YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS
MASTER OF APPLIED STATISTICS PROGRAMME

FORECASTING OF RAINFALL IN CENTRAL DRY ZONE IN MYANMAR USING SARIMA MODEL

Thiri Aung
MAS-51

DECEMBER, 2019

YANGON UNIVERSITY OF ECONOMICS
DEPARTMENT OF STATISTICS
MASTER OF APPLIED STATISTICS PROGRAMME

FORECASTING OF RAINFALL IN CENTRAL DRY ZONE IN MYANMAR USING SARIMA MODEL

This thesis is submitted to the Board of Examination as partial fulfilment of the requirement for the Degree of Master of Applied Statistics

Approved by Board of Examiners

Supervised by	Submitted by
Prof. Dr. Maw Maw Khin	Thiri Aung
Professor/Head	Roll No. (51)
Department of Statistics	MAS. (Batch-1)
Yangon University of Economics	Yangon University of Economics

December, 2019

ACCEPTANCE

Accepted by the Board of Examiners of the Department of Statistics, Yangon University of Economics in partial fulfillment for the requirement of the Master degree, Master of Applied Statistics.

…………………………………………………

(Chairman)

Prof. Dr Tin Win

Rector

Yangon University of Economics

..

(Chief Examiner)

Prof. Dr. Maw Maw Khin

Professor/Head

Department of Statistics

Yangon University of Economics

………………………………..

(Examiner)

Prof. Dr. Mya Thandar

Professor

Department of Statistics

Yangon University of Economics

………………………………..

(External Examiner)

Daw Aye Aye Than

Associate Professor (Retd.)

Department of Statistics

Yangon University of Economic

December, 2019

ABSTRACT

Rainfall is one of the most important sources of water on earth supporting the existence of the majority of living organisms. A time series analysis, modeling and forecasting constitutes a tool of importance with reference to a wide range of scientific purposes in Meteorology. Examples is precipitation, humidity, temperature, solar radiation, floods and draught. This study research applies the Box-Jenkins methodology, employing SARIMA (Seasonal Autoregressive Integrated Moving Average) model to perform modeling past rainfall time series components structure and predicting further rainfall in according to the past. This methodology are capable of representing stational as well as nonstationary time series. The model is mostly fit to both show the past rainfall data and thus generate the most reliable further forecasted is selected by the R2, RMSE and BIC for model evaluation criteria. This paper explores the application of the Box-Jenkins approach to Rainfall data series in Chauk and Shwebo Townships in Myanmar. As the statistical characteristic for stochastic seasonal models are also investigated and the fitted model for monthly rainfall data series in Chauk and Shwebo Township in Central Dry Zone of Myanmar from January, 2006 to December, 2018. The monthly rainfall date are found by the following the four states of model building, identification, estimation, diagnostic checking and forecasting. The statistical package, SPSS software and EViews software was used to build models for the above two series.

ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude to Professor Dr Tin Win, Rector of Yangon University of Economics, for allowing me to pursue this study, which is essential in the course of Master of Applied Statistics (MAS).

I wish to express my very special thanks must go to Professor Dr Maw Maw Khin, Head of Statistics Department, Yangon University of Economics, not only as a Program Director, but also as a supervisor for her encouragement, invaluable input, comments and suggestion throughout this survey steering me in the right direction.

I also would like to express my special thanks to all professors, associate professors and lectures for imparting of a great variety knowledge and concept of development, their expert opinion and suggestion during the study period of two years under MAS (1st Batch) Programme of Yangon University of Economics. I would like to express my thanks to all library staff from Yangon University of Economics.

It would not be possible without much help from Department of Meteorology and Hydrology, to let me use their data as main source of research and I would also like to mention another special thanks to all staff in that department for their available resources and detailed explanation.

I offer deep gratitude and gratefulness to all the persons including my family who contributed in different ways to my thesis.

Finally, I would like to thank all my classmates from MAS to share knowledge, provide care and support towards each other for writing this paper in time.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURES

LIST OF ABBRATION

ACF	Autocorrelation Function
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Averages
ADF	Augmented Dickey-Fuller Unit Root TEs
BIC	Bayesian Information Criterion
LCL	Lower Confidence Limit
MA	Moving Average
PACF	Partial Autocorrelation Function
RMSE	Residual Mean Square Error
SARIMA	Seasonal Autoregressive Integrated Moving Average
SPSS	Statistical Package for Social Science Software
UCL	Upper Confidence Limit

CHAPTER I
INTRODUCTION

1.1 Rationale of Study

Rainfall is one of the most important components of water resource management for decision making and planning especially in agricultural sectors and industry. The influence of rainfall on flooding with downstream implication of erosion, water quality, agriculture, the operation of sewage system and tourism among others cannot be over emphasized. For this reason, early warning of rainfall is very important in the use and management of water resources.

As an important resource in all life and transport processes, water is also an energy source that serves many other useful purposes, such as household consumption, agriculture and industry. Plants are watered through natural sources and irrigation. But today, global climate change is affecting water resources around the world. Climate change is no longer Santa Claus, but negatively affects the climate pattern and at the same time increases the frequency of natural disasters. These adverse effects appear to be small on a small scale but looking at a wider range of patterns is threatening the lives and wealth of the world's population. This is the case when the climate crisis has a dangerous effect on the water cycle. In this regard, we must investigate how global climate change negatively affects the water cycle. In other words, the reason why climate change sometimes means more rain is causing floods and sometimes reducing rain, causing drought.

On the other hand, we need to explore how climate change is affecting Myanmar in general and its water resources in particular. Being a tropical country, Myanmar is highly vulnerable to climate change and climatic variances because Myanmar is an agro-based economy where almost 70% of its population is living on farming activities and agricultural productions. In Myanmar, dry land area is located as its central region which occupies 10% of the total area of Myanmar (54,390 sq.km) and contains 57 townships and 13 districts and is home to sixteen million people (one-third of the total population of Myanmar). Dry Zone area is defined by the annual rainfall amount being less than 40 inches (101.6 cm) is known as the dry zone (L.D Stamp, 1964, cited in Saw Myint Tun, 1989-90).

The Central Dry Zone is one of the most climate sensitive and natural resource poor regions in Myanmar and vulnerable to growing food insecurity and serve environmental degradation. It covers about 13% of the country's land area but is home to an estimated 14.5 million, 85% of which are farmers or farm laborers dependent on rainfed agriculture. Central Dry Zone is situated in central Myanmar covering parts of Magwe, Mandalay and Sagaing Regions. Compared to other regions of the country, Central Dry Zone has very harsh climatic conditions with very low rainfall and extremely high temperature. Annual precipitation in the Dry Zone is, in average, less than 30 inches (750 mm), with marked variations occurring within the region. Over the last decade, there has been more frequently less rainfalls. It has a long dry season of six months from December to April with highest temperature rising over 42' C in day time and lowest temperature dropping to about 12' C during night time.

The average rainfall in the Dry Zone is only an estimated 500-1000 mm compared to the national average annual rainfall during rainy season which is usually from May to October. The rainfall of Dry Zone has two distinct peaks as opposed to the unimodal rainfall behavior in other parts of the country. In addition, in July, the rainfall drops in Dry Zone while the month is characterized as part of the peak monsoon period in many other areas in Myanmar. Due to various physical influences, the Dry Zone is also the warmest part of Myanmar with high diurnal variation (i.e, difference in day and night temperatures).

Agriculture is the dominant livelihood in Myanmar's Dry Zone, absorbing over 80% of the working population. The average farm size in the Dry Zone is about 4.5 acres. Subsistence farming (i.e., growing paddy, sesame, beans, groundnuts and rearing small-scale livestock) is the major economy in the region. In fact, the Dry Zone, the most water-stressed area in Myanmar, contributes significantly to national agricultural production. Wet season rainfall is an important water resource for households and livelihoods in many areas of the Dry Zone and therefore, the climate shocks (i.e., extreme rainfall, rainfall insufficiency, and extreme temperature) are considered serious hazards with potential adverse effects.

In fact, changes in rainfall are one of the most critical indicators that can show the overall impact of climate change. Although, at a detailed level, rainfall is much more difficult to predict than other factors, recent studies explore if, in a warmer climate, heavy rainfall likely increases longer dry spells and a higher risk of floods. In addition, recent studies tend to examine if increases in heavy rainfall during winter become more noticeable in the 2020s.

For these reasons, this need to calculate normal monthly rainfall values for the climatic regions of the country. Out of these regions, I have chosen Central Dry Zone to study the new normal period over years in the zone or region. Here, Chauk is a town in Magway and Shwebo is a town in Sagaing region, about 110km northwest of Mandalay between the Irrawaddy and Mu rivers which was selected as study area because the townships were selected on the basis of the observed temperature extremes, frequency of drought per year and the impacts of climatic parameter on food security.

Chauk's local economy is driven by its oil field refinery which processes crude oil. An estimated 42% (57,804) of Chauk's working population of 137,122 depend on agriculture and 55.39% (135,703lka) of the town's land area of 245, 013 lka is irrigated. Chauk has tropical climate with little rainfall throughout the year. Shwebo township is a trade center for agricultural products like beans, rice and sesame and the township has a working population of 261,875 of which 72% (71,157) are farmers then it has a land area of 185, 320lka with 19.38% (35,915lka) non-irrigated.

The purpose of the proposed thesis is to forecast the future rainfall series or there will be more or less rain in Dry Zone in the future and thus, enhancing the capacity o1f farmers to plan for and response to future impacts of Climate on food security. Under this light, this study attempts to fill this gap by forecasting and analyzing the various observed rainfall parameters by using the statistical models of Time Series Analysis, ARIMA Model and Forecasting Method based on secondary data from 2006 to 2018.

1.2 Objectives of the Study

The overall objective of the study is to calculate monthly rainfall data of Central Dry Zone in Myanmar for the new normal period over years.

