International Journal of Advanced Statistics and Probability, 3 (1) (2015) 115-125
www.sciencepubco.com/index.php/IJASP
©Science Publishing Corporation
doi: 10.14419/ijasp.v3i1.4635
Research Paper

Selection of forecast model for consumption (four sectors) and transmission (two Piplines) of natural gas in Punjab (Pakistan) based on ARIMA model

Maliha Butt

Department of Statistics, GC University, Kachery Road, Lahore, Pakistan
E-mail: maliha.hotiana14@gmail.com

Copyright © 2015 Maliha Butt. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

The main purpose of this study is to select an appropriate forecast model for Natural Gas Consumption and Transmission System. For ARIMA model, Box-Jenkins Approach (1976) has been adopted i.e. Stationarity of the series has been checked for each data set, correlogram has been estimated for identification of order of ARIMA model and a class of models has been estimated. Then, most adequate and appropriate model is selected by analyzing diagnostics checks. Later on, by comparing values of Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), and Standard Error (S.E.) of Regression, Root Mean Square Error (RMSC) and Theil Inequality Coefficient (TIC) for each model, forecast model has been finalized. In the end, forecasts have been made using models and compared these forecast values with the actual values for 2010 in order to check the accuracy of the model.

Keywords: *ARIMA Model; Box Jenkins [3] Approach; Consumption of Natural Gas; Time Series Forecasting; Transmission of Natural Gas.*

1. Introduction

Natural gas is an important source of energy and it has many-sided ability. It is commonly used in domestic applications, in power plants for electricity, in automobiles as fuel, by commercial users and for manufacturing purpose etc. Natural Gas Consumption means total amount of gas consumed from all sources in a year. Transmission System is that in which we receive natural gas from the producers and by using transmission network we forward it to different city stations so that it further supply to distribution departments from where it is sold and used by different sectors such as industry, household, power generation, commercial etc.

As, the demand for natural gas is increasing day by day because of which now-a-days Pakistan is facing a shortfall. This shortfall in Pakistan is also expected in coming years too. Pakistan's daily needs of Natural Gas is 6.5 Bcf (approx.) but the overall supply is of 4 Bcf (approx.), this shows that Pakistan is facing shortfall of 2.5 Bcf (approx.). Bailly [2] said in Islamabad in his report that Pakistan is going to face a natural gas shortage from 2007, and it will happen till 2025. The report also highlighted that if proper recovery steps will not be taken by the country then this shortage will become worse in next two decades. The report was correct as this is true that we are facing that shortfall now-a-days.

Turkmenistan, Afghanistan, Pakistan and India signed an agreement with Asian Development Bank (ADB) on December 11th, 2010, namely; Tapi Gas Project, in order to overcome their shortfalls. In 2010, Pakistan signed agreement entitled Indo-Pak-Iran Pipeline Project with Iran regarding pipeline as well. If this project will be completed on time in 2014 (which is not possible according to present political situations) than it will help us to recover from our shortfall of natural gas, although, shortfall will still exist but that will be less than today's.

Indo-Pak-Iran pipeline is economically beneficial for Indian's energy demand as Indian policy makers are making efforts to concentrate on Import of natural gas. But for this purpose, pipeline would have to pass through Pakistan land and India is not ready to accept the role of Pakistan in their project as in case of military fight Pakistan can stop their gas supply. These three countries did not share goals so the project has not been come to reality. Pandian [11] stated that aim of India in this pipeline project is high but its economic interests in this pipeline project are not according to the

political, economic and strategically goals of Iran and Pakistan, due to which this project is in delay. Sen [13] informed that growth of natural gas in both the countries would depend on import of gas but cooperation has not been appearing in India and Pakistan, exploration tasks and success of their projects depends on cooperation between these two countries. Possible effects of environmental value of natural gas and transformation of natural gas to market is also important to consider (Radetzki [3]).

Demand of natural gas has been increased in many other countries as well. Wadud et al. [15] gave a dynamic econometric model in order to understand the demand system of natural gas of Bangladesh and concluded that natural gas in Bangladesh will be high in future. Huntington [8] developed a statistical model of consumption of natural gas in Industry in U.S. He concluded on the basis of his demand framework that over the next 20 years, industrial natural gas consumption will be slow if U.S will not start any project. Xu and Wang [17] used Polynomial Curve and Moving Average Combination Projection model for forecasting 2009-2015 years for consumption of natural gas in China. According to his proposed model, on the average, it was concluded that annual growth rate will increase in coming 7 years. Forouzanfar et al. [6] used a logistic approach in order to make forecast by using yearly and seasonal data of consumption of natural gas of previous 10 years of residential and commercial sectors in Iran. Using the model, he made prediction for next three years. Erdogdu [5] focused on the features of demand of natural gas and estimated short and long-run price and income elasticity of demand of natural gas (sector-wise) in Turkey. By focusing such demand, ARIMA modeling procedure was used for forecasting and then results were matched with official plans.

As far as transmission of natural gas is concern, maintenance for Transmission System and forecast of Transmission Network is also required if consumption is increasing day by day. Han and Weng [7] showed their work in which for safe functioning of supply system of natural gas (as there is always a chance of accident in such cases), for pipeline network of natural gas, they used method of Integrated Quantitative Risk analysis. In this technique, they worked on probability assessment of accidents, analyzed previous results (included both of outside and inside gas pipeline) and evaluated risks. And in result, for pipeline network of natural gas, this method or technique considered to be applicable. Simonoff et al. [14] investigated risk measures and scenario to know about the relation between reasons and happening features with consequences of failures of pipeline. Risk management actually reduces pipeline incidents. It was found in conclusion that for risk management, the important feature of incidents can be different and may depend on the incident which includes the model of type of distribution or transmission pipeline. This method helps the decision makers to give ways in order to get good know-how regarding cost consequence measures which depends on reasons and type of incident.

Based on the past data, in time series analysis approach, ARIMA modeling using Box-Jenkins [3] methodology/approach can be useful to shape such a model, which can be used to estimate upcoming values. We choose the ARIMA model for predicting and forecasting purpose (Edward et al. [4]). Many researchers have applied this methodology (see Nwosa and Oseni [10]; Lee et al. [9]; Alan and Ahiakpor [1]; Wong et al. [16]).

