

---

## **Modeling and Forecasting Residential Natural Gas Demand in IRAN**

Fatemeh Daei Jafari

Master of Industrial Engineering, School of Industrial engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran  
f.d.jafari.88@gmail.com

Raissi Sadigh

Ph.D., Associate Professor -School of Industrial engineering, Islamic Azad University, South Tehran Branch, Tehran, Iran  
f.d.jafari.88@gmail.com

Editor Científico: José Edson Lara  
Organização Comitê Científico  
Double Blind Review pelo SEER/OJS  
Recebido em 11.05.2019  
Aprovado em 09.09.2019



Este trabalho foi licenciado com uma Licença Creative Commons - Atribuição – Não Comercial 3.0 Brasil

### Abstract:

The main focus of this paper is to provide an appropriate mathematical model to predict the natural gas demand for the next six months in the household sector using technical method regardless of influential variables. For this purpose, the most important and most widely used modeling methods of natural gas demands were used as modeling options and four different candidate families were analyzed based on 234 months of historical data of actual consumption and model parameters were estimated using appropriate methods. Then, by using the precision indicators such as average absolute error, average absolute percentage error, waste diffraction model and inequality coefficient of Thiel time series, predicted for a period SARIMA (1,1,2) (1,1,0)<sub>12</sub> accepted as the most appropriate fitness function were Identified and of the next 6 months. The results showed that the proposed method has minimum residual in terms of MAE, MAPE, TIC and error variance.

**Keywords:** Natural Gas Demand, Forecasting, Box-Jenkins' Time Series, Fourier Series, Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

## Modelagem e Previsão de Demanda Residencial de Gás Natural no Irã

### Resumo

O foco principal deste artigo é fornecer um modelo matemático apropriado para prever a demanda de gás natural para os próximos seis meses no setor doméstico usando o método técnico, independentemente de variáveis influentes. Para este propósito, os métodos de modelagem mais importantes e mais amplamente utilizados de demanda de gás natural foram usados como opções de modelagem e quatro diferentes famílias candidatas foram analisadas com base em 234 meses de dados históricos de consumo real e os parâmetros do modelo foram estimados usando métodos apropriados. Em seguida, utilizando os indicadores de precisão, como erro absoluto médio, erro percentual absoluto médio, modelo de difração de resíduos e coeficiente de desigualdade de séries temporais de Thiel, previsto para um período SARIMA (1,1,2) (1,1,0)<sub>12</sub> aceito como a função de aptidão mais apropriada foi identificada e dos próximos 6 meses. Os resultados mostraram que o método proposto possui residual mínimo em termos de MAE, MAPE, TIC e variância de erros.

**Palavras-chave:** Demanda de Gás Natural, Previsão, Séries Temporais de Box-Jenkins, Séries de Fourier, Heterocedasticidade Condicional Autorregressiva Generalizada (GARCH)

## Modelage y pronóstico de la demanda residencial de gas natural en IRÁN

### Resumen:

El objetivo principal de este artículo es proporcionar un modelo matemático apropiado para predecir la demanda de gas natural para los próximos seis meses en el sector de los hogares utilizando un método técnico independentemente de las variables influyentes. Para este propósito, los métodos de modelado más importantes y más utilizados de las demandas de gas natural se utilizaron como opciones de modelado y se analizaron cuatro familias candidatas

diferentes en base a 234 meses de datos históricos del consumo actual y se estimaron los parámetros del modelo utilizando los métodos apropiados. Luego, mediante el uso de indicadores de precisión como el error absoluto promedio, el error porcentual absoluto promedio, el modelo de difracción de residuos y el coeficiente de desigualdad de las series temporales de Thiel, pronosticados para un período SARIMA (1,1,2) (1,1,0) 12 aceptado como se identificaron las funciones físicas más apropiadas y de los próximos 6 meses. Los resultados mostraron que el método propuesto tiene un mínimo residual en términos de MAE, MAPE, ICT y varianza de error.

**Palabras clave:** Demanda de gas natural, pronósticos, series de tiempo de Box-Jenkins, series de Fourier, heterocedasticidad condicional autorregresiva generalizada (GARCH).

## 1. Introduction

Natural gas has thousands of years of history and in the countries which own this energy, it is considered as a relatively abundant and cheap energy. Natural gas as a good substitute for oil has been introduced in the 90s. It is also more efficient in terms of thermal efficiency than oil and in terms of pollution is less than oil. Usually the domestic sector is the most consumer of natural gas energy in the countries that own this resource. Natural gas in the domestic sector due to their availability and low cost and high calorific value and lack of its requirement to sophisticated and expensive technology to use, is one of the most important fuels in the domestic sector. In these countries, this God-given resource is used to a large-scale for industrial and household use and severity use is expected to increase in the coming decades. Now, natural gas has become one of the most important sources of energy supply in the world.

Natural gas is the most important native energy resource in Iran. Iran's natural gas demand in the domestic sector has been increasing Iran's natural gas demand in the domestic sector growth has been increasing growth rate. For development planning it is necessary to understand the evolution of its demand especially in household sector for future. Non-renewable nature of this resource, natural monopoly of the gas industry and government policies in domestic gas section supply in Iran causes that the prediction of demand process of this product to be more in center of attention of planners in the field of macroeconomics in general and the energy sector in particular.

In the literature, there are two main approaches to modeling gas demand. They nominated as fundamental and technical approach. In fundamental modeling, the gas demand considered as a function of a set of significant predictors such as population size, price, GDP and etc. Here

a tuple of historical data set is required and the candidate fitness function coefficients estimated based on them. This approach is more suitable when all predictors have independent of each other and a reliable prediction of them is available. Contrary, in the technical approach a time series is evaluated independent of its associated affecting predictors. In these cases it is assumed that associated predictors was hidden in the nature of the data. So regardless of influential variables to predict the future is looking. In both approach modeling and forecasting domestic gas demand is an issue that is highly dependent on the type of data available. This means that an approved prediction model which adjusted based on a given time series may not be expected suitably for other data sets. This has led many articles have been written to model the demand for natural gas around the world yet. Recent examples of such article which is subject to gas demand modeling and forecasting will be noted.

In order to forecast natural gas demand in the household sector a lot of studies have been done, including the (Akmal & David, 2001) examined energy demand in the household sector in Australia. In their paper, the “Almost Ideal Systematic Demand” models are used and cross-price attractions were estimated for electric energy, natural gas and other energy. Results indicated that it is a substitution between natural gas and other energy resources; on the other hand natural gas and other energy carriers are considered luxury in this model. In other words, an increase of one percent of household income in Australia will increase the natural gas demand by more than one percent.

(Liu & Kaboudan, 2003) predicted short-term demand for natural gas by multivariate regression equations for the US, and have estimated gas consumption for four parts of household, commercial, industrial and power. (Arac & Aras, 2004) Predicted natural gas consumption in the household sector in a city in Turkey. They divided years in two sections of heating and cooling during a year, then modeled for each period separately, and estimated that in those models the relationship between gas consumption and days heating degree and time are checked. (Huntington, 2007) in a survey studied the industrial natural gas consumption in the United States of America to reach an experimental model to assess the future course. This study shows a statistical model of America's industrial natural gas consumption based on data from 1958 to 2003. In particular it considers the possibilities of alternative fuels and changes in the industrial economy's basis. This study clarifies statistical model of natural gas consumption in America based on historical data since the Second World War and one of the results of this study is to help planners and policymakers of companies to pursue more important factors that could affect future trends in industrial consumption.

(Ediger & Akar, 2007) anticipated the expectancy of demand for energy, particularly natural gas of Turkey in the long-term outlook for 2020 as the primary and basic fuel that have been carried out by seasonal time series models. Forecast of total energy demand has higher reliability than individual forecasts of energy resources.

(Forouzanfar, Doustmohammadi, Bagher Menhaj, & Hasanzadeh, 2009) have estimated natural gas for domestic and commercial sectors in Iran with logistic equation-based modeling approach and have predicted it for the 11-year outlook. They have tested Natural gas consumption on a monthly basis. The use of this method in this study compared to other methods seems simple, modeling is provided in two ways. (Filik, Gerek, & Kurban, 2010) have forecasted demand for electrical energy for the moment for a long period as a 20-year long-term outlook for Turkey by econometric methods and neural networks. This study includes a long-term approach, and unlike other modeling it includes real-time results in long-term forecasts accurate and correct.

