Journal of Computations & Modelling, vol.6, no.2, 2016, 1-14
ISSN: 1792-7625 (print), 1792-8850 (online)
Scienpress Ltd, 2016

Modelling and forecasting infant mortality using Autoregressive Moving Average (ARMA) Model

Akintunde M.O.[1], Kgosi, P.M.[2], Olalude G.A.[3] and Ajala, A.G.[4]

Abstract

The study focuses on Modelling and Forecasting Infant Mortality using Time Series methodology called ARMA. The data used for the study was obtained from the hospital record of the General Hospital, Asubiaro, Osun State, south western, Nigeria. The data was subjected to various stationarity tests (graphical and unit root test) and the data was stationary at the second difference. Thereafter, various ARMA model was fitted from which ARMA(1,1) was chosen. The model fitted has a very powerful forecasting ability as Theil-U index value of (0.722313) obtained was moderate, bias proportion (0.0076) and variance proportion (0.00342) almost tends to zero, and covariance proportion (0.99608) is very high. All these combined together enhanced good forecast performance.

Mathematics Subject Classification: 62

---

[1] Department of Statistics, The Federal Polytechnic, Ede, Osun State, Nigeria.
[2] Department of Statistics, University of Botswana, Gaborone, Botswana.
[3] Department of Statistics, The Federal Polytechnic, Ede, Osun State, Nigeria.
[4] Department of Statistics, The Federal Polytechnic, Ede, Osun State, Nigeria.

Article Info: *Received* : November 12, 2015. *Revised* : March 12, 2016.
Published online : June 10, 2016.

Keywords: infant mortality; ARMA model; Stationarity; Theil'U; bias; variance and covariance proportion

1 Introduction

Globally, over 10 million infants die each year before their birthday. Infant mortality takes away society's potential physical, social and human capital. Infancy is seen as a period of accelerated growth and required a lot of intake of calories and proteins. For almost four decades or there about, government and the authority concern has made tremendous progress in the reduction of infant mortality rates in the third world countries. In South Eastern Asian it declined from 146 to 42 per 1000 live births and in Latin America from 107 to 45 per 1000 live births. In most developing countries mortality rate among children under the age of five years reduced from 243 to less than 100 and once more. In Nigeria, Infant Mortality is a major public health concerns as debilitating picture of poverty, diseases and malnutrition still constitutes an unholy decimal in the country landscape. Presently, demographic data on infant mortality rate are still hugely inadequate as many deaths occur at home and are not recorded in official statistics. However, it is estimated that over 157 children per 1,000 live births or approximately one (1) child out of 6 dies before reaching age five. The devastating and long standing health care crunch is influenced by the combination of interrelated factors which includes high numbers of births per mother with short spacing between births, poor weaning foods, use of infants formulas (cow milk), inadequate healthcare delivery system, unhygienic practices and sanitations, poor feeding practices and low educational attainment.

For emphasizing please use italics and do not use underline or bold. Please do not change the font sizes or line spacing to squeeze more text into a limited number of pages.

2 Mathematical preliminaries

2.1 ARMA Model

Sometimes, models like Autoregressive or Moving average alone do not give parsimonious results when fitting the data. Therefore, autoregressive moving average models are combination of autoregressive and moving average models with order p and q respectively.

Where $AR(p)$ is of the type

$$X_t = c + \sum_{i=1}^{p} \theta_i X_{t-1} + \varepsilon_t$$

and

$MA(q)$ is

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i \varepsilon_{t-1}$$

Where in both cases ε_t is white noise process with zero mean and constant variance under the implication of iid normality assumption. Combining the two models, we have $ARMA(p,q)$ where p is the order of the AR and q is the MA part. For ARMA to function very well the roots of $AR(p)$ should be stationary and the $MA(q)$ roots should be invertible. The $ARMA(p,q)$ is given as:

$$X_t = \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-1} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-1}$$

The $ARMA(p,q)$ model above can be expressed as

$$(1-\varphi_1\mathrm{B}-\varphi_2\mathrm{B}^2-\ldots-\varphi_p\mathrm{B}^p)y_t = \theta_0 + (1+\theta_1\mathrm{B}+\theta_2\mathrm{B}^2+\ldots+\theta_q\mathrm{B}^q)\varepsilon_t$$

where B is the backward shift operation, that is $\mathrm{B}^k Y_t = Y_{t-k}$ for k +ve integer

$\phi(\mathrm{B}) = 1-\varphi_1\mathrm{B}-\varphi_2\mathrm{B}^2-\ldots-\varphi_p\mathrm{B}^p$

The process is stationary if the roots of the equations $\phi(\mathrm{B})=0$ lie within the unit circle.