We conducted this investigation in order to see that what forecast model we can use for the purpose of forecasting Consumption pattern and Transmission System of Natural Gas in Punjab (Pakistan). It is necessary to forecast the Consumption and Transmission System of Natural Gas so that we can have an applicable forecast model which helps us to know that what will be the coming target to achieve related to issue of short fall.

2. Research methodology and data description

Source of data is Bureau of Statistics (Punjab). Data has been gathered by all volumes of the publication namely; Punjab Development Statistics by Bureau of Statistics (Punjab). Yearly data of Punjab (Pakistan) on Natural Gas Consumption (in thousand Decca cubic meter) and Natural Gas Transmission System (in kilometer) has been taken. This data is of Sui Northern Gas Pipeline Company, Ltd. The data of Sui Northern Gas Pipeline Company Transmission System (total of two types of pipelines; Mainline and Loopline) includes years; 1979-1980 to 2009-2010 and the data of Sui Northern Gas Pipeline Company Consumption (total of four sectors; Industrial, Commercial, Household and Others) pattern includes years; 1971-1972 to 2009-2010. Consumption group of "Other" include gas carried out for other companies through Sui Northern Gas Pipeline Company too.

Forecast means prediction for any value which can be expected for future time. In time series forecasting, there is an assumption that past pattern and behavior of the variable will be continued in future. So, a model has been used to give prediction for future. Different methods can be used for forecasting in time series; ARIMA (p, d, q) model using Box Jenkins Approach (1976) has been used for this purpose.

2.1. Time series analysis

It is such an analysis in which a stochastic process (when the outcome of variable Y cannot be predicted with certainty than variable Y is called a random variable. If we have random variable Y than set of random variables $(y_1, y_2, ...)$ is a probabilistic series or in general a stochastic series. Set of variables related with its highlighted probability distribution is called probabilistic process or in general called stochastic process) is used to fit to time series data. There are different types of stochastic processes commonly use for model building and forecasting. Uni-variate time series models which can be used for forecasting and represent different stochastic processes are given as; Auto-Regressive $AR(p)$ model and

Moving Average MA(q) model. Combination of Auto-Regressive and Moving Average model gives Auto-Regressive Moving Average ARMA (p, q) model and combination of Auto-Regressive, Integrated and Moving Average model gives Auto-Regressive Integrated Moving Average ARIMA(p, d, q) model. After converting non-stationary data into stationary using difference transformations (if necessary), particularly ARIMA technique for model building and forecasting is used.

2.2. Auto-regressive AR(p) model

In Auto-Regressive (AR) model, the present values are related to previous values itself varying by time lag. So, AR model shows the forecast as a function of previous values of time series. let ε_t is a purely random process with variance σ_ε and zero mean then the process Y_t of higher order regressive model or Auto-Regressive model of order p is given as;
$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$ Where, Y_t is the dependent variable, ϕ is a parameter, $\phi_1, \ldots, \phi_p$ are regression coefficient, ε_t is the residual term or white noise. Y_{t-p} Is an independent variable and treated as the previous value at specific time period.
The process is stationary, if parameter $|\phi_i|(\text{for } i = 1,2, \ldots, p)$ is less than 1 or graphically the root should lie inside the unit circle.

2.3. Moving average MA(q) model

There are two concepts of Moving Average. First, mean of last q observations in time series is known as the moving average of order q. Second, Moving Average method given by Box and Jenkins (1976) is define as; the values of time series are influenced by the current error term and weighted error terms in past. Let ε_t is purely random process with variance σ_ε^2 and zero mean then the process Y_t of higher order Moving Average MA(q) model or Moving Average MA(q) model of order q is given as;
$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$ Where, Y_t is the dependent variable, θ is a parameter, $\theta_1, \ldots, \theta_q$ are weights, ε_t is the residual term or white noise. $\varepsilon_{t-1} \ldots \varepsilon_{t-q}$ Are previous residual terms or white noise. q is the order of model. Moving average MA(q) can also be defined as; $Y_t - \mu = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$ This is model is most commonly used.
The process is invertible, if parameter $|\theta_i|(\text{for } i = 1,2, \ldots, q)$ is less than 1 or graphically the root should lie inside the unit circle.
The transformation of MA to AR model is called invertibility condition and AR to MA is called stationarity condition.

2.4. Auto-regressive moving average ARMA(p, q) model

If AR and MA are not separately suitable for forecasting and modeling than a combination of it called Auto-Regressive Moving Average Model or mixed model can be used. This model with q Moving Average terms and p Auto-Regressive terms can be given as;

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

Furthermore;
If $q = 0$ then $\text{ARMA}(p, q) = \text{AR}(p)$
If $p = 0$ then $\text{ARMA}(p, q) = \text{MA}(p)$

2.5. Auto-regressive integrated moving average ARIMA(p, d, q) model

When time series is not stationary and data is not seasonal then difference transformation can be used to convert the non-stationary data into stationary in order to proceed for model building and forecasting. If Y_t is changed into $\Delta^d Y_t$ by taking first, second or higher order difference and if aim is to build model for such data then $ARMA(p, q)$ will be changed into $ARIMA\ (p, d, q)$ model; d is the order of difference. The objective was to estimate such a model which can be interpreted when sample data will be generated. If this model is used for forecasting then it should assume that the structure of model is constant over time and particularly over future time period. This ARIMA model was then used for forecasting of data. ARIMA (p, d, q) Was given as;

$$\Delta^d Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

2.6. Box-jenkins approach (1976)

Box-Jenkins Approach (1976) for model building and forecasting has three steps.

2.6.1. Identification of order

In this initial step, identification of values p, d, q Which are orders of ARIMA model is required. This can be done by using correlogram. First of all, for stationarity checking, there are different methods available. Some methods are based on visual inspection called subjective approach. This includes either the time plot (line graph) or correlogram by examining Autocorrelation (AC) and Partial Autocorrelation (PAC) at different lags. On the other hand, some procedures are based upon mathematical approach, which is called objective approach. For this purpose, Augmented Dickey-Fuller (Unit Root) Test has been used. If series is not stationary, then we make it stationary by using commonly used "difference" transformation and again apply the above methods for checking stationarity.

