FORECASTING OUTPATIENT VISITS AT MARIMANTI LEVEL 4 HOSPITAL: TIME SERIES ANALYSIS USING SARIMA MODEL

KENNETH KIPCHOGE RONO

A Thesis Submitted to Graduate School in Partial Fulfillment of the Requirements for the Award of Degree of Masters of Science in Applied Statistics of Tharaka University

THARAKA UNIVERSITY

NOVEMBER 2024

DECLARATION AND RECOMMENDATION

DECLARATION AND RECOMMENDATION

This thesis is my original work and has not been presented for an award of a diploma or conferment of degree in any institution.

Declaration

Signature: ____________________ Date: 19/11/2024

Kenneth Kipchoge Rono
SMT18/05469/21

Recommendation

This thesis has been examined, passed and submitted with our consent as University Supervisors.

Signature: ____________________ Date: 19/11/2024

Prof. Dennis K. Muriithi, Ph.D.
Department of Physical Sciences
Chuka University

Signature: ____________________ Date: 19/11/2024

Dr. Daniel Muriithi Mwangi, Ph.D.
Department of Basic Sciences
Tharaka University

COPYRIGHT

DEDICATION

I dedicate this work to my parents and siblings.

ACKNOWLEDGEMENT

This work's accomplishment is largely due to the Almighty God's Grace and Mercies. A special thanks to Tharaka University for the chance to further my education in the field of mathematics. I also want to thank my supervisors, Dr. Daniel Mwangi and Prof. Dennis K. Muriithi, for their insightful comments, suggestions, and great support throughout the research. I would especially want to thank the Medical Superintendent of the Marimanti level 4 Hospital for letting me do research there. A special thanks goes out to the Faculty of Physical Science, Engineering, and Technology, under the direction of Dean Dr. Fidelis Ngugi, for the advancement of my mathematical knowledge. Special appreciation to the University Library for the provision of the much-needed learning resources used during this study. May God bless you!

ABSTRACT

The Outpatient Department (OPD) is the initial point of contact for all patients, and the level of treatment received at OPD influences their perception of the hospital. Turnaround time substantially influence the quality of service offered to the patients and thus patients satisfaction increased. The Kenyan government has implemented several actions and programs to strengthen the health sector, including the Kenya Health Policy 2014-2030 and the Sustainable Development Goals. To achieve these interventions, establishing an accurate and effective forecasting model is necessary. The study aimed to fit a Statistical model to forecast outpatient attendance at Marimanti Level 4 Hospital using SARIMA modeling. The objectives were to; establish the trend and seasonal effects in Outpatient attendance at Marimanti Level 4 Hospital in the past 10 years, fit a SARIMA model using outpatient attendance data at Marimanti Level 4 Hospital and forecast the Outpatient Visits at Marimanti Level 4 hospital for the next 2 year. The study used monthly outpatient visit data from Marimanti level 4 hospital (January 2013 to December 2023). The study contributes to the literature on SARIMA modeling, aids in balance work schedule, increased patient safety and resource allocation. Data analysis was carried out using R and R Studio version 4.4.1. The outpatient visits series at Marimanti level 4 hospital is seasonal. The study used the Box-Jenkins technique in modelling. SARIMA (0,1,2) $(2,1,1)_{12}$ emerged as the most plausible model (AIC = 1139.56, BIC = 1154.52). The forecasting accuracy of the model was assessed using MAPE = 1.66% and MASE = 0.47%. Overall, the two-year prediction showed an increasing number of outpatient visits at Marimanti level 4 hospital. As a result, the hospital Management ought to take into consideration the forecasts to enable them prepare adequately in terms of resource allocation and planning purposes, continuous data collection and analysis to keep the model up-to date. The study recommends; further studies on OPD using machine learning models, continuous updating of the model to ensure any emerging pattern in the data series is captures, finally, the hospital management should rely on the forecasts for planning activities.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	:	Autocorrelation Function
ADF	:	Augmented Dickey-Fuller
AED	:	Accident and Emergency Department
AI Cc	:	Akaike Information Criteria Corrected
AIC	:	Akaike Information Criteria
AR	:	Autoregressive
ARIMA	:	Autoregressive Integrated Moving Average
ASALs	:	Arid and Semi-Arid Lands
BIC	:	Bayesian Information Criteria
ED	:	Emergency department
HK	:	Hyndman-Khandakar
KPSS	:	Kwiatkowski-Philips-Schmidt-Shin
MA	:	Moving Average
MAPE	:	Mean Absolute Percentage Error
MK	:	Mann-Kendall
MSE	:	Mean Squared Error
OPD	:	Outpatient Department
PACF	:	Partial Autocorrelation Function
SARIMA	:	Seasonal Autoregressive Integrated Moving Average
UHC	:	Universal Health Coverage

CHAPTER ONE
INTRODUCTION

1.1 Background of the Study

According to Jonathan & Kung, (2008), time series data is a sequential arrangement of observations over time. Recently, time series models have received more attention, in health sector, it is used to predict medical services demand, including patient attendance, hospital release, bed demand or turnaround time (Kadri *et al*., 2014). According to Aboagye *et al*., (2015) time series modeling is crucial for health care administrators to better understand the characteristics and the nature of the variation. Because the ARIMA model captures linear patterns in data with least computational effort, most research utilize it to characterize the relationship among variables or as a benchmark to measure the efficacy of combination models. ARIMA models have a wide range of application in projecting hospital daily visits to Outpatient and Emergency Departments due to its efficiency in capturing linear properties of trend (Bergs *et al*., 2014).

Based on monthly data, Permanasari *et al*., (2013) examined and applied SARIMA modelling to project the prevalence of Malaria. The ability to forecast diseases is crucial for management decision-making about resource allocation and workload scheduling. Dabral & Murry, (2017) utilized SARIMA model to simulate and predict rainfall time series. SARIMA Model is important in modelling and forecasting because of its ability to extract linear pattern from complex data and simplicity in its computations to achieve desired outcome, saving time and effort.

According to Slawomirski *et al*., (2017) low quality of healthcare leads to increase in disease burden, unmet medical need and has enormous cost repercussions for local communities and the global health system. Trillions of dollars are spent each year on patient harm from life-time impairments, disabilities, and productivity loss. The underprivileged groups in society are disproportionately affected by low quality treatment.

The model's capacity to successfully handle seasonality and trends is shown by empirical assessments of SARIMA modeling's performance in forecasting hospital outpatient visits. To predict outpatient visits and optimize resource allocation, SARIMA models have been used in a variety of healthcare settings around the world. Studies conducted in nations such as USA, China and India have shown that SARIMA effectively captures the cyclical and seasonal patterns in outpatient data, enhancing staffing and supply planning. SARIMA was used to an electronic health record dataset for outpatient predictions at India's L.V. Prasad Eye Institute, producing accurate forecasts that are essential for effective resource management in the provision of eyecare services. Seasonal changes were effectively addressed by the SARIMA model, which had parameters adjusted to match the observed monthly and annual visit patterns at this sizable tertiary network (Sai *et al.*, 2021)

One of the most important objectives that is guaranteed in the Sustainable Development Goals (SDGs) is universal health coverage (UHC). UHC seeks to enable a transition towards more equitable and efficient countries and economies by offering safe healthcare to individuals of all ages and ensuring that everyone has access to essential medical care without facing financial hardship (WHO, 2018). A great opportunity exists for nations with low or middle-incomes to address the quality while constructing UHC. It is possible to influence, guide, and foster a developing health system in the way that is wanted. Quality can become embedded in the system's structure, protocols, and regulations as it grows and evolves. Although providing everyone with high-quality medical care might seem unattainable, it is feasible in all situations with the right administration, careful planning, and financial support. For instance, due to a policy that incorporates communities and people in the design of the healthcare service, Uganda has seen improvements in all indicators, including a 33% fall in child mortality (OECD *et al*., 2018). As a result of the UHC roll out, patients seeking healthcare increases and therefore, having a precise model informs the hospital management's decision making and planning to guarantee effective, efficient and economic resource mobilization and services provided.

Procedures and therapies that do not require patients to remain in the procedure site for further monitoring or care are referred to as outpatient care. Pharmacies, specialty outpatient centers, and emergency rooms are some of the establishments that offer outpatient therapy. Outpatient care services account for a number of different procedures, tests, and treatments which a patient can receive. They include; Services for prevention and well-being including psychological support, instruction on nutrition, a weight reduction program and counseling on diets, diagnostic procedures such as (scanning and lab examinations) and rehabilitation services: Physical and occupational therapy.

Hospitalists experience daily workload changes, some of which are unpredictable. The hospital administration employs a variety of reactive strategies to match workforce level to workload. While overstaffing is expensive and unsustainable, understaffing can have a negative influence on patient safety and hospitalists' job satisfaction. A better understanding of when patient will visit might help with scheduling and staffing decisions, resulting in a better workload, increased job satisfaction for the hospitalists, and improved patient safety.

Marimanti level 4 hospital is a public health facility in Kenya. It is a primary care hospital located in Marimanti ward, Tharaka Constituency, Tharaka South Sub-County, Tharaka Nithi County along Chiakariga – Marimanti road near Marimanti Market. There are 39058 women and 36,190 men in the hospital's 75,250-person catchment area, with a fertility rate of 2.9 births per woman. Compared to the overall crude death rate of 10.5 deaths per 1000 population in Kenya, Tharaka Nithi County crude death rate of 10.6 deaths per 1000 population is higher. The majority of people reside in rural areas.

There is few empirical research on application of SARIMA model in healthcare settings in Kenya, comparable approaches have been investigated for healthcare metrics forecasting, suggesting that local adoption may be possible. SARIMA's capacity to model recurrent seasonal variations should be advantageous for Kenyan healthcare facilities dealing with varying outpatient volumes. By matching resources to anticipated outpatient visit numbers, SARIMA enables local facilities to manage costs effectively. This is especially true for facilities that have digitized patient data accessible for model training.

1.2 Statement of the Problem

Patients always fashion out the hospital by the service they receive at the Outpatient Department. The quality of service largely depends on the time a patient takes to receive the service (Turn-around time). Long waiting time is perceived by the patient as the hindrance to getting the service and may hurt the patient's attitude towards the hospital. Currently, health services received by patients in the many public hospitals are not encouraging because of long queue, inadequate supplies and under staffing. The recently launched Universal Health Coverage (UHC) program and the establishment of Tharaka University within the catchment population of the Marimanti Level 4 hospital, there is an anticipated increase in outpatient visits. Therefore, having a precise model that can project the expected number of outpatient visits in future is important. The forecasts aids Hospital management in proper planning in terms of staffing, workload scheduling, training and allocation of resources, this necessitated the endeavor.

1.3. Purpose of the Study.

The study's goal was to forecast outpatient visits at Marimanti Level 4 Hospital by employing the SARIMA model for time series analysis.

1.4. Specific Objective.

The execution of this study was based on the following stated objectives.

i. Establish the trend and seasonal effects in Outpatient visits at Marimanti Level 4 Hospital in the past 10 years.
ii. Fit a SARIMA model using Outpatient attendance data at Marimanti Level 4 Hospital.
iii. Forecast the Outpatient Visits at Marimanti Level 4 hospital for the next 2 year.

