Australian Journal of Basic and Applied Sciences
2019 July; 13(7): pages 70-77
DOI: 10.22587/ajbas.2019.13.7.11
Original paper
AENSI Publications
Journal home page: www.ajbasweb.com

Forecasting the number of outpatient visits in tertiary hospital using time series based on ARIMA and ES models

Hanaa Elgohari, Ashraf M. Bakr, Mohammed Abdul Majeed

Department of Applied Statistics, Mansoura University, Egypt
Professor of Pediatrics, Mansoura University, Egypt
Faculty of Administration and economics, Salahaddin University, Iraq

Correspondence Author: Hanaa Elgohari, Faculty of commerce, Mansoura University, Algomhoria St.60, Mansoura, Dakahlia, Egypt.
E-mail: - hanaa_elgohary@mans.edu.eg

Received date: 18 May 2019, Accepted date: 26 July 2019, Online date: 29 July 2019

Abstract

This study aims to compare the ARIMA and Exponential models to predict the number of outpatient visits Accurate forecasting of number of outpatient visits in tertiary hospital is beneficial for determining the future staff need and to locate hospital's budget. Most of the studies are focused on emergency department more than the outpatient visits. On the other hand there are controversies on the severity of ARIMA and exponential smoothing models in predicting the number of outpatient visits. This study was done in Mansoura University children's hospital to compare ARIMA and exponential (single, holt's and dampened trend) models by using (BIC, RMSE, MAE and MAPE) Criteria of accuracy check to choose the best model from the selected models in predicting the number of patients attending outpatient clinic in 2018 using the observation data from the last 10 years. The results show that the smallest errors in the models with the estimation are obtained by ARIMA (3,1,3). This model is the best, which gives us the lowest values for each RMSE and BIC and approximately lowest value for MAE and the largest value for the coefficient of determination. After that we used (Ljung- Box) test the parameter of visits by the model ARIMA (3,1,3) and it's showed significance difference in the parameters of lag1 and lag2 by auto regressive also it isn't significance difference for the parameters of moving average in the model. We concluded that ARIMA(3,1,3) model is more sensitive than exponential smoothing models of forecasting the outpatient visits in tertiary hospital.

Keywords: ARIMA(p,d,q), Exponential smoothing models, outpatient visits, Bayesian Information Criteria, The Box-Ljung test.

INTRODUCTION

Time series analysis is considered an essential topic in statistics, where it is usually used to study the future behavior of the series at a determined time. It may be stationary or non-stationary. In the time series model, Autoregressive integrated moving average (ARIMA) model is used. This model applied in various research fields (economics, earth sciences, engineering and technology (Yilan Lin et al., 2015) (Paulo R et al.,2014) (Shirvani A, 2015). Also, exponential smoothing models. Furthermore, Exponential smoothing (ES) methods are used for predicting procedures in healthcare as well as industry and commerce because it is a seasonal model in the data and descriptive trends. (Jones SS et al., 2008). On the outset, simple annual ES model and Automated ES approach (Calegari R, 2016) was used to predict medical care requirement in every day and month in ED visits in hospitals. The prior study illustrated that Exponential smoothing (ES) model combined with the knowledge of time correlation, utilized from the historical information (Zhang G at.al,2016) and achieved high performance in designing ED model demand because it is simple, robust and accurate.

On the other hand, all hospital suffers from increasing the number of patients that is the reason for stress all the year and, the increasing complexity of health conditions, so outpatient department (OD) presents actual operation from the external service of the hospital from that occurs in the hospital. In health care system, many studies focus on the forecasting of patient arrival and the Emergency department(ED)admission in hospital institute that is different from the prediction of outpatient admission or visits (Zhu T, 2017). Chen at.al, 2011 illustrated the implementation of ARIMA model in predicting patient's admission every month. Mai et al., 2015 used a multivariate vector- ARIMA method to explain that the emergency department demand is predicted to exceed population growth (Mai et al., 2015). Most studies depend on linear patterns and for describing the relation between the

