

# The Use of Some of the Information Criterion in Determining the Best Model for Forecasting of Thalassemia Cases Depending on Iraqi Patient Data Using ARIMA Model

Rana Sabeeh Abbood Alsudani<sup>1,2</sup>, Jicheng Liu<sup>1</sup>

<sup>1</sup>School of Mathematics and Statistics, Huazhong University of Science and Technology, Wuhan, China

<sup>2</sup>School of Mathematics College of Basic Education, Maysan University, Maysan, Iraq

Email: rana19792015@hotmail.com

**How to cite this paper:** Alsudani, R.S.A. and Liu, J.C. (2017) The Use of Some of the Information Criterion in Determining the Best Model for Forecasting of Thalassemia Cases Depending on Iraqi Patient Data Using ARIMA Model. *Journal of Applied Mathematics and Physics*, 5, 667-679.

<https://doi.org/10.4236/jamp.2017.53056>

**Received:** January 10, 2017

**Accepted:** March 21, 2017

**Published:** March 24, 2017

Copyright © 2017 by authors and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Thalassemia is a major health problem in Iraq, and despite a prevention programme. There has been no decrease in the prevalence of the disease, due to a lack of awareness, implying that genetic counseling was a failure. This failure has been attributed to a lack of recognition of problems related to Thalassemia, unorganised teamwork and services, lack of knowledge and insufficient numbers of extension workers, lack of Thalassemia support groups, and inadequate research in Thalassemia prevention and control. Autoregressive Integrated Moving Average (ARIMA) model and forecasting has become a major tool in different applications. The ARIMA model introduced by Box and Jenkins (1971) is among the most effective approaches for analysing time-series data. In this study, we used Box and Jenkins methodology to build an ARIMA model to forecast the number of people with Thalassemia, for the period from 2016-2018, from the data base from Maysan Health Center specific for Thalassemia the Maysan Province, Iraq. After the model selection, the best model for forecasting was ARIMA (0, 1, 1) and of models were used for forecasting Thalassemia.

## Keywords

Thalassemia, The Information Criterion, Box-Jenkins Methodology

## 1. Introduction

The Arabisation (The Arabic meaning) of Thalassemia is Mediterranean disease, where the Middle East and the Gulf Arab states are included, Thalassemia, which is spread throughout Iraq and other Middle East countries, is one of the most

famous genetic diseases that lead to severe anaemia and other complications in the long term. It requires quoting a regular blood transfusion every three to four weeks, accompanied by pain to the patients and great suffering to their families. On top of this, the transfer of continuous blood also leads to the accumulation of iron in the vital organs of the body, such as the liver and heart, which leads to serious complications. Infected cases of this disease are divided into two types: Thalassemia Minor—characterises people who have the disease, but are living a normal life and not complaining of any symptoms. The second type is Thalassemia major, which can occur in offspring of two carriers of Thalassemia Minor, with 25% likelihood that their children are not infected with the disease, 25% that they suffer a severe form of the disease, and 50% that they have Thalassemia Minor [1]. Iraq is witnessing a monthly birth-rate of between (100 - 150) children with the disease, which is the same as getting approximately (2000) cases per year and these numbers are disastrous compared to neighbouring countries. Another classification of Thalassemia is according to the site of the defect in the protein molecule of the red blood cells, according to which:  $\alpha$ - and  $\beta$ -Thalassemia,  $\beta$ -Thalassemia is a common hereditary disorder caused by mutations in one or more of the  $\beta$ -globin gene loci that result in reduced  $\beta$ -globin production [2]. Recently, more than 200 different mutations have been detected affecting a variety of levels of  $\beta$ -globin gene expression causing  $\beta$ -Thalassemia. These mutations are not uniformly distributed, but have a geographical particularity and racial origin, as each is characterised by the presence of a few common mutations and variable numbers of rare ones [3]. Mutations such as nucleotide substitutions and/or frame shifts of the insertion/deletion kind have been reported to interfere with the transcription of the  $\beta$ -globin gene, splicing procedures, and translation of  $\beta$ -globin gene mRNA, which causes either absence or reduction of synthesis of  $\beta$ -globin chains [4]. One of the most important elements of building health is the prevention of all diseases, including serious diseases, such as Thalassemia, that cause a high percentage of deaths compared to other diseases, due to the increasing number of people infected with this disease in recent times. This study was conducted in order to forecast the prevalence of this disease in future, which has been increasing in all areas of Iraq, especially in the province of Maysan. This study was based on monthly data for the patients with Thalassemia for the period between 2005-2015, and used data of patients as a series of time for the purpose of analysis for optimal modelling using the ARIMA Methodology, the forecasting to predict the numbers of people with this disease in subsequent, periods in order to take the necessary measures and substitutions to reduce morbidity in the future. A time-series typically consists of a set of observations of a variable taken at evenly spaced intervals of time [5]. The most comprehensive of all popular and widely known statistical models which have been used in the last four decades for time series forecasting are the Box-Jenkins method. However, the ARIMA model is only a class of linear model and, thus, it can only capture linear feature of data time series [6]. Many of standards determining the rank of model have been proposed by researchers [7] [8] [9]. There are