The process is invertible only when the roots of θ(B) lie outside the unit circle. Furthermore, a process is said to be causal when the root of Φ(B) lie outside the unit circle. To have $ARMA(p,q)$ model, both ACF and PACF should show the pattern of decaying to zero. The eventually the ACF consists of mixed damped exponentials and sine terms. Similarly, the partial autocorrelation of an $ARMA(p,q)$ process is determined at greater lags by the MA (q) part of the process. Eventually the partial autocorrelation function will also consist of a mixture of damped exponentials and sine waves.

2.2. Character exhibited by *ACF* and *PACF* for pure seasonal $ARMA(P,Q)s$

	$AR(P)s$	$MA(Q)s$	$ARMA(P,Q)s$
ACF	Tails off at lag ks $k=1,2,3,\cdots$	Cuts off after lag Qs	Tails off at lag ks
$PACF$	Cuts off after lag Ps	Tails off at lag ks $k=1,2,3,\cdots$	Tails off at lag ks

3 Forecast performance

Commonly used measures for comparison of the forecasting performance are Mean absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Theil-U inequality. The ARMA model to be selected in the course of this study will be evaluated using all these measures for evaluating the forecasting performance. The mentioned statistics are defined as:

$$\textit{Mean Absolute Error} = \frac{\sum_{t=T+1}^{T+k} \left|\hat{Y}_t - y_t\right|}{h}$$

$$\textit{Root Mean Squared Error} = \sqrt{\frac{\sum_{t=T+1}^{T+k} \left(\hat{Y}_t - y_t\right)^2}{h}}$$

These two forecast error statistics depend on the scale of the dependent variable. These should be used as relative measures to compare forecasts for the same series across different models. The smaller the error, the better is the forecasting ability of that model according to a given criterion. The remaining two statistics are scale invariant. The Theil inequality coefficient always lies between zero and one, where zero indicates a perfect fit.

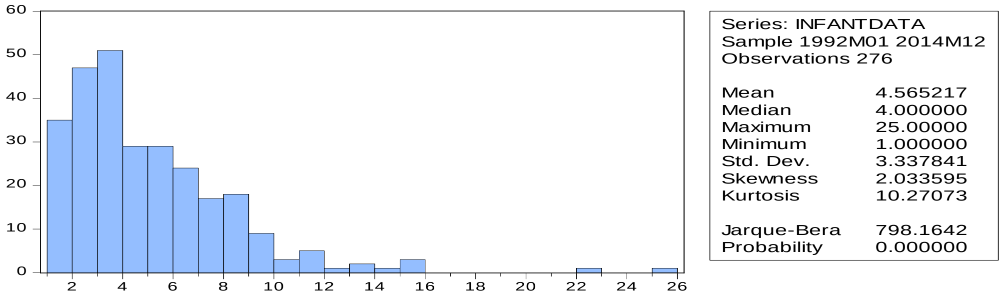
$$\textit{Mean Absolute Percentage Error} = 100 \times \frac{\sum_{t=T+1}^{T+k} \frac{\left|\hat{Y}_t - y_t\right|}{y_t}}{h}$$

$$\text{Theil U Inequalities Coefficient} = \frac{\sqrt{\frac{\sum_{t=T+1}^{T+k} \left(\hat{Y}_t - y_t\right)^2}{h}}}{\sqrt{\frac{\sum_{t=T+1}^{T+k} \hat{y}_t^2}{h}} + \sqrt{\frac{\sum_{t=T+1}^{T+k} y_t^2}{h}}}$$

3.1 Empirical illustration

3.1.1 Descriptive statistics of the infant mortality data

For the assessment of the distributional properties of infant mortality data used in this study, various descriptive statistics are reported in figure 1 below:-



Interpretation:

Figure 1

Figure 1 above revealed that the standard deviation is high which indicates high degree of fluctuations in infant mortality data used; there is evidence of skewness with long right tail indicating that the data used is non-symmetric. From the histogram, the mortality data are leptokurtic or fat tailed because of its large kurtosis value. Jarque-Bera test with p-value less than zero shows that the data is non-normal so that the hypothesis of normality is rejected. From here the study proceeded to determination of stationarity of data used for the study before embarking on the analysis of the data.

