Modelling and Prediction of Outpatients Department on Hospital attendance at the Cape Coast Teaching Hospital using the Box-Jerkins ARIMA model

ABSTRACT:

Aims: Outpatient department is one of the first points of contact for patients accessing health care and provide patients with their primary healthcare as they seek services at the facility. With the introduction of community-based health planning and services, there seems that the outpatient departments have witnessed corresponding progressive and significant increase in attendance at the various health facilities in Ghana of which the research seeks to investigate.

Materials and Methods: The data collected were outpatient hospital attendance of patients on a monthly basis from 2012 to 2019 obtained from the Cape Coast Teaching Hospital. Box Jenkins's methodology of time series analysis was used to analyse the data. The modified Box Pierce (Ljumg-Box) Chi-square statistic criteria of the largest $p-value$ and minimum Chi-square statistic value was in selecting the best fitted model for outpatient department attendance.

Results: The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots suggested an autoregressive (AR) process with order 2 and moving average (MA) process with order 1 which was used in selecting the appropriate model. Candidate models were obtained using the lowest Chi-square value and highest p −value to select adequate models and the best model. The best non-seasonal model for the data was ARIMA (2, 2, 1) for the outpatient department attendance. Model diagnostics test was performed using Ljung-Box test.

Conclusion: The findings of the forecast showed that OPD visits will increase in the next five years. Specifically, continued use of the outpatient department in accessing health care at all levels will experience an increase in hospital visits across the months from June 2020 to December 2025. Recommendations from this research included among others that, the health authorities should continue to expand the outpatient department services to increase access to healthcare by all as it services goes to the core people in the community.

Keywords: [Time series analysis, ARIMA model, Outpatient hospital attendance, forecasting trends]

1. INTRODUCTION

The application of time series has been used in many fields including insurance, healthcare, economics, finance, weather forecasting and etc. Using time series in studying the general behaviour of the outpatient department on hospital attendance by patients in the healthcare system is important in identifying the fluctuations in health attendance indicators on the distribution of resources and specific disease incidence with time among others . [1] Their report on the outpatient department (OPD) usage explained on the essence, how it has become an essential part of all health facilities in Ghana due to the fact that it is one of the first step of the treatment system and point of contact between a hospital and the community. However, it is often considered as the window to health facility services. The patient's history and vital signs of blood pressure, heart rate, respiratory rate and temperature among others, are obtained and documented at the outpatient unit. Similar to other units at the hospital, the OPD offers a 24-hour service and is open throughout the week [2]. The functions of the outpatient department make it an important facet in the admission protocols of all health facilities either contributing to it increase or decrease in attendance.

[2] Applied ARIMA model of time series on the monthly inflation rates from July 1991 to December 2009 in Ghana. The study was done using monthly inflation rates from July 1991 to December 2009. The research indicated that Ghana faces a macroeconomic problem of inflation for a long period of time. The selected model was ARIMA $(1,1,1)(0,0,1)_{12}$ which represents the data behavior of inflation rate in Ghana. Seven months forecast on inflation rates of Ghana outside the sample period (i.e. from January 2010 to July 2010) was done. The forecasted results indicated a decreasing pattern and a turning point of Ghana inflation in the month of July.

[3] Used time series in modelling the autoregressive integrated moving average part of the time series and forecasted hospital attendance. Their research work used a secondary data and interview schedule as the main sources of data. The secondary data focused on monthly outpatient unit attendance from January 2008, to December 2011, using the Obuasi hospital as the case study. ARIMA (2, 1, 0) was the best selected model based on the AIC value of 420.33. Their findings forecasted a steadied trend of ODP attendance for the forecast period and turning point at the month of January 2012.

[4] Using time series analysis, which they applied ARIMA for the outpatient visits forecasting. The data used comprised of one year daily visits of outpatient visits data of two specific departments (internal medicine departments) in the urban area of a hospital in Chengdu. A formulated seasonal ARIMA model focused on the daily time series and also, a single exponential smoothing model of the week time series, thereby establishing a new forecasting model which factors the cyclicity and the day of the week effect into consideration. The results concluded that the use of combinational models, achieves better forecasting performance than the single model.

[5] Their research work used seasonal ARIMA model in a 10 year time period (2008-2017) for hospital attendance in the Cape Coast Teaching Hospital for both insured and uninsured patients on a monthly basis across age groups and gender. The data used was a secondary source. Selected models were SARIMA (1,0,0) $(0,1,0)_{12}$ model for insured (NHIS) and SARIMA (1,1,1) $(2,0,1)_{12}$ model for uninsured (Cash and Carry system) based on their minimum AIC values of 15.66537 (insured) and 13.94181(uninsured). [5] Using Chi-square test also concluded on dependence between insured and uninsured patients in hospital attendance on gender and the years. The research results on it summary concluded that generally,all age groups in the insurance category will increase in attending hospital to seeking health care. Also, [5] patients who are uninsured will have exactly attendance from 0-28 days to 15-17 years, increasing for the next 24 months in seeking healthcare.

This research is conducted to ascertain how the outpatient department has impacted on attendance in seeking health care by patients with time using time series analysis]

2. MATERIAL AND METHODS / EXPERIMENTAL DETAILS / METHODOLOGY

[A retrospective data from the Cape Coast Teaching Hospital in the Central Region of Ghana was used for this research with patients using the outpatient department of the hospital in seeking healthcare. A list of eight years outpatient department of hospital attendance records was considered for the period 2012 to 2019. The research seeks to investigate whether the outpatient department attendance in the hospitals and the introduction of community-based health planning and services have increased health visits in seeking health care. ARIMA model of time series was used in analysing the data in order to make future predictions.