different models of precision can juggle in time series analysis to clarify the given set of data, not be easy to choose the better model, in many cases, it has developed several criteria to compare models in the process of selecting the rank of model, and derive the importance of selecting the rank of the model from the fact that choosing the lowest rank of the actual rank of the model leads to inconsistency of the model parameters, while choosing a higher rank than the actual rank of the form to increase the model variance; this leads to a loss of accuracy due to the increase in the number of model parameters chosen.

## 2. Method

For realizing the forecast of the analysed time-series we use modern methods, such as ARIMA models, because they are among the models that can analyse large time-series data and forecast future cases

### 2.1. Models of Box & Jenkins

The pioneers in this area were Box and Jenkins, who popularized an approach that combines the moving average and the autoregressive models (1971). An ARMA ( $p, q$ ) model is a combination of AR ( $p$ ) and MA ( $q$ ) models, and is suitable for univariate time-series modelling. In an AR ( $p$ ) model the future value of a variable is assumed to be a linear combination of  $p$  past observations and a random error, together with a constant term. Mathematically, the AR ( $p$ ) model can be expressed as [10]:

$$Y_t = c + \sum_{i=1}^p \varphi_i y_{t-i} + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t \quad (1)$$

Here  $y_t$  and  $\varepsilon_t$  are respectively the actual value and random error (or random shock) at time period  $t$ ,  $\varphi_i$  ( $i = 1, 2, \dots, p$ ) are model parameters and  $c$  is a constant. The integer constant  $p$  is known as the order of the model. Sometimes the constant term is omitted for simplicity. Usually, for estimating parameters of an AR process using the given time series, the Yule-Walker equations are used. Just as an AR ( $p$ ) model regresses against past values of the series, an MA ( $q$ ) model uses past errors as the explanatory variables. The MA ( $q$ ) model is given by [11]

$$y_t = \mu + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t = \mu + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Here  $\mu$  is the mean of the series  $\theta_j$  ( $j = 1, 2, \dots, q$ ) are the model parameters and  $q$  is the order of the model. The random shocks are assumed to be a white noise process, *i.e.* a sequence of independent and identically distributed (i.i.d) random variables with zero mean and a constant variance  $\sigma^2$ . Generally, the random shocks are assumed to follow the typical normal distribution. Thus, conceptually, a moving average model is a linear regression of the current observation of the time series against the random shocks of one or more prior observations. Fitting an MA model to a time series is more complicated than fitting an AR model because in the case of the former the random error terms are not fore-seeable. Autoregressive (AR) and moving average (MA) models can be ef-

fectively combined together to form a general and useful class of time series models, known as the ARMA models. Mathematically an ARMA  $(p, q)$  model is represented as

$$y_t = c + \varepsilon_t + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad (3)$$

Usually ARMA models are manipulated using the lag operator notation. The lag or Backshift operator is defined as  $Ly_t = y_{t-1}$ . Polynomials of lag operator or lag polynomials are used to represent ARMA models as follows [11]:

$$AR(p) \text{ model: } \varepsilon_t = \varphi(L) y_t \quad (4)$$

$$MA(q) \text{ model: } y_t = \theta(L) \varepsilon_t \quad (5)$$

$$ARIMA(p, q) \text{ model: } \varphi(L) y_t = \theta(L) \varepsilon_t \quad (6)$$

Here  $\varphi(L) = 1 - \sum_{i=1}^p \phi_i L^i$  and  $\theta(L) = 1 + \sum_{j=1}^q \theta_j L^j$ .

It is shown in that an important property of  $AR(p)$  process is invertibility, *i.e.* an  $AR(p)$  process can always be written in terms of an  $MA(\infty)$  process. Whereas for an  $MA(q)$  process to be invertible, all the roots of the equation  $\theta(L) = 0$  must lie outside the unit circle. This Condition is known as the Invertibility Condition for an  $MA$  process.