(Xu & Wang, 2010) have modeled the China's natural gas consumption based on combined models for outlook of 2015. In this paper, modeling is based on a combination of polynomial curve model and moving average. (Wadud, Dey, AshfanoorKabir, & Khan, 2011) also have forecast the demand for natural gas in Bangladesh for a long period as a primary energy based on dynamic econometric modeling for outlook of 2025. Their effort was on fundamental studies and among influencing factors in this modeling can be mentioned of consumers' income, energy prices, environmental factors (temperature, temperature and climate changes), change in policymaking, change in economic structure, population growth and gross domestic production. (Azadeh, et al., 2011) by using a method of analysis system boundary adaptive fuzzy inference forecasted natural gas consumption in Bahrain, Saudi Arabia and the United Arab Emirates for outlook of 2015 and came to the conclusion that this model is more favorable with regard to the low error rate is useful for future prospects.

(Dagher, 2011) estimated domestic demand for natural gas in Colorado by econometric models and almost ideal systems in which time-series Excel data was on a monthly basis from January 1994 to September 2006. Modeling in this study is tested generally based on price and income elasticity in long and short terms. . In fact, the modeling and the measurement models, strains is less in short-term than long-term period that in short-term factors, the existing stock of durable equipment and in long-term the impact of the rate of change in the existing stock of durable equipment are being studied that in both cases, the price of gas, personal consumer

income, consumer behavior, time, temperature and geographical factors are also considered as affecting factors. (Zia Wadud et al. 2011) applied a dynamic econometric model to determine natural gas demand in Bangladesh, both in the national level and for two end-use sectors. Their model supported the gas price, GDP and population predictors in the form of a Cobb-Douglas demand function and allowed them to offer some comments for the Bangladesh policymakers. (Majazi Dalfard, Nazari Asli, Asadzadeh, Sajjadi, & Nazari Shirkouhi, 2012) forecasted and estimates natural gas consumption based on the final increase of energy in the Iran's long period by regression modeling phase based on fuzzy network. The reason of using this method is compatible conditions and its flexibility compared to other models. (Taspinar, Celebi, & Tutkun, 2012) have estimated daily gas consumption in Turkey for a short period by artificial neural network and regression models. Data from the General Directorate of Meteorology of Turkey and central bank, based on historical data have involved 1800 days between 2007 and 2011, respectively. Also the natural gas consumption data on a daily basis in this paper are involved as input variables. (Melikoglu, 2013) have modeled demand variation for natural gas in Turkey based on Outlook of 2030. In this study, for long-term demand he used two logistic equations econometrics, time series models and exponential of the logistical econometrics, time series and exponential models. And for the medium term, linear econometric equations are considered. In this modeling, according to data from the exponential combined model with time-series models of ARIMA and seasonal models of SARIMA are used.

(Salehnia, Falahi, Seifi, & Mahdavi Adeli, 2013) estimated natural gas prices for Iran by comparative non-linear gamma statistic modeling, linear dynamic regression and neural network during 1997 to 2012, weekly, daily and monthly. (Pourazarm & Cooray, 2013) estimated household electricity demand in the long-term and short-term of Iran based on econometric modeling. In this study, it is modeled according to time-series data of 1967 to 2009 in a fundamental way; it means that the effect of influencing factors on domestic electricity demand is checked in the model.

(Wang & Lin, 2013) have estimated demand for natural gas consumption in the household, commercial and industrial sectors in China, based on econometric regression linear-log models for vision of 2020. Effective variables in this study are household income, per capita income, estimated price, income elasticity, price elasticity and climate changes, respectively. (Kani, Abbasspour, & Abedi, 2013) have estimated demand for natural gas on the basis of econometric regression modeling for long period in Iran, in this research basic approach was used. Influencing factors are such as GDP, temperature in Centigrade and natural gas

consumption in million cubic meters are considered as independent variables and the real price of natural gas is considered as dependent variable. (Kovacic & Sarler, 2014) have predicted daily gas consumption in the Czech Republic steel plant for a short period based on genetic algorithm modeling and regression modeling, and regression method has a better advantage. (Dilaver, Dilaver, & C. Hunt, 2014) have forecasted natural gas demand in Europe for long-term, and solving challenges during 1978 to 2011 based on econometric and regression methods. Data is estimated based on a time series process. (Yu, Zheng, & Han, 2014) have predicted natural gas demand in China by using panel data based on econometric models. In this study, collected data was based on 2006 to 2009 in the cities of China.

Using a technical approach (Szoplik, 2015) forecasted gas demand a selected city of Poland with the use of artificial neural networks. He showed that a given multilayer perceptron model can be successfully used to predict gas consumption on any day of the year and any hour of the day. A similar study on the Pakistan natural gas demand modeling through applying an econometric fundamental model. (Wei Zhang, 2015) Their forecasting method supported a tuple of GDP, urban population, energy consumption structure, industrial structure, energy efficiency and exports of goods and services as covariates. They compare the proposed method with the Gray prediction model, linear regression and artificial neural network. A recent paper focused on modeling natural gas demand in China. (Faheemullah Shaikh, 2016) They forecasting horizon supported for a 20 years long range period using logistic and logistic-population model. (Faheemullah Shaikh, Qiang Ji, Pervez Hameed Shaikh, Nayyar Hussain Mirjat, Muhammad Aslam Uqaili, 2017). Based on population growth in china, natural gas consumption increases. Therefore, the energy is an important factor. Moreover, forecasting demand natural gas in this country is made based on the data obtained from 2014 to 2020 optimized nonlinear gray modeling. On the other hand the outcomes obtained from the model demonstrate and authenticate that the average consumption of natural gas increases 11% during each year. (Ying Chen, Wee Song Chua, Thorsten Koch, 2018). The forecast of supply and demand of natural gas for special and effective benefit of gas distribution in Germany is so necessary. In this paper for forecasting short-time flow of natural gas based on daily data an advanced modeling autoregressive of a function having variables has been used. In this model fewer parameters are required in comparison with other modeling approaches and time series modeling. Therefore the mentioned model is suggested for precise approximation of short-time forecasting natural gas demand in Germany.

Although many articles have been done on how to model and forecast time series associated with energy sector, but since the applied methods are heavily based on collected data and cannot be used in other areas, therefore in this research, research methodology is based on four families and are pursued according to historical data of our country and values and accurate matching and estimation of model parameters. And by using the most validate comparative methods the most appropriate mathematical models are presented to forecast natural gas demand in the residential sector.

The structure of this paper is as follows. In the second section, research methodology is presented and the overview of the most important theoretical principles and practical methods of modeling and forecasting natural gas consumption are done. How fitting mathematical models run on historical data of domestic gas demand are represented in the third part. For all candidate models, apart from determining the most appropriate form of fitting that the residues are white noise, the parameters of the mathematical model have been fitted. The fourth section of this article is dedicated to quantitative comparison of fit patterns in order to reveal the most appropriate pattern of experimental data. Conclusion and further research are presented in chapter five.

## **2. Research Methodology**

The literature showed that the methods of modeling and gas consumption forecasting in different countries have been with different views according to the type, volume, past reliable historical data and research objectives. In general, two common views in this case applied. They are fundamental and technical analysis methods. About the first one, the role of affecting variables in the consumption of natural gas is also cited. In the second view the effect of affecting factors did not enter directly in studies and their affects are being studied indirectly. Since the purpose of this study is to use technical analysis approach and mathematical models to describe behavior of natural gas and predict their future values behavior are the main consideration, therefore, in the choice of curve fitting and forecasting these cases has been on main focus.

Technical time series models are used in order to study and analyze governing patterns of data, forecasting future behaviors and to estimate the missing information. In this method, the assumption is on future survival of factors that affect the past behavior of time series. One important feature in the analysis of time series, in comparison to other statistical methods is

explicit recognition of unchanging data. In this model, unlike econometric models the prediction of a variable based on behavior is not done by linking them to a set of other variables on the basis of a causal relationship, but predictions are done solely based on the behavior of that variable or variables on the basis of their past behavior. In time series modeling and forecasting approach, the most common techniques are: the Box-Jenkins autoregressive integrated moving average (ARIMA) models, Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models, artificial neural network (ANN), curve fitting, the Fourier series and wavelets analysis. The first three methods try to model the time series using inside autocorrelation behaviors, while other methods (except ANN) survey fitness function based on time covariate.

