IOSR Journal of Mathematics (IOSR-JM)
e-ISSN: 2278-5728, p-ISSN: 2319-765X. Volume 12, Issue 5 Ver. IV (Sep. - Oct.2016), PP 22-28
www.iosrjournals.org

Modeling Medical Oxygen Demand Using Box-Jenkins Methodology Applied Study on Soba University Hospital

Walid Sayed Mohamed Hamad Karar[1], Ahmed Abdallah Hamdi [2]
Sudan University of Science & Technology

Abstract: *This paper aim to find a suitable model to predict the demand of medical oxygen. The study data contained a number of 120 observations of oxygen consumption(in m3) were taken from Soba University Hospital in Khartoum state, since January 2005 to December 2014. We have used Box-Jenkins, Auto Regressive Integrated Moving Average (ARIMA) methodology for building forecasting model. Results suggest that ARIMA(0,1,1) is the most suitable model to be used for predicting the demand of medical oxygen. For testing the forecasting accuracy Root Mean Square Error, Mean Absolute Error, and Mean Absolute Percentage Error are calculated.*

Keywords: *Modeling Medical Oxygen Demand , Box-Jenkins Methodology, ARIMA ,forecasting oxygen consumption*

I. Introduction

The importance of medical oxygen as one of the most important medical needs necessary to remedial sections which is used in hospitals, especially in the treatment of acute medical patients in the short term in emergency departments such as respiratory diseases, acute pneumonia, asthma and pulmonary congestion due to heart failure or chronic conditions which are required in some cases. The research takes Box-Jenkins methodology to formulate the demand of the medical oxygen as a method commonly used in time-series models for forecasting analysis. Specifically since the time series is an important statistical topics dealing with the behavior of the phenomena analyzes and interprets over specific periods. The objectives of time series analysis can be summarized to obtain an accurate description of the special features of the process which is generate time series. And build a model to explain the behavior of the time series, and use the results to predict the behavior of the series in the future, in addition to control the process that generated the time series which examines what can happen when you change some of the parameters of the model. This study will contribute to provide accurate information for decision-makers which can helps them in the development of policies leading to the solution of the problem. And to know the real amount of medical oxygen for hospitals to avoid a shortage of the oxygen supply in the future.

II. Methods

Objectives:

This study aims to find the suitable ARIMA model which will use to predict the demand of medical oxygen for Soba University hospital as applied study.

Data collection:

The research data was taken from Soba University hospital which follow to Khartoum University represents one of biggest hospitals in Khartoum State , such as a availability of data are deal the researcher to choose the Soba University hospital as applied study of this research .The collected data contained (120) series of observations of monthly consumption of Oxygen, since January 2005 to December 2014.

III. Statistical Methods

Box-Jenkins Methodology (Theoretical Background):

In 1976 GEORGE E .P . BOX & GWILYM.JENKINS are revised the edition of their book entitled "Time Series Analysis Forecasting and Control ". The book gave an explanation of the time series models stable and unstable including an auto-regressive integrated moving average (ARIMA) process, of order (p,d,q). The process is defined by:

$$w_t = \emptyset_1 w_{t-1} + \cdots + \emptyset_p w_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

The identification stage is the most important and also the most difficult: it consists to determine the adequate model from ARIMA family models. The most general Box-Jenkins model includes difference operators, autoregressive terms, moving average terms, seasonal difference operators, seasonal autoregressive terms, and seasonal moving average terms. This phase is founded on the study of autocorrelation and partial

autocorrelation. The first step in developing a Box-Jenkins model is to determine if the series is stationary and if there is any significant seasonality that needs to be modeled.

Testing for Non-Stationarity:

1. Autocorrelation function (Box-Jenkins approach)-if autocorrelations start high and decline slowly, then series is non stationary, and should be differenced.
2. Dickey-Fuller test

yt = a + b yt-1 + ut would be a non stationary random walk if b = 1. So to find out if y has a "unit root" we regress: Δyt = a + cyt-1 + ut where c = b-1 and test hypothesis that c = 0 against c < 0 (like a "t-test").