2.1 TIME SERIES ANALYSIS

Time series uses past behaviour of the variable in order to predict its future behaviour. A time series usually changes with the passage of time and there are many reasons which bring changes in the time series [5]. These changes are called components, variations movements or fluctuations. There are four types of time series components which are:

i. Trend (Secular or General)
ii. Seasonal Variation
iii. Cyclical Variation
iv. Irregular / Random Variation

Two ways to put the four components together in Time Series Models are:

i. Additive Model
ii. Multiplicative Model

[6] Box and Gwilyn Jenkins developed the ARIMA methodology of time series thus the Box-Jenkins methodology. The data were plotted against time (months) in order to identify features such as trend, seasonality, and stationarity of the dataset. Also the Augmented Dickey–Fuller unit root test [7] was used to further ascertain the stationarity of the data. Box and Jenkins recommend the differencing approach to achieve stationarity Differencing was used to transform the data in order to attain the stationarity assumption.
There are three basic components of an ARIMA model mainly, auto-regression (AR), differencing or integration (I), and moving-average (MA) [8]. Notational, all AR (p) and MA (q) models can be represented as ARIMA (1, 0, 0) that is no differencing and no MA part. The general model is ARIMA (p,d,q) where p is the order of the AR part, d is the degree of differencing and q is the order of the MA part. The ARIMA process can be written as $Y_t = V^d Y_t = (1-B)^d Y_t$. The general ARIMA process is of the form:

$$Y_t = \sum_{i=1}^{p} \alpha_i Y_t - i + \sum_{i=1}^{q} \theta_i e_t - i + \mu + et \qquad (1)$$

An example of ARIMA (p, d, q) process is the ARIMA (1, 1, 1) which has one autoregressive parameter, one level of differencing and one MA parameter and is given by

$$Y_t = \alpha_t Y_{t-1} + \theta_1 e_{t-1} + \mu + e_t$$

$$(1-B)Y_t = \alpha_1 (1-B) Y_{t-1} + \theta_1 e_{t-1} + \mu + e_t \qquad (2)$$

which can be simplified further as

$$Y_t - Y_{t-1} = \alpha_1 Y_{t-1} + \alpha_1 Y_{t-2} + \theta_1 e_{t-1} + \mu + e_t$$

$$Y_t - Y_{t-1} = \alpha_1 (Y_{t-1} - Y_{t-2}) + \theta_1 e_{t-1} + \mu + e_t \qquad (3)$$

2.2 MODEL IDENTIFICATION AND IT ORDER

After achieving the stationarity assumption, the next task was to select the appropriate model and the order of the model. The behaviour of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) were used to identify the model and the order that describes the stationary time series data. Theoretically, we expect 95% of the values of the partial autocorrelation coefficients, to fall within the limits $\pm \frac{2}{\sqrt{N}}$ and values outside the range are significantly different from zero. The implication is that the sample partial autocorrelation function PACF) of an AR (p) model 'cuts off' at lag p so that the values beyond p are not significantly different from zero. However, the order of a MA (q) model is usually clear from the sample autocorrelation function (ACF). The theoretical autocorrelation function of an MA (q) process 'cuts off' at lag q and values beyond q are not significantly different from zero. The general behaviors of the ACF and PACF for ARMA/ARIMA models are summarized in the table below according to [9] as:

Table 1. Behaviour of the ACF and PACF for ARMA Models

	AR(p)	**MA (q)**	**ARMA(p, q),p > 0, and q**
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

2.3 Diagnosis Checking

The Ljung-Box statistic, also called the modified Box-Pierce statistic, is a function of the accumulated sample autocorrelations, r_j, up to any specified time lag m. As a function of m, it is determined according to [5] as

$$Q(m) = n(n+2)\sum_{j=1}^{m}\frac{{r_j}^2}{n-j}, \qquad (4)$$

Where n is the sample size after any differencing operation, and the test statistic follows the Chi-square distribution with degrees of freedom $(df) = m - p$. A small p-value (say p-value < 0.05) indicates the possibility of non-zero autocorrelation within the first m lags [10, 11]. The distribution of Q(m) is based on the following two cases:

(i) If the r_j are sample autocorrelations for residuals from a time series model, the null hypothesis distribution of Q(m) is approximate to a χ^2 distribution with df = m – p, where p = number of coefficients in the model. (Note: m = lag to which we are accumulating, so in essence the statistic is not defined until m > p).

(ii) When no model has been used, so that the ACF is for raw data, *p* = 0 and the null distribution of *Q(m)* is approximately a χ^2 distribution with degrees of freedom (df) = *m*.

The Ljung-Box test can be defined as follows:

H_O:: The data are independently distributed
H_A: The data are not independently distributed

The choice of a plausible model depends on its p-value for the modified Box-Pierce if is well above 0.05, indicating "non-significance." In other words, the bigger the p-value, the better the model [5].

Note: Review paper may have different types of subsections.