2.6.2. Model estimation

After finding the suitable values of p, d, q from correlogram, we estimate models by using different combinations of defined orders of MA(q) and AR(p). Eviews has been used for this purpose. In estimating one combination, when all the AR and MA terms using in that combination shows a significant result then that combination will be finalized as a model to proceed further.

2.6.3. Diagnostic checks

Different diagnostic checks are used to see whether the selected model is adequate or not. Diagnostic checks include; Analyze the actual values and fitted values (graphically) of estimated model. Check that process is stationary and invertible or not. Correlogram of residuals and squared residuals are used for checking of constant variance and autocorrelation, respectively. In both the cases, Q-stat of residuals and squared residuals also use to be inspected as well. Normality test is that in which it is determine that whether the residuals of the model is following Normal Distribution or not. Anderson- Darling test has been used for this purpose. Breusch-Godfrey serial correlation Lagrange Multiplier (LM) Test gives conclusion for autocorrelation among the residual terms. This test is used especially when higher-order differences exist. Autoregressive Conditional Heteroscedasticity (ARCH) Test used to test about the homoscedasticity of residual terms.
From above, analysis of correlogram of residual and squared residuals, normality test, ARCH test and LM test are residual tests. All the diagnostic checks should be fulfilled in order to select an adequate model.

2.7. Forecasting

In this step, forecast model is finalized using final estimated ARIMA models which were analyzed and finalized through diagnostic checks mentioned above.

Selection of Forecast model is based on values of Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC) and Standard Error (S.E.) of Regression. Furthermore, In order to know that how well our forecast model is as compare to other models, forecast evaluation criteria (graphical) is used. Static forecast method has been used using Eviews software. It calculates series of one-step ahead forecasts using actual values. In this, we used to analyze Root Mean Square Error (RMSE) and Theil's Inequality Coefficient (TIC) and graphical look as well.

3. Result and discussion

Final forecast models MA (1) MA (3) MA (4) at 2nd difference and ARIMA (0, 2, 6) for Natural-Gas Consumption by Four Sectors (Household, Industry, Commercial and Other) and Natural-Gas Transmission of Two Types of Pipeline (Loopline and Mainline), respectively, has been finalized, which gave closer forecast values when compared to actual ones so these were considered to be suitable forecast models for their corresponding time series (data set). Eviews and Minitab software package for the analysis purpose have been used. Detail of this analysis is given as follows.

3.1. Tests for stationarity

The time plot at actual data is not stationary for Natural-Gas Consumption and Natural-Gas Transmission System. So, Figure 1 & 2 shows that by using 2nd difference transformation for Natural-Gas Consumption and Natural-Gas Transmission System, series become stationary. Then with the help of correlogram, Figure 3 & 4, it has been realized the same as its spikes are not showing any pattern, which means series or data becomes stationary. In Table 1 & 2, Augmented Dickey-Fuller (Unit Root) Test is showing that p-value is smaller than α 1% which illustrations that null hypothesis of unit root test, i.e., series is non-stationary or series has unit root, is rejected.

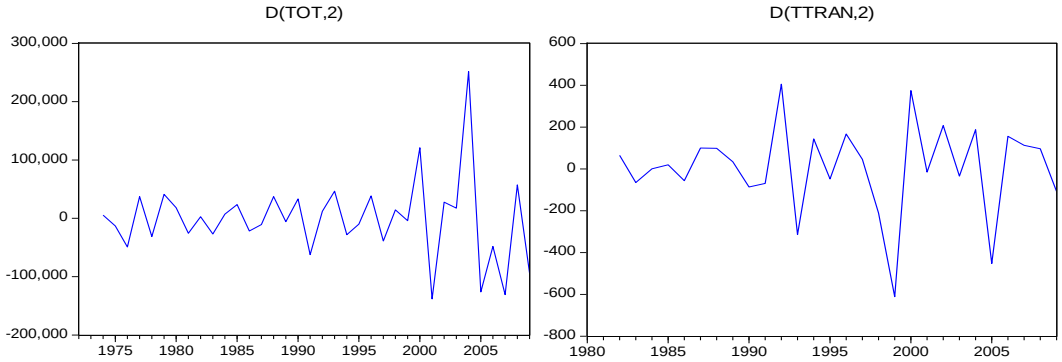
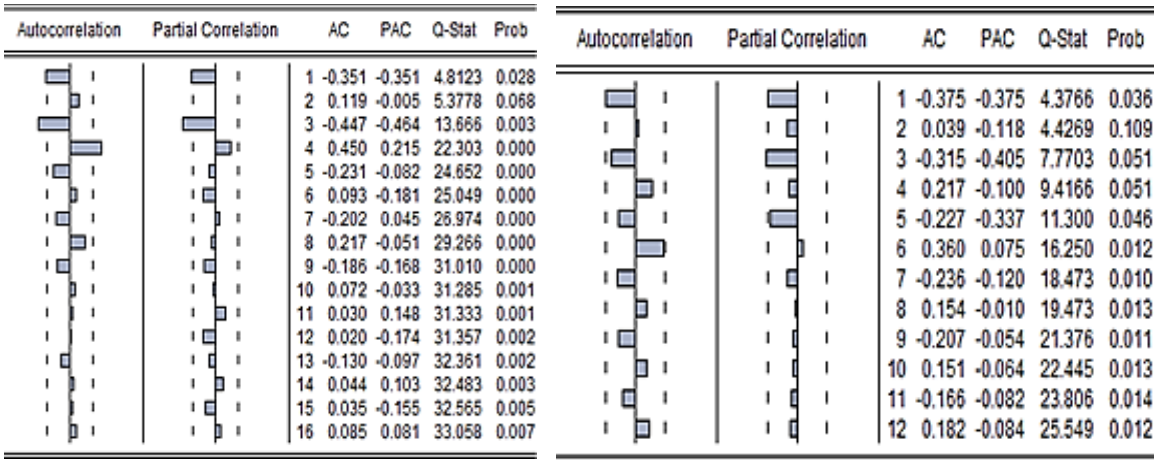


Fig. 1-2: Time Plots for Natural Gas Consumption and Natural Gas Transmission Series at 2nd Difference