Iterative stages in the selection of the model :

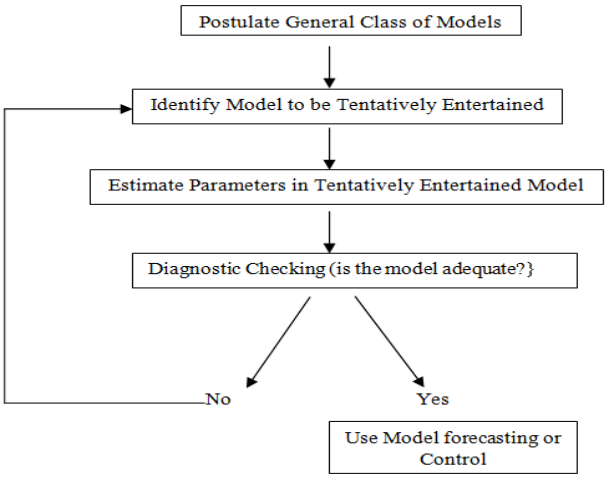


FIG.1 Stages in the iterative approach to model building

IV. Results

The data:

The variable used in the analysis is the medical oxygen consumption from 2005 to the end of 2014 and its available monthly. The source of data is the Monthly demand of oxygen by Soba University Hospital In Khartoum.

Stationary Test:

Table: 1

Sample: 2005M01 2014M12
Included Observations: 120

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
.|**** |	.|**** |	1	0.556	0.556	37.996	0.000
.|**** |	.|** |	2	0.504	0.282	69.519	0.000
.|*** |	.|* |	3	0.426	0.107	92.234	0.000
.|**** |	.|** |	4	0.515	0.282	125.66	0.000
.|*** |	.|. |	5	0.377	-0.050	143.73	0.000
.|*** |	.|. |	6	0.367	0.031	161.03	0.000
.|** |	.|. |	7	0.347	0.066	176.64	0.000
.|** |	.|. |	8	0.340	-0.003	191.73	0.000
.|** |	.|* |	9	0.328	0.076	205.89	0.000
.|** |	.|* |	10	0.341	0.077	221.35	0.000
.|** |	.|. |	11	0.293	-0.033	232.88	0.000
.|** |	.|. |	12	0.257	-0.025	241.85	0.000
.|** |	.|. |	13	0.256	0.014	250.82	0.000
.|** |	.|. |	14	0.279	0.045	261.53	0.000
.|** |	.|. |	15	0.268	0.047	271.52	0.000
.|* |	**|. |	16	0.119	-0.215	273.50	0.000

.|** |	.|* |	17	0.239	0.172	281.59	0.000
.|** |	.|. |	18	0.248	0.072	290.41	0.000
.|** |	.|* |	19	0.299	0.080	303.36	0.000
.|* |	*|. |	20	0.129	-0.130	305.79	0.000
.|* |	*|. |	21	0.152	-0.095	309.22	0.000
.|* |	*|. |	22	0.103	-0.066	310.80	0.000
.|* |	*|. |	23	0.085	-0.095	311.88	0.000
.|. |	.|. |	24	0.018	-0.060	311.93	0.000
.|. |	.|. |	25	0.013	-0.034	311.96	0.000
.|. |	.|. |	26	-0.004	0.007	311.96	0.000
.|. |	.|. |	27	0.011	0.014	311.97	0.000
.|. |	.|. |	28	0.034	0.057	312.16	0.000
.|. |	.|. |	29	0.061	0.061	312.77	0.000
.|. |	.|. |	30	0.005	-0.011	312.78	0.000
*|. |	*|. |	31	-0.093	-0.198	314.21	0.000
.|. |	.|. |	32	-0.034	-0.016	314.40	0.000
*|. |	*|. |	33	-0.112	-0.129	316.53	0.000
*|. |	.|. |	34	-0.125	-0.026	319.17	0.000
**|. |	.|. |	35	-0.211	-0.031	326.87	0.000
*|. |	.|. |	36	-0.091	0.039	328.31	0.000