3. RESULTS

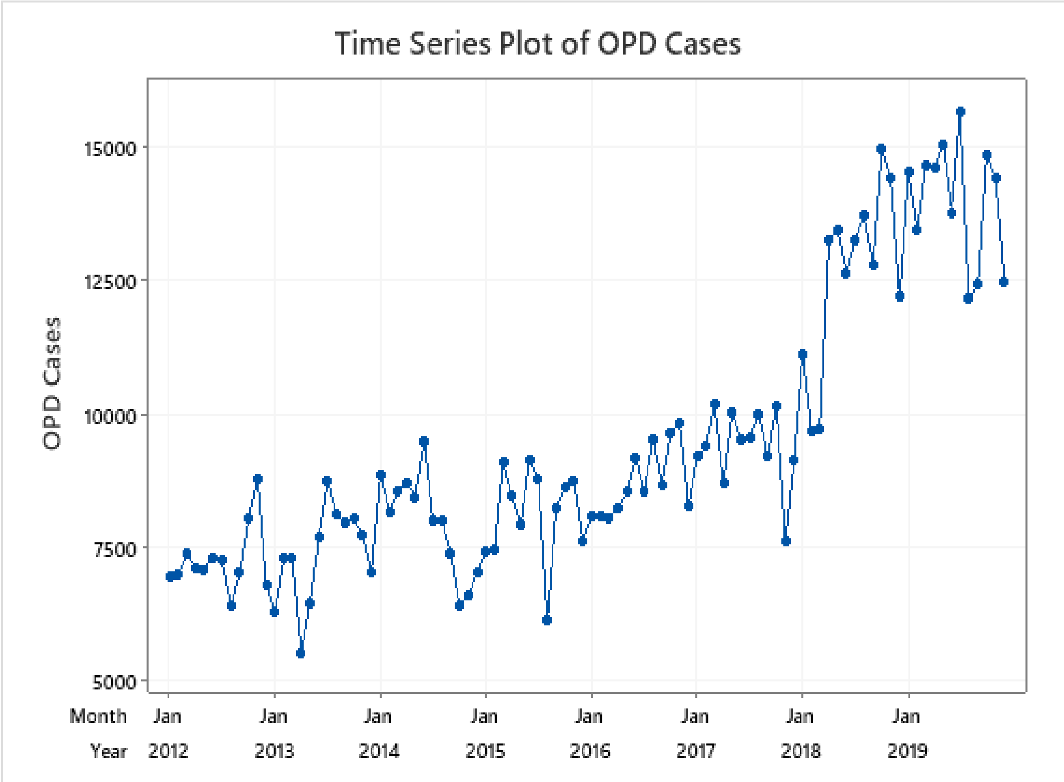

Figure 1. Time Series Plot of OPD Cases for raw data

From Figure 1. It can be observed that the time series plot has no seasonal variation in the number of hospital attendance per month. Again, it can be observed also that the series exhibit additive property as the random fluctuations are roughly constant in size over time and do not seem to depend on the level of the time series. It can be observed that the attendance, exhibit for the raw data from 2018 to 2019 recorded the highest number of hospital visits at the outpatient department.

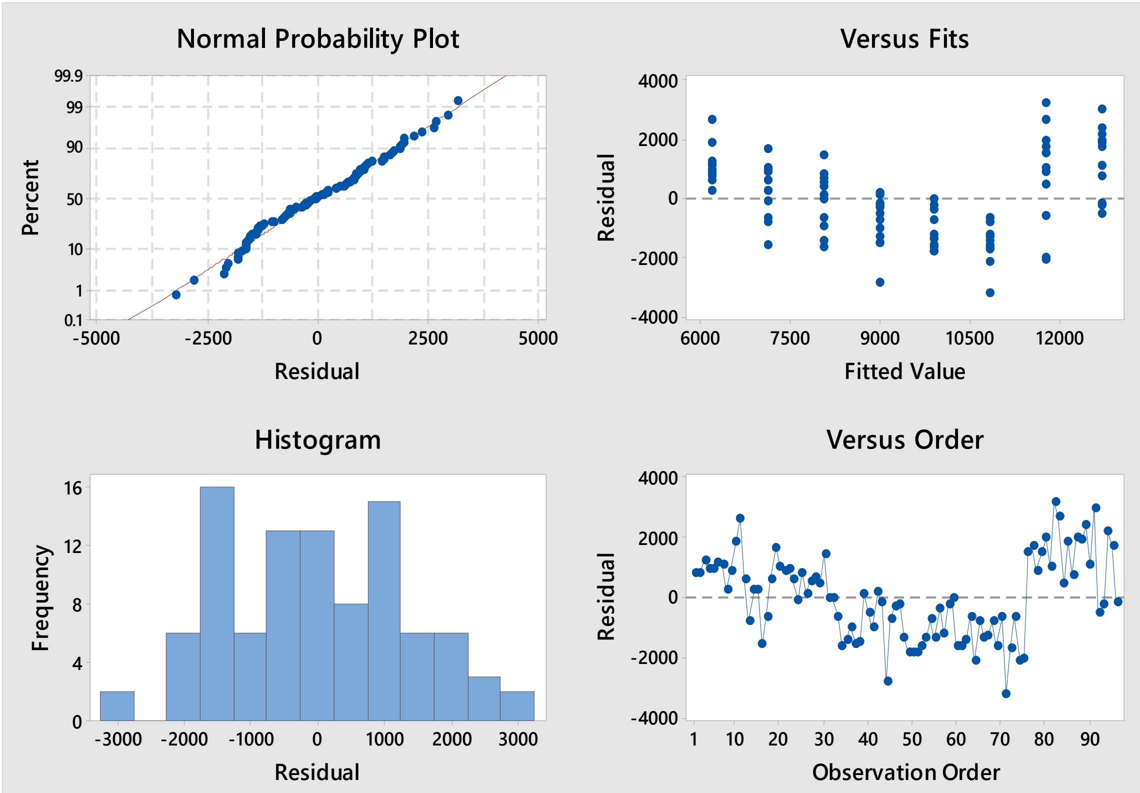

Figure 2. Residual Plot for OPD Cases

3.1 Checking For Normality, Constant Variance Assumption, Independent Assumption and Uncorrelated of the Data Set

From Figure 2, the normal plot of residual of the OPD cases, it can be seen that the residuals do not deviate much from the straight line. This indicates that the errors are quite close to normal with no clear outliers. Thus, the normality assumption holds. The histogram of residuals confirms this assumption. The plots of residuals versus the fitted values exhibit no trend in dispersion. This indicates that the data satisfies the constant variance assumption. The plot of residuals versus the order of the data suggests that the residuals are uncorrelated. Thus the independent assumption is not violated. Since the assumptions hold the data can be seen as valid to carry out the analysis.