## 2.2. ARIMA Model

In both statistics and econometrics, time series analysis of an autoregressive integrated moving average, an ARIMA model is the integration of an autoregressive moving average (ARMA) model. These models are fitted to time-series data, either to better understanding the data or to forecast future points in the series (forecasting). It is applied, in some cases where the figures show proof that they are not stationary, where an initial differencing step (corresponding to the integrated fraction of the model) can be applied to reduce the non-stationarity [12], non-periodical ARIMA models which are generally denoted by ARIMA  $(p, d, q)$  where parameters  $p$ ,  $d$ , and  $q$  are non-negative integers,  $p$  is the order of the Autoregressive model,  $d$  is the degree of differencing, and  $q$  is to arrange of the Moving-average model. ARIMA models form an important part of the Box-Jenkins approach to time-series modeling [13]. ARIMA models can be seen as a chain of two models. The first is not fixed:

$$x_t = \theta_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_t - q \quad (7)$$

where  $X_t$  and  $e_t$  are the actual values and random error at time  $t$ , respectively,  $\phi_i (i = 1, 2, \dots, p)$  and  $\theta_j (j = 1, 2, \dots, q)$  are model parameters.  $p$  and  $q$  are integers and often referred to as orders of autoregressive and moving average polynomials respectively.

## 2.3. Steps of ARIMA Methodology [14]

- **Analysis of the series:** The first step in the process of modelling is to check for the stationary of the time series data.
- **Identification of the model:** This step aimed to detect periodically and to

identify the order of seasonal autoregressive terms and seasonal moving average terms. This stage includes calculation of the estimated autocorrelation function (ACF) and estimation of partial autocorrelation function (PACF) these functions measure the statistical dependence between observations of data outputs.

- **Estimation of ARIMA parameters:** The estimation of ARIMA parameters is achieved by the nonlinear least squares method. The values of the model coefficients are determined in relation to a particular criterion; one of these may be the maximum likelihood criterion. It can be shown that the likelihood function associated with a correct ARIMA model, used to determine the estimates of maximum likelihood of the parameters, contains all the useful information from the data series about the model's parameters.
- **Diagnostic checking:** In this stage it is assumed that the errors represent a stationary process and the residues are white noise (or independent if the distribution is normal), a normal distribution with mean and variance stable. The tests used to validate the model are based on the estimated residues. It is checked that the components of this vector are autocorrelated. If there is autocorrelation, the checked model is not correctly specified. In this case, the dependencies between the components series are specified in an incomplete manner, and we have to return to the model identification step and try another model. Otherwise, the model is good and can be used to make predictions for a given time horizon.
- **Forecasting.**

### 3. Criteria for Selection of the Rank of the Model

#### 3.1. Bayesian Information Criteria (BIC) [8]

$$BIC = n \ln \hat{\sigma}_a^2 - (n - M) \ln(1 - M/n) + M \ln(n) + M \ln \left[ \left( \frac{\hat{\sigma}_y^2}{\hat{\sigma}_a^2} \right) / M \right] \quad (8)$$

where:

$P$ : model rank,

$n$ : Views,

$M$ : The number of parameters,

$\hat{\sigma}_y^2$ : Estimator series variance,

$\hat{\sigma}_a^2$ : Estimator error variance,

$$\hat{\sigma}_a^2 = \sum_{t=1}^n (y_t - \hat{y}_t)^2 / (n - p) \quad (9)$$

#### 3.2. Akaike Information Criterion [7]

$$AIC(M) = n \ln \hat{\sigma}_a^2 + 2M \quad (10)$$

or

$$AIC(p, q) = \ln \hat{\sigma}_a^2 + 2(p + q)/n \quad (11)$$

where:

$M: p + q$ ;  $p, q$ : model rank;  $n$ : views;  $\hat{\sigma}_a^2$ : Estimator error variance

### 4. The Application of Data

This study is based on the time-series data provided by the hereditary blood disease centre in Iraq (Maysan) from people diagnosed with Thalassemia for the period 2005-2015.

#### 4.1. Analysis of the Series

The first step in the process of modelling is to check for the stationary of the time series data. This is done by observing the graph of the data or autocorrelation and the partial autocorrelation functions [2]. It notes, through a graph of the time series, that there are high rates of Thalassemia as compared with previous years (Figure 1).

#### 4.2. Stationary

The first stage of ARIMA model building is to identify whether the variable, which is being forecasted, is stationary in the time-series or not. By stationary, with autocovariance functions, we can define the covariance stationarity, or weak stationarity. In the literature, usually stationarity means weak stationarity, unless otherwise specified. The time series  $(x_t, t \in Z)$  s.