Citation: Hanaa Elgohari, *et al.*, Forecasting the number of outpatient visits in tertiary hospital using time series based on ARIMA and ES models. Australian Journal of Basic and Applied Sciences, 13(7): 70-77. DOI: 10.22587/ajbas.2019.13.7.11

actual variables and using these variables as the benchmark for testing the sensitiveness of integrated models together with blended results, so in a time series, linear patterns combined with little computational efforts. Sun et al., 2015 evolved ARIMA models to predict any attendance at the emergency department in hospitals every day for confirming that time series analysis is a helpful and available tool to forecast emergency department workload. In this study will test the stationary of series to know the stationary and equity of mean and variance by using a t-test and leveness test.
Moreover, as we showed, most of the studies are focused on the emergency department more than the outpatient visits. So, this research involved in time series analysis to predict the daily rate of visitors periods for the future months by using non- seasonal time series model ARIMA(p, d,q) and exponential smoothing (ETS) models. By using root mean square error (RMSE), Coefficient of determination (R^2), Mean Absolute Percentage Error (MAPE) Mean Absolute Deviation (MAE), and Bayesian Information Criteria (BIC) criteria of accuracy check to choose the best model from the selected models.

2. METHODOLOGY

2-1 Non-seasonal series (Autoregressive integrated moving average (ARIMA)

The Box-Jenkins methodology refers to the set of procedures for ARIMA model building, identification, estimation, examining ARIMA models combined with time series data and forecasting. This illustrates in figure (1) as follows:

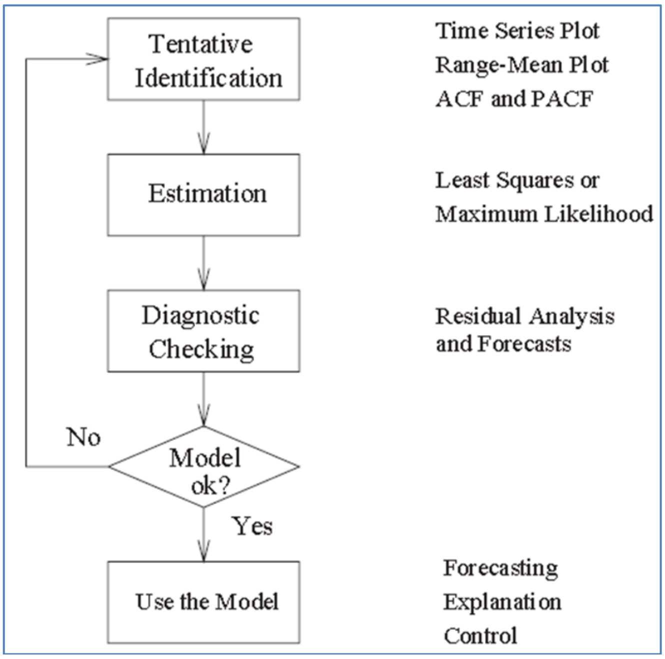

Figure 1: steps for Constructing Model in time series

Step one: Identification: contains determining the order of the model required (p,d,q). (p) Refers to the order of autoregressive, (d) the amounting of variation and (q) refers to the order of moving average. In this step, we can know if the series is stationary or not by the graphical procedure of (ACF). An ARIMA model is defined by:

$$\Delta z_t = z_t - z_{t-1} = z_t - Bz_t = (1-B)z_t \quad (1)$$

Where: Δ: is the linear differential operator.

The stationary series w_t obtained as the d-th difference (Δ^d) of z_t.

$$W_t = \Delta^d z_t = (1-B)^d z_t \quad (2)$$

ARIMA(p,d,q) include the general form:

$$\emptyset_p(B)(1-B)^d z_t = \mu + \theta_q(B)\varepsilon_t \quad (3)$$

Step two: Estimation of all parameters for an interim model that was selected.

Step three: Diagnostic examining: after estimate the parameters of the model will make some of the tests. For instance, ACF and PACF of the residuals to emphasis that model is efficient to prediction.
Step four: Predicting for one period or a number of periods of the future with the parameters for a tentative model that was selected.