1.5. Research Questions

i. Are there regular patterns in Outpatient attendance at Marimanti Level 4 hospital?
ii. Is it possible to empirically model the outpatient visit data?
iii. What is the expected number of outpatient visits in the next 2 years?

1.6. Significance of the Study

The fitted model was used to explain the hospital Outpatient attendance and to project the expected patient volumes for 2 years ahead, which provided a basis of improving staffing and scheduling decision to provide a better workload balance and improve job satisfaction of the hospitalist as well as improving patients' safety. The forecasts aids in decision making on allocation of resources and improving hospital's processes.

This research work adds to a pool of information available on modelling hospital OPD attendance and the study finding forms the foundation for future research. The research provided recommendations for further improvement in future.

This research work provides an evident based decision making upon which policies are built on. This ensures quality is entrenched in regulations, procedures, and policies to guarantee high quality of healthcare provided.

The primary beneficiaries of this research are the patients. The findings offer insights to hospital management for planning purposes such as resource allocation, training, recruiting, and workload balancing, which impact positively on the services provided to patients, leading to increased satisfaction and patient safety.

1.7. The Scope of the Study

Secondary data on monthly outpatient attendance at Marimanti Level 4 Hospital was used in this study. The data covered eleven-year span, from January, 2013 to December, 2023. SARIMA model was developed and used to predict outpatient attendance for the next two years. R and R- Studio software version 4.4.1 was used in the analysis. The research period was chosen to assess how devolution has impacted the health sector in the country.

1.8. Assumption of the Study

The Outpatient visit is stochastic process. Since the study used the secondary data on Outpatient attendance, it is assumed that the data was accurate and authentic to make inference.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Time series models is covered in this chapter, along with research that has effectively used SARIMA to forecast outcomes. The SARIMA model was used to model the outpatient attendance data at Marimanti Level 4 Hospital and make forecasts.

2.2. Time Series Model

2.2.1. Autoregressive Process AR (p)

A p^{th}order autoregressive process X_t satisfy the equation.

$$X_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_p x_{t-p} + e_t \tag{2.1}$$

Where e_t is Gaussian White Noise and is independent of $x_{t-1}, x_{t-2} \cdots, x_{t-p}$ and $\beta_1,\ldots,$ β_p are the coefficients.

The present value of the sequence X_t is a regression of the latest p values on itself with the innovation of an error component e_t, which includes anything new in the process at t not accounted for by the previous values. The statistical features of the AR (p) process are as follows;

Mean:

$$E(X_t) = 0$$

Variance:

$$var(X_t) = \frac{\delta^2}{(1-\beta^2)} \tag{2.2}$$

Autocorrelation:

$\rho(h) = \frac{R(h)}{R(0)}$,

Where $R(h)$ is the autocovariance and $R(0)$ is the variance.

$$\rho(h) = \beta_1 \rho(h-1) + \beta_2 \rho(h-2) + \cdots + \beta_p \rho(h-p) \tag{2.3}$$

For $k \geq 1$

The Yule Walker equation, an even difference equation, is represented as equation (2.3). We use $\rho(h) = \mathrm{Z}^h$ as the trial solution to get the matching solutions.

$$Z^h = \beta_1 Z^{h-1} + \beta_2 Z^{h-2} + \cdots + \beta_p Z^{h-p}$$

$$Z^h - \beta_1 Z^{h-1} - \beta_2 Z^{h-2} - \cdots - \beta_p Z^{h-p} = 0$$

$$Z^{h-p}\left(Z^p - \beta_1 Z^{p-1} - \beta_2 Z^{p-2} - \cdots - \beta_p\right) = 0$$

$Z^{h-p} \neq 0$, therefore

$$Z^p - \beta_1 Z^{p-1} - \beta_2 Z^{p-2} - \cdots - \beta_p = 0 \qquad (2.4)$$

Which is the characteristic equation.

The AR (p) process is stable if all the solutions of equation (2.4) lie outside the unit circle.

Employing the backward shift operator (G) on the equation (2.1), the AR process can be expressed as;

$$X_t = \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_p x_{t-p} + e_t$$

$$X_t - \beta_1 \mathrm{G} X_t - \beta_2 \mathrm{G}^2 X_t - \cdots - \beta_p \mathrm{G}^p X_t = e_t$$

$$(1 - \beta_1 \mathrm{G} - \beta_2 \mathrm{G}^2 - \cdots - \beta_p \mathrm{G}^p) X_t = e_t$$

$$X_t = \frac{e_t}{\left(1 - \beta_1 \mathrm{G} - \beta_2 \mathrm{G}^2 - \cdots - \beta_p \mathrm{G}^p\right)} \qquad (2.5)$$

Case of AR (1) process

AR (1) is given as;

$$X_t = \frac{e_t}{(1 - \beta \mathrm{G})}$$

$$= \varepsilon_t (1 + \beta \mathrm{G} + \beta^2 \mathrm{G}^2 + \cdots)$$

$$= \sum_{i=0}^{\infty} \beta^i e_{t-i} \qquad (2.6)$$

The mean;

$$E(X_t) = E\left(\sum_{i=0}^{\infty} \beta^i \, e_{t-i}\right)$$

$$= \sum_{i=0}^{\infty} \beta^i \, E(e_{t-i})$$

$$= 0 \tag{2.7}$$

The variance;

$$Var\,(X_t) = Var\left(\sum_{i=0}^{\infty} \beta^i \, e_{t-i}\right)$$

$$= \sum_{i=0}^{\infty} \beta^{2i} \; Var(e_{t-i})$$

$$= \; \delta^2 \sum_{i=0}^{\infty} \beta^{2i}$$

$$= \; \delta^2 (1 + \beta^2 + \beta^4 + \beta^6 + \cdots)$$

$$= \frac{\delta^2}{(1 - \beta^2)} \tag{2.8}$$

Autocovariance function (ACVF);

$$Cov(X_t, X_{t+h}) = R(h) = E\left(\sum_{i=0}^{\infty} \beta^i \, e_{t-i}\right)\left(\sum_{l=0}^{\infty} \beta^l \, e_{t+h-l}\right)$$

$$= \sum_{i=0}^{\infty} \sum_{l=0}^{\infty} \beta^i \, \beta^l \, E(e_{t-i}, e_{t+h-l})$$

By letting $l = i + h$ we have;

$$= \sum_{i=0}^{\infty} \beta^{2i} \, \beta^h E(e_{t-i}, e_{t-i})$$

$$= \delta^2 \beta^h \sum_{i=0}^{\infty} \beta^{2i}$$

$$= \delta^2 \beta^h (1 + \beta^2 + \beta^4 + \beta^6 + \cdots)$$

$$= \frac{\delta^2 \beta^h}{(1 - \beta^2)} \tag{2.9}$$

Autocorrelation function (ACF);

$$\rho(h) = \frac{R(h)}{R(0)}$$

$$= \frac{\frac{\delta^2 \beta^h}{(1 - \beta^2)}}{\frac{\delta^2}{(1 - \beta^2)}}$$

$$= \beta^h \tag{2.10}$$

Where $R(h)$ is the autocovariance and $R(0)$ is the variance, (Nyamao, 2014).

2.2.2. Moving Average Process

Consist of sequence of Gaussian e_t. Where $e_t \sim (0, \delta^2)$

MA (q) is denoted by;

$$X_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q}$$

$$= \sum_{l=0}^{q} \theta_l \, e_{t-l} \tag{2.11}$$

Mean:

$$E(X_t) = E\left(\sum_{l=0}^{q} \theta_l \, e_{t-l}\right)$$

$$= \sum_{l=0}^{q} \theta_l \, E(e_{t-l})$$

$$= 0 \tag{2.12}$$

Variance:

$$Var\,(Xt) = \; Var\left(\sum_{l=0}^{q} \theta_l \, e_{t-l}\right)$$

$$= \delta^2 \sum_{l=0}^{q} {\theta_l}^2 \qquad (2.13)$$

Autocorrelation:

$$\rho(\mathrm{k}) = \begin{cases} 1, & k = 0 \\ \dfrac{\sum_{l=0}^{q-k} \theta_l \theta_{l+k}}{\sum_{l=0}^{q} \theta_l}, & 0 \leq k \leq q \\ o, & k > \mathrm{q} \end{cases}$$

Where $\rho(\mathrm{k})$is the autocorrelation. Using the backshift Operator to equation (2.11) we have;

$$X_t = e_t + \theta_1 \mathrm{G} e_t + \theta_2 \mathrm{G}^2 e_t + \cdots + \theta_p \mathrm{G}^q e_t$$

$$= \left(1 + \theta_1 \mathrm{G} + \theta_2 \mathrm{G}^2 + \cdots + \theta_q \mathrm{G}^q\right) e_t$$

$$= \theta(\mathrm{G}) e_t \qquad (2.14)$$

Where $\theta(\mathrm{G}) = 1 + \theta_1 \mathrm{G} + \theta_2 \mathrm{G}^2 + \cdots + \theta_q \mathrm{G}^q$is a polynomial of order q in G.

2.2.3. Autoregressive Moving Average (ARMA) Process

The following is the generic ARMA (p, q) equation;

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_p X_{t-p} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} \qquad (2.15)$$

Where e_t is Gaussian White Noise.

Employing backshift operator (G) to equation (2.15) it reduces to;

$$X_t - \beta_1 \mathrm{G} X_t - \beta_2 \mathrm{G}^2 X_t - \cdots - \beta_p \mathrm{G}^p X_t = e_t + \theta_1 \mathrm{G} e_t + \theta_2 \mathrm{G}^2 e_t + \cdots + \theta_q \mathrm{G}^q e_t$$

$$\left(1 - \beta_1 \mathrm{G} - \beta_2 \mathrm{G}^2 - \cdots - \beta_p \mathrm{G}^p\right) X_t = \left(1 + \theta_1 \mathrm{G} + \theta_2 \mathrm{G}^2 + \cdots + \theta_q \mathrm{G}^q\right) e_t$$

$$\beta(\mathrm{G}) X_t = \theta(\mathrm{G}) e_t \qquad (2.16)$$

Where,

$\beta(\mathrm{L}) = 1 - \beta_1 \mathrm{G} - \beta_2 \mathrm{G}^2 - \cdots - \beta_p \mathrm{G}^p$, and

$$\theta(B) = 1 + \theta_1 \mathrm{G} + \theta_2 \mathrm{G}^2 + \cdots + \theta_q \mathrm{G}^q$$

The Autoregressive (AR) component determines the process's stationarity, while the Moving Average (MA) component is taken into consideration for the process's invertibility, (Velicer & Molenaar, 2012).

2.2.4. Autoregressive Integrated Moving Average (ARIMA) Process

AR and MA models are combined using the integration approach in ARIMA (p, d, q). The letter I (integrated) indicates that an initial-order difference to make the series stationary because ARIMA modelling requires that the time series data be either stable or capable of becoming stationary, (Hyndman & Athanasopoulos, 2018).

ARIMA (p, d, q) is represented;

$$\beta(\mathrm{G})(1-\mathrm{G})^d X_t = \theta(\mathrm{G})e_t \tag{2.17}$$

Where $\beta(\mathrm{G})$ is an AR polynomial, $(1-\mathrm{G})^d$is difference to detrend the outpatient series, $\theta(\mathrm{G})$ is a MA polynomial and e_t is a White Noise, G is the backshift operator and d is the degree of differencing to render the series stationary.

