Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 8, No. 3, August 2020, pp.1525-1534

A comparison of some forecasting models to forecast the number of old people in Iraqi retirement homes

Lamyaa Mohammed Ali Hameed, Hind Waleed Abdulrahman, Iqbal Mahmood Alwan
Department of Statistics, College of Administration and Economics, Baghdad University, Iraq

ABSTRACT

Statistical methods of forecasting have applied with the intention of constructing a model to predict the number of the old aged people in retirement homes in Iraq. They were based on the monthly data of old aged people in Baghdad and the governorates except for the Kurdistan region from 2016 to 2019. Using Box-Jenkins methodology, the stationarity of the series was examined. The appropriate model order was determined, the parameters were estimated, the significance was tested, adequacy of the model was checked, and then the best model of prediction was used. The best model for forecasting according to criteria of (Normalized BIC, MAPE, RMSE) is ARIMA (0, 1, 2).

Keywords: ARIMA Models, Elderly Care, Forecasting, Time Series

Corresponding Author:

Lamyaa Mohammed Ali Hameed
Department of Statistics,
College of Administration and Economics,
Baghdad University, Iraq
E-mail: lamiaa.mohammed@coadec.uobaghdad.edu.iq

1. Introduction

Caring about the elderly is represented by paying attention to their social, health and services' aspects. Iraq has made distinctive efforts for its elderly to ensure a good life for them as the Iraqi government took a special care of them and has adapted a unique strategy that always submit to evaluation constantly, and the government-run retirement homes have already begun to receive both genders. Living in these retirement homes is free and does not result in any financial expenses. Caring about the elderly in such manners is a result of the modern era which has particularized in the increase of the ages in life expectancy rates which resulted in the rising population of elderly among the nation. These retirement homes aim to host and ensure a housing space for elderly people and provide all the health and social caring they need to achieve psychological harmony to helps them to socially adapt and provide them rest and tranquility about their lives. Caring for the elderly is not easy, so the person who has an elderly person in his family should take care of him, and if he does not provide him with full care because of their preoccupation, the elderly must have an elderly sitter or be transferred to live in an elderly home. Aging is the period during which weakness and collapse occur in the body with a disorder of mental function. The individual becomes less efficient, and has no specific role, with poor compatibility, low motivation, and socially isolated [1].in this paper we studying the Box-Jenkins methodology [2]and using it to determine the best statistical model which will be used to estimate elderly people in Baghdad and other governorates except the Kurdistan region. Considerable studies discussing the forecast in time series. Hsu, et al. [3] presented algorithm for predicting death among older adults in the home care setting. Harasheh [4] attempted to establish applicable forecasting model for electricity prices with respect to Italian power market, using neural networks. Nyoni and Nathaniel [5] used ARMA, ARIMA, and GARCH models to forecast inflation in Nigeria. Kingston et al. [6], forecasted the care needs of the older population in England over the next 20 years based on population ageing and care simulation modeling study. The overall findings are that the proportion of independent older people will increase between 2015 and 2035. Zhang et, al. [7] applied ARIMA model to forecast $M_{2.5}$ concentrations in Fuzhou, China. Victoria et al. [8], forecasted the rare earth elements price by means of transgenic time series developed with ARIMA models. Mohanasundaram et al. [9], used the seasonal autoregressive integrated moving average SARIMA models to modeling ground water level data with improved way for estimating the seasonal component by adopting 13- month moving average trend, Henry et al. [10] used data from Nigeria exchange rate to predict base on in – sample information criteria ARIMA(0,1,1). Dumitru and Gligor[11] used two different models ARIMA model and a model based on neural networks to forecast the production of electrical energy. Marta et al.[12], examined the performance in forecasting coking coal prices of traditional time series models, using robust models, ARIMA models, generalized regression neural networks and multi-layer feed forward networks. Finally, Wei in [13] predicted population age structures of China, India, and Vietnam by 2030 based on compositional data. This paper is organized as follows: Section 2 deals with models of time series and the methodology of Box and Jenkins, while section 3 contains the practical side. Lastly, section 4 includes the most important conclusions.