2-2 Model order selection criteria

Citation: Hanaa Elgohari, *et al.*, Forecasting the number of outpatient visits in tertiary hospital using time series based on ARIMA and ES models. Australian Journal of Basic and Applied Sciences, 13(7): 70-77. DOI: 10.22587/ajbas.2019.13.7.11

Bayesian Information Criteria (BIC) is an estimate of a function of the posterior probability of a model being true, under a certain Bayesian setup, so that a lower BIC means that a model is more likely to be the true model:

$$\mathrm{BIC} = 2\log \mathrm{n} - 2\log\left(\mathrm{L}(\hat{\theta}/\mathrm{y})\right) \quad (4)$$

A Box-Ljung test is a diagnostic tool used to test the lack of fit of a time series model. The test is applied to the residuals of a time series after fitting an ARMA(p,q) model to the data. The test examines m autocorrelations of the residuals. If the autocorrelations are very small, we conclude that the model does not exhibit a significant lack of fit.
For forecasting values, (where n is the number of forecasted errors):

Mean Square Error $\mathrm{MSE} = \frac{\sum e_i^2}{n}$, where $e_i = y_i - \hat{y}_i$ (5)

Root Mean Square Error $\mathrm{RMSE} = \sqrt{\mathrm{MSE}}$ (6)

Mean Absolute Percentage Error $\mathrm{MAPE} = \frac{1}{n}\sum \frac{|e_i|}{y} * 100\%$ (7)

Mean Absolute Deviation $\mathrm{MAD} = \sum \frac{|e_i|}{n}$ (8)

2-3 Exponential smoothing models

2-3-1 Simple exponential and Brown Method

Firstly, Robert Macaulay (1931) recognized Exponential smoothing as a method and Robert G. Brown II, developed it and predict from time-series data during World War.

The methods of exponential smoothing examine in isolating seasonality or trends from irregular variation. The more developed methods can model and identify these patterns if these patterns are found. Moreover, the models can combine these patterns with the forecasting. In the case of predicting, (ES) can apply at the averages of weight of the initial data. It is expected that the effect of new observations reduces over time. Concerning the historical time, the less effect of each observation that has on the prediction. To symbolize this geometric incline in the effect, the scheme of exponential weighting should apply in the procedure that represent as elementary exponential smoothing (Robert, 1999).

It is expected that the forecasting may focus on the average of moving that is called MA_{t+1} ,but the initial average of moving is called MA_t. If the average of moving contains10 observations, the moving average need to update and slide over one period at a time. Each period needs the average of the 10 observations to take. Therefore, conceiving this same process need 1:10 observations at a time and to deduct 1:10 the average moving formed from the advanced observations before integrating them to create modern forecasting moving average (Brown, 1963) the difference between the single and double smoothed values can be added to the oothed value and adjusted for trend:

$$MA_{t+1} = (1 - 1/10)MA_t + (1/10)X_t \quad (9)$$

The moving average comprises ten observations, so the moving average may be made up of any number of observations. The proportion of the latest observation taken is called a smoothing constant, α. Between zero and 1, the formula representing this simple smoothing is

$$\begin{aligned} MA_{t+1} &= (1 - \alpha)MA_t + (\alpha)X_t \\ &= (\alpha)X_t + (1 - \alpha)MA_t \end{aligned} \quad (10)$$

Because this moving average is a smoothing function that may be applied for forecasting, a forecast f_t may be substituted for the moving average, MA_t ,in this formula to obtain a formula to forecasting:

$$F_{t+1} = \alpha X_t + (1 - \alpha)F_t \quad (11)$$

In this expression extended two and then n steps into the past, it becomes:

$$F_{t+1} = \alpha X_t + (1 - \alpha)[\alpha X_{t-1} + (1 - \alpha)]F_{t-1} \quad (12)$$

At this point, the smoothing weight refers to the volume of the smoothing constant that is between 0 and 1 — a smaller smoothing constant presents to the oldest observations extra relative weight. In contrast, larger smoothing constant presents extra weight to the most modern observation, but it presents less weight to the oldest observation. In the outset, the symbol of smoothing constant is, α and it controls the process memory.