In statistics experts believed that forecasting should be applied based on an appropriate calibrated fitness function. Here a calibrated model statistically has normal distribution with zero mean and a constant variance. It is noted that if residuals are auto-correlated, then error terms have the capability to express using an appropriate order auto-regressive moving average time series model. Then the most appropriate model could be attained using comparison on statistical goodness of fit metrics such as mean absolute error (MAE), mean absolute percentage error (MAPE), waste dispersion modeling, the Thiel inequality coefficient (TIC) and residual standard coefficient. Finally forecasting could be carried out based on the best alternative which has the minimum amounts of the said metrics.

In this research we considered ARIMA, GARCH, polynomial curve fitting and the Fourier series as candidates for the fitness function and all relevant parameters based on the best orders estimated and the model residuals statistically tested for examining the auto-correlation, partial auto-correlation, randomness, and goodness of normal fit. Then all calibrated models were compared statistically to the best choice of the most appropriate fitness function for forecasting.

### **3. Material and Methods**

In order to forecast gas consumption demand, four types of fitness function candidates are:

#### **a) The Box-Jenkins Time Series models**

The most common technique in univariate time series is on using Box and Jenkins models. Such model focused on time series pattern without reflecting the associated covariates. By using

this approach it is possible to fit appropriate models to show level, trend, seasonality and random patterns on wind speed. Since the present level of wind speed ( $Y_t$ ) depends on its immediate past, we describe this univariate time series  $Y_t$  by the process

$$Y_t = E[Y_t | \Omega_{t-p}] + \epsilon_t \quad (1)$$

where  $E[. | .]$  denotes the conditional expectation operator,  $\Omega_{t-p}$  the information set at time  $t - p$  and  $\epsilon_t$  the innovations or residuals of the time series which are uncorrelated, have mean zero and play the role of the unpredictable part of the time series. The most well-known model is Autoregressive Moving Average  $ARMA(p, q)$  model which following (mean adjusted) form.

$$Y_t - \mu = \sum_{i=1}^p \varphi_i (Y_{t-i} - \mu) + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \quad (2)$$

where  $\mu$  is mean of time series,  $\varphi$  is autoregressive coefficient and  $\theta$  is moving average coefficient. When  $q = 0$  we have a pure autoregressive process and when  $p = 0$  pure moving average process.

It may happen that squared residuals exhibit significant serial correlation. It indicates that errors are not independent although they are serially uncorrelated. Residuals are called then conditionally heteroscedastic and GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models have been proved to be very successful at modeling the serial correlation in the second moment of the underlying time series. The variance equation of the process is expressed as follows:

$$\epsilon_t = z_t \sigma_t \quad ; z_t \sim D(0,1) \quad (3)$$

$$Var(\epsilon_t | \Omega_{t-1}) = E(\epsilon_t^2 | \Omega_{t-1}) = \sigma_t^2 = K + \sum_{j=1}^a \alpha_j \sigma_{t-j}^2 + \sum_{i=1}^b \beta_i \epsilon_{t-1}^2 \quad (4)$$

$$= K + \alpha(B) \sigma_{t-1}^2 + \beta(B) \epsilon_{t-1}^2 \quad (5)$$

where  $z_t$  are iid random variables with zero mean and unit variance and  $D$  is their probability density function. Common choices for density function  $D$  are normal distribution, student- $t$  distribution and generalized error distribution. Thus, under the  $GARCH(a, b)$  model, the conditional variance of  $\epsilon_t$ ,  $\sigma_t^2$  depends on the squared residuals in the previous  $b$  periods, and the conditional variance in the previous  $a$  periods.

In general seasonal models may be applied after de-seasonalizing a given time series. In such cases behaviors of wind speed in repeated time intervals could be presented in the form of a SARMA model after evaluating and determining seasonal effects. For example wind speed behavior may be repeated on every morning.

In general, a seasonal model can be written as follows.

$$\Phi(L)A_p(L^S)Y_t = \theta_q(L)B_Q(L^S)u_t \quad (6)$$

$$\Phi(L)A_p(L^S)(\Delta^d\Delta^D)y_t = \theta_q(L)B_Q(L^S)u_t \quad (7)$$

Where  $S$ , represents monthly, quarterly, weekly and yearly repetition,  $D$  stands for seasonal differencing,  $P$  for seasonal autoregressive order and  $Q$  for seasonal moving average order.

In fact, these models are respectively  $SARMA(p, q)(P, Q)_s$  and  $SARIMA(p, d, q)(P, D, Q)$  and to describe them we should use the efficient algorithms and suitable software, such as Eviews and we can estimated calibrated model parameters. Rather models of Box and Jenkins family ARIMA are typical or seasonal type and describe the behavior average of time series and the usage of them is sufficient in situations where there is no significant variance changes. There are cases that variance time series change is not a minor study. In this case, for more accurate modeling of time series, we use description of conditional variance with the help of ARCH and GARCH. If the ARCH test indicates the existence of significant variability for diffraction time series, the use of this type of modeling is discussed. In this test, Lagrange multipliers are used which is equal to  $nR^2$ . If  $nR^2$  is greater than the desired critical value such as 0.05, then it is necessary to model conditional variance using the GARCH models.

### b) The Fourier Series

An alternate fitness functions as a group of simple oscillating terms of sine, cosine or complex exponential function called as the Fourier series. Therefore the Fourier series can also be a good alternative for modeling in situations that there are repeating behaviors. If the behavior model of a time series  $f$  is a periodic function with periodicity, it means it is  $f(t + T) = f(t)$  and integral able, then this function can be written as Equation (8).

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(\omega_n t) + b_n \sin(\omega_n t)] \quad (8)$$

Where  $\omega_n$  harmonic  $n$  of Fourier series to radians and coefficients of  $b_n$ ,  $a_n$  and  $a_0$  can be obtained from Euler's formula. If the length function is  $2l$ , then parameters can be calculated from the equations (9), (10) and (11).

$$a_0 = \frac{1}{2l} \int_{-l}^l f(x) dx$$

(9)

$$a_n = \frac{1}{l} \int_{-l}^l f(x) \cos\left(\frac{n\pi}{l} x\right) dx \quad (10)$$

$$b_n = \frac{1}{l} \int_{-l}^l f(x) \sin\left(\frac{n\pi}{l} x\right) dx$$

(11)

### c) Exponential Model Curve Fitting

The exponential model is one form of curve fitting that could be applied to historical data. Such curve might be act as a suitable option for describing behavior that has Increasing or decreasing exponentially trend over time. A general candidate form for an exponential curve presented by equation (12) in which the constant  $e$  is called Euler's number and is approximately equal to 2.7182.

$$y_t = ae^{bt} + \varepsilon \quad (12)$$

Here,  $a$  and  $b$  are called as fitness parameters that should be estimated using historical data. Applying statistical software could be facilitate curve fitting process. MATLAB has a professional toolbox for curve fitting process based on a wide collection of the most common models. Here  $\varepsilon$  stands for the model residuals.

#### **d) Polynomial Model Curve Fitting**

If behavioral changes of time series are taken into account as a function of time variable, so the time series may be expressed by using a polynomial functions. In general, a  $k$  degree polynomial curve can be offered by Equation (13).