3.2 Test for Stationarity of OPD Data

In checking for the stationarity of the dataset, Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was employed.

Hypothesis statement

$$\mathrm{H_O}: \text{Data is stationary}$$

$$\mathrm{H_A}: \text{Data is not stationary}$$

Table 2. KPSS Statistic of the OPD Data

Variables	KPSS Level	P-Value	Truncated lag
Before differencing	2.0614	0.020	2
After differencing	0.6612	0.140	2

∝ = 0.05 (significance level)

For the raw outpatient department data, since the *p-value* is 0.020, less than α = 0.05, we reject the null hypothesis. Hence, we conclude that the series of the raw OPD data is not level stationary, therefore needs differencing. For the differenced OPD data, since the $p - vlue = 0.140$ is greater than $\alpha = 0.05$, we fail to reject the null hypothesis and therefore conclude that the series of the differenced OPD data is level stationary. The differenced series can now be used for forecasting.

3.3 Fitting Model and Forecasting For the OPD Data

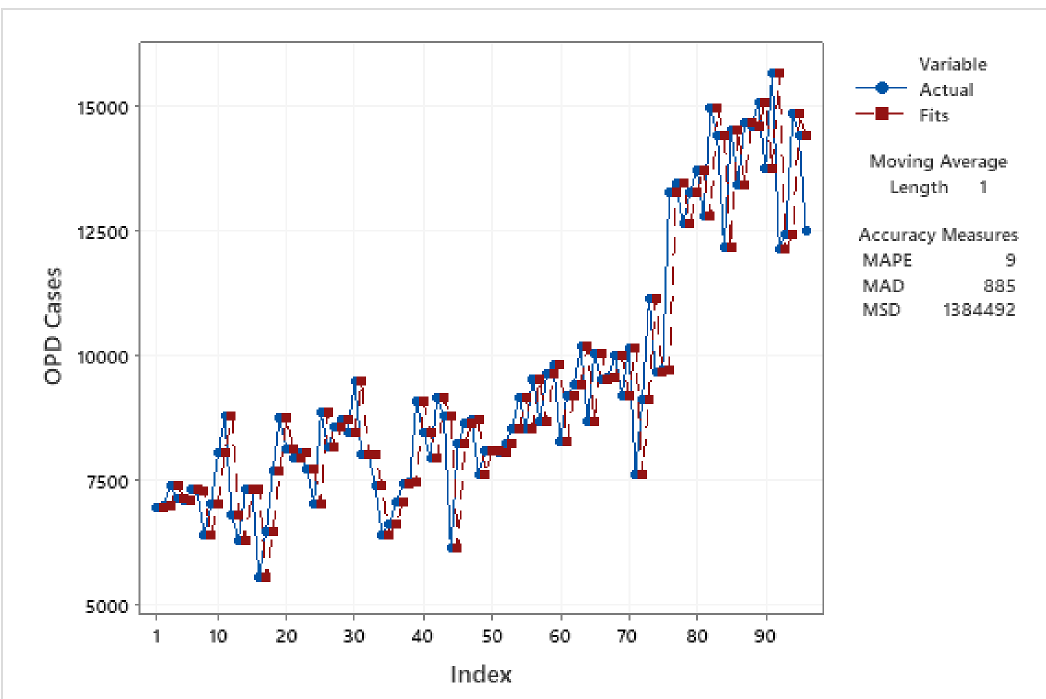

Figure 3. Moving Average (MA) with 1 Average

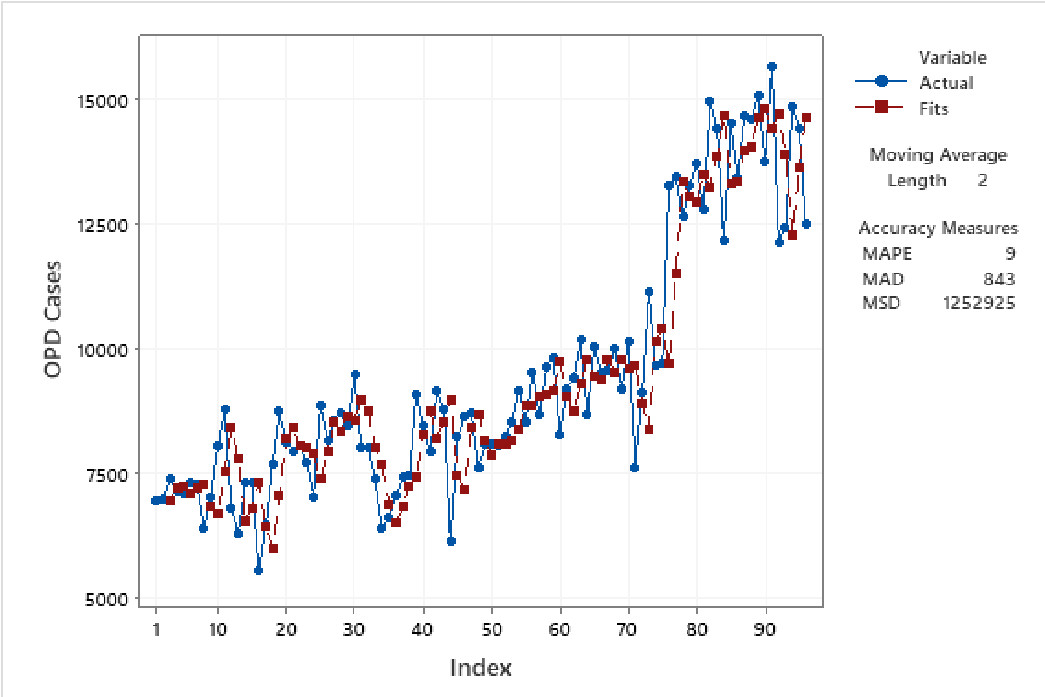

Figure 4. Moving Average (MA) with 2 Averages

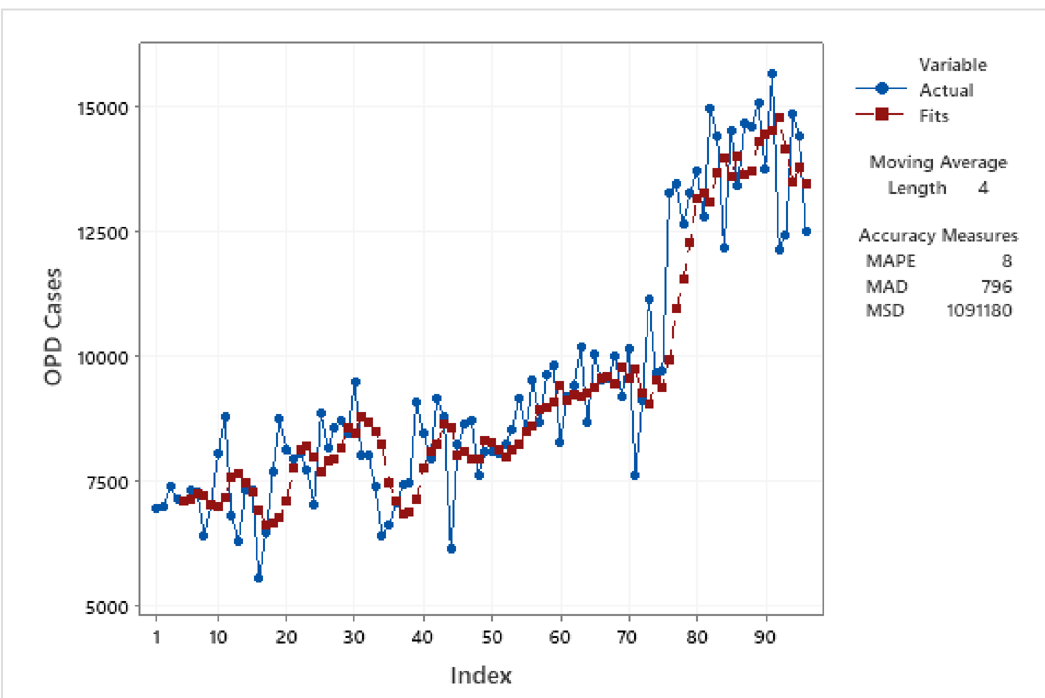

Figure 5. Moving Average (MA) with 4 Averages

The Moving Average (MA) analyses for lags 1, 2 and 4 are in Figures 3, 4, and 5 above. A comparison of their respective Mean Absolute Percentage Error and Median Average Deviation as criterion for selecting, there is a clear indication that MA (1) better fits the OPD attendance data than the others.