2.2.5. Seasonal Autoregressive Integrated Moving Average (SARIMA) Process

When modelling data series that have both regular and periodic components, SARIMA (p, d, q) (P, D, Q) is used. The series' trend (p, d, q) is eliminated using trend difference, and its seasonality (P, D, Q) is eliminated using seasonal difference. The definition of the SARIMA (p, d, q) (P, D, Q)$_s$ model is;

$$\beta_p(\mathrm{G})\beta_P(\mathrm{G}^s)(1-\mathrm{G})^d(1-\mathrm{G}^s)^D X_t = \theta_q(\mathrm{G})\Theta_Q(\mathrm{G}^s)e_t \tag{2.18}$$

s is the seasonality length and e_t is a sequence of Gaussian white noise (Hyndman & Athanasopoulos, 2018).

2.3. Stationarity and Invertibility

2.3.1. Stationarity

A stationary time series has no regular movement in its mean (Trend) or variances.

Weakly Stationary: If a time series' average, variance, and autocovariance are not affected by time, it is stationary in the weak sense. Therefore, if X_t is a time series that is defined for $t = 0, \pm 1, \pm 2, \ldots$ Then, it is weakly stationary if;

$$E(X_t) = \mu < \infty, \quad \forall t \in T$$

The covariance $cov(X_t, X_{t+k}) = R(k) < \infty, k = 1, 2, \ldots$

Strongly Stationary

A time series X_t is stationary in strong sense if its probability structure remains unchanged under a displacement in time.

$$F[X_{t1}, X_{t2}, X_{t3}, \dots, X_{tn}] = F[X_{t1+k}, X_{t2+k}, X_{t3+k}, \dots, X_{tn+k}]$$

$\forall t_1, t_2, t_3, \dots, t_n$ and $t_{1+k}, t_{2+k}, \dots, t_{n+k} \in T$, and F is the cumulative distribution function.

2.3.2. Invertibility

Invertibility is a condition imposed to the constant to ensure that the MA process can be identified uniquely from a given ACF. Using the backshift operator, MA (q) process is expressed as in equation (2.14). To be invertible, all the solutions to the equation $\theta(G) = 0$, exceeds 1 in absolute value. And,

$\theta(BG) = 1 + \theta_1 G + \theta_2 G^2 + \cdots + \theta_q BG^q$, exceeds 1 in absolute value.

2.4. Box-Jenkins Procedure

This prediction process, commonly referred to as the Box-Jenkins forecasting approach, is based on ARIMA models. Model identification, parameter estimates, diagnostic checks, and model validation are the four processes in the Box-Jenkins technique's model fitting process.

2.4.1. Model Identification

The appropriate member of the ARIMA process is determined by examining the data. This is revealed by the behavior of the ACF and the PACF to identify the model and the order. An AR (p) process' hypothetical PACF terminates after lag p, making the values that follow p insignificantly different from zero. This is summarized in the Figure 2.1.

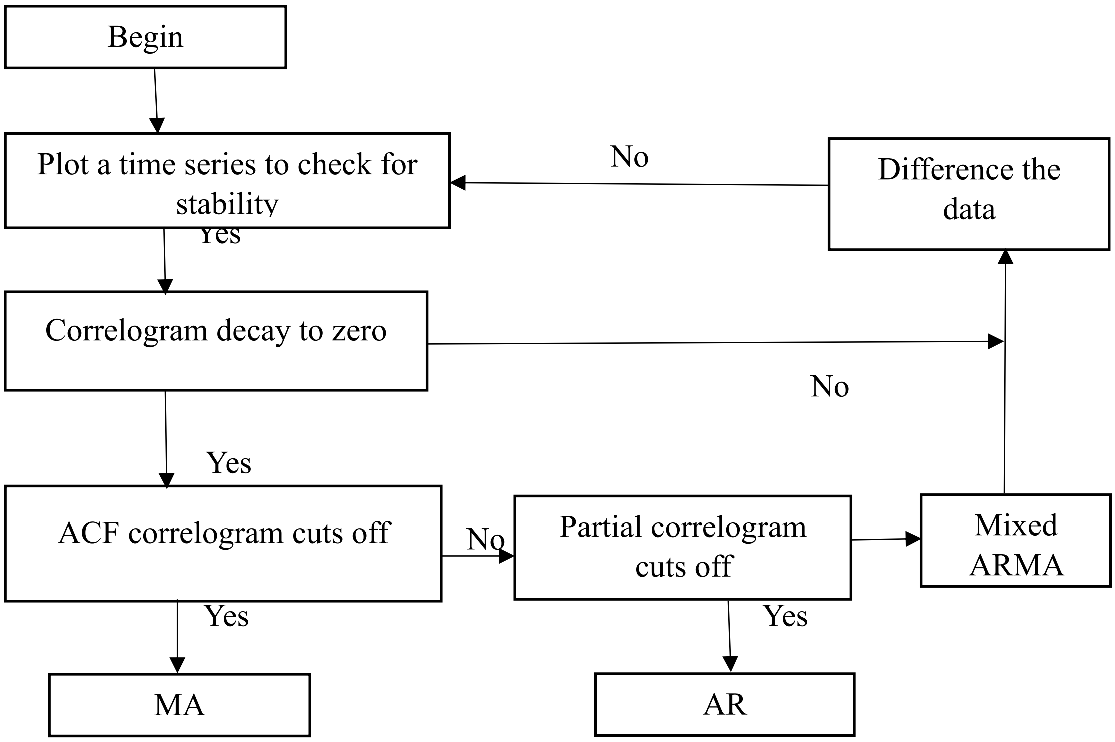

Figure 2.1. Summary of model identification

2.4.2. Estimation of Parameters

After an appropriate model have been chosen, the parameters are estimated. The method of Maximum likelihood provides a unified and a practical approach to parameter estimation for ARMA (p, q) process. The method assumes that the white noise is Gaussian.

2.4.3. Diagnostic Checking

The selected model residuals are examined to assess whether the model is appropriate. This is done by studying the autocorrelation plots of the residual to see if a further structure can be found.

2.4.4. Model Accuracy

The mean square error (MSE) and mean absolute percentage error (MAPE) are measures to access model's projecting reliability. MSE achieves the minimal error accuracy, hence it is commonly used to measure the efficacy and precision of statistical models (Prayudani *et al*., 2019).

2.5. Modeling using Seasonal ARIMA Models

A study was done by Otieno *et al*., (2014) to simulate the demand for hotel rooms among tourists in Kenya. A forecasting model was created using the Box-Jenkins model quarterly statistics on tourist bed use in Kenya between the year 1974 and 2011.

SARIMA (1, 1, 2) $(1, 1, 1)_4$ was selected and used to produce a forecast for the next five years. The investigator advises conducting the same study with monthly data.

For 10-year period data from 2008 to 2017, Borbor *et al*., (2019) modeled historical monthly hospital attendance in Cape coast teaching Hospital for both Insured and Uninsured patients. The preferred models were SARIMA (1,0,0) $(0,1,0)_{12}$ and SARIMA (1,1,1) $(2,0,1)_{12}$ for Insured and Uninsured patients respectively, based on the minimum AIC values of 15.66537 and 13.94181 respectively. The models were used to generate forecast for the respective attendance in future. The research was conducted to ascertain how the Outpatient has impacted on patients' attendance in seeking health care with time using time series analysis.

Borbor, (2020) observed that Outpatient Hospital Attendance instances demonstrated variability of processes produced by several irregular factors that cannot be removed in cases recorded. The study found a candidate model that, on average, best fit the data. The ARIMA (2, 2, 1) model was found to be the most plausible for the Outpatient visits data. There was no seasonal pattern in the number of hospitals visits each month. The result indicated that over the anticipated time period, there was a rise in Outpatients' attendance.

Abu *et al*., (2022) set out a study to fit and forecast patient arrival at the University of Cape Coast Hospital Accident and Emergency Department (AED). The study found that the data on daily patient arrivals at the University hospital's AED had a considerable positive skewness. The study findings identified ARIMA (0, 1, 2) as the most optimal model for day patient visits at the university's hospital AED unit among the forefront twenty-five alternative models carefully picked. The findings revealed that a decreased trend in patients who arrived daily at the University's hospital AED during both the research as well as predicted period.

Kam *et al*., (2010) proposed three models to project daily patient visits to the Emergency Department (ED). (a.) Moving Average MA (2); (b.) Single-variate SARIMA model: SARIMA (1, 0, 1) $(0, 1, 1)_7$; (c.) SARIMA (1,0,2) $(0,1,1)_7$, a multi variate seasonal ARIMA model7. The Univariate and Multivariate Seasonal ARIMA models' residuals are uncorrelated, according to a diagnostic analysis of the models' residuals. Since it may reveal the relative degree of the forecasting error between the predicted and observed values, Mean Absolute Percentage Error (MAPE) was selected as the metric to assess the model's prediction performance. The MAPEs of both SARIMA models are less than 10%, indicating acceptable precision. The ability to adjust seasonal components and incorporate autoregression makes the SARIMA models more accurate than the moving average.

Mohamed & Mohamad, (2020) set out an empirical investigation to identify the best ARIMA model for predicting monthly attendance at orthopaedic clinics. The study made use of monthly data collected at an orthopaedic clinic between January 2013 and June 2018. The stationarity condition of the data was examined using the ACF and PACF graphs. The Box-Jenkins approach was used in the study to model the data. With the lowest mean absolute percentage error, ARIMA (1,0,0) was the most likely model. Mann-Kendall was thus utilized in this investigation to check for trends. Additionally, because SARIMA modelling can model data with both trend and seasonal components, it was used in the study.

Arthur, (2013), conducted a study to predict outpatient department attendance at Salt Pond Municipal Hospital using statistical modelling. The analysis was done using the Box-Jenkins approach. Ten SARIMA models were suggested; a critical analysis using the P-value and Chi-square statistic revealed that three models as the tentative models. SARIMA (1,1,3) $(0,1,1)_{12}$ was judged to be the most plausible of the competing models due to its high p-value and low chi-square value. Further, model diagnostic checks using the PACF and ACF revealed that the residuals are White Noise and the residuals autocorrelations are all zero.

Baharsyah & Nurmalasari, (2020), used ARIMA and ES to predict patient visits to the RSUD Kembangan emergency department. Monthly data from April 2015 through June 2019 were used. The model's forecasting accuracy was assessed using MSE and MAPE. Because ARIMA (1,1,2) had the lowest MSE (22600.3) and MAPE (10.6), it was selected. On the other hand, because the data had a trend component (MSE 26900.6 and MAPE 11.8), exponential smoothing was chosen. ARIMA (1,1,2) was chosen as the most likely model to forecast patient visits to the emergency room of RSUD Kembangan Hospital based on the MSE and MAPE.

Dan, (2014) employed SARIMA modelling to model and predict malaria death rates using time series analysis. From January 1996 to December 2013, the monthly malaria death rate in Imo State was forecast using the Box-Jenkins approach. Estimating the monthly malaria fatality rate for 2014 was the aim of the study. From a group of tentative models, the best model was chosen using the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC). To ascertain whether the residual autocorrelation was zero, the Ljung-Box statistic was employed. The forecasting accuracy of the model was assessed using the mean absolute percentage error (MAPE). The most effective model for malaria mortality rate data was SARIMA (1,1,1) $(0,0,1)_{12}$. MSE was utilized in this study to assess the accuracy of the model predicting.