2. Time series

The Time Series is an essential subject to predict the data trend and the behavior of a specific phenomenon. It is possible to predict the phenomenon's data for years to come as a long-term study or predicting it for the short term thus controlling any problems that could emerge. Time series are a chain of organized and successive observations. Even though this organization or arrangement is usually based on time intervals, as noted through the equal time intervals, it is possible to arrange it using other dimensions such as Distance. It can be mathematically identified by saying that the independent time variable (t) and its corresponding values are in the dependent variable (X) and that each value in (t) has a corresponding value in the dependent variable (X) so that (X) is time function, such that:

$$X = f(t)$$

A time series can be one of two types depending on what the phenomenon's data look like so that it can be a Discrete Time Series, or it can be a Continuous Time Series.

2.1. The Models of Time Series [14]

2.1.1. Autoregressive Model

The equation for this model is:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t \tag{1}$$

Where:

- X_t = the Series' observations.
- p = the order of the model.
- ε_t = error terms that are identically and independently normally distributed with zero mean and variance(σ_ε^2).
- ϕ_i = the model's parameters, such that $i = 1,2,\dots,p$.

2.1.2. Moving Average Model

The equation for this model is:

$$X_t = a_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{2}$$

Where:

- q = the order of the model.
- θ_i = the model's parameters, and that $i = 1,2,\dots,q$.

2.1.3. Autoregressive Moving Average Model

The model will be symbolized as $\mathbf{ARMA(p,q)}$ and the general equation will be:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{3}$$

If the time series was not stationary, then the model will be referred to **ARIMA** instead of **ARMA** which would then be of the orders$(\mathbf{p}, \mathbf{d}, \mathbf{q})$. Non-Stationary Time Series can be made stationary by taking the differences between the values in the series which are known as $(X_t - X_{t-1})$ for several times until it becomes stationary series of the order d in which d represents the order of the differences.

2.1.4. Box – Jenkins approach

For first stage, identification is the first and most important stage by drawing the series and using Autocorrelation Functions(ACF), and Partial Autocorrelation Functions (PACF) in order to determine the order of the model ARIMA (p, d, q)
The Autocorrelation Function (ACF) can be used to measure the relationship between the values of the same series for several periods of time and plays an essential role in identifying the models of the time series. Through the Autocorrelation function and the Partial Autocorrelation Function (PCF) for the model is determined, and the mathematical model can be established, to find the Autocorrelation Function for a random procedure with a difference in the displacement of k as:

$$\rho_k = \frac{Cov(X_t, X_{t+k})}{\sqrt{Var(X_t)}\sqrt{Var(X_{t+k})}} = \frac{E(X_t - \mu)(X_{t+k} - \mu)}{\sqrt{E(X_t - \mu)^2 E(X_{t+k} - \mu)^2}} = \frac{\gamma_k}{\gamma_0} \quad (4)$$

The ACF and PACF functions will help to determine the order of the ARMA model that is suitable for representing data.
The second stage is estimation. In order to achieve the primary goal which is prediction, prediction quality and suitability for the time series must be guaranteed, model's coefficients are estimated which was chosen in the first stage using one of the efficient estimation methods, which are: Exact Maximum Likelihood Method, Least Squares Method, Method of Moment, and Approximated Maximum Likelihood Method.
The third stage is diagnosis of the model**.** The main assumption on which the quality of the model is based on those errors are random variables have zero mean and constant variance. The model is examined and validated by knowing the nature of the error distribution by means of a graph of the standard errors and comparing it with the standard Normal distribution. The suitability of the model and its suitability to represent the data can be tested by analyzing the auto-correlations of the residual in two ways:
Ljung-Box Test is a statistical test used to test any group of autocorrelations in the time series if it is different from zero or not. This test can also be used to verify the total randomization of data using a set of displacements [16]:

$$Q = T(T+2)\sum_{i=1}^{k}\frac{r_i^2(u)}{(T-i)} \quad (5)$$

where,
T is the length of the time series, r_i is the kth autocorrelation coefficient of the residuals. Large values of $Q*$ indicate that there are significant autocorrelations in the residual series. It can be tested against a χ^2 distribution with (k-p-q) degrees of freedom where *i* is the number of parameters estimated in the model. The hypothesis for this test is:

$$H_0: r_i(u) = r_1(u) = r_2(u) \ldots = r_K(u) = 0$$

For residual test, the coefficients of the auto -correlation and the partial auto-correlation for the residuals of the estimated model are extracted and plotted. If all values of the residual self-correlation coefficients are within confidence limits at the 95% order, the chain of residues is random, and the model used is good and appropriate.
The fourth stage is forecasting [17]. The futuristic values of the time series are found using the suitable model which was obtained in the previous stages for the best prediction possible and that the estimation is of a minuscule error and its variance is as little as possible. The prediction of the futuristic values of the observations of the time series is the conditional prediction in the period $T + 1$ at the time$T$. The $(ARIMA)(p, d, q)$ model is written as follow

$$E(X_{T+t}) = \phi_1 E(X_{T+t-1}) + \phi_2 E(X_{T+t-2}) + \cdots + \phi_{p+d} E\left(X_{T+t-p-d}\right) - \theta_1 E(\varepsilon_{T+t-1}) - \theta_2 E(\varepsilon_{T+t-1}) - \theta_{q+d} E\left(\varepsilon_{T+t-q-d}\right) \quad (6)$$

For criteria for diagnosing the model's order, in case both functions fail, **ACF** and **PACP**, to identify the suitable model, there are some criteria to identify the order of the best model without depending on the correlation functions [18]:

A. **Root mean square error**

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (7)$$
$$\text{MSE} = \frac{1}{n}\sum_{t=1}^{n} e_t^{\,2}$$

B. Mean absolute percentage error (MAPE) which is calculated as

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n} |\text{PE}_t| \quad (8)$$

$$PE_t = \left(\frac{Y_t - e_t}{Y_t}\right)*100 \quad (9)$$

C. Bayesian information criterion (BIC)

$$BIC = 2\,Ln\left(\sigma_u^2\right) + V\,Ln(M) \quad (10)$$

(M &V) represent the series' observations and the total number of the model's coefficients respectively.

3. Practical side

3.1. Describing the data

The data of this article based on the numbers of the elderly in retirement homes in Baghdad and the governorates except for the Kurdistan Region for the period of 2016-2019 using 48 observations. After the data were collected, the Box -Jenkins methodology was applied for forecasting by using the statistical program (SPSS 23). Table 1 represents the monthly statistics for the elderly in retirement homes in Iraq.

Table 1. The monthly statistics of the elderly in retirement homes in Iraq for the years (2016, 2017, 2018, and 2019)

Months	2016	2017	2018	2019
January	454	662	488	344
February	454	458	480	345
March	456	449	472	353
April	452	451	471	345
May	454	453	466	347
June	456	456	469	344
July	454	467	449	343
August	462	467	457	344
September	461	469	452	346
October	471	471	442	334
November	462	477	436	332
December	610	488	460	

3.2. Stationarity of the series

After the data collection process, we plot the data series of the monthly statistics of the elderly people in Iraqi retirement homes to identify the behavior of the time series. Figure 1 depicts a general downward trend.