Specifically, the objectives of the study are:

(i) To build up a model of SARIMA, which will be used in forecasting rainfall pattern of Chauk and Shwebo townships in Central Dry Zone with the secondary rainfall data of 2006-2018.

(ii) To predict the future rainfall in Chauk and Shwebo township in Central Dry Zone based on the fitted SARIMA model.

1.3 Method of Study

This study is based on the model building procedure by Box-Jenkins (1976). The analytical method of time series analysis and forecasting is the use of a model to predict future value, based on previously observed values. The data are obtained from Department of Meteorology and Hydrology (2006-2018). It was used the Statistical Packages Social Software (SPSS) and EViews Software partially. Firstly, the observed data series are verified for existence of seasonality and stationery. After that, the approach of Box-Jenkins methodology in order to build ARIMA models is based on the following steps: Model Identification will be made based on autocorrelation (ACF) and partial autocorrelation function (PACF). The parameter are estimated by using Maximum Likelihood Method, depending on the model. Furthermore, the model diagnostic checking was verified by plots of the correlograms and ACF and PACF of the residuals and Ljung-Box test, which is a test for hypotheses of no correlation across a specified number of time lags. Then, the forecast values for Monthly Rainfall Data are computed based on the fitted model.

1.4 Scope and Limitation of Study

The study focuses on two townships – Chauk township in Magway Region and Shwebo townships in Sagaing region in Central Dry Zone which covers 13 districts and 57 townships. It has the region of Mandalay, Sagaing and Magway. The townships were selected on the basis of observed temperature extremes, frequency of drought per year and the impacts of climatic parameters on food security. The Dry Zone is characterized by high spatial climate variability. The statistical data of 156 monthly rainfalls data during the period of January-2006 to December-2018. Likewise, the monthly secondary data are taken from Department of Meteorology and Hydrology. In this study, the generated historical patterns in a time series will be forecasted using SARIMA model.

1.5 Organization of Study

This study is divided into five chapters. Chapter I is introduction which is consisting of five sub-headings: rationale of the study, objectives of the study, method of study, scope of the study and organization of the study. Chapter II is presented with literature reviews on forecasting of rainfall data in Central Dry Zone in Myanmar for time series analysis. Chapter III is with research methodology. Chapter IV presents model building of a time series data analysis of rainfall forecasting the observed data series and forecast evaluation in Central Dry Zone. Conclusion is illustrated in Chapter V.

CHAPTER II
LITERATURE REVIEWS

Autoregressive integrated moving average (ARIMA) is the method first introduced by Box and Jenkins (1976) and until now become the most popular models for forecasting univariate time series data. The data has been originated from the Autoregressive model (AR), the Moving Average model (MA) and the combination of the AR and MA, the ARMA models. A popular approach to forecast rainfall data in the short-run is the seasonal Box-Jenkins model. In the case of including seasonal components in this model, then the model is called as the SARIMA model. Box-Jenkins procedure containing three main stages to build an ARIMA model which is model identification, model estimation and model checking, is usually used for determining the best ARIMA model for certain time series data.

In this chapter, an attempt has been made to critically review the literature of the past research work relevant to the present study. The available literature on the subject has been reviewed and presented under the following: "Iyengar et al." proposed rainfall forecasting model based on the rainfall data itself which is decomposable into six empirical time series.

Harvey et al. (1987) investigated how patterns of rainfall correlated with general weather conditions and frequency of cycles of rainfall in Brazil and they found that cyclical components are stochastic rather than deterministic, and the gains achieved from forecast by taking account of the cyclic components are stochastic rather than deterministic and the gains achieved from forecast by taking account of the cyclic component are small in the case.

Samagaio and Wolters (2010), who applied the SARIMA model and the Holt-Winters method for examining the official forecasts of rainfall patterns between 2006 and 2018; the conclusion was that the forecasting results of the SARIMA model appear to be acceptable in the short-run.

Amha (2010) studied the monthly rainfall and temperature in Northern region based on Bauchi station and he employed univariate Box-Jenkins method to analyze rainfall in the region and found that SARIMA model is suitable for forecasting future value of monthly rainfall and temperature data and used this model to forecast 12-month rainfall pattern in this study area. Further he concluded that there is no tendency of decreasing or increasing pattern of monthly rainfall and temperature over the forecast period from January 2010 to September 2011.

Seyed et al. (2011) studied that time series method to model weather parameter in Iran and recommended ARIMA $(0,0,1) \times (1,1,1)_{12}$ as the best fit for monthly rainfall data and ARIMA$(2,1,0)(2,1,0)_{12}$ for monthly average temperature for the region.

Mahsin et al. (2012) used Box-Jenkins methodology to build seasonal ARIMA model for monthly rainfall ant temperature data taken for Bangladesh, for the period between 1981-2010. In their work, ARIMA $(0,0,1) \times (0,1,1)_{12}$ model was found adequate and the model is used for forecasting the monthly rainfall and temperature.

Al-Ansari et al. (2013) studies that the statistical analysis of the rainfall measurements for three meteorological stations in Jordan: Amman Airport (central Jordan), Irbid (northern Jordan) and Mafraq (eastern Jordan). Normal statistical and power spectrum analyses as well as ARIMA model were performed on the long-term annual rainfall measurements at the three stations. A time series model for each station was adjusted, processed, diagnostically checked and lastly an ARIMA model for each station is established with a 95% confidence interval and the model was used to forecast 5 years annual rainfall values for Amman, Irbid and Mafraq meteorological stations. Further result indicated that there is decreasing trend for forecasted rainfall results in all stations.

Nail, P.E. et al, [2009] used Box-Jenkins methodology to build ARIMA model for monthly rainfall data taken for Amman airport station for the period from 1992-1999 with a total of 936 readings. In their research, ARIMA $(1, 0, 0) (0, 1, 1)_{12}$ model was developed. This model was used to forecasting the monthly rainfall for the upcoming 10 years to help decision makers establish priorities in terms of water demand management. They then recommended an intervention time series analysis to be used to forecast the peak values of rainfall data.

CHAPTER III
RESEARCH METHODOLOGY

A time series is an ordered a sequential set of observations. The ordering is usually through time, particularly in terms of some equally spaced time intervals such as hourly, daily, weekly, monthly or yearly time separations. The ordering may also be taken through other dimensions such as space. Time series occur in a variety of fields. In agriculture, this can observe an annual crop production and prices. In meteorology, this can over serve hourly wind speeds, daily temperature and annual rainfall. The main objective of studying a time series is to forecast or predict the further behavior or movements on the basis of its past and present situations.

However, many applied time series, particularly those arising from economic and business areas, are non-stationary. A very useful class of homogeneous non-stationary time series models is the Autoregressive Integrated Moving Average (ARIMA) Models. In the case where seasonal components are included in this model, then the model is called as the SARIMA model. In this studied, SARIMA $(0,0,0) \times (2,1,0)_{12}$ and $(0,0,0) \times (2,1,0)_{12}$ model constructed for Monthly Rainfall data in Chauk and Shwebo township in Myanmar. The fitted models will be applied to forecast the future values, which can be used in decision-making and policy-making.

3.1 Components of A Time Series

A time series in general is supposed to be affected by four main components which can be separated from the observed data. These components are: Trend, Cyclical, Seasonal and Irregular components. A model represents the underlying process that generates a time series. A mathematical model of a time series may be expressed in functional form. The relationship is usually described by one of two models: the multiplicative model and the additive model.

$$Y = T + S + C + I$$

$$Y = T_t + S_t + C_t + I_t \tag{3.1}$$

Where,

Y = observed value of the variable of interest

T = trend component

S = seasonal component

C = cyclical component

I = irregular component

3.1.1 Trend Component

The general tendency of a time series to increase, decrease or stagnate over a long period of time is termed as Secular Trend or simply Trend. Thus, it can be said that trend is a long term movement in a time series and a major use of trend analysis for long-term forecasting. The trend may either be an upward trend or downward trend. In this case, the trend is a period- to- period increase that follows a straight line, a pattern called a linear trend.

3.1.2 Seasonal Component

The seasonal component describes effects that occur regularly over a period of a year, month, quarter, week or day. Seasonal variations in a time series are fluctuations within a year during the season. The important factors causing seasonal variations are: climate and weathered conditions, customs, traditional habits, etc. Seasonality can occur many different ways, for example, by week of the year, month of the year, day of month, day of the week.

3.1.3 Cyclical Component

The cyclical component of a time series refers to (regular or periodic) fluctuations around the trend, excluding the irregular component, revealing a succession of phases of expansion and contraction. The cyclical component can be viewed as those fluctuations in a time series which are longer than a given threshold, 1½ years, but shorter than those attributed to the trend.

The cyclical variation in a time describes the medium -term changes in the series, caused by circumstances, which repeat in cycles. The duration of a cycle extends over longer period of time, usually two more years. The duration of a cycle depends on the type of business or industry being analyzed. Any pattern showing an up and down movement around a given trend is identified as a cyclical pattern.

3.1.4 Irregular Component

The irregular component of time series is the residual time series after the trend-cycle and the seasonal component of a time series is the residual time series after the trend-cycle and the seasonal components (including calendar effects) have been removed. It corresponds to the high frequency fluctuations of the series. The irregular component results from short term fluctuations in a series which are not systematic and in some instances, not predictable, uncharacteristic weather patterns. In a highly irregular series, these fluctuations can dominate movements, which will make the trend and seasonality.

3.2 Stochastic Processes

A stochastic process is a family of time indexed random variables Z(*w,t*), where *w* belongs to a sample space and *t* belongs to an index set. For a fixed *t*, Z(*w,t*) is a random variable. For a given *w, Z(w,t)*, as a function of *t*, is called a sample function or realization. The population that consists of all possible realization is called ensemble in stochastic process and time series analysis. Thus, a time series is a realization or sample function from a certain stochastic process. Important classes of stochastic processes are stationary stochastic processes and non-stationary stochastic processes.