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.351	-0.351	4.8123	0.028
		2	0.119	-0.005	5.3778	0.068
		3	-0.447	-0.464	13.666	0.003
		4	0.450	0.215	22.303	0.000
		5	-0.231	-0.082	24.652	0.000
		6	0.093	-0.181	25.049	0.000
		7	-0.202	0.045	26.974	0.000
		8	0.217	-0.051	29.266	0.000
		9	-0.186	-0.168	31.010	0.000
		10	0.072	-0.033	31.285	0.001
		11	0.030	0.148	31.333	0.001
		12	0.020	-0.174	31.357	0.002
		13	-0.130	-0.097	32.361	0.002
		14	0.044	0.103	32.483	0.003
		15	0.035	-0.155	32.565	0.005
		16	0.085	0.081	33.058	0.007

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.375	-0.375	4.3766	0.036
		2	0.039	-0.118	4.4269	0.109
		3	-0.315	-0.405	7.7703	0.051
		4	0.217	-0.100	9.4166	0.051
		5	-0.227	-0.337	11.300	0.046
		6	0.360	0.075	16.250	0.012
		7	-0.236	-0.120	18.473	0.010
		8	0.154	-0.010	19.473	0.013
		9	-0.207	-0.054	21.376	0.011
		10	0.151	-0.064	22.445	0.013
		11	-0.166	-0.082	23.806	0.014
		12	0.182	-0.084	25.549	0.012

Fig. 3-4: Correlogram for Natural Gas Consumption and Natural Gas Transmission Series at 2nddifference

Table 1: Augmented Dickey-Fuller (Unit Root) Test for Natural Gas Consumption Series at 2nd Difference

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6.273295	0.0000
Test critical values:	1% level	-3.646342	
	5% level	-2.954021	
	10% level	-2.615817	

*MacKinnon (1996) one-sided p-values.

Table 2: Augmented Dickey-Fuller (Unit Root) Test for Natural Gas Transmission Series at 2nd Difference

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.416408	0.0000
Test critical values:	1% level	-3.699871	
	5% level	-2.976263	
	10% level	-2.627420	

*MacKinnon (1996) one-sided p-values.

3.2. Identification of ARIMA model

For each data set, we can see from correlogram that which spikes are touching the bounds or are outside the bounds, so these orders (spikes) of AR and MA is used to be chosen from AC, and PAC lags, respectively, in order to proceed for estimation of models. From different combinations of selected AR and MA orders, different models (by using the least square method) is used to obtain, and by going through the process of different diagnostic checks, models are finalized to proceed for selection of forecast model. Summary of selected spikes of AR and MA (selected from AC and PAC lags), selected models to final forecast model is shown in Table 3.

We have selected models by analyzing p-value for MA terms and AR terms at their respective orders. Table 4 & 5 shows that estimated models are statistically significant i.e., smaller than α 1%. So that's why these models have been selected. Table 4 & 5 gave the models which were estimated and considered as the final forecast models at the end.

Table 3: Selected Spikes, Models and Forecast Model for Natural Gas Consumption

Data Type	Identification of models using Correlogram	Models selected	Models Finalized using Diagnostic Checks	Forecast Model
Natural Gas Consumption (total of four sectors)	MA(1), MA(3), MA(4), AR(1) and AR(3)	ARIMA(3,2,0), ARIMA(0,2,1), MA(1) MA(3) MA(4), ARIMA(0,2,3), MA(3) MA(4) AR(3) and ARIMA(0,2,4)	MA(1) MA(3) MA(4) and MA(3) MA(4) AR(3)	MA(1) MA(3) MA(4)
Natural Gas Transmission System (total of two types of pipeline)	MA(1), MA(3), MA(6), AR(1), AR(3) and AR(6)	ARIMA(0,2,1), MA(1) MA(6), ARIMA(3,2,3) and ARIMA(0,2,6)	ARIMA(0,2,1), MA(1) MA(6) and ARIMA(0,2,6)	ARIMA(0,2,6)

Table 4: Estimated MA (1) MA (3) MA (4) for Natural- Gas Consumption Series at 2nd Difference

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2421.563	1101.954	2.197517	0.0353
MA(1)	-0.584674	0.155495	-3.760079	0.0007
MA(3)	-0.900638	0.056427	-15.96104	0.0000
MA(4)	0.552078	0.148893	3.707889	0.0008
R-squared	0.533690	Mean dependent var		-2004.806
Adjusted R-squared	0.489973	S.D. dependent var		69258.08
S.E. of regression	49461.47	Akaike info criterion		24.56021
Sum squared resid	7.83E+10	Schwarz criterion		24.73616
Log likelihood	-438.0839	Hannan-Quinn criter.		24.62162
F-statistic	12.20794	Durbin-Watson stat		1.728670
Prob(F-statistic)	0.000017			

Table 5: Estimated ARIMA (0, 2, 6) for Natural- Gas Transmission Series at 2nd Difference

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	26.35741	61.64226	0.427587	0.6725
MA(6)	0.882277	0.039222	22.49421	0.0000
R-squared	0.358474	Mean dependent var		5.357143
Adjusted R-squared	0.333800	S.D. dependent var		216.1714
S.E. of regression	176.4414	Akaike info criterion		13.25260
Sum squared resid	809421.0	Schwarz criterion		13.34776
Log likelihood	-183.5365	Hannan-Quinn criter.		13.28169
F-statistic	14.52838	Durbin-Watson stat		2.459977
Prob(F-statistic)	0.000763			

3.3. Diagnostic checks

- In graph of actual and fitted values, actual and fitted value must follow each other.
- In ARMA structure, if all the MA roots and AR roots for the models are lying inside and at the unit circle, then it means that process is invertible and stationery, respectively.
- If spikes are inside the bounds or at the bound in correlogram of residuals then it shows that residual terms have constant variance and are purely random, and vice versa. P-values of Q-stat must be greater than α 1%; it will mean that no autocorrelation exists.
- In correlogram of squared residuals, if spikes are inside the bounds then squared residual terms do not have autocorrelation, on the other hand; p-values of Q-stat must be greater than α 1%, it will mean that no autocorrelation exists.
- For normality checking of residuals, p-value of Anderson-Darling Test is use to check. If p-value less than 1% which mean null hypothesis i.e. residuals follows normal distribution, is rejected.
- In LM test, p-value of F-statistics must be greater than α 1% then it is use to conclude that there is no autocorrelation exists.
- P-value of F-statistics must be greater than α 1% so it can conclude that there is homoscedasticity in residual terms, which means there is no ARCH effect and residuals have constant variance.