Before simple exponential smoothing begins, two choices should include the value of the smoothing weight and the smoothing constant. Firstly, we should realize the optimal smoothing constant that can be organized by statistical or graphical comparison. Moreover, indicators can be applied to compare one forecast error with another. Therefore, the better smoothing weight is less than 0.5 and greater than 0.10, while this need not be the case.

2-3-2: Holt's linear, exponential smoothing

In failing to account for trends in the data, simple exponential smoothing remains unable to handle interesting and important Nonstationary processes. E. S. Gardiner expounds on how C. C. Holt, whose early work was sponsored by the Office of Naval Research, developed a model that accommodates a trend in the series (Gardiner, 1987). The final model for a prediction contains a mean and a slope coefficient along with the error term:

$$Y_t = \mu_t + \beta_t t + e_t \qquad ...(13)$$

This final model consists of two-component equations for updating(smoothing) the two parameters of the equation system, namely, the mean,μ , and the trend coefficient, β. The updating equation for the mean level of the model is a version of the simple exponential smoothing, except that the trend coefficient is added to the previous intercept to form the component that receives the exponential decline in influence on the current observation as the process is expanded back into the past. The alpha coefficient is the smoothing weight for this equation:

$$\mu_t = \alpha Y_t + (1 - \alpha) + (\mu_{t-1} + b_{t-1}) \qquad ...(14)$$

The coefficient is updated by a similar exponential smoothing. To distinguish the trend updating smoothing weight from that for the intercept, γ is used instead. The values for both smoothing weights can range from 0 to 1.0.

$$b_t = \gamma(\mu_t - \mu_{t-1}) + (1 - \gamma)b_{t-1} \qquad ...(15)$$

In the algorithm by which this process works, first the level is updated. The level is a function of the current value of the dependent variable plus a portion of the previous level and trend then. Once the new level is found, the trend parameter is updated based on the difference between the current and previous in percept and a complement of the previous trend.

2-3-3 The dampened trend linear, exponential smoothing model

Although taking a linear trend into account represents an improvement on simple exponential smoothing, it does not deal with more complex types of trends. Neither dampened nor exponential trends are linear. A dampened trend refers to a regression component for the trend in the updating equation. The updating (smoothing) equations are the same as in the linear Holt exponential smoothing model except that the lagged trend coefficients, b_{t-1}, are multiplied by a dampening factor, Φ^i .When these modifications are made, the final prediction model for a dampened trend linear, exponential smoothing equation with no seasonal component follows:

$$Y_{t+h} = \mu_t + \sum_{t=0}^{n} \Phi^i b_t \qquad ...(16)$$

With Φ^i: dampening factor

Alternatively, the model could have an exponential trend, where time is an exponent of the trend parameter in the final equation $Y_{t+h} = \mu_t + b^t{}_t$ Many series have other variations in the type of trend. It is common for a series to have a regular annual variation that also needs to be taken into account. For exponential smoothing to be widely applicable, it would have to be able to model this variation as well.

3. PRACTICAL ASPECT

We take the data of Visits in hospital for the period from Jan. *2011* to Dec. *2017* in the Figure(2) for the original data below, to forecasting the daily rate of visitor periods for the future months, using the non-seasonal time series model *(ARIMA),* From the Figure(2), we notice the increasing in the following of every Jull month's (2011-2017),a spatially in July month of year 2017 and decreasing after 2011 to 2017. Also, we use the SPSS program to analyze data.