3.3 Stationary Stochastic Processes

Stationary stochastic process is based on the assumption that the process is in a particular static of statistical equilibrium. A stochastic process is said to be strictly stationary if its moments are unaffected by a change of time origin. For a discrete process to be strictly stationary, the joint distribution of any set of observations must be unaffected by shifting all the times of observations forward or backward by any integer amount k. The AR (1) process is said to be stationary, if absolute $\emptyset_1 < 1$.

3.3.1 Mean and Variance of Stochastic Process

The stochastic process Z_t has a constant mean,

$$\mu = \mu_t = E(Z_t) = \int_{-\infty}^{+\infty} z_t\, p(z_t) dz_t \tag{3.2}$$

This defines the level about which it fluctuates, and a constant variance.

$$\sigma_z^2 = \sigma_{z_t}^2 = E(z_t - \mu_t)^2 = \int_{-\infty}^{+\infty} (z_t - \mu_t)^2\, p(z_t) dz_t \tag{3.3}$$

which measures its spread about this time level.

The mean μ of the stochastic process can be estimated by the sample mean

$$\bar{Z}_t = \frac{1}{N}\sum_{t=1}^{N} Z_t \tag{3.4}$$

and the variance $\sigma_{Z_t}^2$ can be estimated by the sample variance

$$\hat{\sigma}_{Z_t}^2 = \frac{1}{N}\sum_{t=1}^{N}(Z_t - \bar{Z}_t)^2 \tag{3.5}$$

If the $\mu_t, \sigma_{Z_t}^2$ do not depend on t (or) they are constant values, the stochastic process is called strictly stationary process.

3.3.2 Autocovariance and Autocorrelation Coefficients

The covariance between Z_t and Z_{t+k} is called the auto-covariance at lag k and is defined by

$$\gamma_k = (Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu) \tag{3.6}$$

and the correlation between Z_t and Z_{t+k} as

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_o} \tag{3.7}$$

where $Var(Z_t) = Var(Z_{t+k}) = \gamma_o$. As function of k, γ_k is called the autocovariance function and ρ_k is called the autocorrelation function (ACF) in time series analysis.

For a stationary process, the autocovariance function γ_k and the autocorrelation function ρ_k have the following properties.

1. $\gamma_o = Var[Z_t], \rho_o = 1$
2. $|\gamma_k| \leq \gamma_o, |\rho_k| \leq 1$
3. $\gamma_k = \gamma_{-k}$and$\rho_k = \rho_{-k}$ for all k
4. $\sum_{i=1}^{n} Z_t \sum_{j=1}^{n} \alpha_i\ \alpha_j \left|\gamma_{ij} - t_1\right| \geq 0$

3.3.3 Partial Autocorrelation Function

The following conditional correlation

$$\text{Corr}\ (Z_t, Z_{t+k}|Z_{t+1}, \dots, Z_{t+k-1}) \tag{3.8}$$

is usually referred to as the partial autocorrelation in time series analysis.

Consider a stationary process $[Z_t]$and, this assume that E (Z_t) = 0.

Let the linear dependence of Z_{t+k} on Z_{t+1},Z_{t+2},....., and Z_{t+k-1} be defined as the best linear estimate in the mean square sense of Z_{t+k} as linear function of Z_{t+1}, Z_{t+2}, ..., and Z_{t+k-1}. That is, if $\hat{Z}_{t+k}$ is the best linear estimate of Z_{t+k}, then

$$\hat{Z}_{t+k}=\alpha_1 Z_{t+k-1}+\alpha_2 Z_{t+k-2}+\cdots+\alpha_{k-1} Z_{t+1}, \tag{3.9}$$

Where $\alpha_1(1 \leq i \leq k-1)$ are the mean squared linear regression coefficients obtained from minimizing

$$\mathrm{E}(Z_{t+k}-\hat{Z}_{t+k})^2=\mathrm{E}(Z_{t+k}-\alpha_1 Z_{t+k-1}-\cdots-\alpha_{k-1} Z_{t+1})^2 \tag{3.10}$$

The routine minimization method through differentiation gives the following linear system of equations

$$\gamma_i=\alpha_1 \gamma_{i-1}+\alpha_2 \gamma_{i-2}+.\ldots.+\alpha_{k-1} \gamma_{i-k+1}\ (1 \leq i \leq k-1). \tag{3.11}$$

Hence,

$$\rho_i=\alpha_1 \rho_{i-1}+\alpha_2 \rho_{i-2}+\cdots+\alpha_{k-1} \rho_{i-k+1}\ (1 \leq i \leq k-1). \tag{3.12}$$

In terms of matrix notation, the above system of (3.11) becomes

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_{k-1} \end{bmatrix}=\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots \quad \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots \quad 1 \end{bmatrix}\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_{k-1} \end{bmatrix} \tag{3.13}$$

Similarly,

$$\hat{Z}_t=\beta_1 Z_{t+1}+\beta_2 Z_{t+2}+\cdots+\beta_{k-1} Z_{t+k-1} \tag{3.14}$$

Where $\beta_i(1 \leq i \leq k-1)$ are the mean squared linear regression coefficients obtained by minimizing

$$\mathrm{E}(Z_t-\hat{Z}_t)^2 = \mathrm{E}(Z_t-\beta_1 Z_{t+1}-\cdots-\beta_{k-1} Z_{t+k-1})^2 \tag{3.15}$$

Hence

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \cdots \\ \rho_{k-1} \end{bmatrix}=\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \rho_{k-4} & \cdots & 1 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_{k-1} \end{bmatrix} \tag{3.16}$$

This implies that $\alpha_i=\beta_i(1 \leq i \leq k-1)$.

It follows that the partial autocorrelation between $Z_t$and Z_{t+k} will equal the ordinary autocorrelation between $\left(Z_t - \hat{Z}_t\right)$ and$\left(Z_{t+k} - \hat{Z}_{t+k}\right)$. Thus, letting $P_k$denote the partial autocorrelation between Z_t andZ_{t+k} , this have

$$P_k = \frac{Cov[(Z_t - \hat{Z}_t),(Z_{t+k} - \hat{Z}_{t+k})]}{\sqrt{Var(Z_t - \hat{Z}_t)}\sqrt{Var(Z_{t+k} - \hat{Z}_{t+k})}} \quad (3.17)$$

Now, $Var\left(Z_{t+k} - \hat{Z}_{t+k}\right) = E[(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})^2]$

$$= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})^2]$$

$$- \alpha_1 E[Z_{t+k-1}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \ldots - \alpha_{k-1} Z_{t+1})]$$

$$\ldots - \alpha_{k-1} E[Z_{t+1}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})]$$

$$= E[Z_{t+k}(Z_{t+k} - \alpha_1 Z_{t+k-1} - \ldots - \alpha_{k-1} Z_{t+1})]$$

Since all other remaining terms reduce to zero by virtue of equation (3.11).

Hence,

$$Var\left(Z_{t+k} - \hat{Z}_{t+k}\right) = Var\left(Z_t - \hat{Z}_t\right) = \gamma_0 - \alpha_1\gamma_1 - \ldots - \alpha_{k-1}\gamma_{k-1.} \quad (3.18)$$

Next, using the fact that $\alpha_i = \beta_i (1 \le i \le k-1)$, There have

$$Cov\left[\left(Z_t - \hat{Z}_t\right), \left(Z_{t+k} - \hat{Z}_{t+k}\right)\right]$$

$=E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1} Z_{t+k-1})(Z_{t+k} - \alpha_1 Z_{t+k-1} - \cdots - \alpha_{k-1} Z_{t+1})]$

$=E[(Z_t - \alpha_1 Z_{t+1} - \cdots - \alpha_{k-1} Z_{t+k-1}) Z_{t+k}]$

$=\gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1.$ (3.19)

Therefore,

$$P_k = \frac{\gamma_k - \alpha_1\gamma_{k-1} - \cdots - \alpha_{k-1}\gamma_1}{\gamma_0 - \alpha_1\gamma_1 - \cdots - \alpha_{k-1}\gamma_{k-1}} = \frac{\rho_k - \alpha_1\rho_{k-1} - \cdots - \alpha_{k-1}\rho_1}{1 - \alpha_1\rho_1 - \cdots - \alpha_{k-1}\rho_{k-1}}. \quad (3.20)$$

Solving the system in (3.13) for α_i by Cramer's rule gives

$$\alpha_i = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_1 & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{i-2} & \rho_1 & \rho_i & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{i-3} & \rho_2 & \rho_{i-1} & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & \rho_{k-i} & \rho_{k-i-1} & \rho_{k-i-2} & \cdots & 1 \end{vmatrix}} \tag{3.21}$$

as the ratio of two determinants. The matrix in the numerator is the same as the symmetric matrix in the denominator except for its i^{th} column being replaced by $(\rho_1, \rho_2, \ldots, \rho_{k-1})$. Substituting α_i in (3.21) to equation (3.20) and multiplying both the numerator and denominator of (3.20) by the determinant

$$\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{vmatrix},$$

the resulting P_k in (3.20) can be easily seen to equal the ratio of the expansion of the following expression in terms of the last column,

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}} \tag{3.22}$$

The partial autocorrelation can also be derived as follows. Consider the regression model, where the dependent variable Z_{t+k} from a zero mean stationary process is regressed on k lagged variables Z_{t+k-1}, Z_{t+k-2}, ..., and Z_t, i.e.,

$$Z_{t+k} = \emptyset_{k1} Z_{t+k-1} + \emptyset_{k2} Z_{t+k-2} + \cdots + \emptyset_{kk} Z_t + e_{t+k}, \tag{3.23}$$

Where $\emptyset_{ki}$ denotes the i^{th} regression parameter and e_{t+k} is a normal error term uncorrelated with Z_{t+k-j} for $j \geq 1$. Multiplying Z_{t+k-j} on both sides of the above regression equation and taking the expectation, Those get

$$\gamma_j = \emptyset_{k1} \gamma_{j-1} + \emptyset_{k2} \gamma_{j-2} + \cdots + \emptyset_{kk} \gamma_{j-k} \tag{3.24}$$

and hence,

$$\rho_j = \emptyset_{k1} \rho_{j-1} + \emptyset_{k2} \rho_{j-2} + \cdots + \emptyset_{kk} \rho_{j-k} \tag{3.25}$$