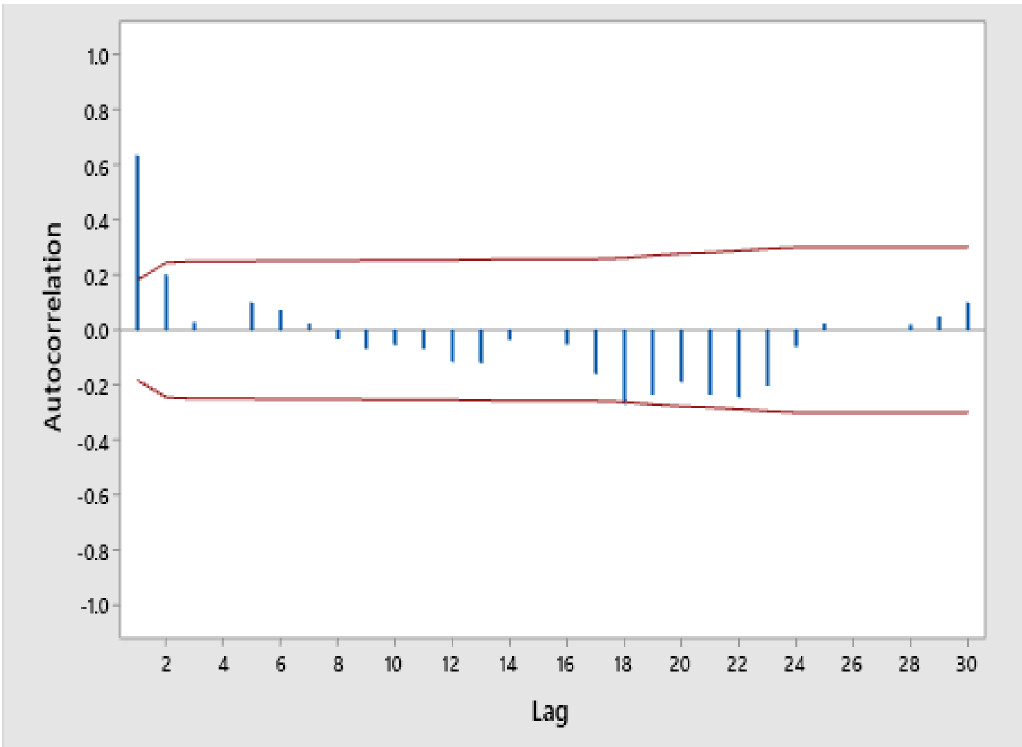

Figure 6. ACF for Second Order Differencing

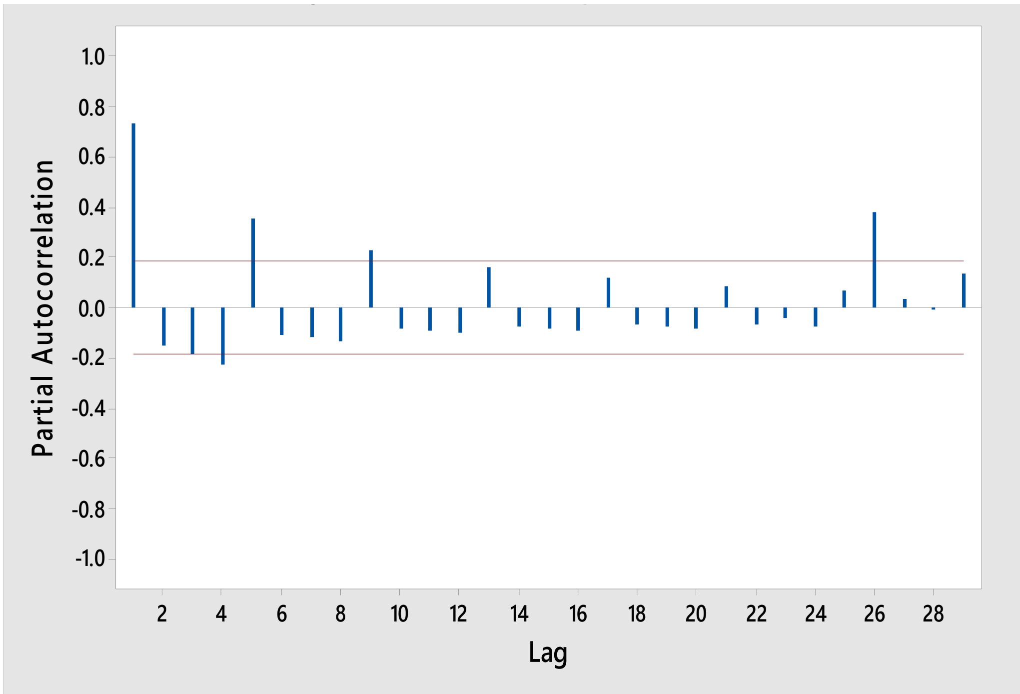

Figure 7. PACF for Second Order Differencing

Figures 6 and 7 present the plot which determines the order of the AR and MA for both seasonal and non-seasonal components. This was suggested by the sample ACF and PACF plots based on the Box-Jenkins approach. From Figure 6, the correlations are significant for a large number of lags, but the autocorrelations at lags 2 or and above are merely due to the propagation of the autocorrelation at lag 1. This is confirmed by the PACF plot in Figure 7. The ACF and PACF plots, respectively suggest that q = 1, and p = 1 would be needed to describe this data set as coming from a non-seasonal moving average and autoregressive process respectively.

3.4 ARIMA Model Estimations

Several non-seasonal ARIMA models are constructed as follows:

Table 1. Summary of Models for OPD Data

Models	Chi Square	Df	P-Value
ARIMA (2,2,1)	5.3	8	0.756
ARIMA (2,2,2)	5.6	8	0.536

ARIMA (1,2,1)	6.3	8	0.528

In comparing the p-values and Chi-square values of the three non-seasonal ARIMA, it can be concluded that model ARIMA (2, 2, 1) has the highest p-value and a relatively low Chi-square values of 0.756 and 5.3. This indicates that it is the best non-seasonal model for the data. The partial autocorrelation and the autocorrelation of the second differences suggest that the original series can be modelled as ARIMA (2, 2, 1).

Table 4. Final Estimates of Parameters

Type	**Coef.**	**SE Coef.**	**T-Value**	**P-Value**
AR 1	-0.5279	0.0981	-5.38	0.000
AR 2	-0.4402	0.0981	-4.49	0.000
MA 1	0.9780	0.0319	30.63	0.000
Constant	4.07	6.00	0.68	0.500

Table 4 presents the final estimate of parameters for the model. The MA (1), AR (1) and AR (2) parameters having *p-value* of 0.000, 0.000 and 0.000, indicating significant model parameters.

Table 2. Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	Chi Square	DF	P- Value
12	8.02	8	0.432
24	14.91	20	0.782
36	25.92	32	0.767
48	41.53	44	0.578

Table 5 presents the modified Box-Pierce Chi-square statistic. It can be seen that all the lags have a *p-value* greater than the level of significant of 0.05. This indicates non-significance implying that the model is appropriate. Again, the Ljung-Box gives a no significant p-values, indicating that the residuals appear to be uncorrelated.

