

# Statistical Modeling of Appendicitis Cases at the Korle-Bu Teaching Hospital, Accra, Ghana

Gideon Mensah Engmann<sup>1,\*</sup>, Grace Adzo Motey<sup>2</sup>, Mohammed Adjei Adjieteh<sup>1</sup>

<sup>1</sup>Department of Statistics, Faculty of Mathematical Sciences, University for Development Studies, Navrongo, Ghana

<sup>2</sup>Department of Applied Biology, Faculty of Applied Sciences, University for Development Studies, Navrongo, Ghana

**Abstract** Appendicitis is inflammation of the appendix. In Ghana, most people affected by this disease subsequently die as a result of late treatment. The study seeks to model the appendicitis cases and also forecast future trends of the disease using Box-Jenkins approach. Data on monthly reported cases of appendicitis from 2008 to 2014 was collected from the Biostatistics department of the Korle-Bu Teaching Hospital. The results shows that  $ARIMA(1,0,0) \times (0,1,1)_{12}$  adequately model the number of appendicitis cases. The forecasted values also revealed that the number of appendicitis cases will increase during the months of April to July which happens to be the major rainy season of the country.

**Keywords** Appendicitis, Box-Jenkins, Korle-Bu Teaching Hospital, Accra

## 1. Introduction

Appendicitis (also called epityphlitis) is inflammation of the appendix. Many cases of appendicitis require removal of the inflamed appendix by laparotomy or laparoscopy due to the high mortality associated with rupture of the appendix [6]. Appendectomy (sometimes also called appendisectomy) is the surgical removal of the vermiform appendix. Appendicitis was first described by Reginald Fitz in 1886, and is today recognized as one of the most common and significant causes of severe acute abdominal pain worldwide.

The causative agents of appendicitis include bezoars, foreign bodies, trauma, intestinal worms, lymphadenitis and calcified fecal deposits. The major symptoms of Appendicitis are dull pain near the navel or the upper abdomen that becomes sharp as it moves to the lower right abdomen, loss of appetite, abdominal swelling, nausea and/or vomiting soon after abdominal pain begins, fever of 99-102 degrees Fahrenheit and inability to pass gas.

Treatment of appendicitis is mainly by surgery however intravenous antibiotics can be used. The average hospital charge for an elderly patient's negative appendectomy in the United States is \$ 47,120.3 [1]. Acute Appendicitis is a common clinical condition worldwide. In Ghana, most people affected by this disease subsequently die as a result of late treatment. Nashiru and Sarpong (2011) used Arima to model and forecast hypertensive cases in Navrongo, Ghana. Sarpong (2013) also used Arima to model and forecast

maternal mortality in Kumasi, Ghana.

This study seeks to model the appendicitis cases and also forecast future trends of the disease. This will help reveal the current pattern/trend of appendicitis cases and go a long way to help public health workers and government to organize educative programmes to keep the disease under control. We therefore employed the Box-Jenkins approach to model and forecast appendicitis cases at Korle-Bu Teaching Hospital, Accra, Ghana.

## 2. Material and Methods

The study was conducted at the Korle-Bu Teaching Hospital in the Greater Accra Region of Ghana which is one of the major referral hospitals in the country. Data on monthly reported cases of appendicitis from 2008 to 2014 was collected from the Biostatistics Department of the hospital. The data was modelled using Seasonal Autoregressive Integrated Moving Average (SARIMA) stochastic model. An  $ARIMA(p,d,q)$  model is a combination of Autoregressive (AR) which shows that there is a relationship between present and past values, a random value and a Moving Average (MA) model which shows that the present value has something to do with the past residuals. If the data has a seasonal component, then  $ARIMA(p,d,q)$  is extended to include the seasonal component. The seasonal autoregressive integrated moving average model of Box and Jenkins (1970) is denoted as an  $ARIMA(p,d,q) \times (P,D,Q)_s$ .