Figure: 2

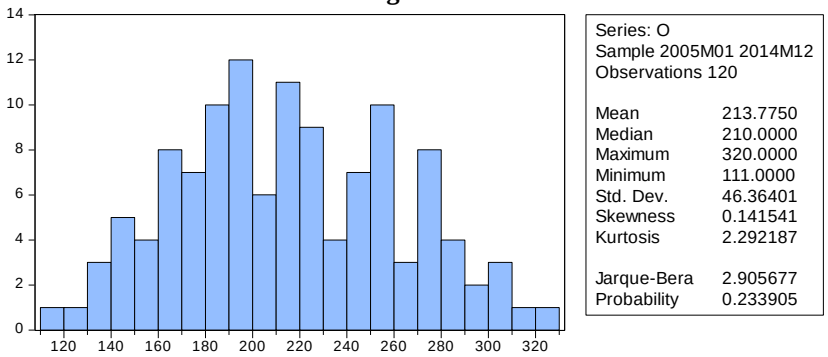


Table 1.& Figure2. Show that the medical oxygen consumption rate from 2005 to 2014.
Source: Monthly registration of oxygen demand of Soba University Hospital In Khartoum.

The first step in developing a Box-Jenkins model is to determine if the series is stationary. For this, we use the autocorrelation function (ACF) and Augmented Dickey-Fuller test (ADF).This section is devoted to the application of Augmented Dickey Fuller of unit root and correlogram tests in testing whether the medical oxygen consumption series is stationary. The correlogram test is one of the most powerful statistical tests used in testing whether the time-series data is stationary, the null hypothesis of the test is that all autocorrelations is equal to zero against the alternative hypothesis at one of the autocorrelations is not equal to zero. The acceptance of the null hypothesis all autocorrelations is equal to zero is an indication of stationarity series. While the Augmented Dickey Fuller test is testing the stationarity of the series under the null hypothesis that the series has a unit root against the alternative hypothesis the series does has a unit root, the acceptance of the null hypothesis is an indication of non-stationarity series.

Table:2 Null Hypothesis: O has a unit root
Exogenous: Constant
Lag Length: 3 (Automatic - based on SIC, maxlag=12)

Prob.*	t-Statistic		
0.3739	-1.810335	Augmented Dickey-Fuller test statistic	
	-3.487550	1% level	Test critical values:
	-2.886509	5% level	
	-2.580163	10% level	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(O)
Method: Least Squares
Sample (adjusted): 2005M05 2014M12
Included observations: 116 after adjustments

Prob.	t-Statistic	Std. Error	Coefficient	Variable
0.0729	-1.810335	0.088909	-0.160955	O(-1)
0.0000	-4.632861	0.108673	-0.503466	D(O(-1))
0.0034	-2.996282	0.107304	-0.321513	D(O(-2))
0.0004	-3.626825	0.089436	-0.324369	D(O(-3))
0.0589	1.908583	19.17717	36.60121	C

1.215517	Mean dependent var	0.367095	R-squared
42.66631	S.D. dependent var	0.344288	Adjusted R-squared
9.964810	Akaike info criterion	34.54949	S.E. of regression
10.08350	Schwarz criterion	132497.1	Sum squared resid
10.01299	Hannan-Quinn criter.	-572.9590	Log likelihood
2.007821	Durbin-Watson stat	16.09547	F-statistic
		0.000000	Prob(F-statistic)

Table2. show that the Augmented Dickey Fuller unit root test is employed for testing whether medical oxygen consumption series is stationary. Also in determining the order of differencing required in performing time series models. The Augmented Dickey-fuller of unit root test (ADF) with trend, intercept and lag difference of 1 result in above figure shows that the ADF value in absolute terms (**1.810335**) is less than the 1%, 5% and 10% critical values in absolute terms (3.487550, 2.886509 and2.580163) respectively, this results conclude that medical oxygen consumption series has a unit root.