3.1.2 Stationarity test

For any study to be valid in time series there is the need to verify the stationarity of the series involve before the analysis. Two important methods used in this study are Graphical analysis (Figures 2 through 4) and Unit root test (Tables 1 through 3). From these two methods, the series was stationary at second difference. Although the probability value of the ADF is lesser than 5% i.e. prob. = 0.0000 < 0.05, but the test can only account for 27.1% fitness and cannot be accepted for proceeding in the course of this study. At first difference, value of the ADF is lesser than 5% i.e. prob. = 0.0000 < 0.05 and gives an account of 90.5% which is assumed to be unfavorable because of the level of significance and at the second difference, the value of the ADF is lesser than 5% i.e. prob. = 0.0000 <

0.05 and gives a coefficient of determination of 98.7% which assumed to be a good fit, hence suggesting that the infant mortality data is stationary.

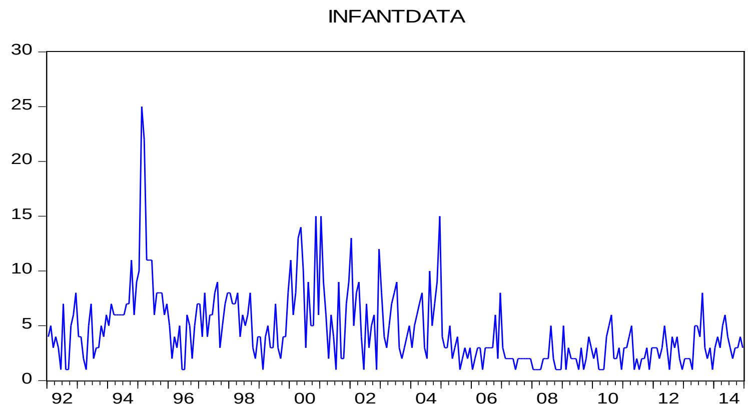


Figure2

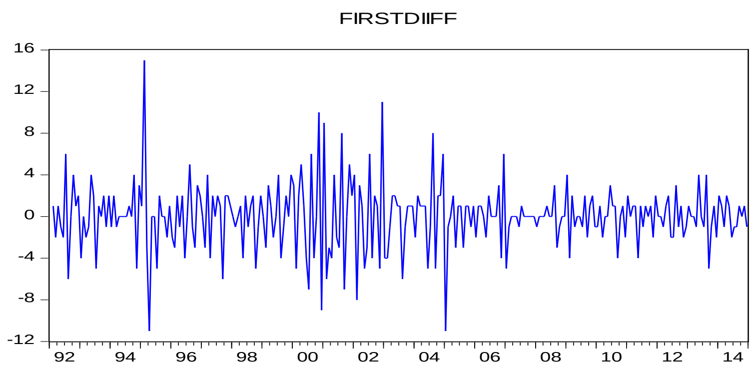


Figure 3

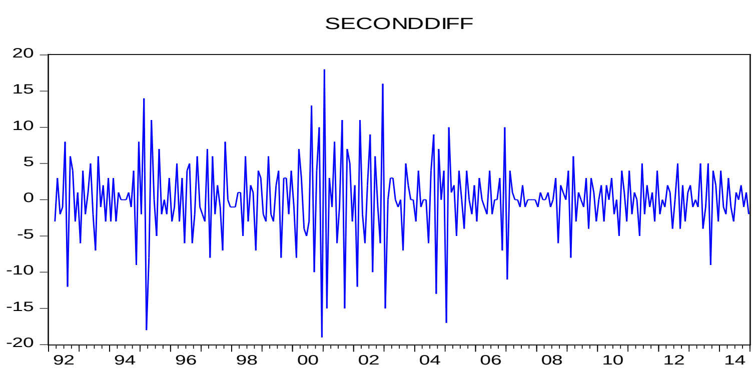


Figure 4

Table 1: Unit root test of the original series

Null Hypothesis: INFANTDATA has a unit root

Exogenous: Constant

Lag Length: 2 (Automatic - based on AIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.978454	0.0000
Test critical values:	1% level	-3.454174	
	5% level	-2.871922	
	10% level	-2.572375	