The orders  $p$  and  $q$  are for the non-seasonal AR and MA component respectively. The orders of the Seasonal AR and MA component are  $P$  and  $Q$  respectively. Also, the order of differencing for the seasonal and non-seasonal

\* Corresponding author:

gengmann@uds.edu.gh (Gideon Mensah Engmann)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

is  $d$  and  $D$  respectively. The seasonal order is  $s$ . The estimation of the model consists of three steps, namely: identification, estimation of parameters and diagnostic checking.

### 2.1. Model Identification

We first determine whether the series is stationary or not formally by conducting Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. If the series is not stationary; it would then be converted to a stationary series by differencing. That is, the original series is replaced by a series of differences. An ARIMA model is then specified for the differenced series.

### 2.2. Model Estimation

Once a model is identified, the next stage for Box-Jenkins approach is to estimate the parameters. The model with the minimum values of Akaike Information Criterion (AIC), Modified Akaike Information Criterion ( $AIC_c$ ) and Normalised Bayesian Information Criterion (BIC) was considered as the best model.

### 2.3. Diagnostic Checking

In this step, model must be checked for adequacy by considering the properties of the residuals whether the residuals from an ARIMA model must have the normal distribution and should be random. An overall check of model adequacy is provided by the Ljung-Box Q statistic.

The test statistic Q is

$$Q_m = n(n+2) \sum_{k=1}^m \frac{r_k^2}{n-k} \quad (1)$$

Where  $r_k^2$  is the residual autocorrelation at lag  $k$ ,

$n$  = the number of residuals and  $m$  the number of times lags includes in the test. If the P-value associated with the Q statistic is small ( $P\text{-value} < \alpha$ ), the model is considered inadequate.

ARCH LM-test for conformity of the presence of, or otherwise ARCH effect is given by the hypothesis;

$H_0$  = There is no conditional heteroscedasticity in the residuals of the model

$H_1$  = There is conditional heteroscedasticity in the residuals of the model

The test statistic is given by;

$$LM = TR^2 \quad (2)$$

Where  $T$  = number of observation

$R^2$  = coefficient of determination computed from the auxiliary regression.

### 2.4. Unit Root Test

Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test are performed to test whether the data is stationary or not. Unit

Root Test was derived in 1979 by Dickey and Fuller to test the presence of a unit root versus a stationary process. The null hypothesis of interest is  $H_0: \phi = 0$  (the series has unit root or is non-stationary) against the alternative  $H_1: \phi < 0$  (the series has no unit root or is stationary).

All statistical tests were controlled at 5% significance level ( $\alpha$ ).

## 3. Results and Discussion

Figure 1 shows the time series plot of the Appendicitis data. The series show a periodic pattern over the period of time. The plot seems to be showing that the series is stationary. The values had regular swings and hence regular variability. From Figure 2, the ACF shows the series to be non-seasonally stationary and seasonally non-stationary because the seasonal lags dies off slowly therefore the need for seasonal differencing. From Table 1, Kwiatkowski - Phillips-Schmidt-Shin (KPSS) test showed that the series is non-seasonally stationary since the p-value of the KPSS test is 0.1 which is greater than the significance level (5%), so we fail to reject the null hypothesis that the data is level or trend stationary. Again from Table 1, the Augmented Dickey-Fuller (ADF) test which is used for verifying whether or not there is unit root suggests that there is unit root since the p-value of the ADF test is smaller than the significance level (5%), so we reject the null hypothesis of unit root and claim that the series is stationary. Figure 2 indicates clearly the evidence of seasonality in the data as the Partial Autocorrelation Function (PACF) has a significant spike at season lag 12.

**Table 1.** Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test and Augmented Dickey-Fuller (ADF) test of appendicitis cases

TEST TYPE	TEST STATISTICS	P-VALUE
KPSS	0.0208	0.1
ADF	-5.2138	0.01

### 3.1. Model Estimation and Evaluation

An  $ARIMA(p, 0, q) \times (P, 1, Q)_{12}$  models will be fitted to the data since the data was seasonal differenced once (Figure 3) and the best model will be selected based on the minimum values of Akaike Information Criterion (AIC), Modified Akaike Information Criterion ( $AIC_c$ ) and Normalized Bayesian Information Criterion (BIC). The identified models are presented in Table 4 below with their corresponding values of AIC,  $AIC_c$  and BIC.