In China's major hospitals, Luo *et al*., (2017) investigated a prediction method to forecast outpatient visits. In order to estimate short-term daily outpatient visits, they combined a SARIMA model with a Simple Exponential Smoothing (SES) model to evaluate the forecast accuracy for each model, as well as that of a combinatorial model. Consequently, all of the choice simulates are relatively easy to implement and have low computational needs, making them appropriate for quick prediction. When predicting daily outpatient visits one week in advance, the combinatorial model outperforms simple models and extracts more precise information with a limited training sample size.

Ibrahim *et al*., (2016), did a study to fit SARIMA model on Peads patient visits at outpatient Medical Laboratory, Mayo Hospital Laboratory using the Box-Jenkins techniques. The study used quarterly data on patients visits for the period September, 2007 to December, 2013. The ARIMA (1,1,1) and SARIMA (1,1,1) $(1,0,1)_4$ after the diagnostic checks. Model validation found SARIMA model as the most plausible model predicting the expected number of patient visits more accurately since it models the seasonality in the data. The forecasts for the month of January, 2014 were 119 and 92 using ARIMA and SARIMA respectively compared to the actual number of patient visits which is 102.

In Des Moines, Lowa, USA, Choudhury & Urena (2020) set out a model for predicting hourly patient visits at a hospital emergency department. The plausible model was SARIMA (3, 0, 0) $(2, 1, 0)_{24}$. The SARIMA model displayed the highest predicting accuracy, despite the fact that the TBATs and Neural Network models produce acceptable results. The study concluded that the ARIMA model can be used as a decision support tool in the emergency healthcare system. However, the study recommends that future research take a broader perspective and incorporate outside forces such as environment and mobility into the model. Therefore, this study used monthly outpatient attendance data.

A complex mathematical model for forecasting rail passenger flow was put out by Milenković *et al*., (2018). The best suitable SARIMA model for simulating rail passenger demand on Serbian railways is SARIMA (0, 1, 0) $(0, 1, 1)_{12}$. The study developed the model from a monthly data from the Serbian railways.

To forecast the number of people with different needs at the Veterans Health Administration clinic (USA), Al-Haque et al. (2015) fitted a SARIMA model in response to the seasonal demand for medical services from travelling patients. The forecasts aligned with the clinic's past patient demand trends. Based on historical patterns, it was projected that the need would peak in January 2013 with 359 expected patients. The number of patients was expected to have dropped to its lowest level by the summer of 2013 between May and June, which was also consistent with historical trends. When compared to historical average forecasts, the SARIMA model's RMSE was shown to be significantly lower.

According to these studies, SARIMA is a flexible and trustworthy model for outpatient forecasting worldwide, and it has a good chance of being used in Kenya to enhance resource management and healthcare delivery.

CHAPTER THREE
METHODOLOGY

3.1. Location of the Study

The study used data from Marimanti Level 4 Hospital Outpatient department, which is a primary care hospital in, Marimanti Ward, Tharaka Constituency, Tharaka South Sub-County, Tharaka Nithi County (Kenya). Located at latitude 0^0 9’ 31” South; longitude 37^0 58’ 31” East. It is situated in an ASAL area with an estimated terrain elevation of 658 meters above the sea level. Marimanti Level 4 Hospital is appropriate for the study due to the increased catchment population as a result of the establishment of Tharaka University, which has student population increasing steadily.

3.2. Research Design

Time Series Research Design was employed in the study. This method of longitudinal research involves analyzing vast amounts of data collected over time on the same variable. The approach aids in identifying trends in the data set by basically plotting the data (time plot). Additionally, it provides the chance to clean the data by facilitating the easy identification of outliers and missing values, as well as the prediction by fitting the data to a model that best describes the observed time series. All the procedures involve simple computation saving on time and effort to achieve the desired result, thus convenient.

3.3. Data Collection

The research used monthly Outpatients visits data from Marimanti Level 4 Hospital over a period of 10 years (2013-2023). The monthly outpatients’ visits data was used to fit a SARIMA model. The period under study has 132 observation which is sufficient to fit the SARIMA model according to Meyler *et al*. (1998), Chatfield (1996) who argued that 50 observations are sufficient to fit ARIMA model.

3.4. Data Analysis

Statistical Software R and R-Studio was used in the analysis, which is a free source software. The equation of SARIMA model as given in equation (2.18). SARIMA model is appropriate since it fit data with both trend and seasonality, it is accurate in short-term forecasting and its procedures involves simple computations.

3.5. SARIMA Model

3.5.1. Procedure for SARIMA Model

The first step in modeling time series data is to ensure the series is stationary either by transformation of data or differencing. This study employed Augmented Dickey-Fuller (ADF) to test whether the outpatient series is stable and to determine the number of trend and seasonal differencing to achieve stationary.

The Hyndman-Khandakar (HK) method was used with the forecast package in R, (Hyndman & Khandakar, 2008). The approach uses an iterative strategy that saves time and make it simple to identify appropriate model without having to compare it to every other model that might be used (Dabral & Tabing, 2020). Figure 3.1 summarizes SARIMA modeling.

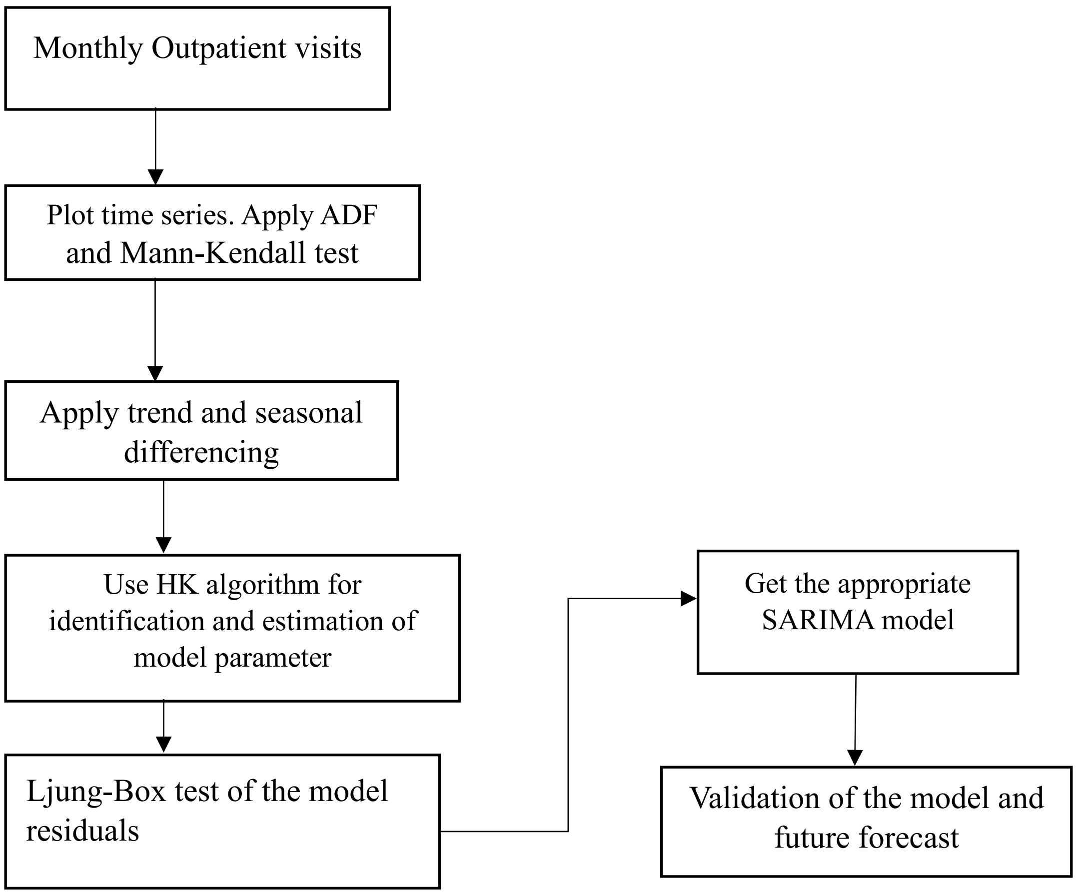

Figure 3.1. Flow chart of SARIMA modeling

3.5.1.1. Test for trend in Time Series Data

The Mann-Kendall (MK) test was used to determine if the outpatient visits series exhibit a monotonic upward or decreasing trend over time. MK, unlike parametric linear regression analysis, does not require the fitted line's residual be normally distributed. Kendall's test ranks all the data by time order, the difference between consecutive values is calculated and the sum of the signs of difference as the Kendall sum, S statistic, (Kendall, 1948). Mann-Kendall test trend component in time series data and has the following hypothesis.

H_0: The monthly outpatient attendance has no trend.

H_1: The monthly outpatient attendance has trend.

The H_0 is rejected if p-value < 5% level of significance (two-tailed), meaning there is a trend in the monthly outpatient attendance

3.5.1.2. Test for Stationarity for Time Series Data

Test for Stationarity is important in understanding the data and selecting the prediction model. This study will test for stationarity using the Augmented Dickey-Fuller (ADF) test. ADF is a unit root test.

ADF test has the following hypothesis.

H_0: The monthly outpatient attendance series non stable.

H_1: The monthly outpatient attendance series is stable.

The H_0 is rejected if p-value < 5% level of significance, meaning the series has no unit root/the series is stationary, (Dickey & Fuller, 1979).

3.5.1.3. Akaike Information Criteria (AIC)

Akaike information criteria was developed by Hirotugu Akaike in 1974. Assuming X_t is an ARMA process, then, the AIC is defined as;

$$AIC = -2\log\left[L\left(\hat{\varphi}, \hat{\theta}\right)\right] + 2k \tag{3.19}$$

Where $L\left(\hat{\varphi}, \hat{\theta}\right)$ is the maximum likelihood function which measures a model fit. k Is the number of parameters; it penalizes overfitting when more terms are added.

Time series model with lowest AIC value is chosen. AIC favors an over fitted model for small samples (Claeskens & Hjort, 2009). Consequently, AICc was generated to address the overfitting for small sample size.

AICc is defined as;

$$AICc = AIC + \frac{2k^2 + 2k}{n - k - 1} \tag{3.20}$$

where n is the number of observation and k is the number of parameters.

Burnham & Anderson. (2016) noted that the extra penalty term converges to zero as the sample size approaches infinity, leading to convergence of AICc to AIC.

3.5.1.4. Bayesian Information Criteria (BIC)

It was developed by Schwarz, (1978). It is defined as;

$$BIC = -2\log[L(\hat{\varphi},\hat{\theta})] + \log(n)\,k \qquad (3.3)$$

Where $\log\ (n)\,k$ is the penalty term.

BIC penalizes model complexity more than AIC. Basing on BIC, the model with the lowest BIC is selected.

3.5.2. Model Diagnostic

The Ljung-Box statistic is a function of the total sample autocorrelation, r_j up to any time lag m. It was developed by Greta Marianne Ljung and George Edward Pelham Box in 1978. The model's residuals were evaluated using the Ljung-Box statistic to determine whether or not all error autocorrelations were zero. The Ljung-Box statistics is defined as;

$$Q(m) = n(n+2)\sum_{j=1}^{m}\frac{{r_j}^{k}}{n-j} \qquad (3.4)$$

where, following differencing, n is the useful number of data points.