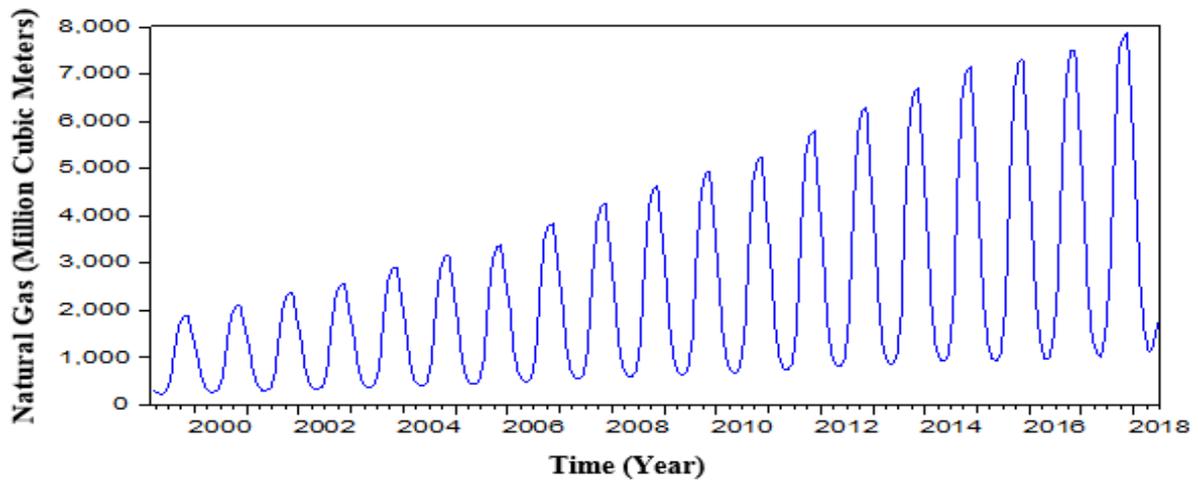
$$y_t = \alpha + \beta t + \gamma t^2 + \delta t^3 + \dots + \theta t^k + \varepsilon \quad (13)$$

Where parameters of  $\alpha, \beta, \gamma, \delta, \dots, \theta$  are called as intercept parameters and coefficients of associated time degree. Here  $\varepsilon$  indicates error terms. If residuals deploy a normal distribution with zero mean and small standard deviation, then the candidate polynomial function assumed as a calibrated model. In the case of  $k=1$ , the polynomial called as a simple linear function of time and also called as a quadratic form of time-dependent and if  $k$  is equal 2. The last one has a parabolic form over time and has good capability to deploy many real time series.

#### **4. Natural gas consumption Time series Modeling**

Monthly demand record of natural gas in household sector in Iran has a 20 years old antiquity and its trustee is the Institute of International Energy Studies. In this research, valid historical data gathered from March 1999 up to September 2018. Figure 1 shows the real gas monthly consumption in the domestic sector in which periodic changes and trend is quite

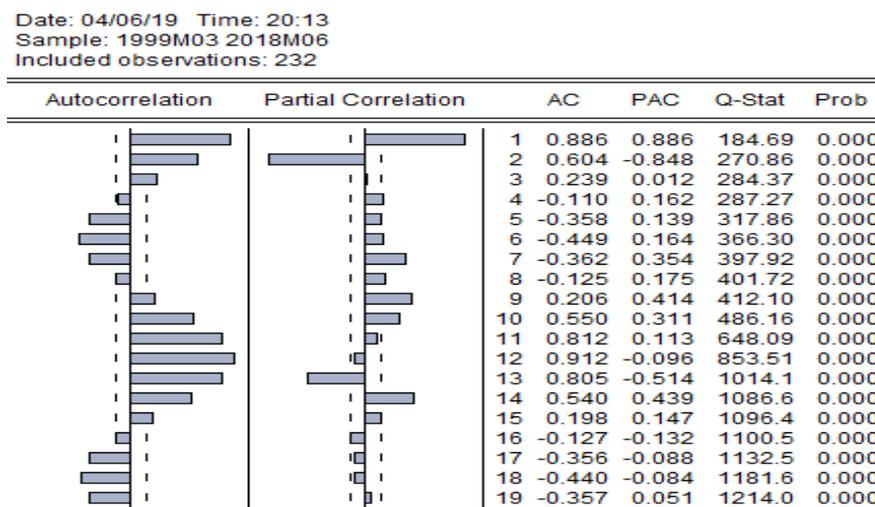
obvious.



**Figure 1.** Monthly natural gas consumption in the domestic sector during a period of 20 years

**a) Fitting a Box-Jenkins Model**

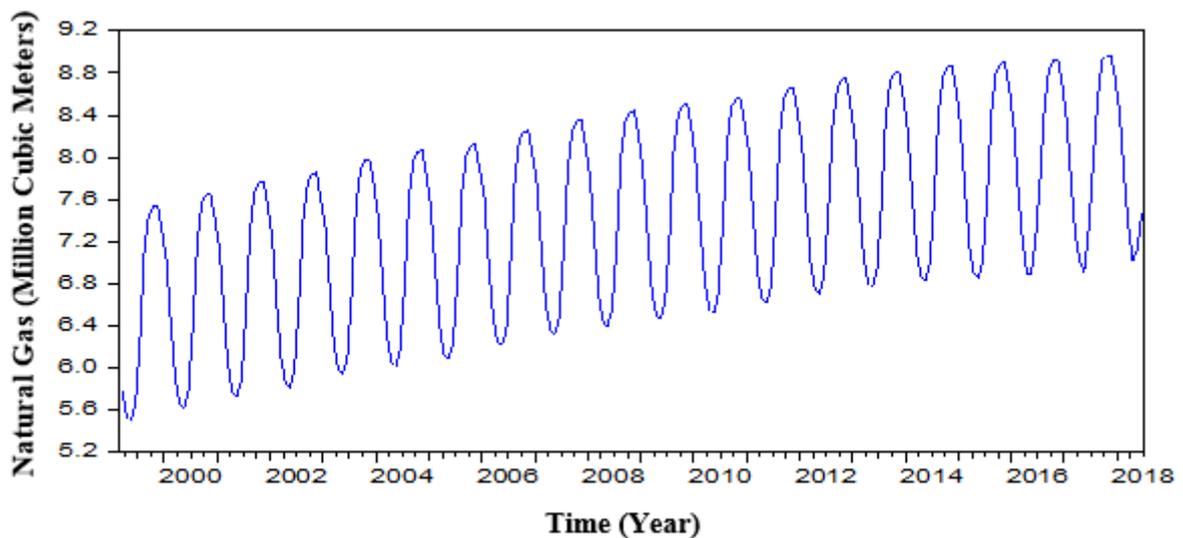
In order to calibrate an appropriate ARIMA model, the autocorrelation function (ACF) and partial autocorrelation function (PACF) illustrated in Figure 2 using the 232 observation entered on Eviews econometrics package. As shown, there are many significant spikes on both graphs that represents both trend and seasonal patterns. A more focus on the pattern reveals a repeated pattern which take place every 12 months.



**Figure 2.** The ACF and PACF diagram of 232 observation of natural gas demand in Iran’s household sector

Also conducting the Dickey-Fuller test with statistic value of  $ADF = 2.23$  show that it is smaller than the absolute value of the t-statistic at the levels of 1%, 5% and 10% which reveals

that the unit root hypothesis is significant. Consequently, it can be stated that natural gas consumption data has a unit root, or in other words is non-viable and it is needed to be seasonalised. For this purpose consecutive differencing is needed. Also the logarithm transformation could be act as a good choice to divert time series in a stationary form. Figure 3 presents the pattern of time series after applying a logarithm transformation which the increasing trend and cyclical pattern are more evident.



**Figure 3.** The logarithmic graph of the time series

In order to remove the seasonal effects, a multiplicative models with differencing repetition period of 12 months was used and to eliminate the effects of positive trend, a simple differencing was applied. Table 1 shows the main indicators on choosing the best form of the Box- Jenkins ARIMA models. Here we presented the Akaike information criterion (AIC), the Schwarz criterion (SC), and the Hannan-Quinn criterion (HQC) as comparison criterion. Result show that a multiplicative Seasonal-ARIMA model in the form of SARIMA (1,1,2)(1,1,0)<sub>12</sub> is the most appropriate fitting metrics which is a multiplicative model of the ARIMA model with minimum standards of AIC, SC and HQC metrics.