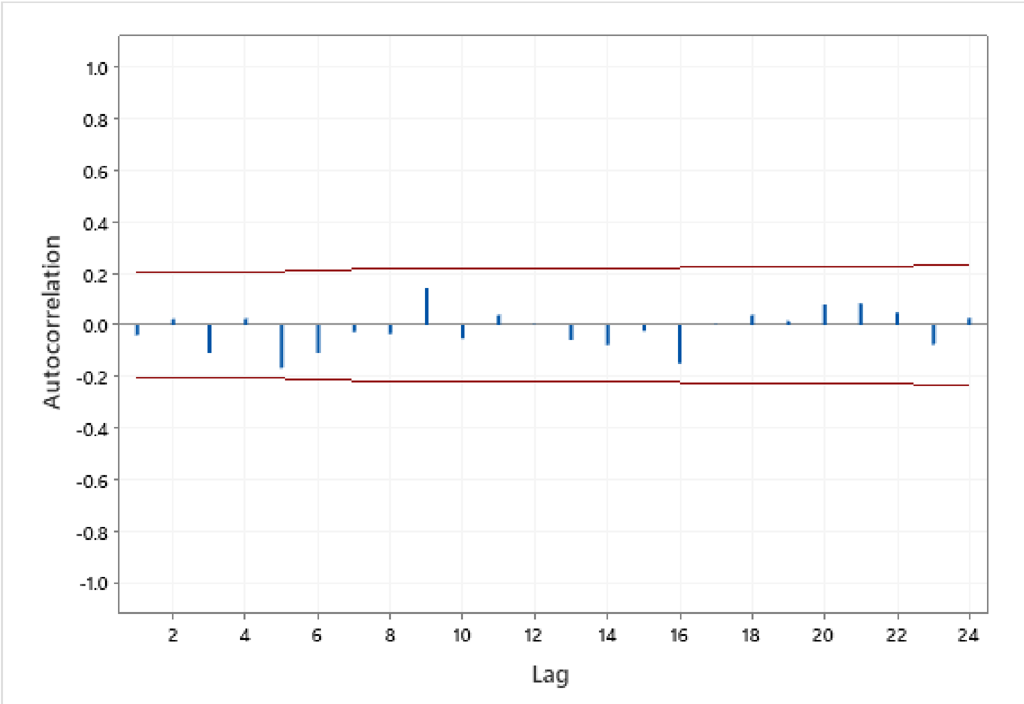

Figure 8. AFC Diagnostic Plot of the Residuals for ARIMA (2, 2, 1) Model

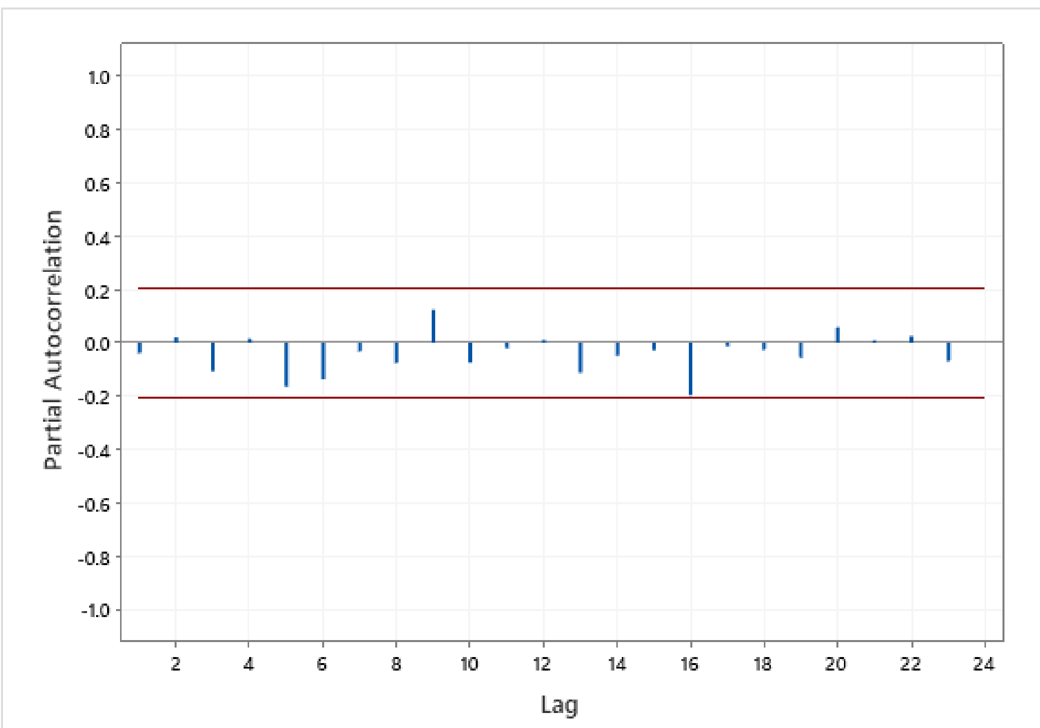

Figure 9. PACF Diagnostic Plot of the Residuals for ARIMA (2, 2, 1) Model

The residual diagnostic test as shown in Figures 8 and 9 is performed for further confirmation of the selected model.

4. DISCUSSION

Table 6. 2020 Forecasted Values for Outpatient Department Hospital Attendance Compared with Actual 2019 Attendance

Years	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
2019	14533	13436	14655	14607	15062	13759	15664	12151	12429	14855	14409	12496
2020	14020	14379	13844	14628	14628	14592	14808	15056	15180	15360	15569	15742

Table 7. Forecasted values of insured hospital attendance.