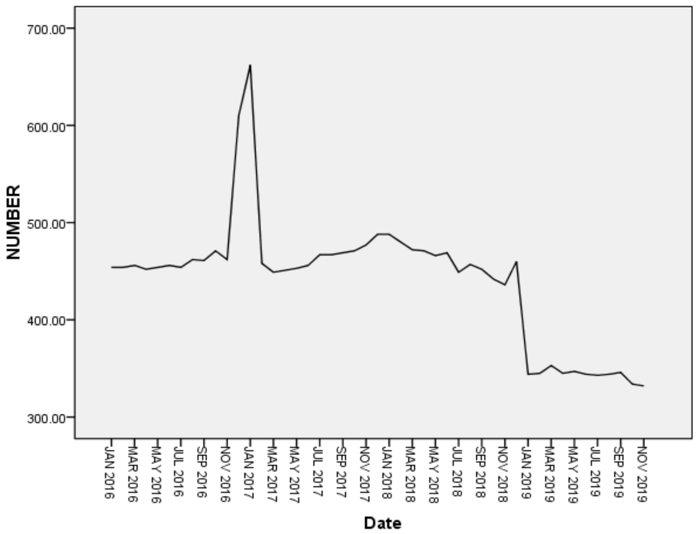

Figure 1. A graph for the monthly statistics for the original series

For more accuracy, the (ACF) and the (PACF) may be drawn as shown in Figures 2 and 3.

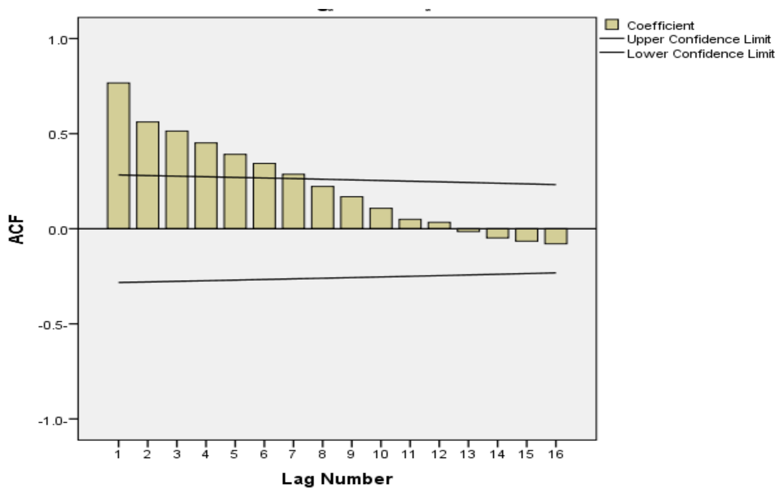

Figure 2. The Graph of autocorrelation function (ACF) for the original series

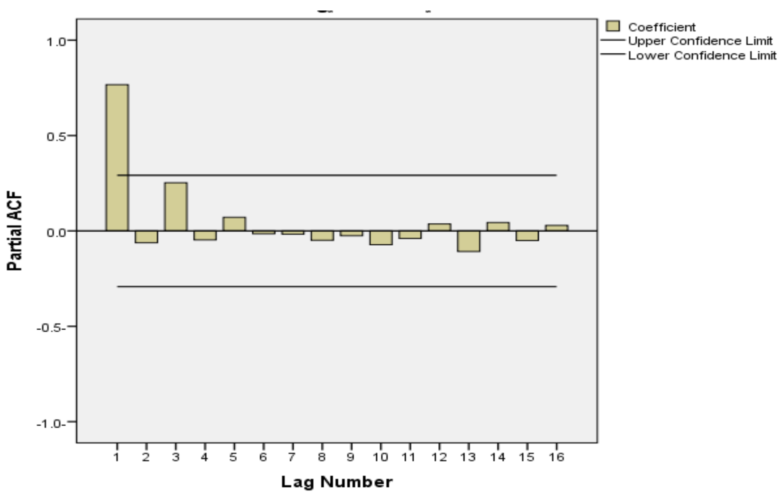

Figure 3. The partial autocorrelation function (PACF) for the original series

From Figures 2 and 3, we notice that the ACF and PACF coefficients are outside the confidence limits which indicate that the series is not stationary. In order to eliminate this non-stationary in the series, in average, we take the first difference. Through graphing the series after the first difference is taken, we notice that the series no longer has a general trend and became stationary as depicted in Figure 4.