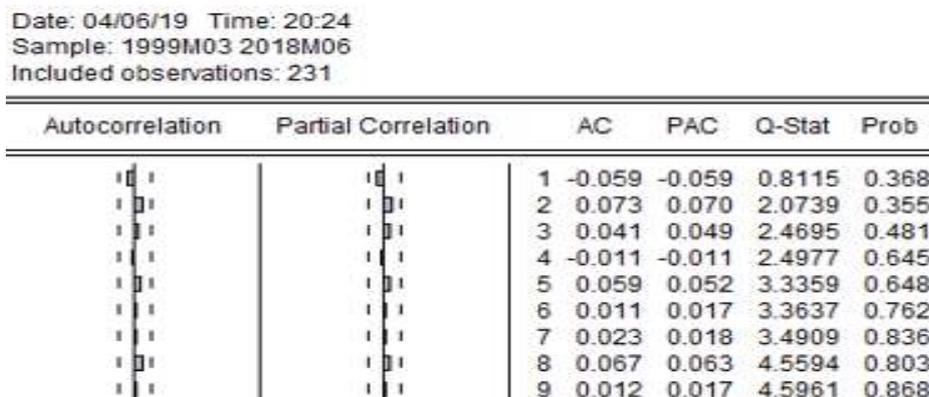
**Table 1.**

Comparison among the best alternative for multiplicative seasonal ARIMA models

Candidate ARIMA models SRIMA(p,i,q)(P,I,Q)12	AIC	SIC	HQIC
SARIMA(1,1,2)(1,1,0)12	-4.62	-4.57	-4.60
SARIMA(1,1,1)(0,1,1)12	-2.48	-2.16	-2.73
SARIMA(1,1,1)(1,1,0)12	-2.73	.227	-2.39
SARIMA(1,1,1)(1,1,1)12	-1.02	-1.38	-1.85
SARIMA(2,1,1)(0,1,1)12	3.71	3.86	3.97
SARIMA(2,1,1)(1,1,1)12	5.29	5.77	5.63
SARIMA(1,1,2)(1,1,1)12	-3.28	-3.01	-3.12

In order to examine volatility in the time series, an ARCH test has been conducted. The amount of the Fisher statistics ( $f = 2.8$ ) and  $\chi^2 = nR^2 = 3.04$  indicate that data has a constant variance and variance heterogeneity was not observed in them. Further analysis on time-series models also indicate that residuals has a normal distribution with zero mean and a constant residual variance of 0.13. Figure 4 presents the ACF and PACF of residuals that reveals no significant spikes (All p-values are greater than 0.05). Consequently there is no need for further modeling efforts and calibrated model attained as equation (14).

$$y_t = -0.084y_{t-1} - 0.329y_{t-1} - 0.479u_{t-2} - 0.479y_{t-12} + \varepsilon_t \tag{14}$$



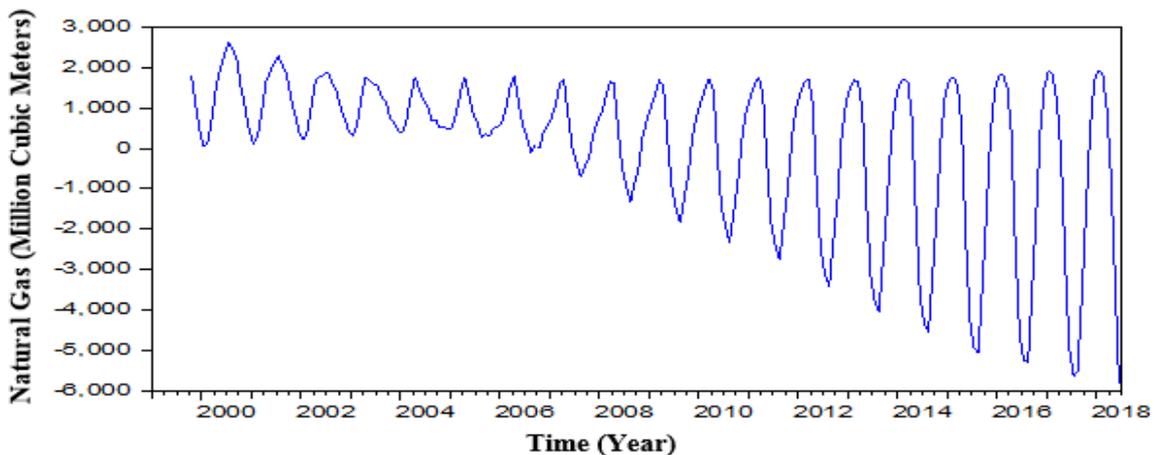
**Figure 4.** The ACF and PACF of residuals for the best SARIMA model

**b) Fitting a Fourier Series Based Model**

In order to fit the most appropriate Fourier model on data collection, graphical interface curve fitting of MATLAB was used. Equation (15) presents the fitted Fourier series for the historical data.

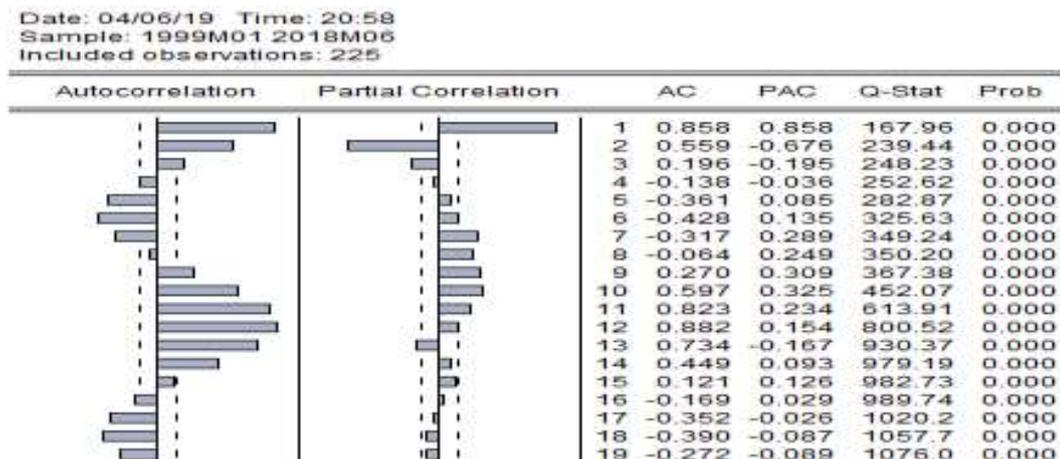
$$y_t = 2412 + 1423 \text{Cos}(0.5235t) - 1645 \text{Sin}(0.5235t) + \varepsilon_t \tag{15}$$

Such curve has R square adjusted of 0.99 that is a very close to one. But evidence that suggests significant internal solidarity is evident in the residuals of the model as depicted in Figure 5.



**Figure 5.** Residuals of gas consumption time series after fitting the best Fourier series

In order to complete the modeling process and calibrating the Fourier model, the error term should be modeled using an ARIMA terms. Consequently, the ACF and PACF graph for residuals of equation 18 is provided in Figure (6) which reveals significant spikes.



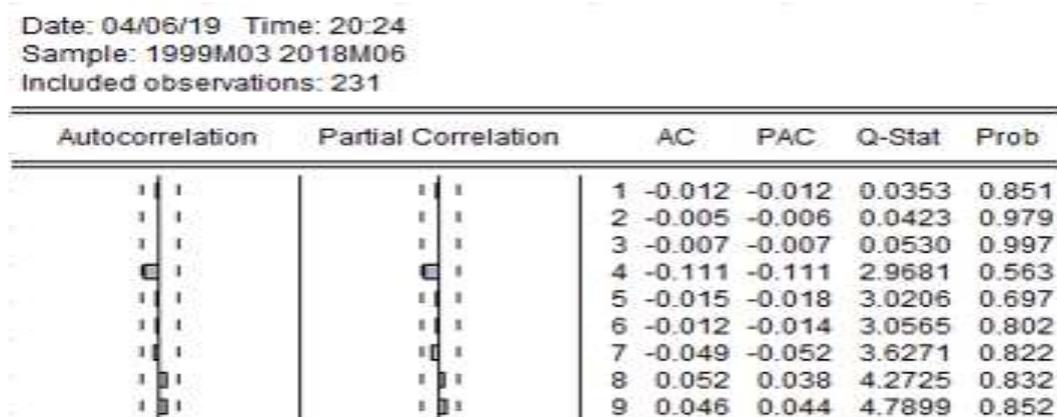
**Figure 6.** The ACF and PACF of gas consumption time series after fitting the best Fourier series

The Dickey-Fuller test with statistic value of  $ADF = 1.26$  shows the need for detrending and de-seasonal process. A similar comparison such as table 1 reveals that the multiplicative seasonal ARIMA model in the form of SARMA(1,2)(0,1)12 has best alternative to examine the Fourier error terms. Also the ARCH test indicated that there are significant volatility in the residuals. Consequently, We could not considered that the residuals has a constant variance. The best model to present the conditional variance introduced by Eviews in the form of GARCH(1,1). Here the ultimate residuals statistically normally distributed with zero mean and a constant variance of 0.5. Based on the ACF and PACF presented in figure (7) we concluded that the calibration model attained in the form of a combination set of the aforementioned Fourier series and SARMA(1,2)(0,1)12 Box-Jenkins model could present the mean of time series and a conditional variance in the form of GARCH(1,1). So another calibrated model for gas consumption presented by the set of equation (16).

$$y_t = 2412 + 1423 \cos(0.5235t) - 1645 \sin(0.5235t) + 0.772y_{t-1} - 0.266u_{t-1} + 0.344u_{t-2} + 0.531u_{t-12} + \varepsilon_t ; \varepsilon \sim N(0,0.5)$$

$$\sigma_t^2 = 325.34 + 1.31u_{t-1}^2 + 2.37\sigma_{t-1}^2 \tag{16}$$

In this case the final remnants are the normal distribution model character of  $\varepsilon \sim N(0,0.5)$ . Figure 7 shows autocorrelation function and partial autocorrelation final waste residues graph that indicates variance is stable.