The identification - the autocorrelation is computed on the first differences series

Figure 3.

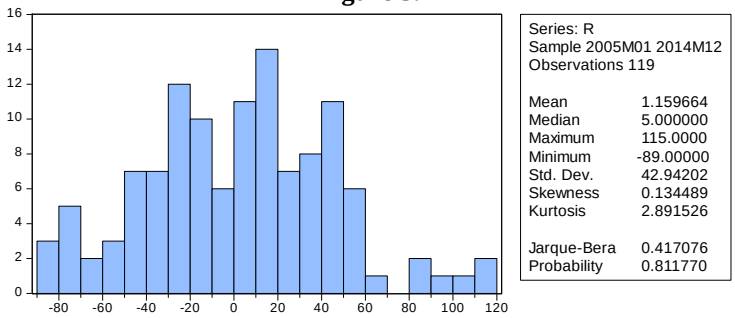


Table:3

Sample: 2005M01 2014M12
Included observations: 119

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
***|. |	***|. |	1	-0.448	-0.448	24.536	0.000
.|. |	**|. |	2	0.031	-0.213	24.653	0.000
**|. |	***|. |	3	-0.206	-0.374	29.944	0.000
.|** |	.|. |	4	0.265	-0.028	38.746	0.000
*|. |	*|. |	5	-0.145	-0.101	41.404	0.000
.|. |	*|. |	6	0.001	-0.153	41.404	0.000
.|. |	.|. |	7	0.024	-0.014	41.476	0.000
.|. |	*|. |	8	-0.023	-0.124	41.546	0.000
.|. |	*|. |	9	-0.029	-0.140	41.656	0.000
.|* |	.|. |	10	0.096	0.038	42.864	0.000
.|. |	.|. |	11	-0.007	0.018	42.870	0.000
.|. |	.|. |	12	-0.051	-0.013	43.226	0.000
.|. |	.|. |	13	-0.030	-0.026	43.347	0.000
.|. |	*|. |	14	0.022	-0.089	43.411	0.000
.|* |	.|* |	15	0.173	0.208	47.552	0.000
**|. |	*|. |	16	-0.312	-0.194	61.157	0.000
.|* |	*|. |	17	0.138	-0.098	63.842	0.000
.|. |	.|. |	18	-0.051	-0.065	64.210	0.000
.|** |	.|* |	19	0.259	0.120	73.837	0.000
**|. |	.|* |	20	-0.220	0.108	80.901	0.000
.|. |	.|. |	21	0.056	0.034	81.365	0.000

.|. |	.|. |	22	-0.039	0.034	81.587	0.000
.|. |	.|. |	23	0.021	-0.021	81.653	0.000
.|. |	.|. |	24	-0.020	-0.020	81.711	0.000
.|. |	.|. |	25	0.012	-0.048	81.732	0.000
.|. |	.|. |	26	-0.033	-0.056	81.900	0.000
.|. |	*|. |	27	-0.035	-0.106	82.087	0.000
.|. |	*|. |	28	0.016	-0.121	82.128	0.000
.|* |	.|. |	29	0.110	-0.034	84.056	0.000
.|. |	.|* |	30	0.049	0.160	84.439	0.000
*|. |	.|. |	31	-0.193	-0.011	90.532	0.000
.|* |	.|* |	32	0.170	0.152	95.332	0.000
*|. |	.|. |	33	-0.110	0.014	97.365	0.000
.|* |	.|. |	34	0.100	-0.007	99.044	0.000
**|. |	.|. |	35	-0.233	-0.047	108.34	0.000
.|* |	.|. |	36	0.210	-0.047	115.97	0.000

Figure 3 &Table 3. Show that the autocorrelation is computed on the first differences series

Table 4.