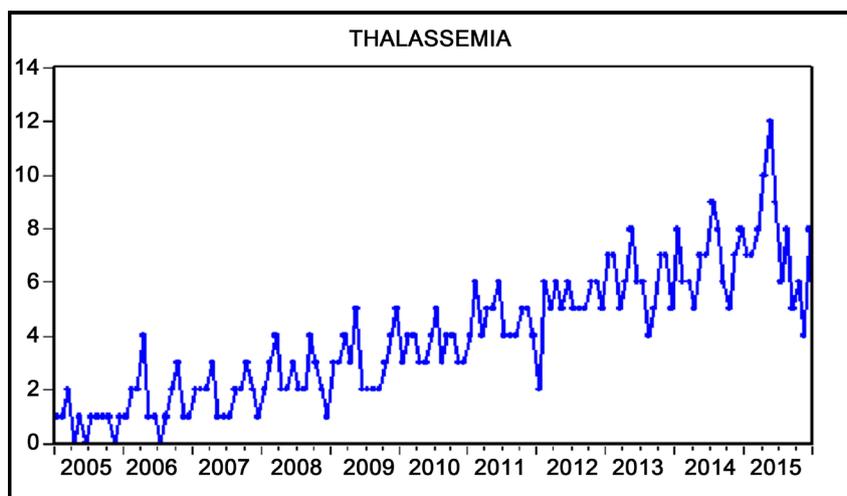
Where  $Z$  is the integer set is said to be stationarity if:

$$\text{var}(x_t) < \infty \forall t \in Z.$$

$$EX_t = \mu \forall t \in Z$$

$$\gamma x(s, t) = \gamma x(s + h, t + h) \forall s, t, h \in Z$$

The time plot of the  $\{x_t\}$  must have three features: finite variation, constant first moment, and that the second moment  $\gamma x(s, t)$  only depends on  $(t - s)$



**Figure 1.** Graph of original series, increase the number of patients suffering from thalassemia year 2005-2015 in southern Iraq, time series before differencing shows the variability of the series appears to be changing with time. Therefore the mean and variance are not constant, suggesting that the series is not stationary.

and not depends on  $s$  or  $t$  In light of the last point, we can rewrite the auto covariance function of a stationary process as

$$\gamma_x(h) = \text{Cov}(x_t, x_{t+h}) \text{ for } t, h \in Z \tag{12}$$

Also, when  $x_t$  is stationary we must have

$$\gamma_x(h) = \gamma_x(-h) \tag{13}$$

When  $h = 0$ ,  $\gamma_x(0) = \text{cov}(x_t, x_t)$  is the covariance of  $x_t$  so the autocorrelation function for stationary time series  $x_t$  is defined to be

$$\rho_x(h) = (\gamma_x(h)) / (\gamma_x(0)) \tag{14}$$

In **Figure 1** the thalassemia data, clearly shows that the data is not stationary (actually, it shows an increasing trend in time series). The ARIMA model cannot be built until we make this series stationary. We first have to differentiate the time series “ $d$ ” times, to obtain a stationary series in order to have an ARIMA ( $p, d, q$ ) model with “ $d$ ” as the order of differencing. Care should be taken in differencing, as over differencing will tend to an increase in the standard deviation, rather than a reduction. The best idea is to start with differencing of the lowest order (of first order,  $d = 1$ ) and test the data for unit root problems 6.

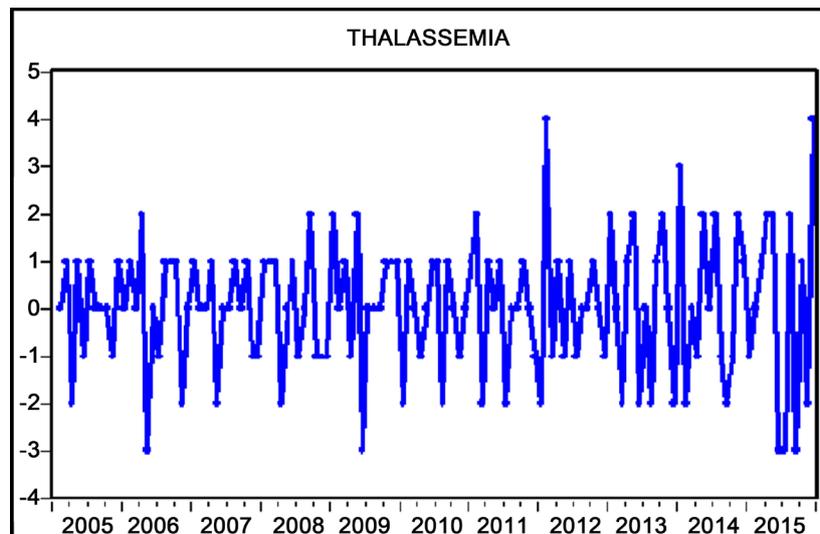
First Difference:  $Z_t = y_t - y_{t-1}$  where  $t = 2, 3 \dots n$

Second Difference:  $Z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$  where  $t = 3, 4 \dots n$

As a result we obtained a time-series of first order differencing and **Figure 2**, below, is the line plot of the first order differenced Thalassemia data. It can easily be inferred from the graph that the time series appears to be stationary both in its mean and variance. Moreover, the time series data was subjected to Dickey-Fuller test to check the number of differenced time-series data for stationarity.