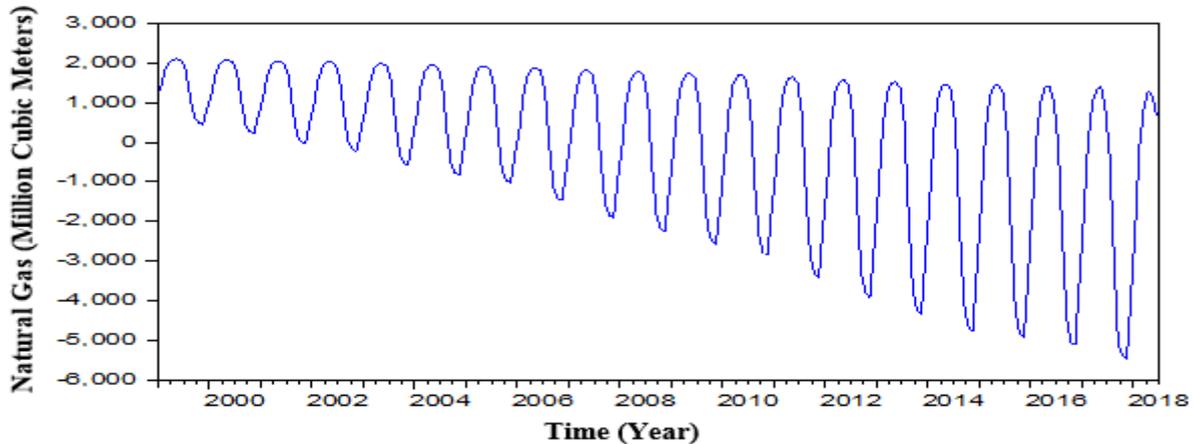
**Figure 7.** The ACF and PACF of ultimate residuals after fitting a combination of Fourier-SARIMA- GARCH model

**c) Fitting an Exponential Based Model**

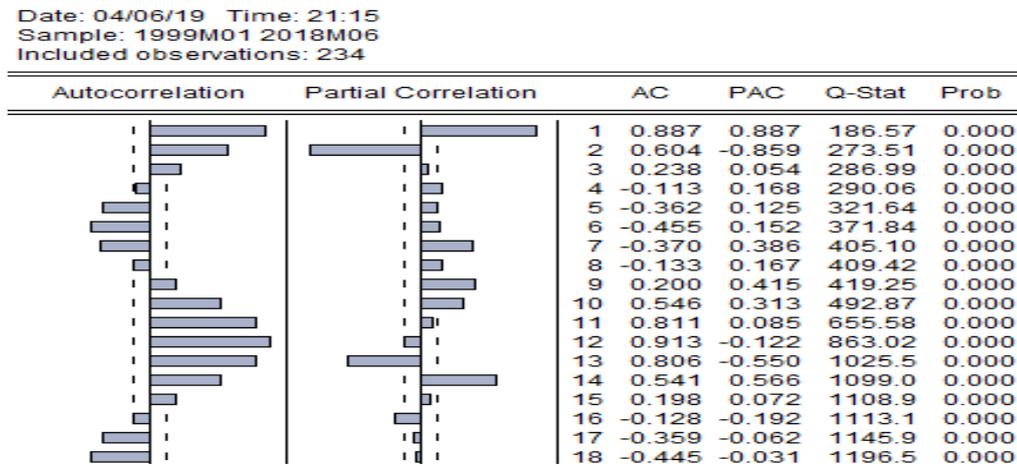
A curve fitting process based on a series of exponential model is followed using the MATLAB package. The best fit estimated by equation (17).

$$y_t = 1149e^{0.000255t} + \varepsilon_t$$

(17) Here adjusted determination factor of 0.97 is very close to one. But significant evidence on residual graph (Figure 8) presents a negative trend beside a cyclic pattern.



**Figure 8.** The residuals of gas consumption time series after exponential curve fitting. The ACF and PACF depicted on figure 11 beside the Dickey-Fuller test with statistic value of  $ADF = 2.22$  shows the need for de-trending and de-seasonal process.



**Figure 9.** The ACF and PACF of gas consumption time series after fitting the best exponential

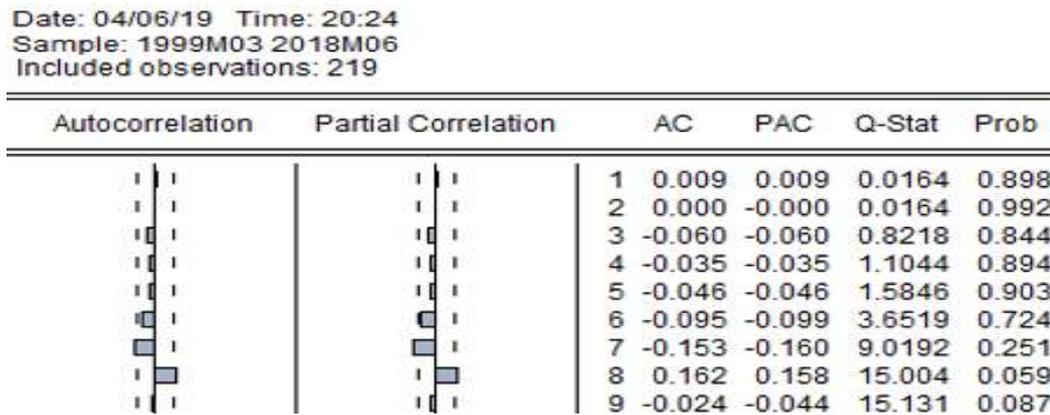
A similar approach conducted us to model the exponential error terms using SARIMA(1,0,1)(1,1,0)12 minimum AIC metrics among other ARIMA models. Also results of ARCH tests presented a significance evidence using the  $F = 13.52$  and the amount of  $\chi^2 = nR^2 = 11.7$  which is significant large. Consequently a GARCH model fitted to present conditional variance. Eviews conducted us for GARCH(1,1) as the best model. Here the final residual analysis exposed a normally distributed residuals with zero mean and a constant variance of 0.6. Therefore another calibrated fitness function based on exponential curve fitting could be presented by equation 18.

$$y_t = 1149e^{0.000255t}(1.07y_{t-1} - 0.99u_{t-1} + 0.679y_{t-12}) + \varepsilon_t \quad ; \quad \varepsilon \sim N(0, 0.6)$$

(18)

$$\sigma_t^2 = -0.265 + 0.79u_{t-1}^2 + 0.63\sigma_{t-1}^2$$

Figure 10 presents that there is no significant evidence to reject the calibrated model.



