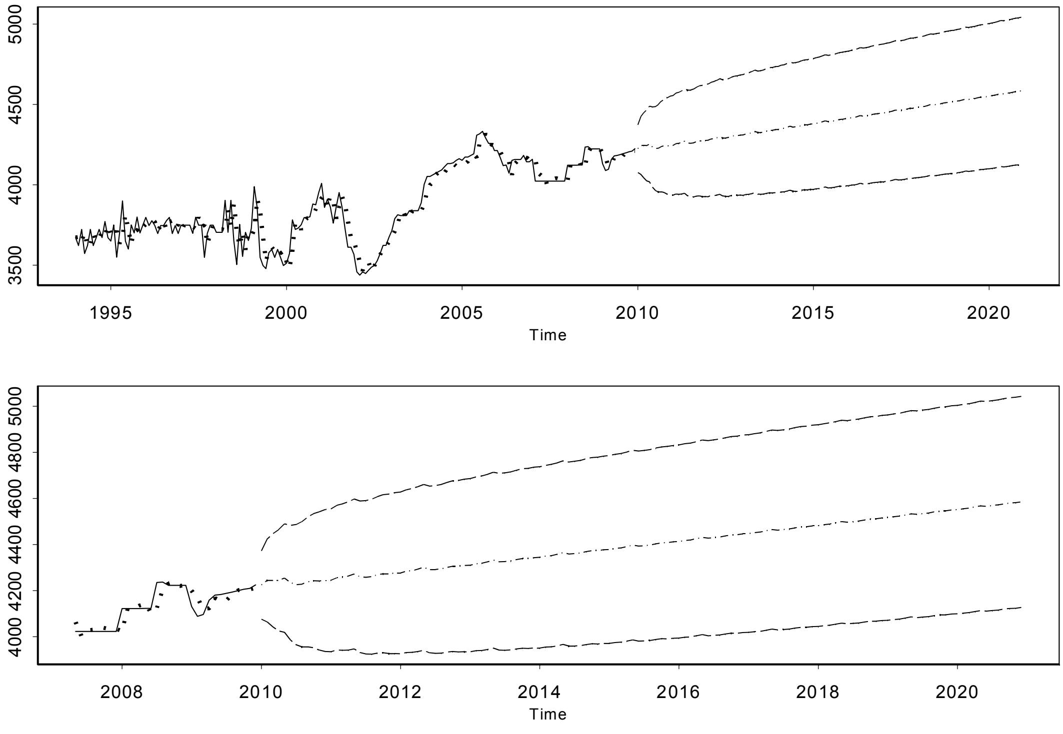


Figure 7. Prediction plot of oil supply in Iran (2010-2020)

Table 3. Out-sample forecasts for the SARFIMA (0, 1, 1) (0, -0.199, 0)$_{12}$ model

upperCL	lowerCL	Prediction	Date
4372.939	4076.793	4224.866	2010-01
4423.095	4064.985	4244.040	2010-02
4446.569	4043.063	4244.816	2010-03
4464.660	4024.765	4244.713	2010-04
4489.326	4018.676	4254.001	2010-05
4483.343	3985.836	4234.589	2010-06
4486.284	3964.796	4225.540	2010-07
4499.173	3955.928	4227.551	2010-08
4519.272	3956.049	4237.660	2010-09
4533.748	3952.008	4242.878	2010-10
4541.347	3942.313	4241.830	2010-11
4551.095	3935.812	4243.453	2010-12
4555.127	3934.799	4244.936	2011-01
4569.854	3942.150	4256.002	2011-02
4575.812	3940.624	4258.218	2011-03
4584.655	3941.999	4263.327	2011-04
4597.244	3947.204	4272.224	2011-05
4588.720	3931.412	4260.066	2011-06
4589.843	3925.400	4257.622	2011-07
4595.965	3924.526	4260.245	2011-08
4607.346	3929.054	4286.200	2011-09
4615.601	3930.595	4273.098	2011-10
4618.773	3927.192	4272.982	2011-11
4624.912	3926.889	4275.901	2011-12

lowerCL: lower confidence limits of forecasts
upperCL: upper confidence limits of forecasts