Autocorrelations of zero are the optimal ACF for residuals. For the first m lags, a p-value < 0.05 suggests the likelihood of non-zero autocorrelation.

3.5.3. Model Accuracy Evaluation

Mean squared error (MSE) and root Mean absolute percentage error (MAPE) values was used to provide insight into the model’s performance with the goal of assessing forecasting accuracy and make adjustments as deemed.

3.6. Ethical Consideration.

To introduce the researcher to the Marimanti Level 4 Hospital Authorities, a letter of introduction was requested from Tharaka University. The Tharaka University Research and Ethics Committee was consulted for ethics permission and clearance. The National Commission of Science, Technology, and Innovation (NACOSTI) in Kenya was then consulted for a research permit, which is included in the appendix. All information sources used in this study were cited.

CHAPTER FOUR
RESULTS AND DISCUSSIONS

4.1. Introduction

This chapter discusses the results and the findings of the analysis in line with the objectives of the study.

4.2. Descriptive Statistics

Table 4.1 presents the summary statistics for the outpatient visits series. The significant difference between the highest and lowest number of outpatient visits supports an upward trend. The negative value of skewness indicates that the data has a left tail, whereas the small value suggests that the data series is relatively symmetric about the mean. The negative kurtosis value indicates that the observations are not following a normal distribution.

Table 4.1. Descriptive statistics of the series

Descriptive	**Statistics**
N	132
Mean	2563.17
Standard deviation	215.47
Skewness	-0.23
Kurtosis	-0.61
Maximum	3000
Minimum	2000

4.3. The trend of Outpatient Attendance at Marimanti Level 4 Hospital

To start time series modelling, the series needs to be stationary. A time series is deemed stable if there is no discernible variation in either the mean (trend) and the variance (season). The stationarity of the series is ascertained using the ACF and PACF correlograms, the unit root test (ADF test), the time plot, and series decomposition. The series is made stationary by taking the difference.

4.3.1 Time Plot

Figure 4.1, illustrates that the outpatient visit data fluctuates continuously during the study period, with noteworthy peaks and troughs occurring. The low volume of outpatient visits in 2013 can be related to the shift from a national to a devolved government. Outpatient attendance was low during 2016 and 2017, due to hospitalist strikes throughout the same time period. In general, outpatient visits have increased over time, with the lowest number in 2013 at 2000 and the highest in 2023 at 3000.

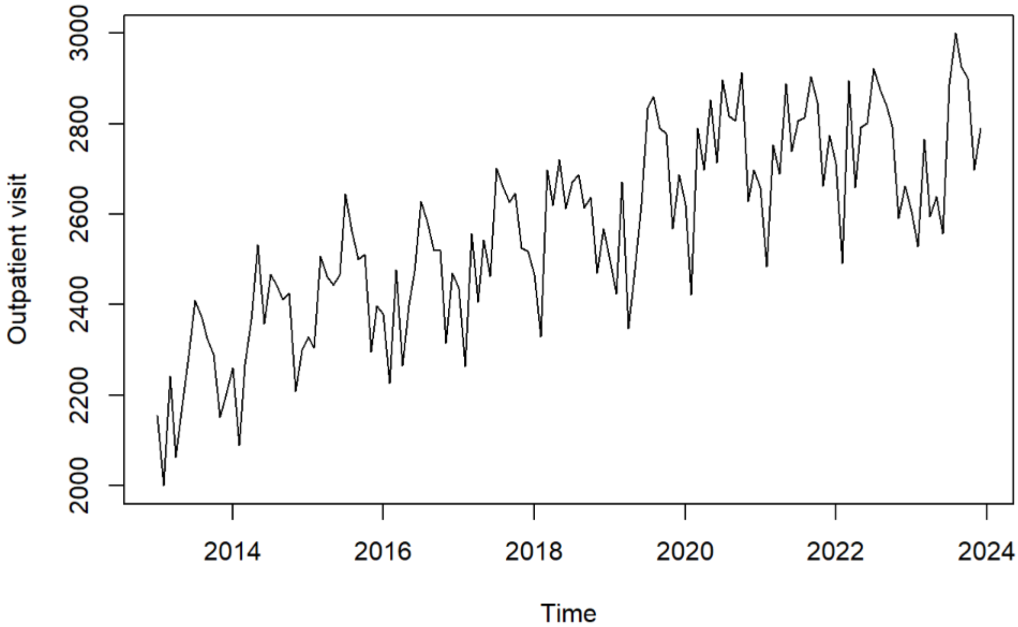

Figure 4.1. The time plot of the outpatient visits at Marimanti level 4 hospital

4.3.2 ACF and PACF Correlogram

ACF and PACF plots are critical checking if the data series is stationary. Randomness is checked using the plot; if it is, the autocorrelation should be near zero.

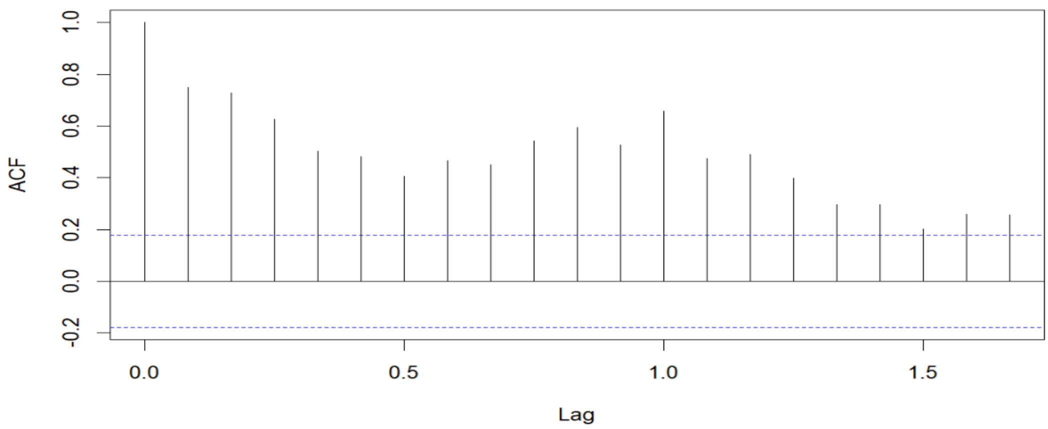

Figure 4.2. ACF correlogram

From figure 4.2, the correlogram of the outpatient visits series gradually declines, suggesting correlation of the past values and the current values. This indicates that the series is non-stationary. The partial autocorrelation function (PACF) shows spikes at the lower lag which exceeds the significant bound as shown in Figure 4.3.

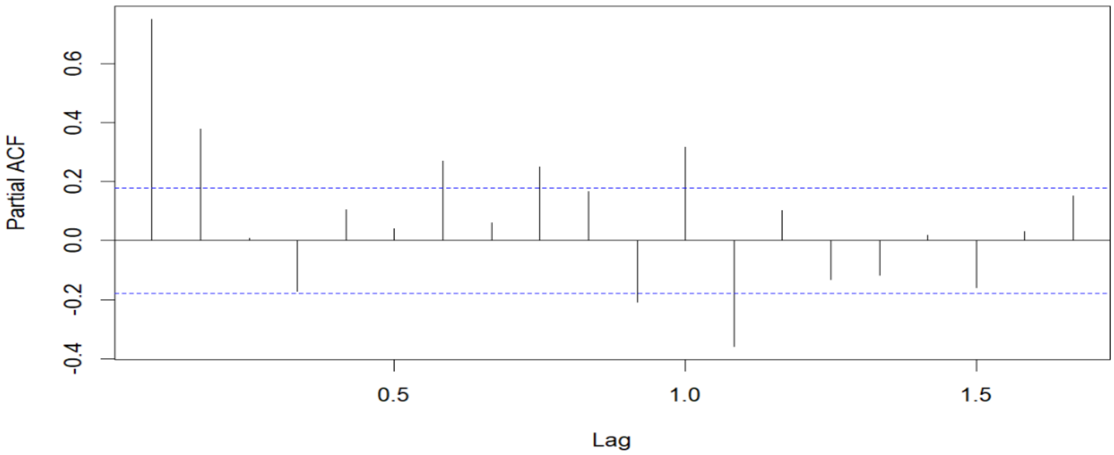

Figure 4.3. PACF correlogram

4.3.3. Augmented Dickey Fuller (ADF) Test

The outpatient visits series' stationarity was examined using the ADF test. The outpatient visit series is non-stationary is supported by the p value of 0.4134, which is higher than the significance level of 0.05. Additionally, trend analysis utilizing the Mann Kendal test verified that the outpatient visits have a positive trend. The series requires differencing since the p-value is $2.22e^{-16}$, which is below the 5% standard of significance. The current study also used the ADF unit root test and Mann Kendal test to check for stationarity and trend, in contrast to the study of Mohammed & Mohamad (2020), which used the ACF and PACF plots to check the stationarity of the series.

4.3.4. Series Decomposition

Breaking down a series in order to separate its constituent parts is known as series decomposition. In accordance with research by Arthur (2003), Kam et al. (2010), Otieno et al. (2014), and Borbor et al. (2019), which found the SARIMA model suitable for predicting patient attendance, the decomposition of the outpatient visits series reveals that the series is composed of trend, periodic, and random components, as illustrated in Figure 4.4. As a result, modelling the series using the SARIMA model is appropriate. January through December were the months for which the seasonal influence was investigated. The number of outpatient visits is expected to rise in July and fall in February, with the biggest periodicity occurring in July (about 147) and the lowest occurring in February (around -224).

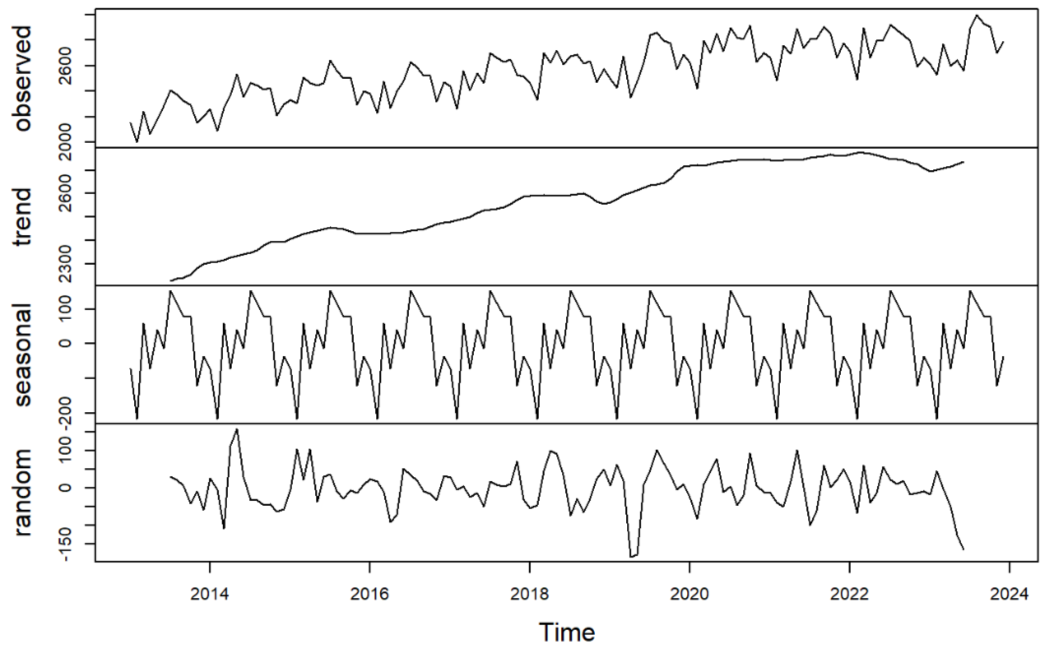

Figure 4.4. Decomposition of outpatients' visit

4.3.5. Differencing the Outpatient Visits Series

After taking the first seasonal difference, Figure 4.5 indicates some significant trends in the series; so, the first trend difference is taken in order to eliminate trends in the series. The absence of a discernible trend in the outpatient visit series across time indicates that the series is stable (see Figure 4.6). The first differenced series was subjected to trend analysis using the Mann Kendall test; the p-value, which is higher than the 5% level of significance, is 0.85897, indicating that the series is trend stationary following the first difference. Additionally, an ADF test was conducted to see whether the series had a unit root following the initial change. The test yielded a p-value of 0.01 below the 0.05 level of significance, indicating that the null hypothesis of non-stationarity is rejected and the series is prepared for modelling.