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

Table 2: Unit root test of the first difference

Null Hypothesis: D(FIRSTDIIFF) has a unit root

Exogenous: Constant

Lag Length: 12 (Automatic - based on AIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.061143	0.0000
Test critical values:	1% level	-3.455289	
	5% level	-2.872413	
	10% level	-2.572638	

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

Table 3: Unit root test of the second difference

Null Hypothesis: D(SECONDDIFF,2) has a unit root

Exogenous: Constant

Lag Length: 15 (Automatic - based on AIC, maxlag=15)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-11.06776	0.0000
Test critical values:	1% level	-3.455786	
	5% level	-2.872630	
	10% level	-2.572754	

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

3.1.3 Model fitting

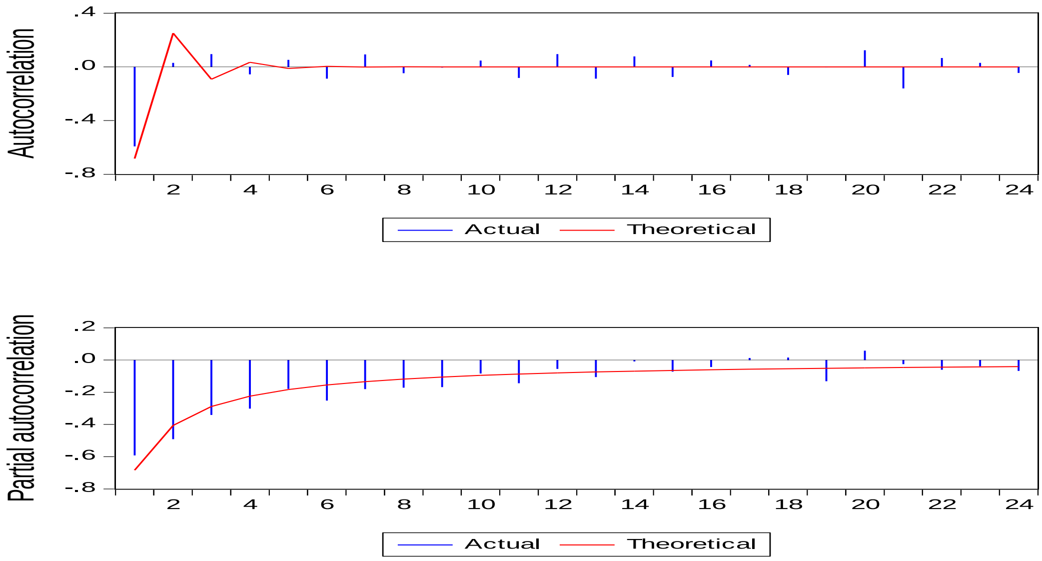
As the descriptive statistics given in Figure 1 reflects that the distribution of the infant mortality series is a non-normal distribution. It means volatility is present. Suitable econometric modeling techniques are required for the infant mortality series of this study. To start with, we model the conditional mean process by autoregressive process AR(1) and moving average MA(1) and to do this, we draw a gird search table which is a table designed to search for the order that give parsimonious value. We looked at various autoregressive value up to order six, cross tabulated with Moving average up to the same order as shown in the table 4 below. Both processes demonstrate correlation residuals. Among the different models applied to the data, ARMA (1,1) appears to be relatively better fit on the basis of Akaike Information Criterion.

Table 4: Grid search table

MA(q) \ AR(p)	1	2	3	4	5	6
1	5.009524	5.143203	5.157664	5.160930	5.151525	5.138613
2	5.418786	6.152532	6.138817	6.156295	6.151977	6.130488
3	5.684018	6.129364	5.684018	6.149620	6.143103	6.122627
4	5.719172	6.150022	6.147545	6.143165	6.151165	6.128588
5	5.719397	6.150702	6.144840	6.156683	6.133356	6.130461
6	5.717215	6.145883	6.138738	6.150499	6.149156	6.124317

The graph of the ARMA(1,1) residuals is given as in the figure 11 below:

From Figure 5 above, the correlogram of both autocorrelation and partial autocorrelation give the impression that the estimated residuals are purely random. Hence, there is no need to search for another ARMA model.