For *j*=1,2,…,*k* , Those have the following system of equations: it can be written as follows

$$\rho_1 = \emptyset_{k1}\rho_0 + \emptyset_{k2}\rho_1 + \cdots + \emptyset_{kk}\rho_{k-1}$$

$$\rho_2 = \emptyset_{k1}\rho_1 + \emptyset_{k2}\rho_0 + \cdots + \emptyset_{kk}\rho_{k-2}$$

$$\vdots$$

$$\rho_k = \emptyset_{k1}\rho_{k-1} + \emptyset_{k2}\rho_{k-2} + \cdots + \emptyset_{kk}\rho_0$$

Using Creamer’s rule successively for k = 1,2, …,

$$\emptyset_{11} = \rho_1$$

$$\emptyset_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

$$\emptyset_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}$$

$$\vdots$$

$$\emptyset_{kk} = \frac{\begin{vmatrix} 1 & P_1 & P_2 & \cdots P_{k-2} & P_1 \\ P_1 & 1 & P_1 & \cdots P_{k-3} & P_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{k-1} & P_{k-2} & P_{k-3} \cdots & P_1 & P_k \end{vmatrix}}{\begin{vmatrix} 1 & P_1 & P_2 & \cdots P_{k-2} & P_{k-1} \\ P_1 & 1 & P_1 & \cdots P_{k-3} & P_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ P_{k-1} & P_{k-2} & P_{k-3} \cdots & P_1 & 1 \end{vmatrix}} \quad (3.26)$$

Comparing Equation (3.26) with (3.22), it can be seen that $\emptyset_{kk}$ equals ρ_k. Thus, the partial autocorrelation between Z_t and Z_{t+k} can also be obtained as the regression coefficient associated with Z_t in (3.23). Because $\emptyset_{kk}$ has becomes a standard notation for the partial autocorrelation between Z_t and Z_{t+k} in time series. As a function of k, $\emptyset_{kk}$is usually referred to as the partial autocorrelation function (PACF) in time series analysis.

3.3.4 White Noise Processes

In process $\{a_t\}$is called a white noise process if it is a sequence of uncorrelated random variables from a fixed distribution with constant mean E(a_t) = μ_a , assumed to be 0, constant variance Var $(a_t) = \sigma_a^2$ and γ_k = Cov $(a_t, a_{t+k}) = 0$ for all $k \neq 0$. A white noise process $\{\alpha_t\}$ is stationary with the autocovariance function

$$\gamma_k = \begin{cases} \sigma_{a,}^2 & k=0, \\ 0, & k\neq 0, \end{cases} \tag{3.27}$$

The autocorrection function

$$\rho_k = \begin{cases} 1 & k=0, \\ 0, & k\neq 0, \end{cases} \tag{3.28}$$

The partial autocorrection function $\emptyset_{kk} = \begin{cases} 1 & k=0, \\ 0, & k\neq 0, \end{cases}$ (3.29)

By definition ρ_0 = $\emptyset_{00}$ =1 for any process, when the autocorrection and partial auto correlations, refer only to ρ_k and $\emptyset_{kk}$ for $k \neq 0$. The basic phenomenon of the white noise process is that ACT and PACT are identically equal to zero.

3.4 Autoregressive Processes

Based on a finite number of available observations, a finite order parametric model was constructed to describe a time series process. In this chapter, the autoregressive modes were described as a special case. These models are useful in describing a wide variety of time series. The characteristics of each process in terms of autocorrelation and partial autocorrelation functions will also be discussed in this section.

In time series analysis, there are two useful representations to express a time series process. The first one is to write a process $\hat{Z}_t$ in an autoregressive (AR) representation, in which regress the value of $\hat{Z}$ at time t on its own past values plus a random shock, i.e.,

$$\hat{Z}_t = \pi_1 \hat{Z}_{t-1} + \pi_2 \hat{Z}_{t-2} + \cdots + a_t \tag{3.30}$$

or equivalently,

$$\pi(B)\hat{Z}_t = a_t \tag{3.31}$$

where $\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j$, and $1+\sum_{j=1}^{\infty} |\pi_j| < \infty$. The autoregressive representation is a very useful model for the mechanism of forecasting. In the autoregressive representation of a process, if only a finite number of π weights are nonzero, i.e., $\pi_1 = \emptyset_1, \pi_2 = \emptyset_2, \dots, \pi_p = \emptyset_p$ and $\pi_k = 0$ for $k > p$, then the resulting process is said to be an autoregressive process (model) of order p, which is denoted as AR (p). It is given by

$$\hat{Z}_t = \emptyset_1 \hat{Z}_{t-1} + \emptyset_2 \hat{Z}_{t-2} + \cdots + \emptyset_p \hat{Z}_{t-p} + a_t \tag{3.32}$$

$$\emptyset_p(B)\hat{Z}_t = a_t, \tag{3.33}$$

or

where $\emptyset_p(B) = (1 - \emptyset_1 B - \cdots - \emptyset_p B^p)$, $\hat{a}_t = a_t - \mu$ and a_t is random error or disturbance term. Since $\sum_{j=1}^{\infty}|\pi_j| = \sum_{j=1}^{\infty}|\emptyset_j| < \infty$, the process is always invertible. To be stationary, the roots of $\emptyset_p$ (B) = 0 must lie outside of the unit circle.

The AR processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock. First, consider the following simple autoregressive models.

3.4.1 The General p^{th} Order Autoregressive AR (p) Process

The p^{th}order autoregressive process AR (p) is

$$\left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right)\hat{Z}_t = a_t \tag{3.34}$$

or

$$\hat{Z}_t = \emptyset_1 \hat{Z}_{t-1} + \emptyset_2 \hat{Z}_{t-2} + \cdots + \emptyset_p \hat{Z}_{t-p} a_t \tag{3.35}$$

(a) Autocovariance Function of General AR (p) Process

To find the autocovariance function, both sides of equation (3.33) is multiply by $\hat{Z}_{t-k}$

$$\hat{Z}_{t-k}\hat{Z}_t = \emptyset_t \hat{Z}_{t-k}\hat{Z}_{t-1} + \cdots + \emptyset_p \hat{Z}_{t-k}\hat{Z}_{t-p} + \hat{Z}_{t-k}\alpha_t$$

and take the expected value

$$\gamma_k = \emptyset_1 \gamma_{k-1} + \cdots + \emptyset_p \gamma_{k-p'} \qquad k > 0,$$

where E $(a_t Z_{t-k})$ = 0 for $k > 0$.

(b) Autocorrelation Function of General AR(p) Process

The following recursive relationship for the autocorrelation function;

$$\rho_k = \emptyset_1 \rho_{k-1} + \emptyset_2 \rho_{k-2} + \cdots + \emptyset_p \rho_{k-p'}, \qquad k > 0, \tag{3.36}$$

From (3.32), the ACF ρ_k is determined by the difference equation $\emptyset_p(B)\rho_k = \left(1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p\right)\rho_k = 0$ for k > 0. Hence, it can be written as

$$\emptyset_p(B) = \prod_{i=1}^{m}(1 - G_iB)^{d_i},$$

where $\sum_{i=1}^{m} d_i = p$, and G_i^{-1}(i = 1, 2, ...,m) are the roots of multiplicity d_i of $\emptyset_p(B) = 0$. Using the difference equations results, as follows,

$$\rho_k = \sum_{i=1}^{m} G_i^k \sum_{j=0}^{di-1} A_{ij}k^j \qquad (3.37)$$

If $d_i = 1$ for all i, then G_i^{-1}, are all distinct and the above reduces to

$$\rho_k = \sum_{i=1}^{p} A_iG_i^k \text{ k} > 0 \qquad (3.38)$$

For a stationary process, $\left|G_i^{-1}\right| > 1$ and $|G_i| < 1$. Hence, the ACF ρ_k tails off as a mixture of exponential decays or damped sine waves depending on the roots of $\emptyset_p(B) = 0$. Damped sine waves appear if some roots are complex.

(c) **Partial Autocorrelation Function of General AR (*p*) Process**

By using the fact that $\rho_k = \emptyset_1\rho_{k-1} + \emptyset_2\rho_{k-2} + \cdots + \emptyset_p\rho_{k-p}$ for $k > 0$, it can obviously be seen that when $k > p$ the last column of the matrix in the numerator of $\emptyset_{kk}$ in (3.26) can be written as linear combination of previous column of the same matrix. Hence, the *PACF* $\emptyset_{kk}$ will vanish after lag p.

3.5 Moving Average Processes

In the moving average representation of a process, if only a finite number of ψ weights are nonzero, that $\psi_1 = -\theta_1, \psi_2 = -\theta_2, \ldots, \psi_q = -\theta_q$ and $\psi_k = 0$ for $k > q$, then the resulting process is said to be a moving average process or model of order q and is denoted as MA (*q*). It is given by

$$\hat{Z}_t = a_t - \theta_1 a_{t-1} - \theta_1 a_{t-2} - \cdots - \theta_q a_{t-q} \qquad (3.39)$$

Or

$$\hat{Z}_t = \theta(B)a_1;$$

Where

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \emptyset_q B^q).$$

Because $1 + {\theta_1}^2 + {\theta_2}^2 + \cdots + {\theta_q}^2 < \infty$, a finite moving average process is always stationary. This moving average process is invertible if the roots of $\theta(B) = 0$ lie outside of the unit circle. Moving average processes are useful to describe a phenomena in which events produce an immediate effect that only lasts for only short periods of time. To discuss other properties the MA(q) process, let us first consider the following simpler cases.