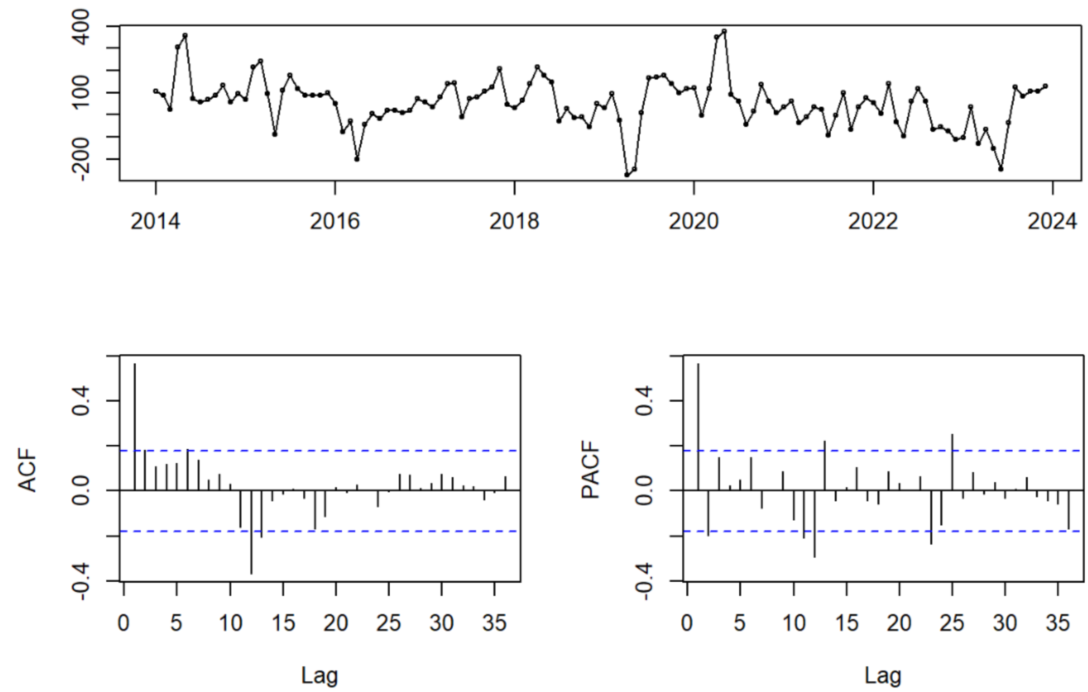

Figure 4.5. Graph of first seasonal differenced data

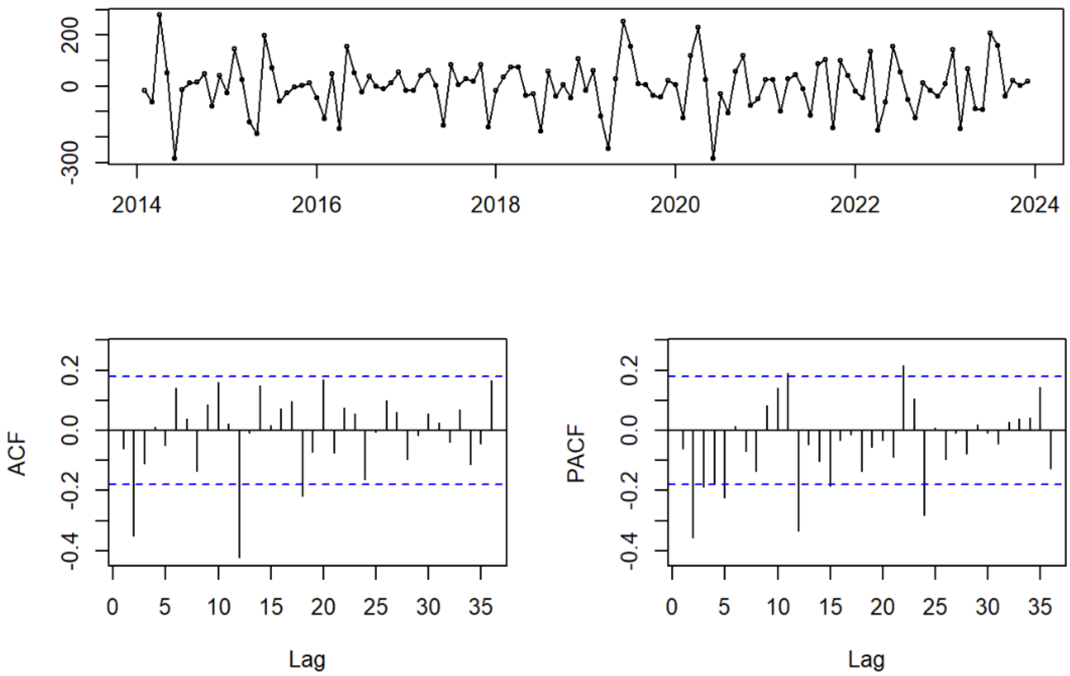

Figure 4.6. Graph of trend differenced outpatient visits

4.4. Fitting SARIMA Model using Outpatient Attendance Data

Once the series is stationary, modeling follows. In SARIMA model development, Box-Jenkins approach was used which involves; model identification, parameter estimation and diagnostic checking.

4.4.1. Model Identification

The objective was to use the ACF and PACF of the different outpatient series in Figure 4.6 to generate a suitable SARIMA model. A non-seasonal MA is suggested by the ACF's notable rise at lag 2 (2). A seasonal MA is suggested by the ACF's notable rise at lag 12 (1). Therefore, the seasonal ARIMA (0,1,2) $(0,1,1)_{12}$ model was our starting point. The Seasonal ARIMA (0,1,2) $(2,1,1)_{12}$ model was found using the PACF plot, in which the seasonal component of the model was chosen using the ACF and the non-seasonal part using the PACF. Among the competing set of models, the model with the lowest AIC and BIC values is regarded as the best model. Table 4.2 shows the set of competing SARIMA models alongside their AIC and BIC.

Table 4.2. Set of competing SARIMA models

Model	AIC	BIC
SARIMA (0,1,1) $(1,1,2)_{12}$	1150.63	1163.34
SARIMA (0,1,1) $(2,1,1)_{12}$	1142.71	1155.42
SARIMA (0,1,2) $(1,1,2)_{12}$	1147.44	1162.7
SARIMA (0,1,2) $(1,1,1)_{12}$	1173.52	1188.78
SARIMA (0,1,2) $(2,1,1)_{12}$	1139.56	1154.82
SARIMA (0,1,2) $(2,1,0)_{12}$	1170.16	1182.87
SARIMA (1,1,1) $(2,1,2)_{12}$	1176.86	1192.12

The most likely model for outpatient visits at Marimanti Level 4 Hospital with the lowest AIC and BIC values was found to be SARIMA (0,1,2) $(2,1,1)_{12}$.

4.4.2. Model Parameter Estimation

Using the Box-Jenkins methodology and time series modelling, the method of maximum likelihood was used to estimate the model parameters. Equation 2.18 illustrates the Sarima model's generic form. SARIMA (0,1,2) (2,1,1)12 is the most effective model for predicting outpatient visits at Marimanti Level 4 Hospital. The estimated coefficients are shown in Table 4.3.

Table 4.3. Table of coefficients

Variable	Coefficient	Standard Error
MA (1) [θ_1]	-0.3965	0.0926
MA (2) [θ_2]	-0.3456	0.0926
SAR (1) [β_1]	-0.5925	0.1512
SAR (2) [β_2]	-0.4579	0.1148
SMA (1) [Θ_1]	-0.3407	0.1645

Further, Table 4.3 shows that θ_1 and θ_2 are negatives, this indicate that the fraction of the shock two period ago is still felt in the current period. β_1 and $\beta_2$are the seasonal AR coefficients, these suggest that the value of the outpatient visits at time t is strongly impacted by the number of outpatient visits two periods ago that is, values of period 12 and 24. The negative shows an inverse association. $\Theta_1$is the coefficient of the seasonal MA, this shows the impact of the error term from the same season in the previous year on the current number of outpatient visits. The negative value suggests that a positive shock one year ago reduces the current number of outpatient visits by approximately 34.07% of the shock's size.

The SARIMA (0,1,2) $(2,1,1)_{12}$ is given as;

$$(1 + 0.5925B^{12} + 0.4579B^{24})(1 - B)^d(1 - B^S)^D y_t = (1 + 0.3965B + 0.3456B^2 + 0.3407B^{12} + 0.1351B^{13} + 0.1177B^{14})\varepsilon_t \quad (4.1)$$

where, $\Delta_{12}y_t = y_t - y_{t-12}$(Seasonal difference), $\Delta y_t = y_t - y_{t-1}$ (non-seasonal differencing) and y_t is the number of outpatient visits at time t

4.4.3. Diagnostic test and Model Accuracy Evaluation

The residues of the model are checked to determine if it follows normal distribution and are uncorrelated. These residuals are the difference between the observed values of the series and the forecasted values. Figure 4.7 shows that residual series is a sequence of White Noise since the time plot has no patterns, further, the autocorrelations for first 36 lags all lie within the significant bound. Therefore, the residues are uncorrelated and follow normal distribution as shown by the normality plot. This suggest that the residues are sequence of White Noise with mean zero and a constant variance which is an ideal condition for the model's residuals.