Summary of the all estimated models is given in Table 6 and 7. This Table shows that which model satisfied the diagnostic check and finally after fulfilling all diagnostic checks which model is considered as final.

Table 6: Diagnostic Check Summary for Natural Gas Consumption Series

Category		ARIMA(3,2,0)	ARIMA(0,2,1)	MA(1) MA(3) MA(4)	ARIMA(0,2,3)	MA(3) MA(4) AR(3)	ARIMA(0,2,4)
Time Plot	2nd						
Correlogram	MA(1) MA(3) MA(4) AR(1) AR(3)						
Unit Root Test	2nd						
Graph of Actual, Fitted and Residuals			√	√	√	√	
ARMA Structure		√	√	√	√	√	√
Correlogram of Residuals				√	√	√	√
P-value of Q-stat		√	√	√	√	√	√
Correlogram of Squared Residuals			√	√		√	
P-value of Q-stat		√	√	√	√	√	√
Normality Test		√		√	√	√	√
LM Test		√	√	√	√	√	√
ARCH Test		√	√	√	√	√	√

Table 7: Diagnostic Check Summary for Natural Gas Transmission Series

Category		ARIMA(0,2,1)	MA(1) MA(6)	ARIMA(3,2,3)	ARIMA(0,2,6)
Time Plot	2nd				
Correlogram	MA(1) MA(3) MA(6) AR(1) AR(3) AR(6)				
Unit Root Test	2nd				
Graph of Actual, Fitted and Residuals		√	√		√
ARMA Structure		√	√	√	√
Correlogram of Residuals		√	√	√	√
P-value of Q-stat		√	√	√	√
Correlogram of Squared Residuals		√	√	√	√
P-value of Q-stat		√	√	√	√
Normality Test		√	√	√	√
LM Test		√	√	√	√
ARCH Test		√	√	√	√

Table 6 shows that by using Diagnostics Checks MA (1) MA (3) MA (4) and MA (3) MA (4) AR (3) has been finalized for Consumption of Natural Gas series as all the diagnostic checks has been fulfilled for these two models. Similarly, for Transmission of Natural Gas series ARIMA (0, 2, 1), MA (1) MA (6) and ARIMA (0, 2, 6) models were selected (Table 7).

3.4. Forecast model

After selection of models through diagnostic checks, last step is to finalize the single model which can be used for forecast purpose. Graphically, forecast evaluation is used to be observed. Forecast values must be within the bounds. Additionally, Akaike Information Criterion (AIC), Schwarz Information Criterion (SIC), Standard Error (S.E.) of Regression, Root Mean Square Error (RMSE) and Theil Inequality Coefficient (TIC) of selected models should have smaller values as compare to other models. Table 8 and 9 are showing the summary that which model has the smallest value as compare to other selected models.

Table 8: Summary for Final Forecast Model for Natural-Gas Consumption Series

Category	MA(1) MA(3) MA(4)	MA(3) MA(4) AR(3)
Forecast Graph Evaluation	√	√
AIC	√	
SIC	√	
Standard Error of Regression	√	
Root Mean Square Error	√	
Theil Inequality Coeffeicent	√	

Table 9: Summary for Final Forecast Model for Natural -Gas Transmission Series

Category	ARIMA (0,2,1)	MA(1) MA(6)	ARIMA (0,2,6)
Forecast Evaluation Graph	√	√	√
AIC			√
SIC			√
Standard Error of Regression			√
Root Mean Square Error			√
Theil Inequality Coeffeicent			√

Table 8 shows that all the conditions are fulfill for MA (1) MA (3) MA (4) of Natural Gas Consumption series. So this model with be the final forecast model for estimating Consumption of Natural Gas, given as;

$$\Delta^2 Y_t = c + \theta_1 \varepsilon_{t-1} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4}$$

$$\Delta^2 Y_t = 2421.563 - 0.584674 \varepsilon_{t-1} - 0.900638 \theta_3 \varepsilon_{t-3} + 0.552078 \varepsilon_{t-4}$$

The coefficients of the model have been picked from Table 4.
Similarly, Table 9 shows that all the conditions are fulfill for ARIMA (0, 2, and 6) of Natural Gas Transmission series. So this model with be the final forecast model for estimating Transmission System of Natural Gas, given as;

$$\Delta^2 Y_t = c + \theta_6 \varepsilon_{t-6}$$

$$\Delta^2 Y_t = 26.35741 + 0.882277 \varepsilon_{t-6}$$

The coefficients of the model have been picked from Table 5.
The results of the diagnostic checks and graphical forecast evaluation for the final forecast model of Natural Gas Consumption have been shown in Figure 5-10 and Table 10-11.

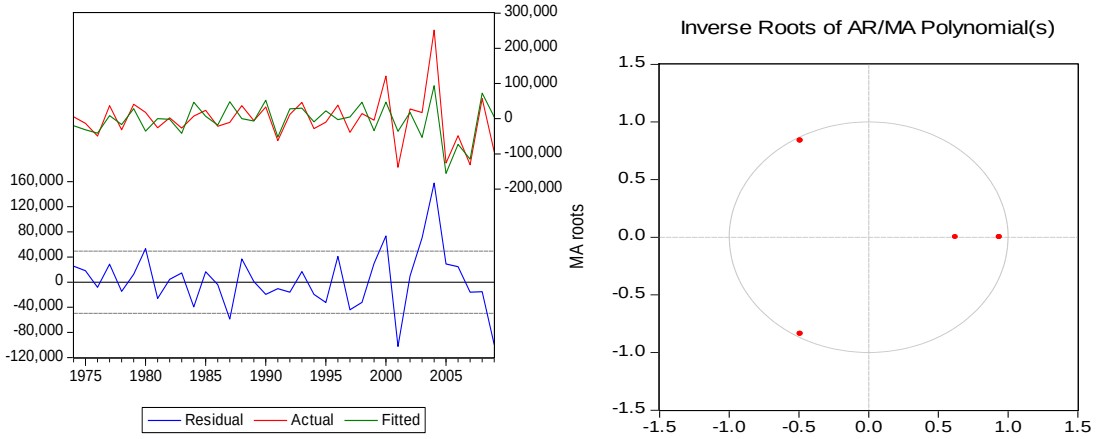


Fig. 5-6: Graph of Actual, Fitted, Residual Values and ARMA Structure for MA (1) MA (3) MA (4) of Natural- Gas Consumption Series