**Using the adjusted (ADF) test [15]:**

Our null hypothesis ( $H_0$ ) in the test is that the time series data is non-stationary; while the alternative hypothesis ( $H_a$ ) is that the series is stationary. The



**Figure 2.** Time series for after first difference  $d = 1$  become stationary.

hypothesis is then tested by performing appropriate differencing of the data in  $d$ th order and applying the ADF tests to the differenced time series data. First order differencing ( $d = 1$ ) means we generate a table of differenced data of current and immediately previous time  $\Delta x_t = x_t - x_{t-1}$ . The ADF test result, as obtained upon application, is shown below (using Eviews program) in **Table 1**.

We, therefore, fail to accept  $H_0$  and, hence, can conclude that the alternative hypothesis is true *i.e.* the series is stationary in its mean and variance. Thus, there is no need for further differencing of the time series and we adopt  $d = 1$  for our ARIMA ( $p, d, q$ ) model. This test enables us to go further in steps for ARIMA model development *i.e.* to find suitable values of  $p$  in AR and  $q$  in MA in our model. For that, we need to examine the correlogram and partial correlogram of the stationary (first order differenced) time series. The dickey fuller test depends on three simple equations and assumes a random the context of a pattern of downhill autocorrelation of (1) these equations are:

$$\Delta x_t = \alpha_1 x_{t-1} + e_t \tag{15}$$

$$\Delta x_t = \alpha_0 + \alpha_1 x_{t-1} + e_t \tag{16}$$

$$\Delta x_t = \alpha_0 + \alpha_1 x_{t-1} + B_{t+} e_t \tag{17}$$

Whereas:

$\Delta$  : The first difference factor, which:  $\Delta x_t = x_t - x_{t-1}$ ,

$e_t$  : White noise process.

**Table 1.** The augmented dickey-fuller test results.

Augmented Dickey –Fuller test statistic	t-statistic	Prob.		
	-11.38261	0.0000		
1% level	-3.482035			
Test critical values	5% level	-2.884109		
	10% level	-2.578884		
Augmented Dickey-Fuller Test Equation				
Dependent variable :D(DTHALASSEMIA)				
Method Least Squares				
Sample(adjusted)2005 M05 2015M12				
Variable	Coefficient	Std. error	t-statistic	Prob.
DTHALASSEMIA(-1)	-2.176864	0.191245	-11.38261	0.0000
D(DTHALASSEMIA(-1)	0.699137	0.145316	4.811158	0.0000
DTHALASSEMIA(-2)	0.381127	0.085537	4.455672	0.0000
C	0.094297	0.109530	0.860923	0.3909
R-squared	0.719809	Mean dependent var		0.046875
Adjusted R-squared	0.713030	S.D dependent var		2.310058
S.E. of regression	1.237487	Akaike info criterion		3.294795
Sum squared resid	189.8905	Schwarz criterion		3.383920
Log likelihood	206.8669	Hannan-Quinn criter		3.331007
F-Statistic	106.1852	Dubin-Watson stat		1.998406
Prob(F-statistic)	0.000000			

### 4.3. Correlogram and Partial Correlogram

Figure 3, below, represents the plot of correlogram (auto-correlation function, ACF) from lags 1 to 20 of the first order differenced time-series of the Thalassemia patients (Figure 3). The above correlogram infers that the auto-correlation and partial autocorrelation between lag 1 and 20 does not exceed the significance limits and auto-correlations tail off to zero the autocorrelation at rest.