Table 4. In-sample forecasts for the SARFIMA (0, 1, 1) (0, -0.199, 0)$_{12}$ model

Error	Forecasts	Actual	Date
-67.884	4199.122	4131.228	2009-01
-78.823	4167.687	4088.864	2009-02
-23.022	4119.831	4096.809	2009-03
35.125	4123.085	4158.210	2009-04
12.984	4167.386	4180.370	2009-05
27.312	4155.733	4183.045	2009-06
30.811	4158.006	4188.817	2009-07
13.106	4181.294	4194.400	2009-08
-1.363	4201.535	4200.172	2009-09
0.694	4205.248	4205.942	2009-10
5.358	4204.310	4209.668	2009-11
13.334	4211.703	4225.037	2009-12

4. Conclusions

This paper has examined a seasonal long memory process, denoted as the SARFIMA model. As an illustration of the use of SARFIMA model, we considered monthly oil supply in Iran. We fitted the data with SARIMA and SARFIMA models, and estimated the parameters using CSS method. The results indicated the best model was SARFIMA (0, 1, 1) (0, -0.199, 0)$_{12}$ model which was used to predict the data. On the basis, we conclude that the SARFIMA model is effective and can be usefully employed as a substitute for the SARIMA model when fitting Iran's oil supply data.

References

Agiakloglou, C., Newbold, P., 1994. Lagrange Multiplier Tests for fractional difference. *Journal of Time Series Analyisis*, 15, 253-262.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, 73, 5-59.

Bisognin, C., Lopes, R.C., 2009. Properties of seasonal long memory processes. *Mathematical and Computer Modeling*, 49, 1837-1851.

Box, G.E.P., Jenkins, G.M., 1976. Time series analysis forecasting and control, 2nd ed. Holden-Day. San Francisco.

Brockwell, P.J, Davis., 1991. Time series: Theory and Methods. Springer, New York.

Chung, C.F., 1996. Estimating a generalized long memory process. *Journal of Econometrics*, 73, 237-259.

Chung, C.F., Baillie, R.T., 1993. Small Sample Bias in Conditional Sum-of-Squares Estimators of Fractionally Integrated ARMA Models. *Empirical Economics*, 18, 791-806.

Doukhan, P, Oppenheim, G., M.S. Taqqu, M.S., 2003. Theory and applications of long-range dependence. Birkheuser, Boston.

Giraitis, L., Leipus, R., 1995. A generalized fractionally differencing approach in long memory modeling. *Lithuanian Mathematical Journal,* 35, 65-81

Godfrey, L.G., 1979. Testing the adequacy of a time series model. *Biometrika*, 66, 67-72.

Granjer, C.W., Joyeaux, R., 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, 1, 15-29.

Gray, H.L., zhang, N.F., Woodward, W.A., 1989. On generalized fractional processes. *Journal of Time Series Analysis*, 10, 233-257.

Hosking, J.R.M., 1981. Fractional differencing. *Biometrika*, 68, 165-176.

International Energy Outlook 2010. www.eia.gov/oiaf/ieo/index.html. July 2010

Katayama, N., 2007. Seasonally and fractionally differenced time series. *Hitotsubashi Journal of Economics*, 48, 25-55.

Porter-Hudak., 1990. An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of American Statistical Association*, 84, 410, 338-344.

Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional Heteroskedasticity in multiple regressions. *Journal of Econometrics*, 47, 67-84.

Robinson, P.M., 1994. Efficient tests of non-stationary hypotheses. *Journal of American Statistical Association*, 89, 1420-1437.
Tanaka, K., 1999. The non-stationary fractional unit root. *Econometric theory*, 15, 549-582.
Woodward, W.A., Cheng, Q.C., Gray, H.L., 1998. A k-factor long memory model. *Journal of Time Series Analysis*, 19, 485-504.