Null Hypothesis: R has a unit root
Exogenous: Constant
Lag Length: 2 (Automatic - based on SIC, maxlag=12)

Prob.*	t-Statistic		
0.0000	-11.43476	Augmented Dickey-Fuller test statistic	
	-3.487550	1% level	Test critical values:
	-2.886509	5% level	
	-2.580163	10% level	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(R)
Method: Least Squares
Date: 06/27/16 Time: 18:02
Sample (adjusted): 2005M05 2014M12
Included observations: 116 after adjustments

Prob.	t-Statistic	Std. Error	Coefficient	Variable
0.0000	-11.43476	0.210293	-2.404647	R(-1)
0.0000	4.936855	0.158149	0.780761	D(R(-1))
0.0000	4.319492	0.086248	0.372548	D(R(-2))
0.4659	0.731668	3.245993	2.374990	C

0.689655	Mean dependent var	0.778962	R-squared
73.25541	S.D. dependent var	0.773042	Adjusted R-squared
9.976666	Akaike info criterion	34.89897	S.E. of regression
10.07162	Schwarz criterion	136409.1	Sum squared resid
10.01521	Hannan-Quinn criter.	-574.6466	Log likelihood
2.030951	Durbin-Watson stat	131.5671	F-statistic
		0.000000	Prob(F-statistic)

Table 4. Shows the ADF test were also applied to the first difference of medical oxygen consumption series from above figure the result illustrate that the absolute value of the ADF test (11.43476) is grater than the 1%, 5% and 10% critical values in absolute terms (3.487550, 2.886509 and 2.580163) respectively, this result conclude that the first difference of medical oxygen consumption rate series is stationary.

Model identification and Coefficient Estimates:

After the test of stationary, we conclude that the data is stationary at first difference. The repressor that would be chosen from the model is selected from various iteration for AR(p) and MA(q), the selection is based on observing the ACFs and PACFs. We used E-views for estimating the coefficients and testing the goodness of fit of the model. The search algorithm tried number of different coefficient values, after several iterations, and based on comparing Akaike Information Criteria (AIC), and Schwarz Information criteria (SIC), the best model to forecast medical oxygen demand is ARIMA (0,1,1) since it contains the least AIC and SIC ratios. Table 5 shows the AIC and SIC value for various ARIMA (p,d,q) iterations:

Table 5

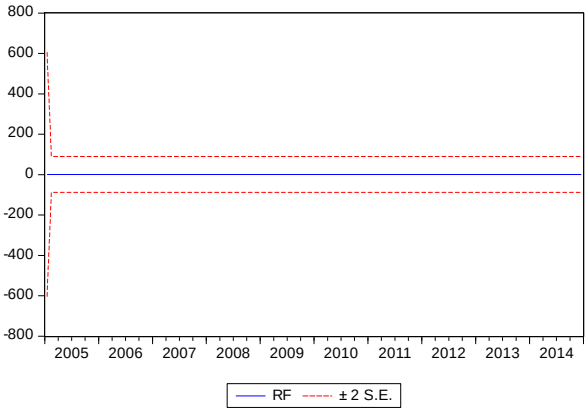
ARIMA (p,d,q).	AIC	SIC
ARIMA (1,1,0)	10.15862	10.205
ARIMA (0,1,1)	**9.968449***	**10.015***
ARIMA (1,1,1)	9.984488	10.054
ARIMA (1,1,2)	9.986844	10.024
ARIMA (2,1,1)	9.985561	10.079
ARIMA (2,1,2)	9.999428	10.117

Forecasting Accuracy: There are several methods of measuring accuracy and comparing one forecasting method to another, we have selected Root Mean Square Error (RMSE). Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The RMSE, MAE and MAPE are as follows :

	ARIMA (0,1,1)
RMSE	42.76187
MAE	34.54954
MAPE	97.84004

The above table shows that the Root Mean Squared Error and Mean Absolute Error are less in ARIMA(0,1,1) as compared to other ARIMA models .

Forecast Result Analysis: Therefore, the estimation of ARIMA (0,1,1) model is validated, the time series can be described by an ARIMA(0,1,1) process.