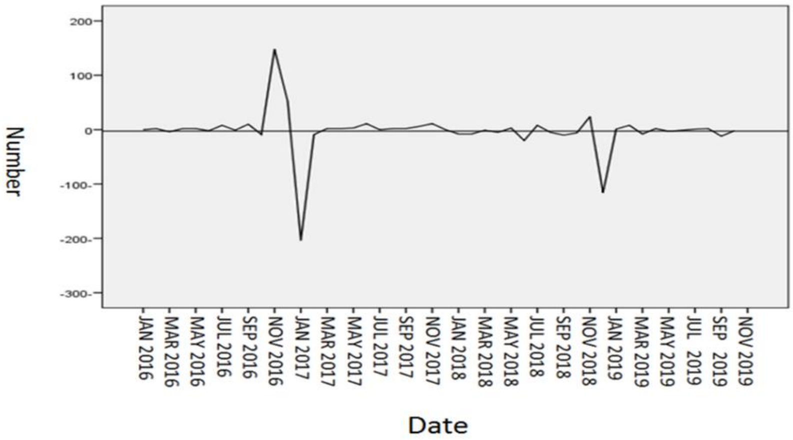

Figure 4. The time series after the first difference is taken for the series

From Figure 4, we notice that the series does not contain a general trend, and for more accuracy, the autocorrelation function (ACF) and the partial autocorrelation function (PACF) will be drawn.

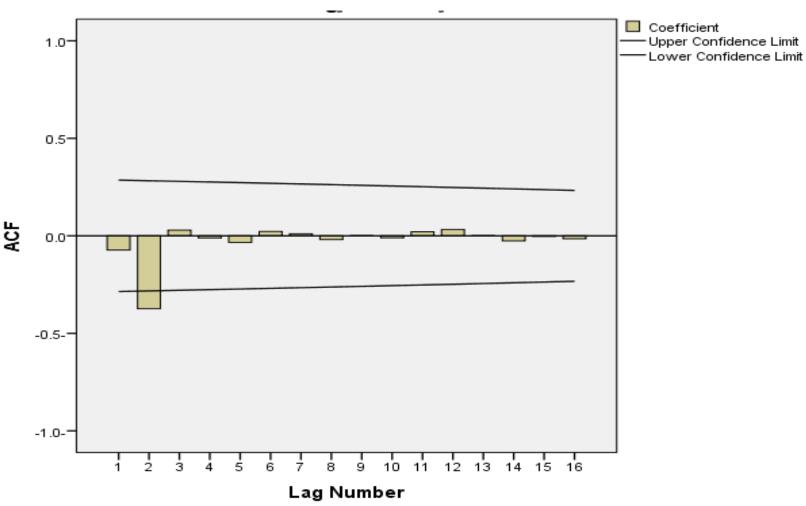

Figure 5. The autocorrelation function (ACF) after the first difference is taken for the series

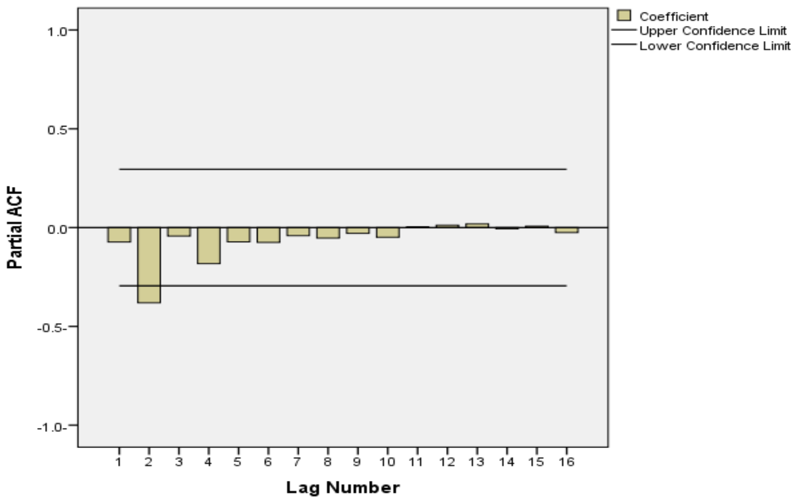

Figure 6. The partial (ACF) after the first difference is taken for the series

3.3. Choosing the best model

After taking the first difference of the values of the time series and achieving stationary in the time series, we identify the suitable model to represent the time series. To find this suitable model to represent the time series, the criteria (BIC), (MAPE), and (RMSE) have been adopted to choose the best among a group of models. Several models were approved and the best among these was chosen as shown in Table 2.