From table 2, among these possible models,  $ARIMA(1, 0, 0) \times (0, 1, 1)_{12}$  was chosen as the appropriate fitted models for the data since its AIC,  $AIC_c$ , BIC are smaller as compared to the other models. Parameters of the model were estimated. From table 3, at 95% confidence level, we conclude that the parameters are significant.

The residuals also follow a white noise process that is residuals have zero mean, constant variance and also uncorrelated as show in figure 4. Also, from the diagnostic

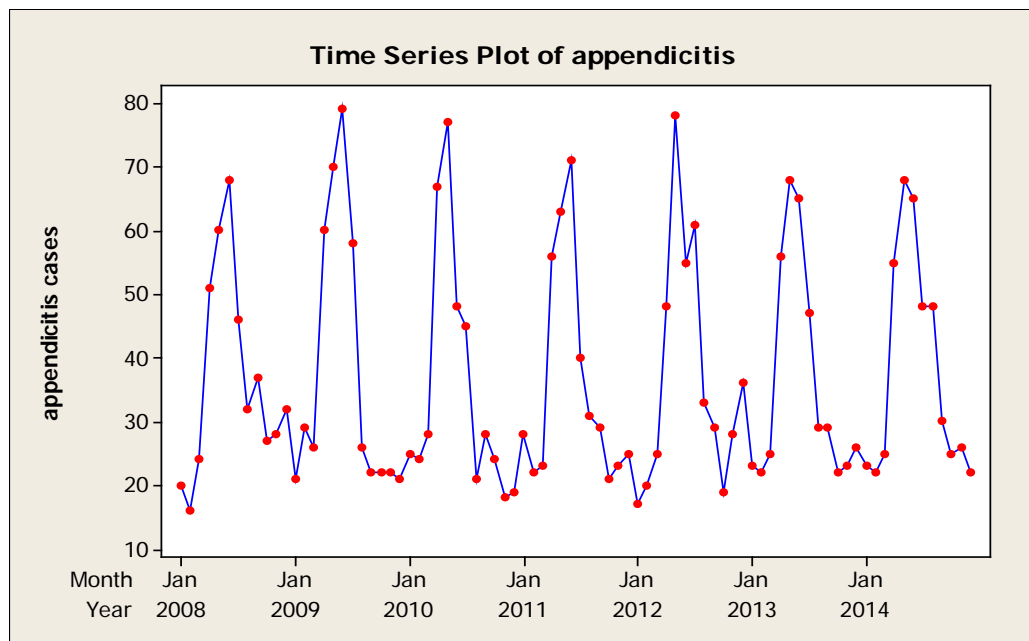
plots in figure 4, the Ljung-Box test statistic had p-values greater than the alpha ( $\alpha$ ) value of 0.05. We therefore fail to reject the null hypothesis and conclude that there is no serial correlation in the residuals of the model. Hence the model is adequate. The ARCH LM test also shows no heteroscedasticity in the model residuals since the p-values are all greater than the alpha ( $\alpha$ ) value of 0.05 as shown in table 4.

Finally we made a two year forecast with our model as shown in Table 5. The forecast shows slight decline in the appendicitis cases and then followed by an increment (cyclic order). These fluctuations occur within the major rainy season of the study location.