3.5.1 The General q^{th} Order Moving Average MA (q) Process

The general q^{th}order moving average process is

$$\hat{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\, a_t, \tag{3.40}$$

For this general MA(q) process, the variance is

$$\gamma_0 = \sigma_a^2 \sum_{j=0}^{q} {\theta_j}^2,$$

where $\theta_0 = 1$, then the additional autocovariances are

$$\gamma_k = \begin{cases} \sigma_a^2(-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q) & k = 1, 2, \ldots, q, \\ 0, & k > q \end{cases} \tag{3.41}$$

Therefore, the autocorrelations function are

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1\theta_{k+1} + \cdots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \cdots + \theta_q^2} & k = 1, 2, \ldots, q, \\ 0, & k > q \end{cases}$$

The autocorrelation function of an MA(q) process cuts off after lag q. This important property enables us to identify whether a given time series is generated by a moving average process.

3.5.2 Dual Relationship Between AR (p) and MA(q) Process

Given that stationary AR(p) process,

$$\emptyset_p(B)\hat{Z}_t = a_t, \tag{3.42}$$

Where $\emptyset_p(B) = 1 - \emptyset_1 B - \cdots - \emptyset_p B^p$, it can write

$$\hat{Z}_t = \frac{1}{\emptyset_p(B)} a_t = \psi(B) a_t \tag{3.43}$$

With $\psi(B) = (1 + \psi_1 B + \psi_2 B^2 + \cdots)$ such that

$$\emptyset_p(B)\psi(B) = 1 \tag{3.44}$$

The ψ weights can be derived by equating the coefficients of B^j
on both sides of (3.40).
An example, this can write the AR (2) process as per

$$\hat{Z}_t = \frac{1}{(1 - \emptyset_1 B - \emptyset_2 B^{2)}} a_t = (1 + \psi_1 B + \psi_2 B^2 + \cdots)\, a_t, \tag{3.45}$$

Which implies that

$$(1 + \psi_1 B + \psi_2 B^2)(1 + \psi_1 B + \psi_2 B^2 + \psi_2 B^3 + \cdots) = 1,$$

Thus, it obtain the $\psi_j's$ as follows:

$$\begin{aligned}
&B^1: \quad \psi_1 - \emptyset_1 = 0 \longrightarrow \psi_1 = \emptyset_1 \\
&B^2: \quad \psi_2 - \psi_1\, \emptyset_1 - \emptyset_2 = 0 \longrightarrow \psi_2 = \psi_1 \emptyset_1 + \emptyset_2 = \emptyset_{1+}^2 \emptyset_2 \\
&B^3: \quad \psi_3 - \psi_2\, \emptyset_1 - \psi_1\, \emptyset_2 - \emptyset_2 = 0 \longrightarrow \psi_3 = \psi_2 \emptyset_1 + \psi_1 \emptyset_2 \\
&\qquad \vdots
\end{aligned}$$

In fact, for $j \geq 2$, it has

$$\psi_j = \psi_{j-1}\emptyset_1 + \psi_{j-2}\emptyset_2, \tag{3.46}$$

Where $\psi_0 = 1$. In a special case when $\emptyset_2 = 0$, it has $\psi_j = \emptyset_1^j$, $j \geq 0$. Therefore,

$$\hat{Z}_t = \frac{1}{(1 - \emptyset_1 B)} a_t = (1 + \emptyset_1 B + \emptyset_i^2 B^2 + \cdots)\, a_t, \tag{3.47}$$

This equation implies that a finite order stationary AR process is equivalent to an infinite order MA process.

Given that its general invertible MA*(q)* process,

$$\hat{Z}_t = \theta_q(B) a_t \tag{3.48}$$

with $\emptyset_q(B) = (1 - \emptyset_1 B - \cdots - \emptyset_q B^q)$, this can rewrite it as

$$\pi(B)\hat{Z}_t = \frac{1}{\theta_q(B)} \hat{Z}_t = a_t, \tag{3.49}$$

Where, $\pi(B) = (1 - \pi_1 B - \pi_2 B^2 - \cdots) = \frac{1}{\theta_q(B)}$ (3.50)

Therefore, the π weights can be derived by equation the coefficients of B^j as follows:

$$\begin{aligned}
&B^1: \quad \pi_1 - \emptyset_1 = 0 \longrightarrow \pi_1 = \emptyset_1 \\
&B^2: \quad -\pi_2 - \pi_1 \emptyset_1 - \emptyset_2 = 0 \longrightarrow \pi_2 = \pi_1 \emptyset_1 + \emptyset_2 = -\emptyset_{1+}^2 \emptyset_2 \\
&B^3: \quad -\pi_3 - \pi_2\, \emptyset_1 - \pi_1\, \emptyset_2 - \emptyset_2 = 0 \longrightarrow \pi_3 = \pi_2 \emptyset_1 + \pi_1 \emptyset_2
\end{aligned}$$

In general,

$$\pi_j = \pi_{j-1}\theta_1 + \emptyset_2 = \pi_{j-2}\theta_1 \text{, for } j \geq 3. \tag{3.51}$$

When $\theta_2 = 0$ and the process becomes the MA (1) process, this have θ_1^j for $j \geq 1$,

$$(1 + \emptyset_1 B + \emptyset_i^2 B^2 + \cdots)\hat{Z}_t = \frac{1}{(1-\emptyset_1 B)}\hat{Z}_t = a_t \tag{3.52}$$

Hence, in term of the AR representation, a finite-order invertible MA process is equivalent to an infinite-order AR process.

In summary, this dual relationship between the AR(p) and the MA(q) processes also exits in the autocorrelation and partial autocorrelation functions. The AR(p) process has its autocorrelations tailing off and partial autocorrelations cutting off, but the MA(q) process has its autocorrelations cutting off and partial autocorrelation tailing off.

3.6 Autoregressive Moving Average ARMA $(\boldsymbol{p}, \boldsymbol{q})$Processes

The following useful mixed autoregressive moving average ARMA (p, q) processes

$$\emptyset_p(B)\hat{Z}_t = \theta_q(B)a_t, \tag{3.53}$$

where,

$$\emptyset_p(B) = 1 - \emptyset_1 B - \cdots - \emptyset_p B^p,$$

and

$$\theta_q(B) = 1 - \theta_1 B - \cdots - \theta_q B^q.$$

For the process to be invertible, it requires that the roots of $\theta_q(B) = 0$ lie outside the unit circle. To be stationary, it requires that the roots of $\emptyset_p(B) = 0$ lie outside the unit circle. Assuming that $\emptyset_p(B) = 0$ and $\theta_q(B) = 0$ share no common roots, this process refers to an ARMA (p, q) process or model, in which p and q are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively.

The stationary and invertible ARMA process can be written in a pure autoregressive representation discussed in Section (3.5), i.e.,

$$\hat{Z}_t = \psi(B)a_t$$

where,

$$\psi(B) = \frac{\theta_q(B)}{\emptyset_p(B)} = (1 + \psi_1 B + \psi_2 B^2 + \cdots).$$

Autocorrelation Function of the ARMA (p, q) Process: To derive the autocorrelation function, the equation (3.37) is rewritten as

$$\hat{Z}_t = \emptyset_1 \hat{Z}_{t-1} + \cdots + \emptyset_p \hat{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}$$

and multiplied by $\hat{Z}_{t-k}$ on both sides

$$\hat{Z}_{t-k}\hat{Z}_t = \emptyset_1 \hat{Z}_{t-k}\hat{Z}_{t-1} + \cdots + \emptyset_p \hat{Z}_{t-k}\hat{Z}_{t-p} + \hat{Z}_{t-k}a_t - \theta_1 \hat{Z}_{t-k}a_{t-1} - \cdots - \theta_q \hat{Z}_{t-k}a_{t-q}$$

Now we take the expected value to obtain

$$\gamma_k = \emptyset_1 \gamma_{k-1} + \cdots + \emptyset_p \gamma_{k-p} + E(\hat{Z}_{t-k}a_t) - \theta_1 E(\hat{Z}_{t-k}a_{t-1}) - \cdots - \theta_q E(\hat{Z}_{t-k}a_{t-q}).$$

Because

$$E(\hat{Z}_{t-k}a_{t-i}) = 0 \qquad for\ k > i,$$

This have

$$\gamma_k = \emptyset_1 \gamma_{k-1} + \cdots + \emptyset_p \gamma_{k-p} \qquad k \geq (q+1) \tag{3.54}$$

and hence,

$$\rho_k = \emptyset_1 \rho_{k-1} + \cdots + \emptyset_p \rho_{k-p} \qquad k \geq (q+1) \tag{3.55}$$

Equation (3.55) satisfies the p^{th} order homogeneous difference equation given for the AR(p) process. Thus, the autocorrelation function of an ARMA (p, q) model tails off after lag q just like an AR(p) process, which depends only on the autoregressive parameters in the model. However, the first q autocorrelations $\rho_q, \rho_{q-1}, \dots, \rho_1$ depend on both autoregressive and moving average parameters in the model and serve as initial value for the pattern. This distinction is useful for model identification.

Partial Autocorrelation Function of the ARMA (*p, q*) Process: As the ARMA process contains the MA process as a special case, its PACF will also be a mixture of exponential decays and/or damped sine waves depending on the roots of $\emptyset_p$ (B) = 0 and θ_q (B) = 0.

Non-Stationary Processes

In previous section the stationary processes have been discussed. However, many applied time series, particularly those arising from economic and business areas, are non-stationary. With respect to the class of covariance stationary processes, non-stationary time series can occur in many different ways. They could have non-constant means μ_1, time varying second moments such as nonconstant variance σ_t^2, or have both of these properties.