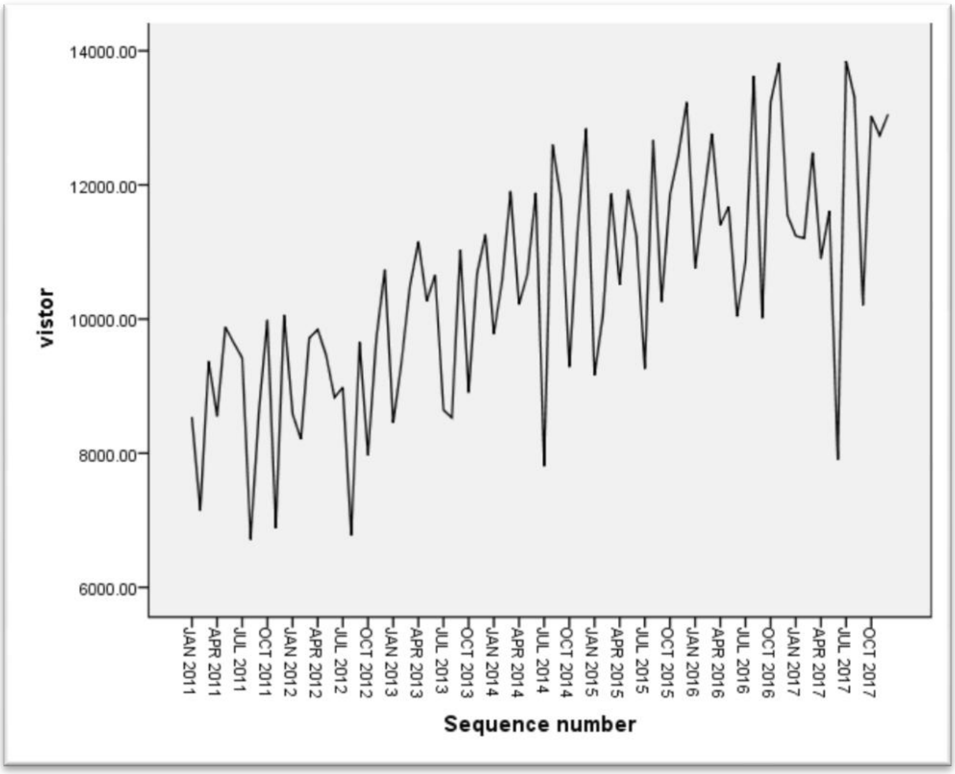

Figure 2: non-seasonal time series of visits in hospital out (2011-2017)

And you have tested the stationary of series to know the stationary and Equality of mean and the variance by t-test with Levenes Test in the table (1) and table(2) :

Table 1: Independent Samples Test

		Levene's Test for Equality of Variances		T-test for Equality of Means	
		F	Sig.	t	Df
visitors	**Equal variances assumed**	.916	.341	-6.625	82
	Equal variances not assumed			-6.625	80.060

Table 2: Independent Samples Test for Equality of Means

		t-test for Equality of Means				
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
					Lower	Upper
visits	Equal variances assumed	.000	-2031.90476	306.68507	-2641.99906	-1421.81046
	Equal variances not assumed	.000	-2031.90476	306.68507	-2642.22048	-1421.58905

The pattern in the ACF is typical of a series that is stationary and non-seasonal. We need at least one order of differencing. If we take a non-seasonal difference (ARIMA(0,1,0)(0,0,0) or ARIMA(0,1,0)) and suggested 16 models with difference equal to one and analyses the models by the program SPSS Package 24 with comparing by R^2,RMSE,MAE,MAPE,BIC for ARIMA Models. The correct order of differencing is a calculation of the error statistics of the series at each level of difference. We can compute these by fitting the corresponding ARIMA models in which only differencing is used in table (3) and using Exponential Smoothing (non-seasonal) Models in the table (4) are as follows:

Table 3: Statistics of ARIMA Models

No.	Model	R^2	RMSE	MAE	MAPE	BIC
1	**ARIMA (0,1,0)**	0.265	1738.769	1409.426	14.093	15.135
2	**ARIMA (1,1,0)**	0.385	1590.025	1279.623	12.728	14.956
3	**ARIMA (0,1,1)**	0.635	1224.495	1003.848	10.087	14.434
4	**ARIMA (2,1,0)**	0.528	1393.235	1070.976	10.737	14.692
5	**ARIMA (2,1,1)**	0.619	1260.243	982.538	10.099	14.544
6	**ARIMA (0,1,2)**	0.648	1210.424	979.975	9.853	14.464
7	**ARIMA (2,1,0)**	0.528	1393.235	1070.976	10.737	14692
8	**ARIMA (1,1,2)**	0.637	1236.613	996.297	10.005	14.560
9	**ARIMA (2,1,2)**	0.681	1166.919	922.358	9.291	14.497
10	**ARIMA (3,1,0)**	0.541	1382.524	1054.596	10.643	14.730
11	**ARIMA (3,1,1)**	0.679	1171.422	914.427	9.296	14.505
12	**ARIMA (3,1,2)**	0.682	1174.044	919.609	9.275	14.562
13	**ARIMA (1,1,3)**	0.678	1173.837	933.692	9.413	14.509
14	**ARIMA (2,1,3)**	0.695	1148.841	911.826	9.154	14.519
15	**ARIMA (3,1,3)**	0.724	1100.832	849.635	8.503	14.487
16	**ARIMA (0,1,3)**	0.664	1189.707	952.676	9.601	14.482

Table 4: Statistics of Exponential Smoothing (non-seasonal) Models

No.	Model	R^2	RMSE	MAE	MAPE	BIC
1	**Simple**	0.522	1367.163	1139.150	11.150	14.494
2	**Holt**	0.868	1279.096	1026.678	10.440	14.413
3	**Brown**	0.852	1366.026	1133.756	11.422	14.492
4	**Damped Trend**	0.593	1276.886	1021.211	10.342	14.494

The smallest errors, in the models with the estimation, are obtained by ARIMA(3,1,3), which uses one difference of each type. This, together with the appearance of the plots, strongly suggests that we should use ARIMA(3,1,3) and nonseasonal variation. And from the table (4) indicate that the Statistics of Exponential Smoothing Models are not good in RMSE, MAE comparison with ARIMA Models.

From the two tables above, we conclude that the model ARIMA (3,1,3) is the best, which gives us the lowest values for each of RMSE, and BIC, and approximately lowest value for MAE and largest value for R^2. So, we will rely on this model to estimate the predictions of the next months of the year 2018.

In the analysis that follows, we will try to improve these model through the addition of non-seasonal ARIMA :

From table (5) The test of the parameters by using Ljung-Box for the model ARIMA (3,1,3) it shows significant difference.

Table 5: Test of the parameters of visits by the model ARIMA (3, 1, 3)

Model	Model Fit statistics						Ljung-Box Q(18)		
	Normalized BIC	Stationary R-squared	R-squared	RMSE	MAPE	MAE	Statistics	DF	Sig.
Visits Mode l_1	14.487	0.724	0.634	1100.832	8.503	849.635	41.973	12	0.000

From the table(6) we can see the significance difference in the parameters of Lag1 and Lag2 by the autoregression(AR) from the part model of ARIMA (3,1,3),also is not significance difference for the parameters of Moving- average in the model and given the Figure(3) stationary in the model ARIMA(3,1,3) for the autoregression (AR) and Moving- Average.

Table 6: Model Parameters in ARIMA (3,1,3)

			Estimate	SE	t	Sig.
No Transformation	Constant		53.286	11101.56	0.005	0.996
	AR	Lag 1	0.053	0.053	-19.918	0.000
		Lag 2	0.055	0.055	-17.546	0.000
		Lag 3	0.049	0.049	0.712	0.479
	visitor-Model_1Difference		1			
	MA	Lag 1	0.722	0.722	-0.142	0.887
		Lag 2	0.719	0.719	0.135	0.893
		Lag 3	0.751	0.751	1.323	0.190
No Transformation	Numerator	Lag 0	5.512	5.512	0.000	1.000

Citation: Hanaa Elgohari, *et al.*, Forecasting the number of outpatient visits in tertiary hospital using time series based on ARIMA and ES models. Australian Journal of Basic and Applied Sciences, 13(7): 70-77. DOI: 10.22587/ajbas.2019.13.7.11