**Figure 10.** The ACF and PACF of ultimate residuals after fitting a combination of Exponential-SARIMA- GARCH model

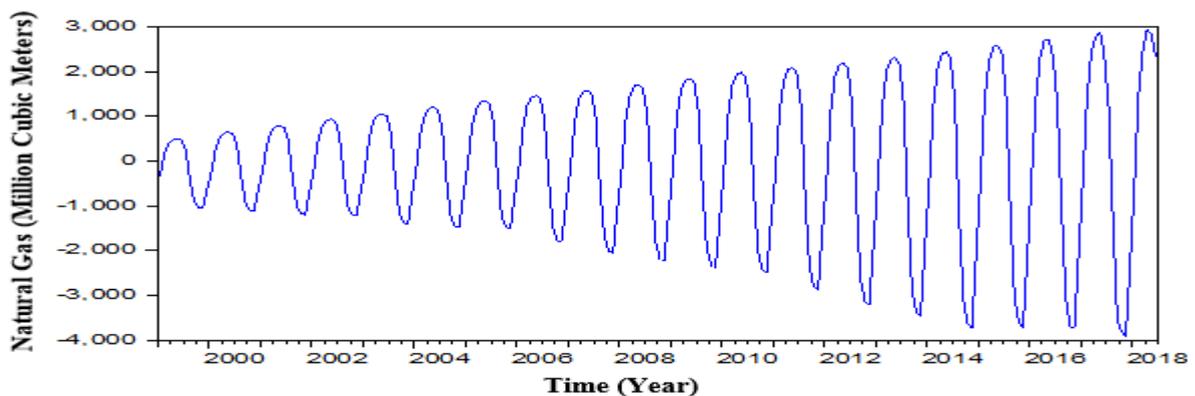
**d) Fitting a Quadratic Based Function Model**

There are many kind of polynomial form in curve fitting toolbox of MATLAB. Among them quadratic function has the highest rate of adjusted R square of 0.977 which is close to one. Here we accept it as a base for fitness function. Equation 19 presents the best fitted quadratic curve for gas consumption time series.

$$y_t = 6731.1 + 14.38x + 0.00646 + \varepsilon_t$$

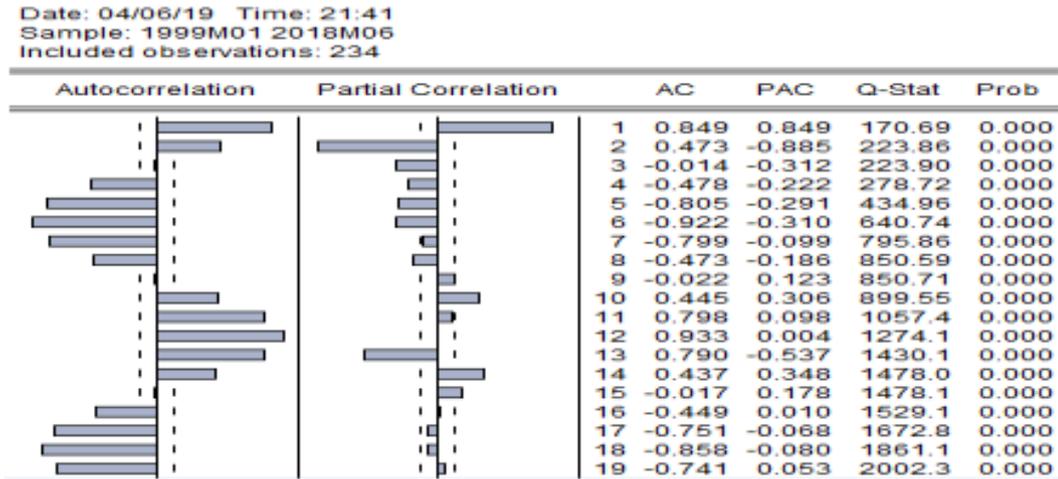
(19)

Figure 10 shows the fitness function error terms over time and reveals evidence on periodicity and trend.



**Figure 11.** The remaining graph binomial model

The ACF and PACF depicted on figure 11 beside the Dickey-Fuller test with statistic value of  $ADF = 0.74$  shows the need for de-trending and de-seasonal process.



**Figure 12.** The ACF and PACF of residual after quadratic function fitting

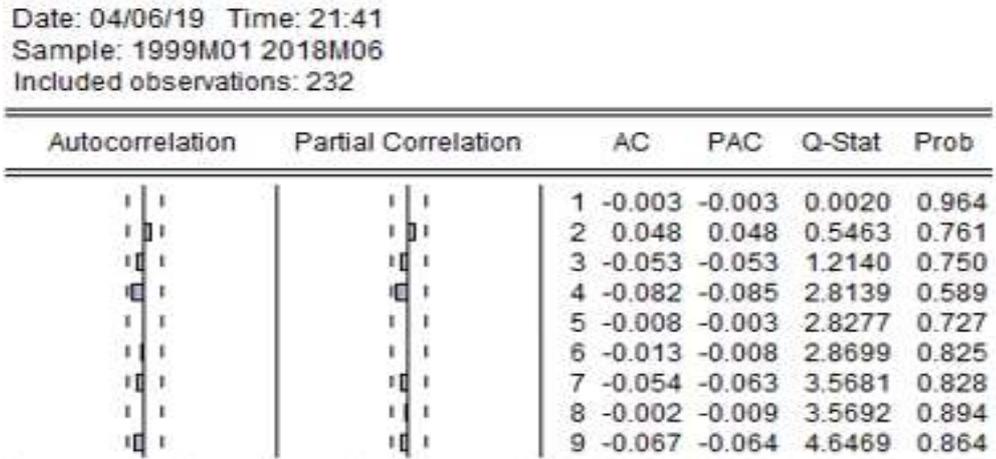
A comparison analysis based on AIC, SC and HQC conducted us to accept SARIMA (1,1,1)(1,1,0)12 as the best Box-Jenkins model. Also ARCH test with  $F = 13.77$  and  $\chi^2 = nR^2 = 13.10$  reveals significant evidence on process volatility. The best estimated model for conditional variance is GARCH(1,1). Also goodness of fit test make known that final residuals has normal distribution around zero with constant variance of 0.5. Consequently one more alternative fitness function based on quadratic function presented by equation 20.

$$y_t = 6731.1 + 14.38x + 0.00646x^2 + 0.16y_{t-1} + 0.9u_{t-1} + 1.04y_{t-12} + \varepsilon_t$$

$$; \varepsilon \sim N(0, 0.5)$$

$$\sigma_t^2 = 110.41 + 1.77u_{t-1}^2 + 0.32\sigma_{t-1}^2 \tag{20}$$

Figure 12 shows a statistical evidence for calibrated model based on a combination set of quadratic function, SARIMA(1,1,1)(1,1,0)12 and GARCH(1,1).



**Figure 13.** The ACF and PACF of last residuals after fitting a combination of Parabolic-SARIMA- GARCH model

### 5. Fitness Functions verification

As presented in section 4, there are four alternative for forecasting the Iranian gas consumption in household sector. All of them statistically calibrated to historical data sets and has good condition for forecasting process. In order to find the most appropriate method in term of best goodness of fit three most common metrics designated. They are Average absolute error (MAE), average absolute percentage error (MAPE) and the Theil inequality coefficient (TIC). Suppose  $y_t$  stands for observed values and  $y_t^f$  for the relevant forecasting values. Here any difference between the actual and estimated value reveals an error term. Here the aforementioned metrics defined by equation 21, 22, and 23.

$$MAE = \frac{(\sum_{t=T+1}^{T+m} |y_t^f - y_t|)}{m} \tag{21}$$

$$MAPE = \left( \frac{\sum_{t=T+1}^{T+m} \frac{|y_t^f - y_t|}{|y_t|}}{m} \right) 100 \tag{22}$$

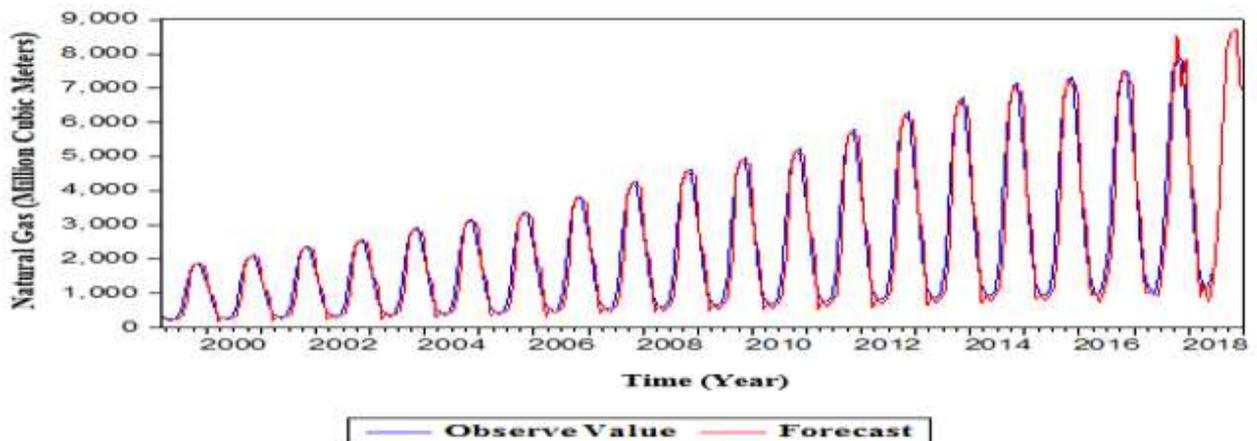
$$TIC = \frac{\sqrt{\frac{\sum_{t=T+1}^{T+m} (y_t^f - y_t)^2}{m}}}{\sqrt{\frac{\sum_{t=T+1}^{T+m} (y_t^f)}{m}} \sqrt{\frac{\sum_{t=T+1}^{T+m} (y_t)}{m}}} \tag{23}$$

All of smaller-the-better criterion could present the accuracy of forecasting model. Table 2 presented the amount of the said criteria beside the last residual variance estimates ( $\hat{\sigma}_e^2$ ). Obviously the best alternative for forecasting has minimum values in rows.