Table 2. The suggested models with their orders and the comparative value

The suggested Model	RMSE	MAPE	Normalized BIC	Notes
ARIMA (0,1,1)	42.617	4.003	7.671	
ARIMA (0,1,2)	**39.145**	**4.146**	**7.584**	**the best model**
ARIMA (0,1,3)	39.608	4.145	7.691	
ARIMA (1,1,0)	42.970	3.732	7.687	
ARIMA (1,1,1)	40.891	4.220	7.672	
ARIMA (1,1,2)	39.608	4.145	7.691	
ARIMA (1,1,3)	40.088	4.144	7.798	
ARIMA (2,1,0)	40.193	3.989	7.637	
ARIMA (2,1,1)	40.290	4.197	7.725	
ARIMA (2,1,2)	40.086	4.147	7.798	
ARIMA (2,1,3)	40.436	4.302	7.899	
ARIMA (3,1,0)	40.631	4.001	7.742	
ARIMA (3,1,1)	40.262	4.345	7.807	
ARIMA (3,1,2)	40.569	4.161	7.905	
ARIMA (3,1,3)	41.084	4.161	8.014	

The above table shows that the best model is $\mathbf{ARIMA}\ (\mathbf{0}, \mathbf{1}, \mathbf{2})$.

3.4. Testing the Model's Precision.

After identifying the model, determining its order and estimating it, the accurateness of its suitability and its efficiency must be tested through the Ljung-Box Test. By applying Ljung-Box statistic to check the fit of the model, through comparing the tabulated value of (χ^2=26.296) to the calculated value (Q=0.852) at lag (k=18) with degree of freedom k-p-q =12, the calculated value is less than the tabulated value, or through the P- value (p- value = 1.000). It is greater than (0.05). This indicates that the errors are random, i.e. we accept the null hypothesis $H_0: r_i(u) = r_1(u) = r_2(u) \dots = r_K(u) = 0$. There are no correlations between the errors) this means the model is good and appropriate. To test of the randomness of the residuals, the coefficients of ACF and PACF were graphed.

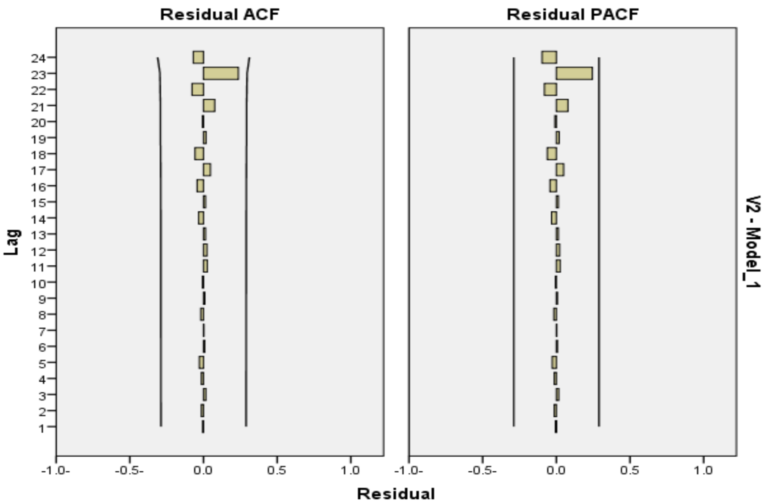

Figure 7. ACF and PACF for the residuals of the chosen model

From Figure 7, we notice that the residuals of the ACF and the PACF for the chosen model are within the confidence limits and this confirms that the model is good and suitable.