3.7 Autoregressive Integrated Moving Average (ARIMA)

A statistical technique that uses time series data to predict future. The parameters used in the ARIMA is (p, d, q) which refers to the autoregressive, integrated and moving average parts of the data set, respectively. ARIMA modeling will take care of trends, seasonality, cycles, errors and non-stationary aspects of a data set when making forecasts. There is no other univariate forecasting method has been more widely discussed than ARIMA model can be defined as

$$\emptyset_P(1-B)^d Z_t = \emptyset(B)\nabla^d Z_t = \theta_0 + \theta(B)a_t \qquad (3.56)$$

Where

$$\emptyset_p(B) = 1 - \emptyset_1 - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$$

$$\theta_q(B) = 1 - \theta_1 - \theta_2 B^2 - \cdots - \theta_q B^q$$

In what follows:

1. $\emptyset_p$ (B) will be called the autoregressive operator: it is assumed to be stationary, that is, the roots of $\emptyset_p$ (B) =0 lie outside the unit circle.
2. $(1-B)^d = \emptyset$ (B) ∇^dwill be called the generalized autoregressive operator: it is a nonstationary operator with d of roots of $(1-B)^d = 0$equal to unity.
3. $\theta_q(B)$will be called the moving average operator; it is assumed to be invertible, that is, the roots of $\theta_q(B)$=0 lie outside the unit circle.

Where d = 0, the model (3.40) represents a stationary process. The requirements of stationary and invertibility apply independently, and in general, the operators $\emptyset_p(B)$and $\theta_q(B)$ will not be of the same order.

3.8 Seasonal Autoregressive Integrated Moving Average, SARIMA$(\mathbf{p}, \mathbf{d}, \mathbf{q}) \times (\mathbf{P}, \mathbf{D}, \mathbf{Q})_s$ Model

The ARIMA model is for non-seasonal non-stationary data. Box and Jenkins have generalized this model to deal with seasonality. The theoretical justification for modeling univariate time series as seasonal ARIMA processes is founded in the time series theorem known as the world decomposition, which applies to discrete time data series that are stationary about their mean and variance. Therefore, it is also necessary to support an assertion that an appropriate seasonal difference will induce stationarity.

The generalized form of SARIMA $(\mathrm{p}, \mathrm{d}, \mathrm{q}) \times (\mathrm{P}, \mathrm{D}, \mathrm{Q})_s$ model can be written as:

$$\emptyset_p(B^s)\emptyset_P(B)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\theta_Q(B^s)a_t \tag{3.57}$$

Where:

$$\emptyset_p(B) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$$
$$\emptyset_P(B^s) = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_P B^P$$
$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$
$$\theta_Q(B^s) = 1 - \theta_1 B - \theta_2 B^{2s} - \cdots - \theta_Q B^{Qs}$$

Where:

$p, d, and\ q$are the order of non-seasonal AR, differencing and MA respectively.

$P, D, and\ Q$is the order of seasonal AR, differencing and MA respectively.

Z_t representstime series data at period *t.*
B represents backshift operator defined by $BZ_t = Z_{t-1}$.

$(1-B)^d$ represents non-seasonal difference.

$(1-B^s)^d$ represents seasonal difference.

s represents seasonal order (s=12 for monthly data)

$a_t$represents white noise process at period t. It is identically and normally distributed with mean zero, variance σ^2; and $cov\ (e_t, e_{t-k}) = 0 \forall k \neq 0$, that is, $\{e_t\} \sim WN\ (0, \sigma^2)$.

From a practical perspective, fitted seasonal ARIMA models provide linear state transition equations that can be applied recursively to produce single and multiple interval forecasts. Furthermore, seasonal ARIMA models can be readily expressed in state space form, thereby allowing adaptive Kalman filtering techniques to be employed to provide a self-tuning forecast model.

3.9 Seasonal time series

A time series have an important seasonal component such as the monthly or quarterly series. A time series is periodic with period s when similarity in the series occurs of after s basic time intervals. Basic time interval means one month in monthly time series and the period is s=12. It means one quarter in quarterly time series and the period is s = 4. In some series, there can be more than one period. Thus, in a weekly time series, there can be a period of s = 4 and a period of s = 12.

Monthly or quarterly time series may show seasonal effects within years. Seasonality means a tendency to repeat a pattern of behavior over a seasonal period, generally one year. Seasonal series are characterized by a display of strong serial correlation at a seasonal lag, that is, the lag corresponding to the number of observations per seasonal lag.

Seasonal time series usually display time to time (for example, month to month) changes over the years, showing also within year variations. It is useful to understand the actual situation and is used for short term planning. For example, if the data represent the daily sales of a large restaurant, a considerable variation may be noticed depending on the day of the week. Another example, the publisher of the monthly magazine, may be concerned with monthly variations in sales throughout of the year. The restaurant owner is concerned with the day of the week that is involved. The publisher of a monthly magazine is concerned with which month of the year is involved. The restaurant's week-end sales are usually higher than the other sales. The seasonal cycle for the restaurant is a week and the days of the week are seasons. For a monthly magazine publisher, months are the seasons and a year is a seasonal cycle.

3.10 Test of Seasonality

There are two main reasons of isolating the seasonal component or element.

(i) to study seasonal variations

(ii) to eliminated them.

In study the second component of time series, namely seasonal component, test of seasonal can be carried out before computing the seasonal indices from the given time series data.

In the study of seasonality, seasonal variation for each month of the year is usually considered. As such, the following statistical model is employed

$$y_{ij} = \mu + a_i + b_j + e_{ij} \, ; \; 1 \le i \le n \, , \; 1 \le j \le k$$

Where , μ = general mean/ unknown constant

a_i = effect of i^{th} year (i = 1, 2, 3, …, n)

b_j = effect of j^{th} month (j = 1, 2, 3, …, n)

e_{ij} = random error

y_{ij} = observed value of y at j^{th} month of i^{th} year.

Hence, it is assumed that eij are independently and normally distributed with mean zero and constant variance σ^2.

i.e. $e_{ij} \sim IN(0, \sigma^2)$

In general, this test,

$H_0 : b_1 = b_2 = \dots = b_{12} = 0$

There exist no monthly effects. I.e. b_j is zero or there is no seasonality.

H_1 : At least one b_j is not equal to zero.

There exists monthly effect and there is seasonality.

These employ test that is obtained from the following computation procedure and analysis of variance (ANOVA) table shown below.

SST = Total Sum of Square = $\sum_i \sum_{ij} y_{ij}^2$ – (C.T)

SSM = Sum of Square due to months = $\frac{1}{n} \sum_{j=1}^{12} R_j^2$ – (C.T)

SSY = Sum of Square due to years = $\frac{1}{12} \sum_{i=1}^{n} C_j^2$ – (C.T)

SSE = Error Sum of Squares = SST – SSM – SSY

Where C.T = G^2/nk, G = $\sum_{i=o}^{n} \sum_{j=1}^{k} y_{ij}$ = Grand Total

After computation of SST, SSM, SSU and SST – SSM – SSY, the following ANOVA table is constructed.

Table (3.10) ANOVA for a Two-way Analysis of Variance

Source	**Sum of Square**	**Degree of Freedom**	**Mean Square**	**F-Ratio**
Due to Months	SSM	k-1	MSM = SSM / k-1	F_1= MSM/MSE
Due to Years	SSY	n-1	MSY = SSY / n-1	F_2= MSY/MSE
Error	SSE	(n-1)(k-1)	MSE = SSE / (n-1)(k-1)	
Total	**SST**	**kn-1**		

At 100 α % level of significant, the critical value from the F table for the degree of freedom (k-1) and (k-1) (n-1) is given by, K= $F_{\alpha(k-1)(k-1)(n-1)}$.

If F ≥ K, reject H_0 and it is decided that there exists seasonality.

If F < K, accept H_0 and it is decided that there exists no seasonality.

3.11 Seasonal Index

Seasonal variation is measured in terms of an index, called a seasonal index. It is an average that can be used to compare an actual observation relative to what it would be if there were no seasonal variation. An index value is attached to each period of the time series within a year. This implies that if monthly data are considered there are 12 separate seasonal indices, one for each month. There exist different methods for measuring the seasonal variation of a time series. The methods have been developed to meet different objectives of estimating seasonal and the assumed models of the time series. The seasonal pattern itself is important in the application of these methods since most of the methods assume that the seasonal pattern is constant or stable. In finding the index of seasonal variation as seasonal measures, it should be noted that the index must (a) Measure all the variation is the series that is seasonal in character, and (b) Measure nothing but the seasonal variation A seasonal index thus consists of a series of percentage figures, averaging 100, which shows the relative level of the series for the various months, quarters or weeks of the year. An index of seasonal variation can be constructed by expressing each item in the time series as a percent of the average monthly or quarterly value for the year.

Ratio to Moving Average Method

Seasonal indices are calculated so that their average is 1. This means that the sum of the seasonal indices equals the number of seasons. The steps for the computation of the seasonal index by the Ratio to Moving Average method are shown as below. (Steiner, 1956) 1. Find the twelve months centered moving averages. This is equivalent to a moving average of thirteen months with weights (1,2,2, … , 2,2,1) By finding twelve months centered moving averages, we eliminate the seasonality, since the seasonal pattern is periodic with a period of twelve months. Also, it will eliminate the random component or irregular movements. Therefore, the centered twelve month moving averages are the approximates of trend and cyclical components. 2. Compute the ratio to moving average values, that is, the original data is divided by its appropriate moving average value. There, the first and last six months may not be obtained. By this step, the trend and cyclical components are removed from the original data and the ratios are the values due to seasonal and random components. They are called specific seasonal. (Steiner, 1956) 3. Compute the averages of these ratios referring to the same months. These averages are the crude seasonal index values. This step involves two different purposes: the elimination of the random components and averaging the seasonal relatives referring to the same months. 4. Adjust the crude seasonal index. In multiplicative mode, the total seasonal index values have to be equal to twelve (or 1200 percent) for monthly series. Therefore, the crude seasonal index is adjusted to get a total of twelve (or 1200 percent).

3.12 Box-Jenkins Methodology

The Box – Jenkins methodology has been expressed steps for model identification, method of estimation of the parameters in the ARIMA models, diagnostic checking and forecasting.