Years	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.
2020	14020	14379	13844	14628	14628	14592	14808	15056	15180	15360	15569	15742
2021	15925	16123	16312	16504	16702	16900	17099	17302	17506	17712	17921	18131
2022	18344	18558	18775	18994	19437	19662	19765	19889	20118	20350	20583	20818
2023	21055	21295	21536	21780	22025	22273	22523	22774	23028	23284	23542	23802
2024	24064	24329	24595	24863	25134	25406	25681	25957	26236	26517	26799	27084
2025	27371	27660	27951	28244	28540	28837	29136	29438	29741	30047	30354	30664

The expected OPD cases for the next five years in Table 7, confirms that the outpatients hospital attendance cases will be increasing with respect to months. Thus the outpatient departments will be witnessing corresponding progressive and significant increase in attendance at the Cape Coast Teaching Hospital of which the research seeks to investigate.

From Table 7, one can observe that the values for the forecast monthly outpatient attendance in 2020 increased for specific months January, February, April, June, August to December than the actual OPD visits across the various months with the year 2019 under review an indication of changing pattern by patients reporting to the facility. The year 2019 recorded an increase in the months of March, May and July compared with the same forcasted months of 2020. Patients accessing the OPD will be increasing in the years under forecast as compared to the number of OPD attendance of the year under review. Continued use of the outpatient department in accessing health care at all levels will see an increase in hospital visits across the months from June 2020 to December 2025.

Also, from Table 7, January to May 2020 exhibited an increasing and decreasing trend, but a stationary trend for the months April to May as compared to the same months in the year 2019.

5. CONCLUSION

In conclusion, it can be said that outpatient department on hospital attendance cases showed variability of processes caused by many irregular factors that cannot be eliminated in cases recorded. This research identified candidate models that generally best fitted the data. Using the modified Box Pierce (Ljumg-Box) Chi-square statistic criteria of the largest $p-value$ and minimum Chi-square statistic value, the selected best fitted model for outpatient department attendance was ARIMA (2, 2, 1). The number of hospital attendance each month showed no seasonal variation. Based on the findings of the time series analysis, outpatient department attendance cases will be increasing for the next five years. Hence the use of the outpatient department in health administration has increased hospital attendance with time.

RECOMMENDATIONS

The government should continue the expansion of community-based health planning and services in all parts of the country to increase access to healthcare by all as it services goes to the core people in the community. The health authorities should continue to expand the outpatient department in order to be able to accommodate the increasing number of patients visit the facility Authorities should support health facilities in terms of personnel and logistics in order to provide quality health care to the increasing OPD patients. There should be an expansion of the existing OPD unit in the Cape Coast Teaching Hospital since the hospital mostly referral in the Central Region.

ETHICAL APPROVAL

It is not applicable

COMPETING INTERESTS

There is no competing interest. This research paper has not been submitted elsewhere to be published

AUTHOR CONTRIBUTION

I Bridget Sena Borbor is the main author of this research, I designed this research work, performed all the statistical analysis, reviewed literature and organized the research paper for publication.

REFERENCES

Anon. 2016 Annual Report of the ABPN. In *The American journal of psychiatry*, 2017:174. Accessed 19 August 2020. Available: https://ajp.psychiatryonline.org/doi/10.1176/appi.ajp.2017.174804

Aidoo, E. *Modelling and forecasting inflation rates in Ghana: An application of SARIMA models* (Dissertation). Börlange, SWEDEN, 2010: Accessed 02 September 2020. Available: http://du.diva-portal.org/smash/record.jsf?pid=diva2%3A518895&dswid=1088

Banor, F., & Gyan, F. "*Modelling Hospital Attendance in Ghana: A case study of Obuasi Government Hospital"* Project work, Garden City University, 2012: Accessed 22 August 2020. Available: https://www.academia.edu/6626411/Modeling_hospital_attendance_in_Ghana_A_case_of_the_Obuasi_government_hospital.

Luo, L., Luo, L., Zhang, X., & He, X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Services Research*,2017: *17*(1), 1–13. Accessed 17 September 2020. Available: https://bmchealthservres.biomedcentral.com/articles/10.1186/s12913-017-2407-9

Borbor, B. S., Bosson-Amedenu S., Daniel Gbormittah D. Statistical Analysis of Health Insurance and Cash and Carry Systems in Cape Coast Teaching Hospital of Ghana, *Science Journal of Applied Mathematics and Statistics*. 2019: 7(3), 36-44. doi: 10.11648/j.sjams.20190703.12. Assessed: 24 Seotember, 2020. Avilable : http://article.sjams.org/pdf/10.116...

George Box P. E and Gwilym Jenkins M. Time series Analysis, Forecasting and control. Holden-Day, Oakland, California, USA, 1976: 2nd edition. Assessed: 25 September 2020. Avilable: http://garfield.library.upenn.edu/classics1989/A1989AV48500001.pdf

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. Time series analysis, forecasting and control (3rd ed.). New Jersey: Prentice Hall, Englewood Cliffs.1994. Assessed: 23 September 2020. Available: https://www.scirp.org/(S(i43dyn45teexjx455qlt3d2q))/reference/ReferencesPapers.aspx?ReferenceID=1936224

Dickey, D., & Fuller, W. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica,* 1981:*49*(4), 1057-1072. doi:10.2307/1912517. Assessed: 24 September 2020. Available: https://www.jstor.org/stable/pdf/1912517.pdf

Qmul. Time series.2018b: Accessed 02 September 2020. Available: http://article.sjams.org/pdf/10.11648.j.sjams.20190703.12.pdf

PSU. Applied time series. 2018b: Accessed 02 September 2020. Available: .http://article.sjams.org/pdf/10.11648.j.sjams.20190703.12.pdf

PSU. Applied time series. 2018c: Accessed 02 September 2020. Available: http://article.sjams.org/pdf/10.11648.j.sjams.20190703.12.pdf