3.5. Forecasting

Depending on the chosen model, the values of the forecasting time series were estimated for the next four years taking into consideration the maximum and minimum confidence limits as shown in Table 3. From this table, we notice that the values of the forecasting time series for the monthly statistics of the elderly people in Iraqi retirement homes are clearly decreasing over time.

Table 3. The forecasting values for the monthly statistic of the elderly people in Iraqi retirement homes and their respective confidence limits

Months	Forecasting Values	Minimums	Maximums
12-2019	333	254.39	412.24
01-2020	330	226.73	434.02
02-2020	328	219.24	435.88
03-2020	325	211.94	437.54
04-2020	322	204.82	439.03
05-2020	319	197.85	440.38
06-2020	316	191.02	441.58
07-2020	313	184.31	442.66
08-2020	311	177.71	443.62
09-2020	308	171.22	444.48
10-2020	305	164.83	445.24
11-2020	302	158.52	445.92
12-2020	299	152.30	446.51
01-2021	297	146.16	447.02
02-2021	294	140.08	447.47
03-2021	291	134.08	447.84
04-2021	288	128.14	448.15
05-2021	285	122.25	448.40
06-2021	283	116.43	448.60
07-2021	280	110.66	448.74
08-2021	277	104.94	448.83
09-2021	274	99.26	448.87
10-2021	271	93.64	448.87
11-2021	268	88.06	448.82
12-2021	266	82.51	448.73
01-2022	263	77.01	448.60
02-2022	260	71.55	448.43
03-2022	257	66.13	448.23
04-2022	254	60.73	447.99
05-2022	252	55.38	447.71
06-2022	249	50.05	447.41
07-2022	246	44.76	447.07
08-2020	243	39.50	446.70
09-2020	240	34.26	446.30
10-2022	237	29.06	445.88
11-2022	235	23.88	445.43
12-2022	232	18.73	444.95
01-2023	229	13.60	444.45
02-2023	226	8.50	443.92
03-2023	223	3.42	443.37
04-2023	221	0.00	442.79

05-2023	218	0.00	442.19
06-2023	215	0.00	441.58
07-2023	212	0.00	440.94
08-2023	209	0.00	440.28
09-2023	207	0.00	439.60
10-2023	204	0.00	438.90
11-2023	201	0.00	438.18
12-2023	198	0.00	437.44

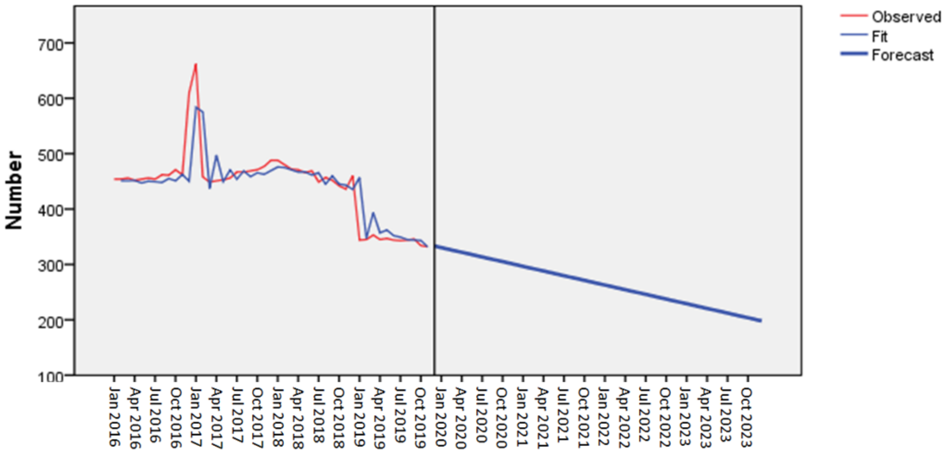

Figure 8. The graph of the forecasting values for the monthly statistic of the elderly people in iraqi retirement homes

From Figure 8, we notice that the convergence of the estimated values with the real values, and we notice that the predictive values have begun to decrease over time, which indicate that numbers of the elderly people in Iraqi retirement homes are constantly declining.