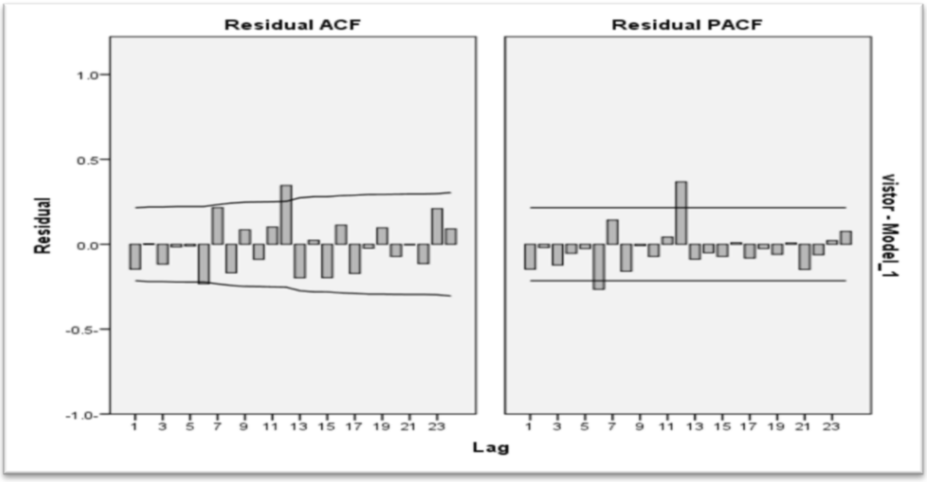

Figure 3: Residual of (ACF) & (PACF) for ARIMA (3, 1, 3)

Therefore in table(7) appears that the forecasting values will increasing for visitors Hospital per month during the year 2018, and by using the model ARIMA(3,1,3) the large forecast of visitors will be (13982) in Oct 2018 as it's shown in table(7) and Figure (4) fitting the model by original data and forecasting for 2018 as follows:

Table 7: Forecasting of No. visits per month of ARIMA(3,1,3) for 2018

Forecast of No. visits		Jan 2018	Feb 2018	**Mar 2018**	**Apr2018**	**May 2018**	**Jun 2018**
Visits Model	Forecast	13253.02	13174.62	11957.17	13483.66	13190.20	12145.33
	UCL	15337.45	15260.07	14042.52	15569.88	15275.84	14230.86
	LCL	11168.60	11089.17	9871.81	11397.44	11104.57	10059.80
		Jul 2018	Aug 2018	Sep 2018	Oct2018	Nov 2018	Dec 2018
Visits Model	Forecast	13742.33	13200.72	12356.86	13982.12	13208.60	12590.17
	UCL	15828.51	15286.47	14442.52	16068.18	15294.39	14675.92
	LCL	11656.16	11114.98	10271.21	11896.05	11122.82	10504.43