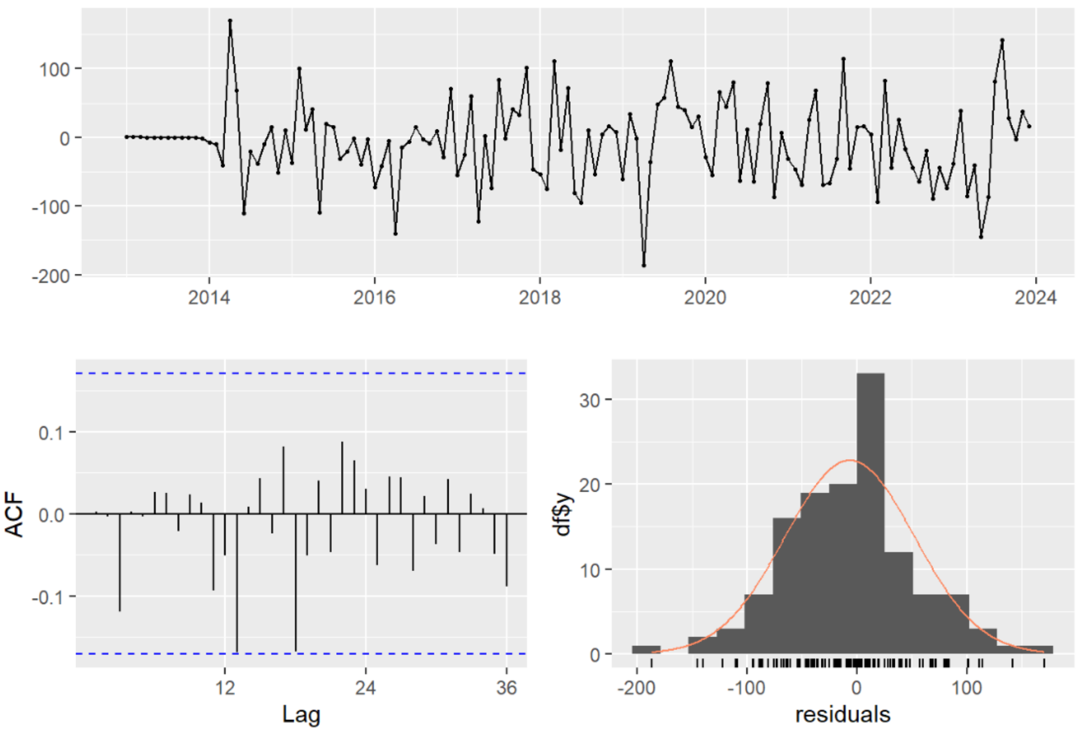

Figure 4.7. Plot of residuals from SARIMA (0,1,2) $(2,1,1)_{12}$ model

Further diagnostic using the Ljung-Box test confirmed that the model's residuals autocorrelation is zero since the test p-value is 0.9409 which is greater than the 5% level of significance. The model's forecasting accuracy was evaluated using the MAPE and the MASE. SARIMA (0,1,2) $(2,1,1)_{12}$ has the least MAPE = 1.664% and MASE = 0.46% and therefore, deemed to accurately forecast the outpatient visit at Marimanti level 4 hospital. This is in agreement with Kam et al., (2010), who assert that a MAPE less than 10% indicate a good level of accuracy and better than Baharsyah & Nurmalasari, (2020), ARIMA model with MAPE = 10.6%.

4.5. Forecast the Outpatient Attendance for the Next Two Year

The graph in Figure 4.8 shows that the number of outpatient visits at Marimanti Level 4 hospital continues to rise in the forecasted period. The forecasts strongly agree with the observed outpatient visits series pattern.

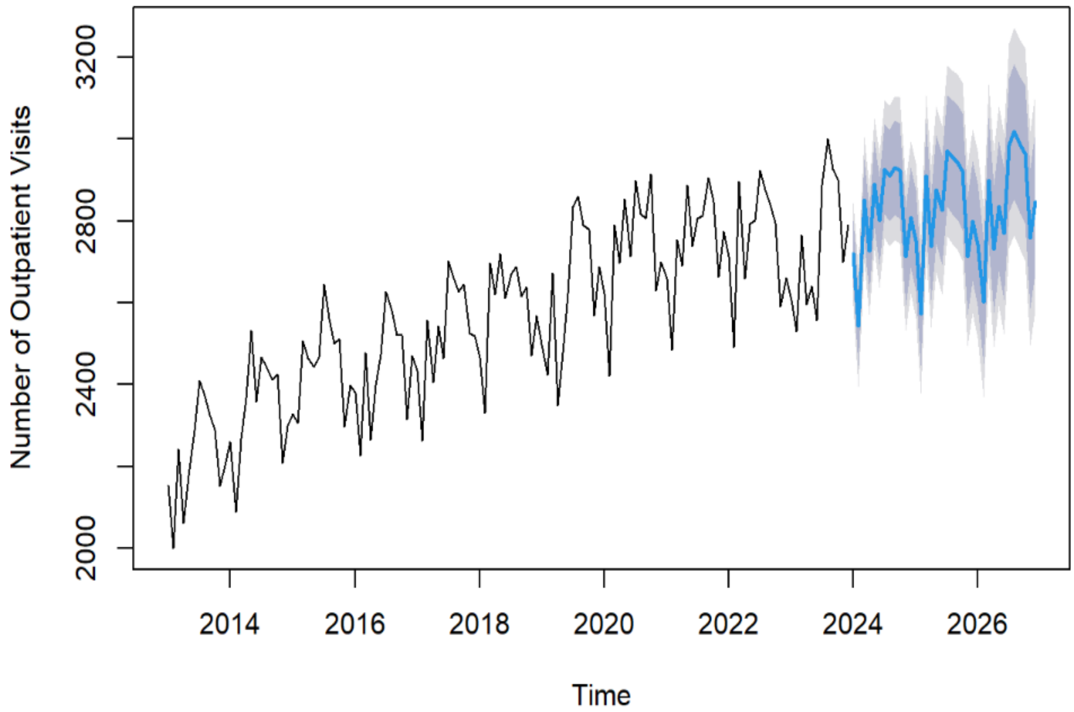

Figure 4.8. Forecast graph

Table 4.4 shows the forecasted monthly outpatient visits at Marimanti level 4 hospital for the year 2025 and 2026 together with their 80% and 95% confident interval.

Table 4.4. Table of forecasts two years ahead

Time	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2025					
January	2762	2621	2904	2546	2978
February	2570	2426	2714	2349	2790
March	2927	2780	3073	2703	3151
April	2752	2603	2900	2524	2979
May	2903	2752	3053	2672	3133
June	2856	2703	3009	2623	3090
July	2980	2825	3134	2743	3216
August	2945	2788	3102	2705	3185
September	2940	2782	3099	2698	3183
October	2919	2758	3080	2673	3165
November	2708	2545	2871	2459	2957
December	2791	2626	2956	2539	3043
2026					
January	2759	2581	2937	2487	3031
February	2591	2406	2776	2308	2874
March	2930	2742	3119	2642	3219
April	2779	2588	2971	2486	3073
May	2933	2738	3129	2635	3232
June	2866	2668	3065	2563	3170
July	3005	2803	3207	2697	3314
August	2970	2765	3175	2657	3284
September	2968	2760	3176	2649	3286
October	2976	2765	3188	2654	3299
November	2752	2538	2967	2425	3080
December	2838	2621	3056	2506	3171

CHAPTER FIVE
SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

5.1. Summary

The study fitted a statistical model to forecast the outpatient visits at Marimanti level 4 hospital using SARIMA modeling. The study used time series research design in modeling the Outpatient visits series using Box-Jenkins techniques. The study used Monthly outpatient data from Marimanti Level 4 Hospital covering 11 years (January, 2013 - December, 2023).

The study found out that the outpatient visit series was seasonal, therefore, the series had trend, seasonal and random variations. The month of February and July had the lowest and highest seasonal effect respective. The study found that the outpatient visits series had trend and strong positive correlation using Mann Kendall test which gave tau = 0.617 and a 2-sided P value = $2.22e^{-16}$, thus the outpatient series was non-stationary. First order seasonal and non-seasonal difference rendered the series stationary.

Basing on AIC and BIC of the set of competing models, three tentative SARIMA models were chosen, further diagnostic check confirmed SARIMA $(0,1,2)\ (2,1,1)_{12}$ as the most plausible model to describe the outpatient visits at Marimanti level 4 hospital having the least MAPE and MASE of 1.664% and 0.46% respectively.

The fitted SARIMA $(0,1,2)\ (2,1,1)_{12}$ model was used to produce forecasts of monthly outpatient visits at Marimanti level 4 hospital two years ahead. The forecasts indicated an increasing trend in the number of outpatient visits at Marimanti Level 4 hospital for the forecasted period.

5.2. Conclusions

The Marimanti Level 4 Hospital's outpatient visits is seasonal. The data became stable with the use of first order seasonal and non-seasonal differencing. The hospital outpatient department sees a high number of outpatient visits in July, August, September, and October while a low numbers in November and a seasonal low in February

SARIMA $(0,1,2)\ (2,1,1)_{12}$ is the most plausible model to describe the patient attendance to Marimanti level 4 hospital Outpatient department of the three tentative models having the least AIC and BIC and least forecasting error.

The volume of outpatient visits at Marimanti Level 4 Hospital is expected to rise during the next two years. To guarantee proper healthcare delivery, hospital authorities should base their planning efforts on the forecasts.

5.3. Recommendations

The study recommends the following;

I. The Management of Marimanti level 4 hospital should ensure availability of adequate supplies and human resource in from the month of June since the number of outpatient visits surges between June and October.

II. To increase the forecast's precision and dependability, data should be gathered and analyzed continuously. This guarantees that the model is kept up to date and that it recognizes any new patterns in the data.

III. The Hospital Management should rely on the forecasts for planning purposes in terms of resource allocation, workload scheduling and staffing to guarantee increased patient safety and satisfaction in service delivery.

5.4. Suggestions for Further Research

The current study looked at the number of patients visiting the hospital's OPD, further research to be done using machine learning models to give more insight to the outpatient visits including the disease prevalence and predisposing factors.

REFERENCE

Aboagye-Sarfo, P., Mai, Q., Sanfilippo, F. M., Preen, D. B., Stewart, L. M., & Fatovich, D. M. (2015). A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in Western Australia. *Journal of Biomedical Informatics*, *57*, 62–73. https://doi.org/10.1016/j.jbi.2015.06.022

Abu, B. M., Kojo Prah, J., Walker, E., Obed, L., Banson, C., & Pinkrah, R. (2022). Modelling and Predicting Daily Arrival of Patients at the Accident and Emergency Department of the University of Cape Coast Hospital, Ghana. *International Journal of Medical Science and Health Research*, *06*(02), 10–25. https://doi.org/10.51505/ijmshr.2022.6202

Al-Haque, S., Ceyhan, M. E., Chan, S. H., & Nightingale, D. J. (2015). Responding to traveling patients' seasonal demand for health care services. *Military Medicine*, *180*(1),111-117. https://doi.org/10.7205/MILMED-D-14-00193

Arthur, J. (2013). Application of statistical models to Outpatient Department (OPD) attendance data in Saltpond Municipal hospital of the Central region of Ghana. Thesis, Kwame Nkrumah University of Science and Technology. Pp 68-69.

Baharsyah, N., & Nurmalasari, M. (2020). *Patient Visit Forecasting at Emergency Department using Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing Method in RSUD Kembangan*. *February*, 234–239. https://doi.org/10.5220/0009590302340239

Bergs, J., Heerinckx, P., & Verelst, S. (2014). Knowing what to expect, forecasting monthly emergency department visits: A time-series analysis. *International Emergency Nursing*, *22*(2), 112–115. https://doi.org/10.1016/j.ienj.2013.08.001

Borbor, B. S. (2020). Modelling and Prediction of Outpatients Department on Hospital Attendance at the Cape Coast Teaching Hospital Using the Box-Jerkins ARIMA Model. *Journal of Advances in Mathematics and Computer Science*, *35*(8), 1–12. https://doi.org/10.9734/jamcs/2020/v35i830309

Borbor, S. B., Senyefia, B.-A., & Gbormittah, D. (2019). Statistical Analysis of Health Insurance and Cash and Carry Systems in Cape Coast Teaching Hospital of Ghana. *Science Journal of Applied Mathematics and Statistics*, *7*(3), 36. https://doi.org/10.11648/j.sjams.20190703.12.