4. Conclusions

1- The time series of the old aged people in retirement homes in Iraq is not stationary about the mean, so the first difference was taken, and the time series has become stationary after this procedure.

2- The best model is ARIMA (0, 1, 2).

3- The number of the old aged people in retirement homes is decreasing over time, which indicates that their families prefer to care the elderly category in home not to put them in such houses and that is a positive state in our society.

References

[1] M. A. Khalifa, "studies in psychology of the elderly", Dar Ghareeb for publishing and distribution, Cairo, 1997.

[2] G.P. Box, and G.M. Jenkins,” time series analysis: forecasting and control, San Francisco: Holden-Day, 1976.

[3] A.T. Hsu, et al., “Algorithm for predicting death among older adults in the home care setting: study protocol for risk evaluation for support: predictions for elder-life in the community tool (REOPE), *BMJ Open*, Vol. 6, No.12, e 01366, 2016.

[4] M. Harasheh,, “Forecasting The Italian Day-Ahead Electricity Price Using Bootstrap Aggregation Method”, *European Journal of Scientific Research*, Vol. 12, No.28, pp.51-76, 2016.

[5] T. Nyoni, and S. Nathaniel, “Modeling rates of inflation in Nigeria: an application of ARMA, ARIMA, GARCH models”, MPRA paper, No.91351, 2018, online at https: //mpra.ub.uni. - muenchem.de/91351/.

[6] A. Kingston, A.C. Herrera, and C. Jagger, “Forecasting the care needs of the older population in England over the next 20 years: estimates from the Population Ageing and Care Simulation (PACSim) modeling study”, *Lancet Public Health*, Vol.3, e447–55, 2018. http://dx.doi.org/10.1016/ S2468-2667(18)30118-X.

[7] L. Zhang , et al., "Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model" Ecological Indicators ,Vol. 95, Part 1, pp. 702-710, 2018.
[8] M. Victoria, et al., "Rare earth elements price forecasting by means of transgenic time series developed with ARIMA models", *Resources Policy*, Vol. 59, pp. 95-102, 2018 . https://doi.org/10.1016/j.resourpol.2018.06.003
[9] S. Mohanasundaram et al., "A novel deseasonalized time series model with an improved seasonal estimate for groundwater level predictions", *H2Open Journal*, Vol.2, No. 1, pp. 25-44. https://doi.org/10.2166/h2oj.2019.022.

[10] A.S. Henry, K. E. Lasisi, E. A. Akpan, and D.A. A. Gwani, "Application out-of-sample forecasting in model selection on Nigeria exchange rate", *WSN*, Vol.127, No.3, pp. 225-247, 2019.
[11] C. D. Dumitru, and A. Gligor, "Wind energy forecasting: a comparative study between a stochastic model (ARIMA) and a model based on Neural Network (FFANN)", Procedia *manufacturing*, Vol.32, pp.410-417, 2019.
[12] M. Marta, et al., "Forecasting coking coal prices by means of ARIMA models and neural networks, considering the transgenic time series theory", *Resources Policy* , Vol. 61, pp. 283-292, 2019. https://doi.org/10.1016/j.resourpol.2019.02.017Get rig .
[13] Y. Wei , et al., "Predicting population age structures of China, India, and Vietnam by 2030 based on compositional data", *PLoS One*, Vol.14, No.4, e0212772, 2019. DOI: 10.1371/journal.pone.0212772
[14] W.W.S. Wei, "Time series analyses univariate and multivariate methods", Pearson Addison Wesley, 2006.
[15] R.j. Hyndman, & G. Athanasopoulos, "Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, 2018.
[16] G. Ljung, and G. Box, "On a Measure of Lack of Fit in Time Series Models", *Biometrika*, Vol. 65, pp.297-303, 1978.
[17] T.C. Mills, "Applied time series analysis a practical Guide to modeling and forecasting "Academic press, 2019.
[18] R. Adhikari, , R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting, Lap Lambert Academic Publishing, 2013.
[19] R.j. Hyndman and G. Athanasopoulos, Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia, 2018.