**Table 2.**  
Comparison of models and identifying an appropriate model to predict

Candidate Model	MAE	MAPE	TIC	$\hat{\sigma}_e^2$
Multiplicative Seasonal ARIMA model	0.01	143.97	0.36	0.13
Fourier based model	23.93	344.42	0.71	0.5
Exponential based model	23.36	263.47	0.82	0.6
Quadratic function based model	17.17	887.35	0.67	0.5

According to the table one can conclude that a multiplicative seasonal ARIMA model in the form of SARIMA (1,1,2)(1,1,0)<sub>12</sub> has the best condition to predict future consumptions. Table 3 presents the last 6 months observation (with gray shade) beside prediction value for a period of the next 6 months based on the best fitness function. Figure 13 presented both actual and predicted values simultaneously.



**Figure 14.**

The forecast for future natural gas demand in home section

Month/Year	Observed value	Prediction value	Deviation
April/2018	4579.68	4581.22	1.54
May/2018	2512.38	2513.24	0.86
June/2018	1552.01	1555.55	3.54
July/2018	1111.01	1121.39	10.38
Aug./2018	1244.44	1244.87	0.43
Sep. /2018	1731.27	1732.88	1.61
Oct./2018	3008.65	3009.85	1.2

Nov./2018	5700.94	5700.04	0.9
Dec./2018	8215.37	8213.97	1.4
Jan./2019	8611.75	8612.4	0.65
Feb./2019	8727.13	8726.15	0.98
Mar./2019	6966.84	6969.94	3.1

## 6. Conclusion

The objective of the presented study was to prepare an appropriate fitness function to examine the residents' gas demand in Iran regardless of the influential variables. Historical monthly data set embedded trend, seasonal and irregular pattern as well volatility. Among the well-known technical forecasting method, four types fitted to data and promoted them to reach to a calibrated model. All of the final fitness functions checked on adequacy to deploy from a Normal statistical distribution with zero mean and a constant variance. Among the candidate model, four model adequacy indices revealed that a multiplicative seasonal ARIMA model in the form of SARIMA (1, 1, 2) (1, 1, 0)<sub>12</sub> had the best condition to predict resident gas demand. It was found that application of such Box-Jenkins time series model represents more reliable forecasting figures. The proposed model decisively covers maximum goodness of fit measures in terms of in terms of MAE, MAPE, TIC and error variance. A precision comparison study for a period of six recent months showed that the nominated model has a positive deviation on the range of [0.90-3.1] million cubic meters from the actual figures. Note that's such small positive bias took place for the past recent six months and embeds a conservator estimation for planners.

## References

Oil Ministry, the Institution of International Energy Studies, a twenty-year master plan for natural gas of Iran.

## Bibliography

- Akmal, M., & David, S. (2001). Residential energy demand in Australia: An application of dynamic OLS. *Australian Bureau of Agricultural and Resource Economics, Australian National University*, WP 0101.
- Arac, N., & Aras, H. (2004). Forecasting residential natural gas demand. *Energy Source* 26, 463-476.
- Azadeh, A., Asadzadeh, S., Saberi, M., Nadimi, V., Tajvidi, A., & Sheikalishahi, A. (2011). A Neuro-fuzzy-Stochastic frontier analysis approach for Long-term natural gas

- consumption forecasting and behavior Analysis: The cases of Bahrain, Saudi Arabia, Syria, and UAE. *Applied Energy* 88, 3850-3859.
- Box G.E.P., Jenkins G.M. (2015). *Time series analysis: forecasting and control* (3rd ed.). USA: Englewood Cliffs, Prentice-Hall.
- Dagher, L. (2011). Natural gas demand at the utility level: An application of dynamic elasticities. *Energy Economics* 34, 961-969.
- Dilaver, O., Dilaver, Z., & C. Hunt, L. (2014). What drives natural gas consumption in Europe? Analysis and projections. *Journal of Natural Gas Science and Engineering* 19, 125-136.
- Ediger, V., & Akar, S. (2007). ARIMA forecasting of primary energy demand by fuel in Turkey. *Energy Policy* 35, 1701-1708.
- Faheemullah Shaikh, Q. J. (2016). Forecasting natural gas demand in China: Logistic modelling analysis. *Electrical Power and Energy Systems*, 25-32.
- Faheemullah Shaikh, Qiang Ji, Pervez Hameed Shaikh, Nanyar Hussain Mirjat, Muhammad Aslam Uqaili. (2017). Forecasting China's natural gas demand based on optimised nonlinear. *Energy*, 941-951.
- Filik, U., Gerek, O., & Kurban, M. (2010). A novel modeling approach for hourly forecasting of long-term electric energy demand. *Energy Conversion and Management* 52, 199-211.
- Forouzanfar, M., Doustmohammadi, A., Bagher Menhaj, M., & Hasanzadeh, S. (2009). Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran. *Applied Energy* 87, 268-274.
- Huntington, G. (2007). Industrial natural gas consumption in the United State: An empirical model for evaluating futur trend. *Energy Economics* 29, 743-759.
- Kani, A., Abbaspour, M., & Abedi, Z. (2013). Estimation of demand function for natural gas in Iran: Evidences based on smooth transition regression models. *Economic Modelling* 36, 341-347.
- Kiani, B., & Pourfakhraei, M. (2010). A system dynamic model for production and consumption policy in Iran oil and gas sector. *Energy Policy* 38, 7764-7774.
- Kovacic, M., & Sarler, B. (2014). Genetic programming prediction of the natural gas consumption in a steel plant. *Energy*, 1-12.
- Liu, Q., & Kaboudan, M. (2003). Forecasting quarterly US demand for natural gas. *Energy* 26, 25-31.
- Majazi Dalfard, V., Nazari Asli, M., Asadzadeh, S., Sajjadi, S., & Nazari Shirkouhi, A. (2012). A mathematical modeling for incorporating energy price hikes into total natural gas consumption forecasting. *Applied Mathematical Modelling* 37, 5664-5679.
- Melikoglu, M. (2013). Vision 2023: Forecasting Turkey's natural gas demand between 2013 and 2023. *Renewable and Sustainable Energy Reviews* 22, 393-400.
- Pourazarm, E., & Cooray, A. (2013). Estimating and forecasting residential electricity demand in Iran. *Economic Modelling* 35, 546-55.
- Salehnia, N., Falahi, M., Seifi, A., & Mahdavi Adeli, M. (2013). Forecasting natural gas spot prices with nonlinear modeling using Gamma test analysis. *Journal of Natural Gas Science and Engineering* 14, 238-249.
- Szoplik, J. (2015). Forecasting of natural gas consumption with artificial neural networks. *Energy*, 208-220.
- Taspinar, F., Celebi, N., & Tutkun, N. (2012). Forecasting of daily natural gas consumption on regional basis in Turkey using. *Energy and Buildings* 56, 23-31.
- Wadud, Z., Dey, H., AshfanorKabir, M., & Khan, S. (2011). Modeling and forecasting natural gas demand in Bangladesh. *Energy Policy* 39, 7372-7380.

- Wang, T., & Lin, B. (2013). China's natural gas consumption and subsidies—From a sector perspective. *Energy Policy* 65, 541-551.
- Wei Zhang, J. Y. (2015). Forecasting natural gas consumption in China by Bayesian Model. *Energy Reports*, 216–220.
- Xu, G., & Wang, W. (2010). Forecasting China's natural gas consumption. *Journal of Natural Gas Chemistry* 19, 493-496.
- Ying Chen, Wee Song Chua, Thorsten Koch. (2018). Forecasting day-ahead high-resolution natural-gas demand and supply in. *Applied Energy*, 1091-1110.
- Yu, Y., Zheng, X., & Han, Y. (2014). On the demand for natural gas in urban China. *Energy Policy* 29, 1-7.