Burnham, K. P., & Anderson, D. R., (2016). *A*voiding Pitfalls When Using Information-Theoretic Methods. *Journal of Wildlife management* 3803155 Accessed: 31-03-2016 14: 20 UTC Y. 66(3), 912–918. URL: http://www.jstor.org/stable/3803155

Chatfield, C. (1996). The Analysis of Time Series. An introduction 5th Edition, Chapman and Hall, London.

Choudhury, A. & Urena, E. (2020). Forecasting hourly emergency department arrival using time series analysis. *British Journal of Healthcare Management*, 26(1), 34-43. https://doi.org/10.12968/bjhc.2019.0067

Claeskens, G., & Hjort, N. L. (2008). Model selection and model averaging. *Cambridge books*.

Dabral, P. P. & Murry, M. Z. (2017). Modelling and forecasting of rainfall time series using SARIMA; *Environmental Processes*, 4(2), 399-419. Https://dio.org/10.1007/540710-017-0226-y

Dabral, P. P., & Tabing, I. (2020). Modelling and Forecasting of Monthly Rainfall and Temperature Time Series Using SARIMA for Trend Detection- A Case Study of Umiam, Meghalaya (India). *International Journal of Environment and Climate Change*, *10*(11), 155–172. https://doi.org/10.9734/ijecc/2020/v10i1130276

Dan, D. E. (2014). Modelling and Forecasting Malaria Mortality Rate using SARIMA Models (A Case Study of Aboh Mbaise General Hospital, Imo State Nigeria). *Science Journal of Applied Mathematics and Statistics*, *2*(1), 31. https://doi.org/10.11648/j.sjams.20140201.15

World Health Organization. (2018). *Delivering quality health services: a global imperative for universal health coverage*. World Health Organization.

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of the American Statistical Association. In *Journal of the American Statistical Association* (Vol. 74, Issue 366, pp. 427–431). http://links.jstor.org/sici?cisi=0162-1459%28197906%2974%3A366%3C427%3ADOTEFA%3E2.0.CO%3B2-3

Hussien, H. H., Eissa, F. H., & Awadalla, K. E. (2017). Statistical methods for predicting malaria incidences using data from Sudan. *Malaria research and treatment.* https://doi.org/10.1155/2017/4205957

Hyndman R. J., & Athanasopoulos G. (2018). *Forecasting: Principles and Practice,* (2nd edition), Otext: Melbourne, Australia. https://otexts.com/fpp2

Hyndman, R. J., & Khandakar, Y. (2008). Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software*, *27*(3), 22. https://doi.org/10.18637/jss.v027.i03

Ibrahim, M., Jamil, M., & Akbar, S. (2016). The fitting of SARIM model on peads patients coming at Outpatient Medical Laboratory (OPML), Mayo hospital, Lahore. *Journal of University Medical and Dental College*, 7(1), 12-18

Jonathan, D. C., & Kung-Sik, C. (2008). Time series analysis with applications in R. http://dspace.agu.edu.vn:8080/handle/agu_library/13381

Kadri F., Harrou F., Chaabane S., & Tahon C. (2014). Time series modeling and forecasting of emergency department overcrowding. *Journal of Medical System*. 38, 1-20. https://doi.org/10.1007/s10916-014-0107-0

Kam, H. J., Sung, J. O., & Park, R. W. (2010). Prediction of daily patient numbers for a regional emergency medical center using time series analysis. *Healthcare Informatics Research*, *16*(3), 158–165. https://doi.org/10.4258/hir.2010.16.3.158

Kendall, M. G. (1948). Rank correlation methods

KNBS. (2022). 2019 Kenya Population and Housing Census: Analytical report on fertility and nuptiality. (vol. vi pp. 43-46). https://www.knbs.or.ke

KNBS. (2022). 2019 Kenya Population and Housing Census: Population by County and Subcounty. (vol. I, PP. 7-10). https://www.knbs.or.ke

Ljung, G. M., & Box, G. E. P. (1978). On a measure of a lack of fit in time series models. *Biometrika*, 65(2), 297-303. https://doi.org/10.1093/biomet/65.2.297

Luo, L., Luo, L., Zhang, X., & He, X. (2017). Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC health services research,* 17, 1–13. https://doi.org/10.1186/s12913-017-2407-9

Meyler, A., Kenny, G., & Quinn, T. (1998). Forecasting Irish inflation using ARIMA models. https://mpra.ub.uni-muenchen.de/id/eprint/11359

Milenković, M., Švadlenka, L., Melichar, V., Bojović, N., & Avramović, Z. (2018). SARIMA modelling approach for railway passenger flow forecasting. *Transport*, *33*(5), 1113–1120. https://doi.org/10.3846/16484142.2016.1139623

Mohamed, B., & Mohamad, M. (2020). Forecasting patient admission in orthopedic clinic at a hospital in Kuantan using autoregressive integrated moving average (ARIMA) models. *Journal of Physics: Conference Series*, *1529*(5). https://doi.org/10.1088/1742-6596/1529/5/052090

Nyamao, R. (2014). Autoregressive integrated moving average (SARIMA) to model and forecast water demand in Kisumu.

Otieno, G., Mung'atu, J., & Orwa, G. (2014). Time series modeling of tourist accommodation demand in Kenya. *Mathematical Theory and Modeling*, *4*(10), 106–115. www.iiste.org.

Permanasari, A. E., Hidayah, I., & Bustoni, I. A. (2013). SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence. *Proceedings - 2013 International Conference on Information Technology and Electrical Engineering: "Intelligent and Green Technologies for Sustainable Development", ICITEE 2013*, *2*, 203–207. https://doi.org/10.1109/ICITEED.2013.6676239

Prayudani, S., Hizriadi, A., Lase, Y. Y., Fatmi, Y., & Al-Khowarizmi. (2019). Analysis Accuracy of Forecasting Measurement Technique on Random K-Nearest Neighbor (RKNN) Using MAPE and MSE. *Journal of Physics: Conference Series*, *1361*(1). https://doi.org/10.1088/1742-6596/1361/1/012089.

Sai P., G., Molugu, N., Kammari, P., Vadapalli, R., & Das, A. V. (2021). Forecast of outpatient visits to a tertiary eyecare network in India using the EyeSmart electronic medical record system. In *Healthcare* (Vol. 9, No. 6, p. 749). MDPI. https://doi.org/10.3390/healthcare9060749

Slawomirski, L., Auraaen, A., & Klazinga, N. (2017). The economics of patient safety: strengthening a value-based approach to reducing patient harm at national level. OECD Health Working Paper No. 96. Paris: *Organization for Economic Cooperation and Development*. https://doi.org/10.1787/5a9858cd-en

Velicer, W. F., & Molenaar, P. C. (2012). Time Series Analysis for Psychological Research. In *Handbook of Psychology, Second Edition* (Issue April 2003). https://doi.org/10.1002/9781118133880.hop202022

OECD, World Health Organization and World Bank Group (2018). Delivering Quality Health Services: A Global Imperative. Switzerland: World Health Organization. https://books.google.co.ke/books?id=LCpjDwAAQBAJ

APPENDICES

Appendix 1: Institutional Scientific Research Ethic Committee Letter

THARAKA UNIVERSITY

P.O BOX 193-60215.
MARIMANTI. KENYA

Telephone: +(254)-0202008549
Website: https://tharaka.ac.ke
Social Media: tharakauni
Email: info@tharaka.ac.ke

INSTITUTIONAL SCIENTIFIC RESEARCH ETHIC COMMITTEE

21st July, 2024

REF: TUN/TUN-ISERC/NSEC/M014

Dear, Kenneth Kipchoge Rono

RE: Forecasting Outpatient Visits at Marimanti Level 4 Hospital: Time Series Analysis using Sarima Model

This is to inform you that ***Tharaka University ISERC*** has reviewed and approved your above research proposal. Your application approval number is ***ISERC04023***. The approval period is ***271st July, 2024-21st July, 2025.***

This approval is subject to compliance with the following requirements;

i. Only approved documents including (informed consents, study instruments, MTA) will be used.
ii. All changes including (amendments, deviations and violations) are submitted for review and approval by ***Tharaka University ISERC.***
iii. Death and life-threatening problems and serious adverse events or unexpected adverse events whether related or unrelated to the study must be reported to ***Tharaka University ISERC*** within 72 hours of notification.
iv. Any changes, anticipated or otherwise that may increase the risks or affected safety or welfare of study participants and others or affect the integrity of the research must be reported to ***Tharaka University ISERC*** within 72 hours.
v. Clearance for export of biological specimens must be obtained from relevant institutions.
vi. Submission of a request for renewal of approval at least 60 days prior to expiry of the approval period. Attach a comprehensive progress report to support the renewal.
vii. Submission of an executive summary report within 90 days upon completion of the study to ***Tharaka University ISERC.***

Prior to commencing your study, you will be expected to obtain a research licence from National Commission for Science, Technology and Innovation (NACOSTI). https://research-portal.nacosti.go.ke and also obtain other clearances needed.

Your Sincerely,

Dr. Fidelis Ngugi, Ph.D.
Chair, ISERC Tharaka University

Appendix 2: NACOSTI Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: **769555**

Date of Issue: **16/July/2024**

RESEARCH LICENSE

This is to Certify that Mr.. KENNETH KIPCHOGE RONO of Tharaka University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Tharaka-Nithi on the topic: FORECASTING OUTPATIENT VISITS AT MARIMANTI LEVEL 4 HOSPITAL: TIME SERIES ANALYSIS USING SARIMA MODEL for the period ending : 16/July/2025.

License No: **NACOSTI/P/24/37776**

769555

Applicant Identification Number

Director General

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix 3: Introduction Letter

THARAKA UNIVERSITY

P.O BOX 193-60215,
MARIMANTI, KENYA

Telephone: +(254)-0202008549
Website: https://tharaka.ac.ke
Social Media: tharakauni
Email: info@tharaka.ac.ke

Education for Freedom

OFFICE OF THE DIRECTOR
BOARD OF POSTGRADUATE STUDIES

REF: TUN/BPGS/PL/03/24

5th June, 2024

To Whom It May Concern

Dear Sir/Madam,

RE: KENNETH KIPCHOGE RONO ADMISSION NUMBER SMT8/05469/21

Mr. Kenneth Kipchoge Rono is a postgraduate student at Tharaka University undertaking a Master degree in **Applied Statistics.** The student has completed his coursework and is expected to proceed for collection of data after successfully defending his proposal at faculty level. The title of the study is "***Forecasting Outpatient Visit at Marimanti Level 4 Hospital: Time Series Analysis Using Sarima Model.***" The proposed study will be carried out in **Tharaka Nithi County.**

Any assistance accorded to him will be highly appreciated.

Thank you in advance.

Yours faithfully,

Dr. Marciano Mutiga, Ph.D.
Director, Board of Postgraduate Studies.

Appendix 4 Tharaka Nithi County Map