Forecast: RF
Actual: R
Forecast sample: 2005M01 2014M12
Included observations: 119

Root Mean Squared Error	42.76187
Mean Absolute Error	34.54954
Mean Abs. Percent Error	97.84004
Theil Inequality Coefficient	0.978581
Bias Proportion	0.000031
Variance Proportion	NA
Covariance Proportion	NA

Forecasting model is: $w_t = a_t - \theta_1 a_{t-1}$

$$MOD = 0.920890 + 0.774676\, a_{t-1}$$

t-Statistic 1.229735 -12.90649
P-value 0.2213 0.0000
DW = 1.876080 , R Square = 0.339200 , SSR= 143786.1

V. Conclusion

In order to develop a univariate Time Series Model, we used medical oxygen consumption of Soba University Hospital in Khartoum state from January 2005 till December 2014.In this paper, we have developed systematic and iterative methodology of Box-Jenkins ARIMA forecasting for medical oxygen demand . A unit root test was applied to the long term monthly consumption of oxygen .This concludes that the oxygen demand series is non stationary. After the test of stationary, we conclude that the data is stationary at first difference, E-views software is used for fitting the coefficient of the model, using graphs, statistics, ACFs and PACFs of residuals and after several iterations, the model selected is ARIMA(0,1,1).There are several ways of measuring forecasting accuracy; we have used Mean Absolute Error, Root Mean, Square Error and Mean Absolute Percentage Error. We may use this model for forecasting the medical oxygen demand for future.

References

[1]. Akaike, H. (1973) "Information theory and an extension of the maximum likelihood principle". In B.N. Petrov and B.F. Csaki, eds, Second International Symposium on Information Theory pp. 267–281, Akademiai Kiad ´ o, Budapest.

[2]. Adriana AnaMaria ALEXANDRU.2008. "MODELLING UNEMPLOYMENT RATE USING BOX-JENKINS PROCEDURE".journal of applied quantitative methods.vol3.p [156-166].

[3]. American National Home Oxygen Patients Association (NHOPA) Reviewed and Revised 2013 (By: Dennis E. Doherty, MD, Physician Member of NHOPA Board of Directors) Source URL :www.homeoxygen.com
[4]. Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis Forecasting and Control, 2nd ed., Holden-Day, San Francisco.
[5]. Box, G and Cox, R. (1964) "An analysis of transformations," Journal of the Royal Statistical Society, Series B 26, pp.211-252
[6]. Box, G. and Jenkins, G. (1970) "Time series analysis: Forecasting and control ," San Francisco: Holden-Day
[7]. Box, G. and Pierce, D. (1970) "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models", Journal of the American Statistical Association , 65: pp.1509–1526.
[8]. Beverly D. Bradley , Stephen R. C. Howie, Timothy C. Y. Chan,Yu-Ling Cheng. 2014. Estimating Oxygen Needs for Childhood Pneumonia in Developing Country Health Systems. Plos One. 9(2).
[9]. Dickey D. and Fuller W.(1979), " Distribution of the estimators for Autoregressive Time Series With a unit Root ", Journal of the American Statistical Association,74: pp .427-431.
[10]. Dickey D. and Fuller W.(1981) 'The likelihood Ratio Statistics for Autoregressive Time Series With a unit Root", Econometric a ,n49: pp .1057-1072.
[11]. FULLER A.W. (1996) "Introduction to Statistical Time Series". JOHN WILEY & SONS, INC, New York.
[12]. FederalMinistryofHealthofSudan1997.SourceURLhttp://www.sho.gov.sd/controller/knb_research_articles.php?lid=1&id=2452&sm_id=&mid=109&query=.
[13]. M. Massarrat Ali Khan.2013. Forecasting of Gold Prices (Box Jenkins Approach). International Journal of Emerging Technology and Advanced Engineering. ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 3.P[662- 670].
[14]. SPYROS MAKRIDAKIS_ AND MICHEÁ LE HIBON.1997. ARMA Models and the Box-Jenkins Methodology. Journal of Forecasting, Vol. 16, P[147-163].