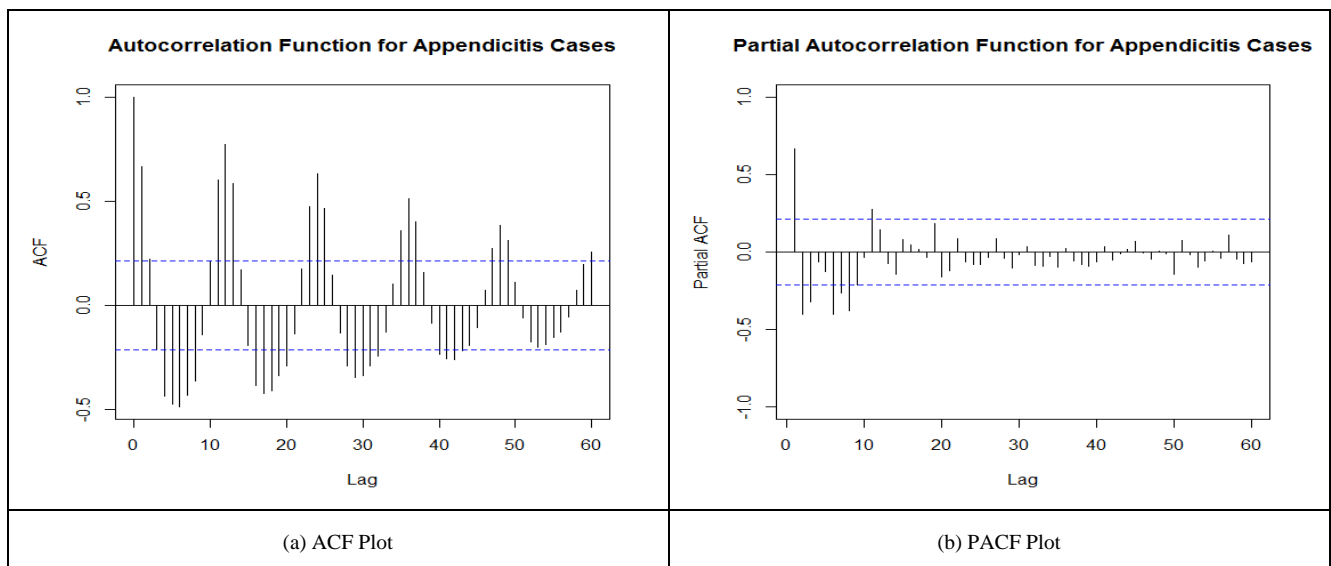
**Table 2.** Tentative SARIMA models

MODELS	AIC	AIC <sub>C</sub>	BIC
ARIMA(1,0,0)×(0,1,1) <sub>12</sub>	489.7*	490.05*	496.53*
ARIMA(1,0,0)×(2,1,1) <sub>12</sub>	491.00	491.91	502.39
ARIMA(1,0,1)×(2,1,0) <sub>12</sub>	493.95	494.86	503.33
ARIMA(1,0,0)×(2,1,0) <sub>12</sub>	492.76	493.36	501.87
ARIMA(0,0,1)×(1,1,0) <sub>12</sub>	504.5	504.86	511.33
ARIMA(0,0,1)×(2,1,0) <sub>12</sub>	492.92	493.52	502.03