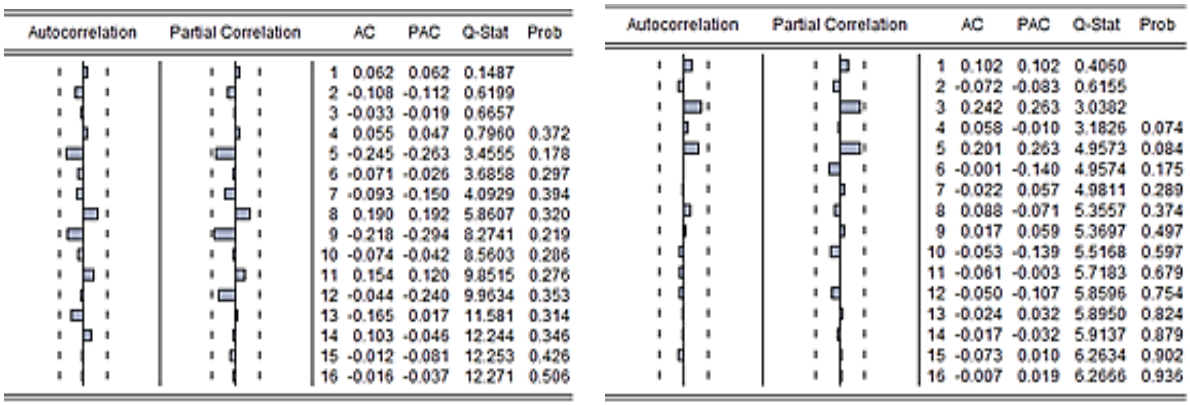


Fig. 7-8: Graph of Correlogram of Residuals and Correlogram of Squared Residuals for MA (1) MA (3) MA (4) of Natural- Gas Consumption Series

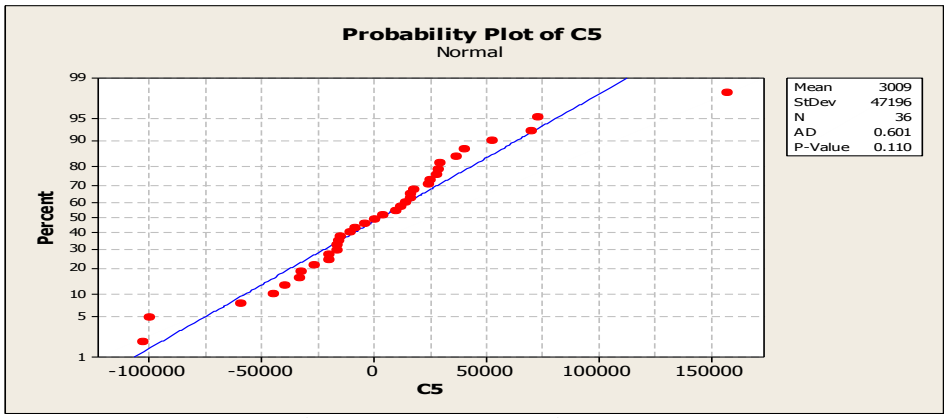


Fig. 9: Normality Test for MA (1) MA (3) MA (4) of Natural Gas Consumption Series

Table 10: Breusch-Godfrey Serial Correlation LM Test For MA (1) MA (3) MA (4) of Natural- Gas Consumption Series

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	0.278524	Prob. F(2,30)	0.7588
Obs*R-squared	0.508515	Prob. Chi-Square(2)	0.7755

Table 11: Autoregressive Conditional Hetroscedasticity (ARCH) Test for MA (1) MA (3) MA (4) of Natural- Gas Consumption Series

Heteroskedasticity Test: ARCH			
F-statistic	0.381977	Prob. F(1,33)	0.5408
Obs*R-squared	0.400491	Prob. Chi-Square(1)	0.5268

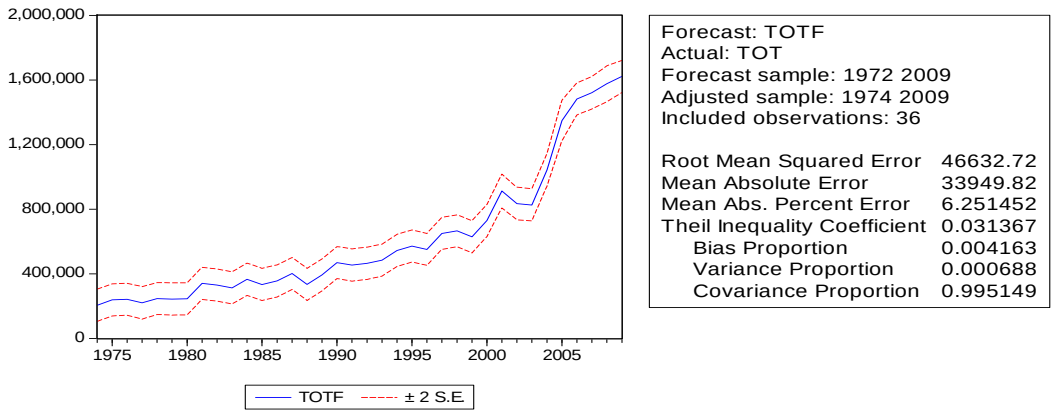


Fig. 10: Graphical Forecast Evaluation For MA (1) MA (3) MA (4) of Natural- Gas Consumption Series

The actual consumption of Natural Gas (total of four sectors) for year 2009-2010 is given as 1521674 thousand decca cubic meter, on the other hand, the forecast value of Consumption of Natural Gas (total of four sectors) with the help of the above given model for year 2009-2010 is 1572640.03 thousand decca cubic meter. By comparing these two values clearly as values are close to each other so estimated forecast model is adequate one.