3.12.1 Model Identification

In time series analysis, the most crucial steps are to identify and build a model based on the available data.

$$(1-\emptyset_1 B-\cdots-\emptyset_1 B^p)(1\text{-}B)^d Z_t=\theta_0+(1\text{-}\theta_1 B-\cdots-\theta_q B^q)a_t$$

Model identification refers to the methodology in identifying the required transformations. The following useful steps will be used to identify a tentative model for a given time series.

(1) Plot the time series data and choose proper transformations.

(2) Compute and examine the sample ACF and sample PACF of the original series to further confirm a necessary degree of differencing. Some general rules are;

If the sample ACF decays very slowly, (the individual ACF may not be large) and the sample PACF cuts off after lag 1, it indicates that differencing is needed.

(3) Compute and examine the sample ACF and PACF of the property transformed and differenced series to identify the orders of p and q.

To identify the order p *a*nd q, the patterns in the sample ACF are matched with the theoretical pattern of known models.

Seasonal patterns require respective seasonal differencing. If the estimated autocorrelation coefficients decline slowly at longer lags, first order differencing is usually needed. However, some time series may require little or no differencing, and that over differenced series produce less stable coefficient estimates.

At this stage it is also need to decide how many autoregressive (p) and moving average (q) parameters are necessary to yield an effective but still parsimonious model of the process (parsimonious means that it has the fewest parameters and greatest number of degrees of freedom among all models that fit the data). In practice, the numbers of the p or q parameters very rarely need to be greater than 2.

Table (3.12) Characteristics of theoretical ACF and PACF for Stationary Processes

Process	ACF	PACF
AR (*p*)	Tails off as exponential decay of damped sine wave	Cuts off after lag *p*
MA (q)	Cut off after lag *q*	Tail off
ARMA (*p,q*)	Tail off after lag (*q-p*)	Tails off after lag*(p-q)*

Source: Univariate and Multivariate Methods (William W.S.Wei)

(4) Test the deterministic trend term $\theta_0$when d >0.

For a non-stationary model, $\emptyset(B)(1-B)^d Z_t = \theta_0 + \theta(B)\, a_t$, the parameter θ_0 is usually omitted so that it is capable of representing series with random changes in the level, slope or trend. To test whether that the differenced series contains a deterministic trend mean, the sample mean $\overline{W}$ of the differenced series$W_t(1-B)^d Z_t$ must be compared with its approximate standard error $S_{\overline{W}}$.

$$\sigma_{\overline{W}}^2 = \frac{\gamma_0}{n}\sum_{j=-\infty}^{\infty} P_j = \frac{1}{n}\sum_{j=-\infty}^{\infty} \gamma_j = \frac{1}{n}\,\gamma(1) \qquad (3.58)$$

Where r (B) is the autocovariance generating function and $\gamma(1)$ is its value at B =1.

3.12.2 Parameter Estimation

After a model is identified for a given time series its is important to obtain efficient estimates of the parameters. In this study, the method of moments, maximum likelihood method will be used to estimate the parameters.

The Method of Moments

The method of moments consists of substituting sample moments such as the sample mean, Z, Sample variance $\hat{\gamma}_0$, and sample ACF $\hat{p}_i$ for their theoretical counterparts and solving the resultant equations to get estimates of unknown parameters.

$$\hat{Z}_t = \emptyset_1\hat{Z}_{t-1} + \emptyset_2\hat{Z}_{t-2} + \ldots + \emptyset_p\hat{Z}_{t-p} + a_t, \qquad (3.59)$$

where, the mean $\mu = E\,(Z_t)$ is estimated by $\bar{Z}$. To estimate $\emptyset$, firstly this use that $p_k = \emptyset_1 p_{k-1} + \emptyset_1 p_{k-2} + \cdots + \emptyset_p p_{k-p}\ for\ k \geq 1$ to get the following system of Yule-Walker equations:

$$P_1 = \emptyset_1 + \emptyset_2 P + \emptyset_3 P_2 + \cdots + \emptyset_p P_{p-1}$$

$$P_2 = \emptyset_1 P_1 + \emptyset_2 + \emptyset_3 P_1 + \cdots + \emptyset_p P_{p-2}$$

$$\vdots$$

$$P_p = \emptyset_1 p_{p-1} + \emptyset_2 P_{p-2} + \emptyset_3 P_{p-3} + \cdots + \emptyset_p. \tag{3.60}$$

Then, replacing P_k by $\hat{P}_k$, these get the moment estimators $\hat{\emptyset}_1, \hat{\emptyset}_2, \ldots, \hat{\emptyset}_p$ by solving the above linear system of equations. That is,

$$\begin{bmatrix} \hat{\emptyset}_1 \\ \hat{\emptyset}_2 \\ \vdots \\ \hat{\emptyset}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{P}_1 & \hat{P}_2 & \hat{P}_{p-2} & \hat{P}_{p-1} \\ \hat{P}_1 & 1 & \hat{P}_1 & \hat{P}_{p-3} & \hat{P}_{p-2} \\ \vdots & & & & \vdots \\ \hat{P}_{p-1} & \hat{P}_{p-2} & \hat{P}_{p-3} & \hat{P}_1 & 1 \end{bmatrix} \tag{3.61}$$

That estimators are usually called Yule-Walker estimators.

Having obtained $\hat{\emptyset}_1, \hat{\emptyset}_2, \ldots, \hat{\emptyset}_p$, these use the result

$$\begin{aligned} \gamma_0 &= E\left(\hat{Z}_t \hat{Z}_t\right) = E\left[\hat{Z}_t\left(\emptyset_1 \hat{Z}_{t-1} + \emptyset_2 \hat{Z}_{t-2} + \ldots + \emptyset_p \hat{Z}_{t-p} + a_t\right)\right] \\ &= \emptyset_1 \gamma_1 + \emptyset_2 \gamma_2 + \cdots + \emptyset_p \gamma_p + \sigma_a^2 \end{aligned} \tag{3.62}$$

then, get the moment estimator for σ_a^2 as

$$\sigma_a^2 = \hat{\gamma}_0 (1 - \hat{\emptyset}_1, \ldots, \hat{\emptyset}_p \hat{P}_1 - \hat{\emptyset}_2 \hat{P}_2 - \cdots - \hat{\emptyset}_p \hat{P}_p \tag{3.63}$$

Regardless of AR, MA, or ARMA models, the moment estimates are very sensitive to rounding errors. The moment estimators are not suggested for final estimation results and should not be used if the process is close to being nonstationary or noninvertible.

Maximum Likelihood Method

The maximum likelihood method has been widely used in estimation.

Conditional Maximum Likelihood Estimation

For the general stationary ARMA *(p,q)* model

$$\hat{Z}_t = \emptyset_1 \hat{Z}_{t-1} + \cdots + \emptyset_p \hat{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}, \tag{3.64}$$

where, $\hat{Z}_t = Z_t - \mu \; and \; (a_t) are$ idependently and idependently normally distributed random variable with mean zero and variance σ_a^2, the joint probability density of a = $(a_1, a_2, \ldots, a_n)$ is given by

$$\mathrm{P}\,(\mathrm{a}|\emptyset, \mu, \theta, \sigma_a^2) = (2\pi\sigma_a^2)^{-n/2} \exp\left[\frac{1}{2\sigma_a^2} \sum_{i=0}^{n} \sigma_a^2\right]. \tag{3.65}$$

and then, a_t can be described as

$$a_t = \theta_1 a_{t-1} + \cdots + \theta_q a_{t-q} + \hat{Z}_t - \emptyset_1 \hat{Z}_{t-1} - \cdots - \emptyset_p \hat{Z}_{t-p}, \tag{3.66}$$

These can write the likelihood function of the parameter $(\emptyset, \mu, \theta, \sigma_a^2)$.

Let Z = $(Z_1, Z_2, \dots, Z_n)'$ and assume that the initial conditions

$Z_* = (Z_{1-p}, \dots, Z_{-1}, Z_0)'$ and $a_* = (a_{1-q}, \dots, a_{-1}, a_0)'$ are known.

The conditional log-likelihood function is

$$\ln L_* (\emptyset, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S_*(\emptyset,\mu,\theta)}{2\sigma_a^2} \tag{3.67}$$

where, $S_*(\emptyset, \mu, \theta) = \sum_{t=1}^{n} a_t^2 \,((\emptyset, \mu, \theta |\, Z_*, a_*, Z)$

is the conditional sum of squares function. The quantities of $\widehat{\emptyset}$, $\hat{\mu}$, and $\hat{\theta}$, which maximum equation (3.67) are called the conditional maximum likelihood estimators. Because of ln L_* $(\emptyset, \mu, \theta, \sigma_a^2)$ contains the data only through $S_*(\emptyset, \mu, \theta)$, these estimators are the same as the conditional least squares estimators got from minimizing the conditional sum of squares function $S_*(\emptyset, \mu, \theta)$, which don't contain the parameter σ_a^2.

By assuming $a_p = a_{p-1} = \dots = a_{p+1-q} = 0$ and replacing Z_t by the sample mean $\bar{Z}$ the conditional sum of squares function $S_*(\emptyset, \mu, \theta)$ can be written as becomes

$$S_*(\emptyset, \mu, \theta) = \sum_{t=p+1}^{n} a_t^2(\emptyset, \mu, \theta |\, Z)$$

After obtaining the parameter estimates $\widehat{\emptyset}$, $\hat{\mu}$, and $\hat{\theta}$, the estimate $\hat{\sigma}_a^2$ is calculated from

$$\hat{\sigma}_a^2 = \frac{S_*(\emptyset,\mu,\theta)}{\text{d.f.}}$$

Where the number of degrees of freedom d.f equals the number of terms used in the sum of $S_*\left(\widehat{\emptyset}, \hat{\mu}, \hat{\theta}\right)$ minus the number of parameters estimated.