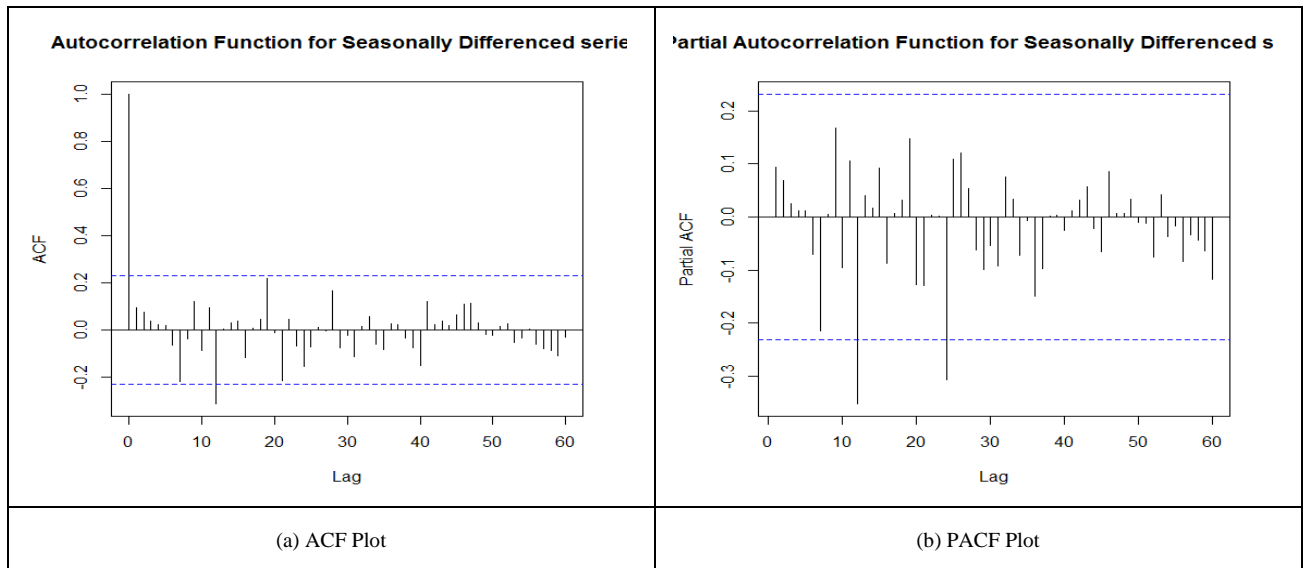
\*Best based on the selected criterion



**Figure 1.** Time Series Plot of the Appendicitis data



**Figure 2.** ACF Plot and PACF Plot of the Appendicitis Data



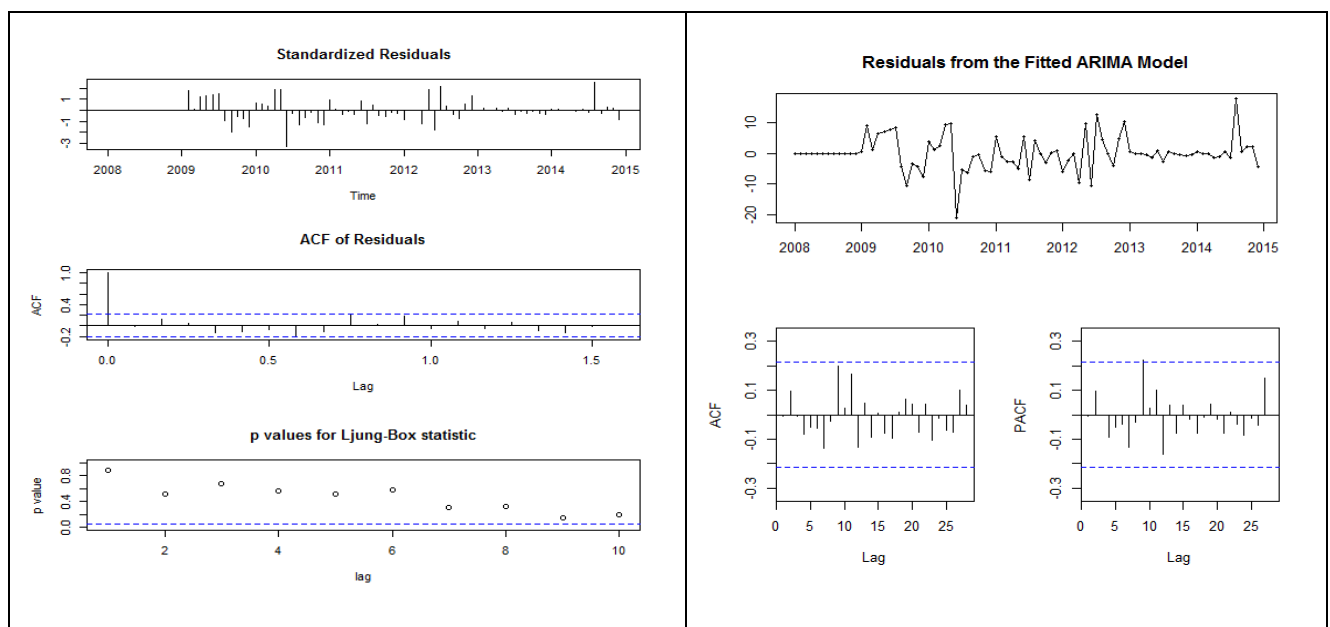
**Figure 3.** Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graph of seasonal differenced series

**Table 3.** Parameter estimates of  $ARIMA(1, 0, 0) \times (0, 1, 1)_{12}$  of the appendicitis cases

TYPE	COEFFICIENTS	STANDARD ERROR COEFFICIENTS	T-VALUE	P-VALUE
AR (1)	0.0400	0.1194	0.3300	0.0461
SMA (12)	1.0361	0.1063	9.7500	0.0000
$\sigma^2 = 0.02513$				