### 4.4. Selecting the Rank of Model [16]

Seen from the Table 2 that the best model for this series is ARIMA (0, 1, 1) for having the lowest values for the standards of information. This is perhaps the most commonly used model in forecasting. It is the exponential smoothing model. The general form of the model is:

$$Z_t - Z_{t-1} = c + a_t - \theta a_{t-1}, \quad |\theta| < 1 \tag{18}$$

Following the common practice, we shall assume  $c = 0$ . Since the model is invertible, the  $\pi$  weights are  $\pi_i = \theta_i - 1(1 - \theta)$  for  $i \geq 1$ . Thus

$$Z_t - Z_{t-1} = 0.05577677 + 0.98928205t - 1$$

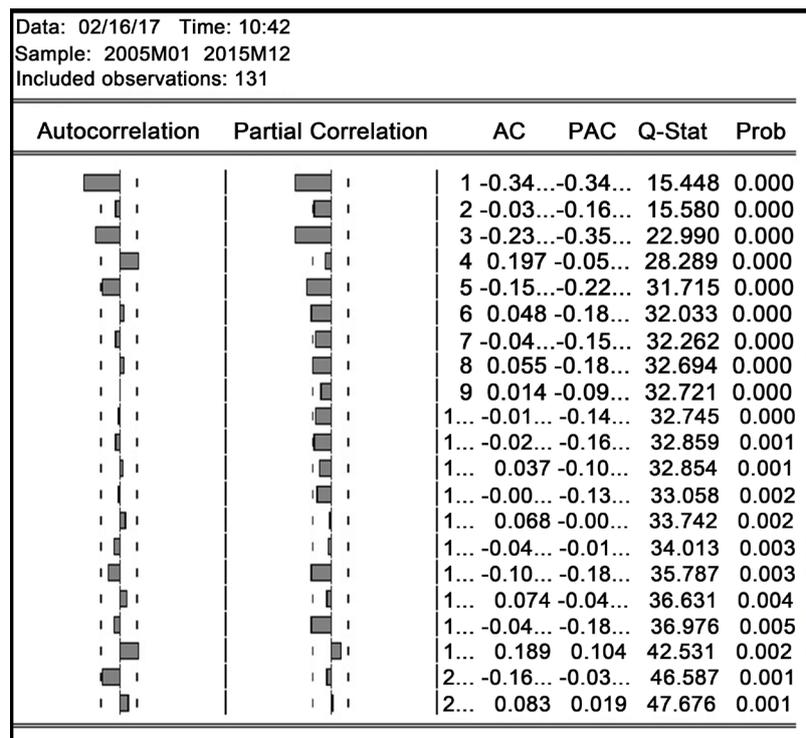


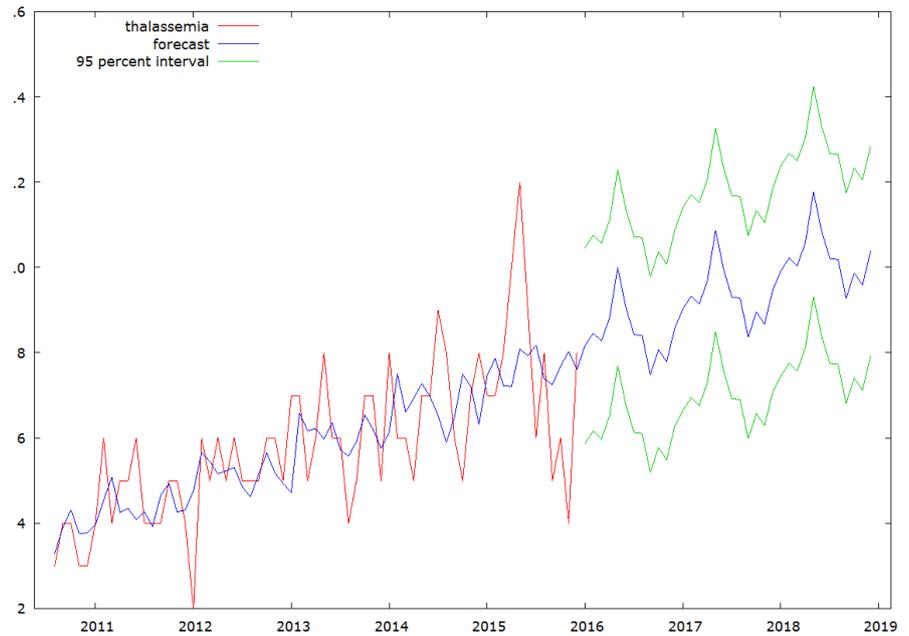
Figure 3. Plot of the autocorrelation and partial autocorrelation functions of the differenced series.

Table 2. Shows the results of information criterion for various models of the time series.