The results of the diagnostic checks and graphical forecast evaluation for the final forecast model of Transmission System of Natural Gas have been shown in Figure 11-15, Table 12-13 and Figure 16.

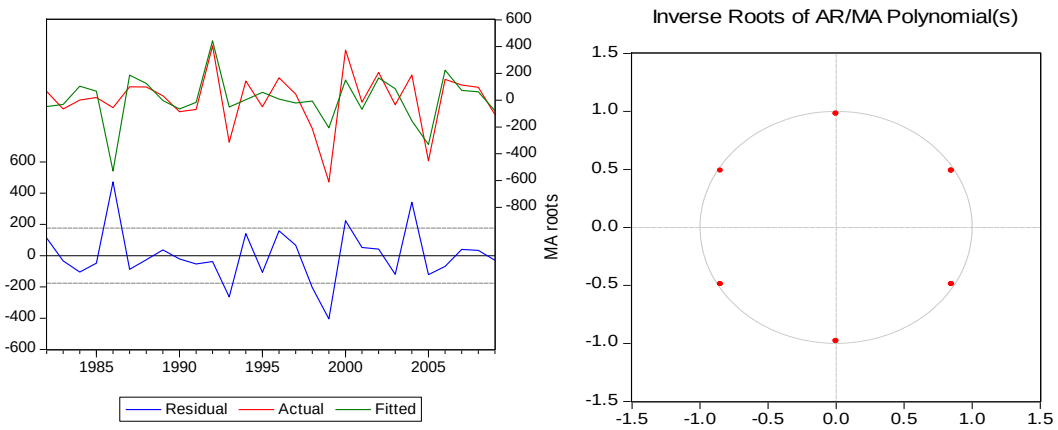


Fig. 11-12: Graph of Actual, Fitted, Residual Values and ARMA Structure for ARIMA (0, 2, 6) of Natural-Gas Transmission Series

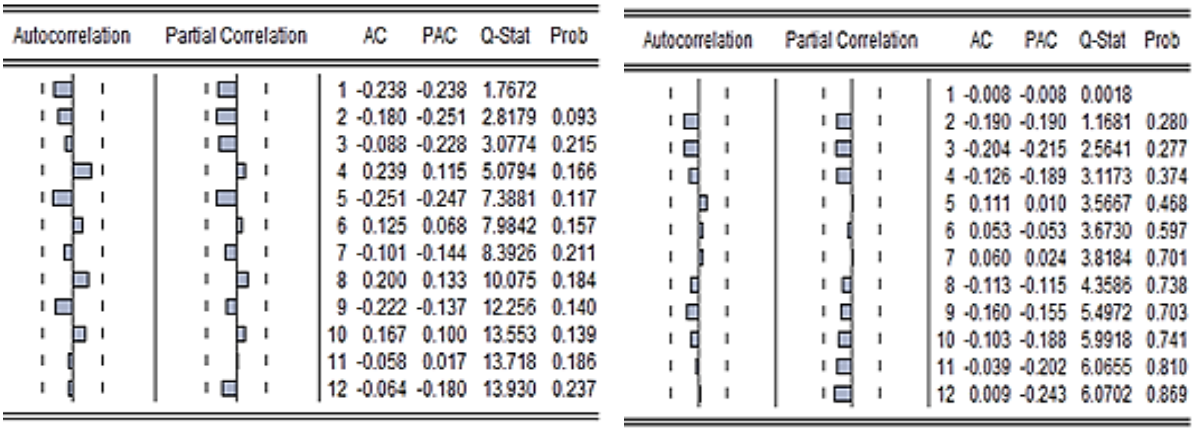


Fig. 13-14: Graph of Correlogram of Residuals and Correlogram of Squared Residuals for ARIMA (0, 2, 6) of Natural -Gas Transmission Series

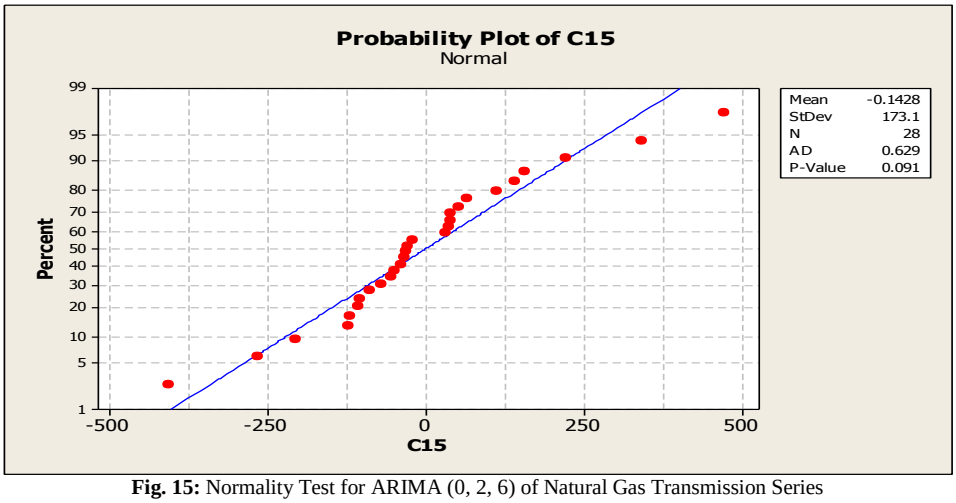


Fig. 15: Normality Test for ARIMA (0, 2, 6) of Natural Gas Transmission Series

Table 12: Breusch-Godfrey Serial Correlation LM Test for ARIMA (0, 2, 6) of Natural -Gas Transmission Series

Breusch-Godfrey Serial Correlation LM Test:			
F-statistic	1.599724	Prob. F(2,24)	0.2227
Obs*R-squared	3.293598	Prob. Chi-Square(2)	0.1927

Table 13: Autoregressive Conditional Hetroscedasticity (ARCH) for ARIMA (0, 2, 6) of Natural- Gas Transmission Series

Heteroskedasticity Test: ARCH			
F-statistic	0.001532	Prob. F(1,25)	0.9691
Obs*R-squared	0.001654	Prob. Chi-Square(1)	0.9676