**Table 4.** ARCH-LM test of residuals  $ARIMA(1,0,0) \times (0,1,1)_{12}$  of the Appendicitis cases

LAG	CHI-SQUARE	DF	P-VALUE
12	4.5402	12	0.9716
24	11.8492	24	0.9816
36	30.4284	36	0.7305
48	36.0000	48	0.8989



**Figure 4.** Diagnostic plot of  $ARIMA(1,0,0) \times (0,1,1)_{12}$  of the appendicitis cases

**Table 5.** Forecasting value of Arima (1, 0, 0)  $\times$  (0, 1, 1)<sub>12</sub> model for monthly reported appendicitis cases for two years

Year	Month	Forecast	Confidence Limits	
			Lower	Upper
2015	January	22.3248	9.9260	34.7236
	February	22.1404	9.7382	34.5426
	March	25.1428	12.7406	37.5450
	April	56.1429	43.7407	68.5450
	May	69.1429	56.7407	81.5450
	June	64.4286	52.0264	76.8307
	July	49.2857	36.8835	61.6879
	August	31.4286	19.0264	43.8307
	September	29.1429	16.7407	41.5450
	October	22.8571	10.4550	35.2593
	November	23.9998	11.5976	36.4020
	December	25.8490	13.4469	38.2511
2016	January	22.4156	10.0135	34.8177
	February	22.1426	9.7404	34.5447
	March	25.1429	12.7407	37.5450
	April	56.1429	43.7407	68.5450
	May	69.1429	56.7407	81.5450
	June	64.4286	52.0264	76.8307
	July	49.2857	36.8835	61.6879
	August	31.4286	19.0264	43.8307
	September	29.1429	16.7407	41.5450
	October	22.8571	10.4550	35.2593
	November	23.9998	11.5976	36.4020
	December	25.849	13.4469	38.2511

## 4. Conclusions

In this study, we model the appendicitis cases at Korle-Bu Teaching Hospital, Accra, Ghana using Box-Jenkins approach. The adequate model identified was ARIMA (1, 0, 0) $\times$ (0,1,1)<sub>12</sub>. The forecasted values also revealed that the number of appendicitis cases will increase during the months of April to July which happens to be the major rainy season

of the country. Future research could focus on multivariate time series analysis of appendicitis cases in Ghana where data from other teaching hospitals in the country could be integrated.

## ACKNOWLEDGEMENTS

Acknowledgements go to all and sundry who in diverse ways contributed to making this research a success especially Mr. Prince Agyire and Miss Irene Ameade.

## REFERENCES

- [1] Ackerman, S., Anastassopoulos, K. and Knight, T. (2005). *Economic Studies for NeutroSpec Imaging Agent*. Covance Market Access Services, Inc.
- [2] Burnham, K. P. and Anderson, D.R. (1998). *Model Selection and Inference*. Springer-Verlag, New York.
- [3] Box, G.E.P. and Jenkins, G.M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- [4] Dickey, D.A. and Fuller, W.A. (1979). *Distribution of estimators for Autoregressive Time Series with Unit Root*. Journal of the American Statistical Association, 74: 427-431.
- [5] Fitz, R.H. (1886). *Perforating inflammation of the vermiform appendix with special reference to its early diagnosis and treatment*. *Am J Med Sci* (92): 321-46.
- [6] Graffeo, C. S. and Counselman, F. L. (1996). "Appendicitis". *Emergency medicine clinics of North America*. 14 (4): 653-71.
- [7] Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). *Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root*. Journal of Econometrics. 54, 159-178.
- [8] Nasiru, S. and Sarpong, S. (2012). *Statistical modeling of hypertension cases in Navrongo, Ghana, West Africa*. American Journal of Social Science and Management Science 2(4): 377-383.
- [9] Sarpong, S. (2013). *Modeling and Forecasting Maternal Mortality; an Application of ARIMA Models*. International Journal of Applied Science and Technology. 3(1), pp: 19-28.