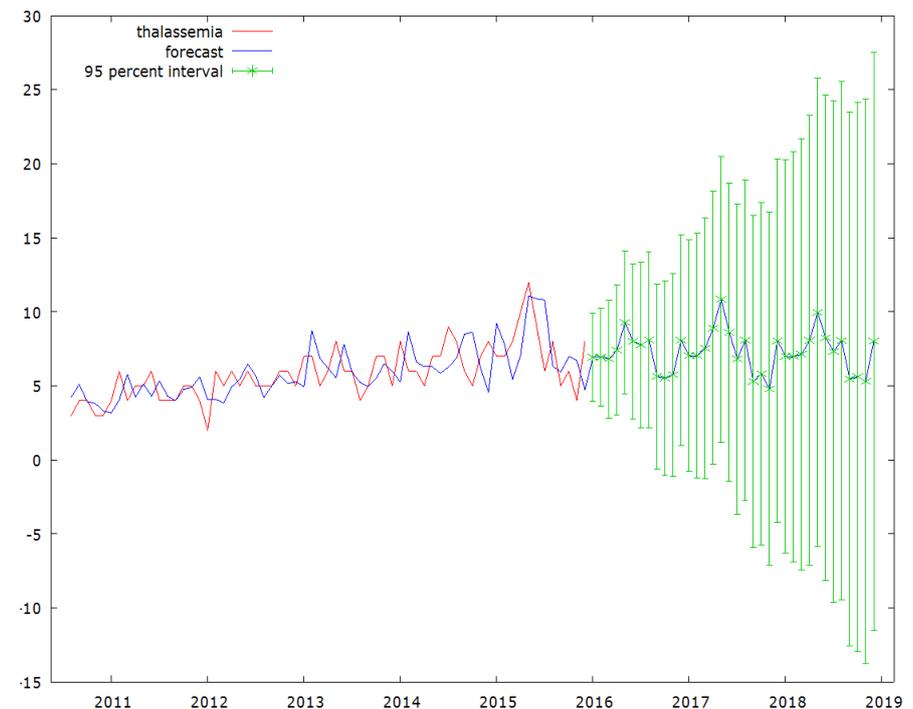
Models	AIC	BIC
ARIMA (1, 1, 0)	89.7921	-65.0102
ARIMA (0, 1, 1)	38.5633	-122.7428

### 5. The Forecasting

The results showed that after we apply (eviews, gretl, minitab) specific statistical programs on our data and charts, there is an increase in cases with Thalassemia in the coming years from 2016 to 2018, according to the **Figure 4**, **Figure 5** and **Table 3**, where the number of patients monthly will be between (7 - 11) patients



**Figure 4.** Forecasting the expected number of patients with thalassemia years (2016-2018) using a form ARIMA (0, 1, 1).



**Figure 5.** Forecasting the expected number of patients with thalassemia years (2016-2018) using a form ARIMA (1, 1, 0).

**Table 3.** Year Forecasting for thalassemia patients The number of people forecaster their thalassemia in 2016-2018 For 95% confidence intervals,  $z(0.025) = 1.96$ .

Obs	Forecast ARIMA (1, 1, 0)	Forecast ARIMA (0, 1, 1)
2016:01	6.94	8.17
2016:02	6.94	8.46
2016:03	6.82	8.28
2016:04	7.41	8.82
2016:05	9.27	10.00
2016:06	8.01	9.07
2016:07	7.77	8.43
2016:08	8.10	8.40
2016:09	5.65	7.48
2016:10	5.53	8.07
2016:11	5.76	7.78
2016:12	8.08	8.58
2017:01	7.04	9.03
2017:02	7.04	9.32
2017:03	7.54	9.14
2017:04	8.91	9.68
2017:05	10.85	10.87
2017:06	8.61	9.94
2017:07	6.82	9.30
2017:08	8.08	9.28
2017:09	5.32	8.37
2017:10	5.82	8.96
2017:11	4.79	8.67
2017:12	8.05	9.46
2018:01	6.99	9.92
2018:02	6.99	10.22
2018:03	7.14	10.04
2018:04	8.07	10.58
2018:05	9.96	11.78
2018:06	8.26	10.85
2018:07	7.33	10.21
2018:08	8.06	10.19
2018:09	5.47	9.28
2018:10	5.62	9.87
2018:11	5.29	10.39
2018:12	8.02	9.969

**Table 4.** The values of criterion of final prediction error.

Model	FPE
ARIMA (0, 1, 1)	0.03256
ARIMA (1, 1, 0)	0.06542

and these numbers are high compared to previous years, where highest forecasting for the patients is during June of 2018. Some months during the period of 2005-2015 did not record any cases while in the forecasting period 2016-2018 every month is accepted to have some patients.

### Final Prediction Error (FPE) [9]

Been using the final prediction error (FPE) a good estimate of prediction error for model with  $n$  parameters is given by the final prediction error see **Table 4**:

$$FPE = \sigma_r^2 (N, \beta^{\wedge}) (N + n + 1) / (N - n - 1) \quad (19)$$

$\sigma_r^2$  =variance of the residuals.