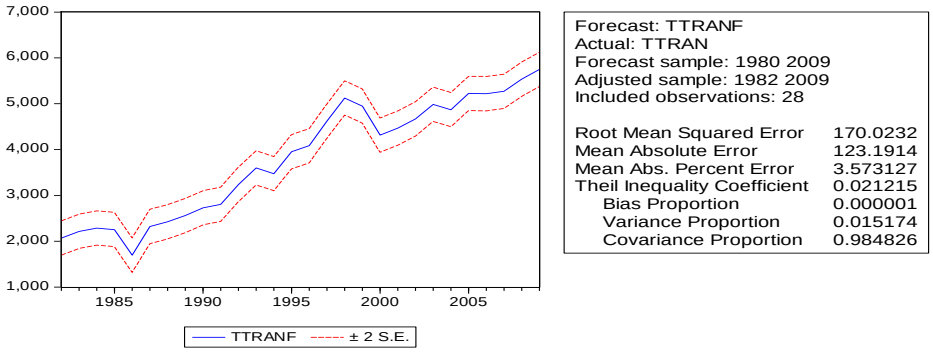


Fig. 16: Graphical Forecast Evaluation for ARIMA (0, 2, 6) of Natural- Gas Transmission Series

The actual transmission of Natural Gas (total of two types of pipeline) for year 2009-2010 is given as 5831kilometer, on the other hand, the forecast value of Transmission of Natural Gas (total of two types of pipeline) with the help of the above given model for year 2009-2010 is 6198. 03kilometer. By comparing these two values clearly as values are close to each other so it can be said that our estimated forecast model is adequate one.

4. Conclusion

In this study, our interest was to select a suitable forecast model from various selected ARIMA models which shows high power for prediction or forecasting.

Among various selected ARIMA models, with the help of diagnostic checks, the best forecast model for each of two data sets have been selected. Forecast model selected for Natural-Gas Consumption (Four Sectors) is MA (1) MA (3) MA (4) at 2nd difference. Forecast model selected Natural-Gas Transmission (Two Types of Pipelines) is ARIMA (0, 2, and 6) at 2nd difference. By using these models, forecast values for year 2010 has been estimated and then compared with the actual values for year 2010.

According to comparison between forecast and actual value, we come to say that MA(1) MA(3) MA(4) at 2nd difference and ARIMA(0,2,6) at 2nd difference for Natural-Gas Consumption (Four Sectors) and Natural-Gas Transmission (Two Types of Pipelines), respectively, are the models which gave closer values as compared to actual ones so these are considered to be appropriate forecast models for their corresponding time series (data set) and can be used for practical implementation.

References

[1] S.E. Alna and F. Ahiakpor, ARIMA approach to predicting inflation in Ghana, Journal of Economics and International Finance 3(5) (2011) 328-336.

[2] H. Bailly, A report by A Global Management Consulting Firm (Islamabad), (2011). Retrieved from: http://www.pakistantoday.com.pk/2011/09/gas-shortage-adversely-impacting-economy/

[3] J.E.P Box and G.M. Jenkins, The time series analysis: Forecasting and control (revised edition), Holden-day, San-Francisco, 2011.

[4] G. Edward, P. Box and G.M. Jenkins, Time series analysis: Forecasting and control, Prentice Hall PTR, USA, 2011.

[5] E. Erdogdu, Natural gas demand in Turkey, Applied Energy 87(1) (2010) 211-219. http://dx.doi.org/10.1016/j.apenergy.2009.07.006.

[6] M. Forouzanfar, A. Doustmohammadi, M.B. Menhaj and S. Hasanzadeh, Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran, Applied Energy 87(1) (2010) 268-274. http://dx.doi.org/10.1016/j.apenergy.2009.07.008.

[7] Z.Y. Han and W.G. Weng, An integrated quantitative risk analysis method for natural gas pipeline network, Journal of Loss Prevention in the Process Industries 23(3) (2010) 428-436. http://dx.doi.org/10.1016/j.jlp.2010.02.003.

[8] H.G. Huntington, Industrial natural gas consumption in the United States: An empirical model for evaluating future trends, Energy Economics 29(4) (2007) 743-759. http://dx.doi.org/10.1016/j.eneco.2006.12.005.

[9] C.K. Lee, H.J. Song and J.W. Mjelde, The forecasting of international expo tourism using quantitative and qualitative techniques, Tourism Management 29(6) (2008) 1084-1098. http://dx.doi.org/10.1016/j.tourman.2008.02.007.

[10] P.I. Nwosa and I.O. Oseni, Monetary policy, exchange rate and inflation rate in Nigeria, Research Journal on Finance and Accounting 3(3) (2012) 62-69.

[11] S. Pandian, The political economy of trans-Pakistan gas pipeline project: Assessing the political and economic risks for India, Energy Policy 33(5) (2005) 659-670. http://dx.doi.org/10.1016/j.enpol.2003.09.011.

[12] M. Radetzki, World Demand for Natural Gas: History and Prospects, the Energy Journal 15(Special Issue) (1994) 219-236. http://dx.doi.org/10.5547/issn0195-6574-ej-vol15-nosi-12.

[13] A. Sen, Natural gas imports into South Asia a study in international relations, Energy Policy 28(11) (2000) 763-770. http://dx.doi.org/10.1016/S0301-4215(00)00061-6.

[14] J.S. Simonoff, C.E. Restrepo and R. Zimmerman, Risk management of cost consequences in natural gas transmission and distribution infrastructures, Journal of Loss Prevention in the Process Industries 23(2) (2010) 269-279. http://dx.doi.org/10.1016/j.jlp.2009.10.005.

[15] Z. Wadud, H.S. Dey, M.A. Kabir and S.I. Khan, Modeling and forecasting natural gas demand in Bangladesh, Energy Policy 39(11) (2011) 7372-7380. http://dx.doi.org/10.1016/j.enpol.2011.08.066.

[16] K. Wong, H. Song, S.F. Witt and D.C. Wu, Tourism forecasting: To combine or not to combine? Tourism Management 28 (2007) 1068-1078. http://dx.doi.org/10.1016/j.tourman.2006.08.003.

[17] G. Xu and W. Wang, Forecasting China's natural gas consumption based on a combination model, Journal of Natural Gas Chemistry 19(5) (2010) 493-496. http://dx.doi.org/10.1016/S1003-9953(09)60100-6.