Interpretation:

Figure 5

4.5 Forecast analysis

Forecast performance of the fitted ARMA (1,1) model of infant mortality series is investigated through the mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and Theil inequality coefficient. The results are shown in the Table 5 below:

Table 5

Root Mean Square Error	4.989259
Mean Absolute Error	3.89432
Mean Absolute Percentage Error	86.27991
Theil Inequality Coefficient	0.722313
Bias Proportion	0.0076
Variance Proportion	0.00342
Covariance Proportion	0.99608

The value of Theil-U inequality obtained is 0.722313 showing that the model fit is good. Looking at the bias proportion (0.0076) and variance proportion (0.00342). These two indices are very close to zero, the implication of this is that the series under study has a little or no bias error. The variance proportion is a bit close to zero implying a better fit. The covariance proportion tends to one implying that this model will be very good for use for forecasting purpose.

5 Summaries of the major findings and conclusion

The study looked at the infant mortality with particular reference to Osun State (General Hospital, Asubiaro). The data used was a secondary data (monthly data from year 1992-2014). From the data, it was discovered that at level and first difference, the series was not stationary but at the second difference, it was stationary. We proceeded to the data analysis stage where ARMA(1,1) model was chosen as it produced the least Akaike Information Criterion (AIC) value. Thereafter, we subjected the result obtained to forecasting evaluation indices where it was discovered that Theil-U produced a better fit, the bias and the variance proportion are very close to zero and covariance proportion is very close to one. All pointing to the fact that the model ARMA(1,1) will have a very good forecasting ability. On the basis of this study we want to recommend the following that:

- A large scale research work should be conducted with the aim of verifying the reliability of result so obtained;
- The government should endure to improve the condition of our various health centers by embarking on the programme that will touch the life of the grass root dwellers by organizing house-to-house vaccines, public enlightenment through television and radio and through the town criers to reduce the

menace of the monster called infant mortality. The reason why the campaign should be carried to the rural dwellers is because they are mostly affected;

- The issue of sanitation should be taken serious as the infants are more prone to diseases which if not resulted into death may adversely affect the wellbeing of the infants involved;

- And lastly, a more advance model like GARCH, Bilinear, SETAR Models and Hybrid Models like BL-GARCH, STAR-GARCH, ANN-STAR, ARMA-STAR and other hybrids may as well be used to investigate the reliability of the result so obtained.

References

[1] Ananth, C.V. and Basso O., Impact of pregnancy-induced hypertension on stillbirth and neonatal mortality, *Epidemiology*, **21**(1), (2010), 118-123.

[2] Chen, A., Feresu, S.A., Fernandez, C. and Rogan, W.J., Maternal obesity and the risk of infant death in the United States, *Epidemiology*, **20**(1), (2009), 74-81.

[3] Durueke. O.O., *The incidence, management and bionomic of malaria in children under 5 years of age*, 2005.

[4] G. Box and G. Jenkins, *Time series analysis: Forecasting and control, San Francisco: Holden-Day*, 1970.

[5] Kitsantas, P. and Gaffney K.F., Racial/ethnic disparities in infant mortality, *J. Perinat Med*, **38**(1), (2010), 87-94.

[6] Ananth, C.V. and Basso O., Impact of pregnancy-induced hypertension on stillbirth and neonatal mortality, Epidemiology, **21**(1), (2010), 118-123.

[7] Reddy, U.M., Laughon, S.K., Sun, L., Troendle, J., Willinger, M., Zhang, J. Pre-pregnancy risk factors for antepartum stillbirth in the United States, **116**(5), (2010), 1119-1126.

[8] Venkatesh, K.K., et al. Morbidity and Mortality among Infants Born to HIV-Infected Women in South Africa: Implications for Child Health in Resource, Limited Settings, 3, (2010).

[9] Wiener N., *The extrapolation, interpolation and smoothing of stationary time series with engineering applications,* Wiley: New York, 1949.

[10] Yeshiwodim and Troendle, J., Spatial analysis of malaria incidence using the method of poisson regression analysis, (2009).