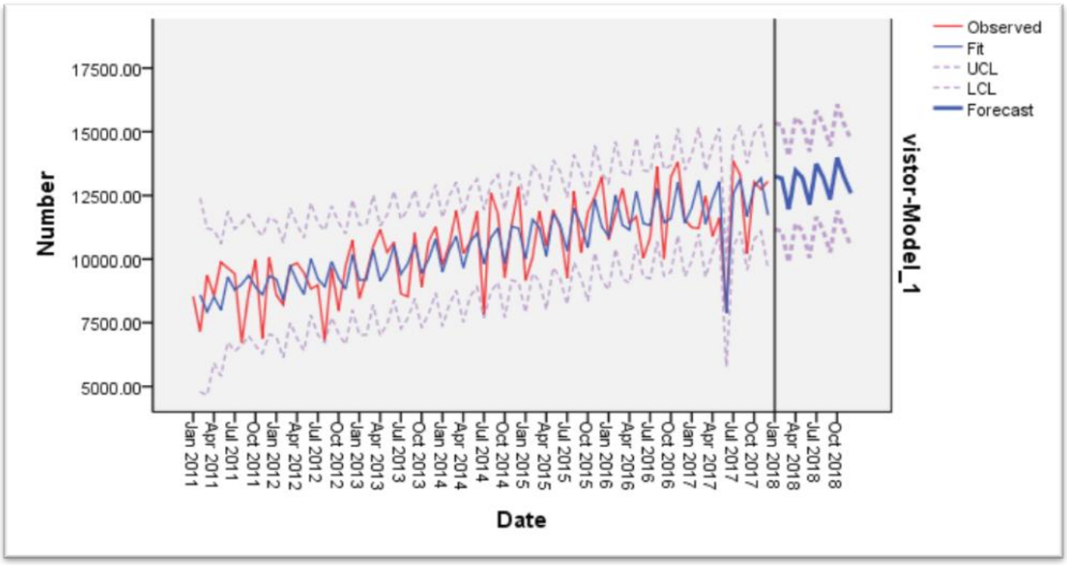

Figure 4: Fitting Model for predictions of visits of ARIMA (3, 1, 3)

4. CONCLUSIONS

This study aims to compare the ARIMA and Exponential models to predict the number of outpatient visits. The results show that by Using the model ARIMA (3,1,3) from (Box-Jenkins) because it was the best model for forecasting from the other models that we used in this research. Moreover, the model ARIMA (3,1,3) shown best results from the other models of Exponential smoothing. In general, forecasting of visits Hospital will increasing for the next months. From Forecasting results in year 2018, it shown that No. of visits patinas greater than in Oct. (13982) and less than in Mar.(11957). In addition to Increase of Number of visits patinas it need more attention and study this case from researchers and governorate.

Citation: Hanaa Elgohari, *et al.*, Forecasting the number of outpatient visits in tertiary hospital using time series based on ARIMA and ES models. Australian Journal of Basic and Applied Sciences, 13(7): 70-77. DOI: 10.22587/ajbas.2019.13.7.11

Acknowledgement

The authors would like to thank the referees for their valuable comments which helped to improve the manuscript.

REFERENCES

Calegari R, Fogliatto FS, Lucini FR, Neyeloff J, Kuchenbecker RS, Schaan BD. (2016) Forecasting daily volume and acuity of patients in the emergency department. *Comput Math Method Med*., 1–8.

Chen CF, HO WH, Chou HY, et al. (2011) Long-term prediction of emergeny department revenue and visitor volume using autoregressive integrated moving average model. *Comput Math methods Med*, 20, 1-7.

Jones SS, Thomas A, Evans RS, Welch SJ, Haug PJ, Snow GL. (2008). Forecasting daily patient volumes in the emergency department. Acad Emerg Med, 15:159–70.

Mai Q, Aboagye-safro P, Sanfilippo FM et al. (2015). predicting the number of emergency department presentation in Western Australia: a population-based time series analysis. *Emerg Med Australas*, 27, 16-21.

Paulo R, Ferrando LRS, EdsondeOp. (2014). ARIMA: an applied time series forecasting model for the bovespa stock index. *Applied mathematics*, 5, 3383-91.

Robert A. Yaffee. (1999) Introduction toTime Series Analysis and Forecastingwith Applications of SAS and SPSS. ACADEMIC PRESS, INC.pp.23.

Shirvani A, Moradi F and Moosavi AA. (2015). Time series modeling of increased soil temperature anomalies during long period." *Int Agrophys* , 29,509-15.

Sun N , Li Y, Wu F, Zheng C, Hou K and Wang K (2015). Predictive analysis of outpatient visits to a grade 3, class a hospital using ARIMA model. *In:proceedings of the 2014 international symposium on information technology(ISIT2014).Dalian:CRC press*, 285.

Yilan Lin, Min Chen, Guowei Chen, Xiaoqing Wu and Tianquan Lin. (2015).Application of an autoregressive integrated moving average model for predicting injury mortality in Xiamen, China. *BMJ open* 5:e0084-91.

Zhang G, Zhang X and Feng H. (2016). Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion. Expert Syst. 2016;33:501–16.

Zhu T, Luo L, Zhang X, Shi Y and Shen W. (2017) Time series approaches for forecasting the number of hospital daily discharged inpatients. *IEEE J Biomed Health Inf*., 21(2):515–26.