$N$  is the number of values in the estimation data set.

From **Table 4** that the model ARIMA (0, 1, 1) has the lowest value of the criterion of (FPE).

## 6. Conclusions

- 1) ARIMA model was suitable for application to Thalassemia data and analysis of other similar medical data.
- 2) By application of ARIMA model, we were able to forecast future cases easily and accurately.
- 3) Cases of Thalassemia will increase within coming years, which means that, currently, no serious efforts are offered to solve or treat this disease in Iraq.

## 7. Recommendations

- 1) Doing more similar studies using data from other provinces of Iraq to overcome the disease.
- 2) Doing pre-marriage tests to detect the carriers to limit the future incidence.
- 3) The bone marrow transplant is the most suitable treatment of the disease but it is so costly for patients that governmental support will be very helpful.
- 4) Should take care to Sanitaryware industry pain less and more safety during blood transfusions to relieve pain and prevent infection from bacterial and viral diseases.

## References

- [1] Pearson, H. (2003) Abnormalities of Hemoglobin Structure and Function.
- [2] Mok, S., Imwong, M., Mackinnon, M.J., *et al.* (2011) Artemisinin Resistance in Plasmodium Falciparum Is Associated with an Altered Temporal Pattern of Transcription. *BMC Genomics*, **12**, 391. <https://doi.org/10.1186/1471-2164-12-391>
- [3] Cao, A. and Galanello, R. (2010) Beta-Thalassemia. *Genetics in Medicine*, **12**, 61-76.

<https://doi.org/10.1097/GIM.0b013e3181cd68ed>

- [4] Lahiry, P., Al-Attar, S.A. and Hegele, R.A. (2008) Understanding Beta-Thalassemia with Focus on the Indian Subcontinent and the Middle East. *The Open Hematology Journal*, **2**, 5-13 <https://doi.org/10.2174/1874276900802010005>
- [5] Popescu, S. (1991) Practica modelarii si predictiei seriilor de timp (Practical Modeling and Prediction of Time Series). Technical Publishing, Bucharest.
- [6] Zhang, G.P. (2003) Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. *Neurocomputing*, **50**, 159-175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- [7] Kadilar, C. and Erdemir, C. (2003) Modification of the Akaike Information Criterion to Account for Seasonal Effects. *Journal of Statistical Computation and Simulation*, **73**, 135-143. <https://doi.org/10.1080/00949650215731>
- [8] Mcquarrie, A.D.R. and Tsai, C. (1998) Regression and Time Series Model Selection. World Scientific Publishing Company, Singapore. <https://doi.org/10.1142/3573>
- [9] Khim, S. and Mahendran, S. (2002) The Performance of AICC as an Order Selection Criterion in ARMA Time Series Models. *Pertanika Journal of Science & Technology*, **10**, 25-33.
- [10] Hipel, K.W. and McLeod, A.I. (1994) Time Series Modelling of Water Resources and Environmental Systems. Amsterdam, Elsevier.
- [11] Cochrane, J.H. (1997) Time Series for Macroeconomics and Finance. Graduate School of Business, University of Chicago, Chicago.
- [12] Meyer, Y. (1993) Wavelets, Algorithm & Applications. SIAM, Philadelphia.
- [13] Chatfield, C. (1999) The Analysis of Time Series: Theory and Practice. Chapman and Hall, London.
- [14] Box, G.E., *et al.* (2015) Time Series Analysis: Forecasting and Control. John Wiley & Sons.
- [15] 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**, 427-431. <https://doi.org/10.1080/01621459.1979.10482531>
- [16] Raicharoen, T., Lursinsap, C. and Sanguanbhoki, P. (2003) Application of Critical Support Vector Machine to Time Series Prediction. *Proceedings of the International Symposium on Circuits and Systems*, Bangkok, 25-28 May 2003, 741-744. <https://doi.org/10.1109/iscas.2003.1206419>

**Submit or recommend next manuscript to SCIRP and we will provide best service for you:**

Accepting pre-submission inquiries through Email, Facebook, LinkedIn, Twitter, etc.

A wide selection of journals (inclusive of 9 subjects, more than 200 journals)

Providing 24-hour high-quality service

User-friendly online submission system

Fair and swift peer-review system

Efficient typesetting and proofreading procedure

Display of the result of downloads and visits, as well as the number of cited articles

Maximum dissemination of your research work

Submit your manuscript at: <http://papersubmission.scirp.org/>

Or contact [jamp@scirp.org](mailto:jamp@scirp.org)