Unconditional Maximum Likelihood Estimation and Back casting Method

A further improvement in estimation, Box, Jenkins, and Reinsel (1994) suggest the following unconditional log-likelihood function:

$$\ln L(\emptyset, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S(\emptyset,\mu,\theta)}{2\sigma_a^2} \tag{3.68}$$

where S$(\emptyset, \mu, \theta)$ is the unconditional sum of squares function given by

$$S(\emptyset, \mu, \theta) = \sum_{t=-\infty}^{n} [E(a_t|\emptyset, \mu, \theta, Z)]^2 \tag{3.69}$$

And E$(a_t|\emptyset, \mu, \theta, Z)]^2$ is the conditional expectation of a_t given $\emptyset, \mu, \theta,$ and Z.

The quantities $\widehat{\emptyset}, \hat{\mu}$ and $\hat{\theta}$ that maximize function (3. 68) are called unconditional maximum Likelihood estimators. Again, Since lnL$(\emptyset, \mu, \theta, \sigma_a^2)$ involves the data only through S$(\emptyset, \mu, \theta)$, these unconditional maximum likelihood estimators are equivalent to the unconditional least squares estimators obtained by minimizing S$(\emptyset, \mu, \theta)$.

3.12.3 Model Diagnostic Checking

Time series model building starts with model identification and parameter estimation. After parameter estimation, it has to assess model adequacy by checking whether the model assumptions are satisfied. The basic assumption is that the $\{a_t\}$ are white noise. That is, the a_t's are uncorrelated random shocks with zero mean and constant variance. For any estimated model, the residuals $\hat{a}_t$`s are estimates of these unobserved white noise a_t's. Hence model diagnostic checking is accomplished through a careful analysis of the residual series $\{\hat{a}_t\}$. Because this residual series is the product of parameter estimation, the model diagnositc checking is usually contained in the estimation phase of a time series package.

(1) To check whether the errors are normally distributed, one can construct a histogram of the standardized residuals $\hat{a}_t/\hat{\sigma}_a$ and compare it with the standard normal distribution using the chi-square goodness of fit test or even Tukey's simple five-number summary.

(2) To check whether the variance is constant and examine the plot of residuals.

(3) To check whether the residuals are approximately white noise, This compute the sample ACF and sample PACF of the residuals to see whether they do not form any pattern and are all statistically insignificant.

Then another useful test is the portmanteau lack of fit test. This test uses all the residual sample ACFs as a unit to check the joint null hypothesis.

Hypothesis H_0: $\rho_1 = \rho_2 = \cdots = \rho_K = 0$, The residual are not autocorrelated.

H_1 : The residual are not autocorrelated.

Test statistic : $Q = n(n+2)\sum_{k=1}^{k}(n-k)^{-1}\hat{\rho}_k^2$ (3.70)

Critical Vale : K= $\mathcal{X}^2_{(\alpha,k\ -\mathrm{m})}$

Decision Rule : Q > K; Reject H_0

This test statistic is the modified Q statistic originally proposed by Box and Pierce (1970). Under the null hypothesis of model adequacy, Ljung and Box (1978) and Ansley and Newbold (1979) show that the Q statistic approximately follows the $\mathcal{X}^2_{(\alpha,k\ -\mathrm{m})}$ distribution, where m = p+q is the number of parameters estimated in the model.

Based on the results of these residual analyses, if the entertained model is inadequate, a new model can be easily derived.

3.12.4 Model Selection Criteria

In time series analysis, there are numerous model selection criteria. In this section, the process of suitable model is illustrated below. Firstly, making test for stationary is made. Statistical inference about the structure of a time series is requirement to make some assumptions about the structure. “The basic idea of the stationary is that, the probability laws that govern the behavior of the process do not change over time. This is to determine that the process is in a statistical equilibrium (Takyi Appiah, S., Otoo, H., Nabubie, I.B., 2015).” If the data are not stationary because the series display a long term pattern and the mean is not zero, it will have to be difference the graph in order to obtain stationary. The stationary is done when there is found an average pattern of the graph. Then, ACF and PACF are used for model identification. Maximum likelihood method is used to estimate the parameters in this study. As there are many feasible adequate models, the selection criteria is normally based on summary statistics from residuals computed from a fitted model or on forecast errors calculated from the out-sample forecast. Some model selection criteria are based on residuals. In this study, coefficient of determination (R^2 or r^2), Root Mean Square Error (RMSE), is a measure of prediction accuracy of a forecasting method in statistics and the Bayesian information criterion (BIC) are used for criteria of tentative ARIMA model. The higher value of R^2, the lower value of RMSE, normalized BIC are used as criteria for selecting the most adequate model among all feasible models.

Minimum Mean Square Error Forecasts for ARIMA Models

The general nonstationary ARIMA (*p, d, q*) model with d $\neq$ 0, i.e.,

$$\emptyset (B)(1- B)^d Z_t = \theta(B)\ a_t,$$

Where $\emptyset (B) = (1- \emptyset B - \dots - \emptyset_p (B^{\emptyset})$ is a stationary AR operator and $\theta(B) = (1- \theta_1 B - \dots - \theta_q (B^q)$ is an invertible MA operator, respectively. Although for this process the mean and the second order moment such as the variance and the autocovariance functions vary over time, the complete evolution of the process is completely determined by a finite number of fixed parameters. Hence, the forecast of the process can be viewed as the estimation of a function of these parameters and obtain the minimum mean square error forecast using a Bayesian argument. It is well known that using this approach with respect to the mean square error criterion, which corresponds to a squared loss function, when the series is known up to time n, the optimal forecast of Z_{n+l} is given by its conditional expectation E(Z_{n+l} Z_n, Z_{n-l},…). The minimum mean square error forecast for the stationary ARMA model discussed earlier is, of course, a special case of the forecast for the ARIMA (*p, d, q*) model with *d* = 0.

To derive the variance of the forecast for the general ARIMA model, we rewrite the model at time t + l in an AR representation that exist because the model is invertible, Thus,

$$\pi(B)\, Z_{t+l} = a_{t+l}, \tag{3.71}$$

Where

$$\pi(B) = 1 - \sum_{i=0}^{n} \pi_j\, Bj = \frac{\emptyset\,(B)(1-B)^d}{\theta(B)} \tag{3.72}$$

or, equivalently

$$Z_{t+l} = \sum_{i=0}^{n} \pi_j\, Z_{t+l\sim j} + a_{t+l} \tag{3.73}$$

Following Wegman (1986), we apply the operator

$$1 + \psi\, B + \ldots + \psi_{l-1}\, B^{t-1} \tag{3.74}$$

To (3.74) and obtain

$$\sum_{j=0}^{\propto} \sum_{k=0}^{l-1} \pi_j\, Z_{t+l-j-k} + \sum_{k=0}^{l-1} \psi_k\ a_{t+l-1} = 0, \tag{3.75}$$

Where π_0 = -1 and ψ_0 = 1. It can be easily shown that

$$\sum_{j=0}^{\propto} \sum_{k=0}^{l-1} \pi_j Z_{t+l-j-k} = \pi_0 Z_{t+l} + \sum_{m=1}^{l-1} \sum_{i=0}^{m} \pi_{m-i}\, \Psi_i Z_{t+l-m}$$

$$+ \sum_{j=1}^{\propto} \sum_{i=0}^{l-1} \pi_{l-1+j-i}\, \Psi Z_{t-j+l} \tag{3.76}$$

Choosing ψ weights so that

$$\sum_{i=0}^{m} \pi_{m-i} \psi_i = 0, \text{ for } = 1,2, \ldots, l\text{-}1 \tag{3.77}$$

We have

$$Z_{t+l} = \sum_{j=1}^{\alpha} \pi_j^{(l)}\, Z_{t-j+l} + \sum_{i=0}^{l-1} \psi_i\ a_{t+l-i}, \tag{3.78}$$

Where

$$\pi_j^{(l)} = \sum_{i=0}^{l-1} \pi_{l-1+j-1}\, \Psi_i. \tag{3.79}$$

Thus, given Z_t, for $t \leq n$, we have

$$\hat{Z}_n(l) = E\,(Z_{n+l} \mid Z_t,\ t \leq n)$$

$$= \sum_{j=1}^{\alpha} \pi_j^{(l)} Z_{n-j+1}, \tag{3.80}$$

Because $E\,(a_{n+j} \mid Z_t,\ t \leq n) = 0$, for $j > 0$. The forecast error is

$$e_n\,(l) = Z_{n+l} - \hat{Z}_n(l)$$

$$= \sum_{i=0}^{l-1} \psi_j\ a_{n+l-j}, \tag{3.81}$$

Where the ψ_j weights, by (3.77), can be calculated recursively from the π_j weights as follow:

$$\psi_j = \sum_{i=0}^{j-1} \pi_{j-i}\, \psi_i,\ \ j = 1, \ldots, l-1. \tag{3.82}$$

Because $E(e_n(l) \mid Z_t,\ t \leq n) = 0$, the forecast is unbiased with the error variance

$$\text{Var}\,(e_n(l)) = \sigma_a^2 \sum_{j=0}^{l-1} \psi_j^2. \tag{3.83}$$

For a normal process, the $(1-\alpha)$ 100% forecast limits are

$$\hat{Z}_n(l) \pm N_{\propto/2}\ [1 + \sum_{j=0}^{l-1} \psi_j^2]^{1/2}\ \sigma_a \tag{3.84}$$

Where $N_{\propto/2}$ is the standard normal deviate such that $P\,(N > N_{\propto/2}) = \frac{\propto}{2}$.

3.12.5 Model Building and Forecasting for Seasonal Model

Seasonal models are special forms of ARIMA models which described the model identification, parameter estimation, diagnostic checking, and forecasting for these models follow the same general methods of ARIMA models which have already discussed in detail for this chapter.

CHAPTERE IV
TIME SERIES ANALYSIS FOR FORECASTING OF RAINFALL DATA

4.1 Data Information

The dataset includes the historical normal rainfall Stations for each month from January 2006 to December 2018. The number of rainfall data in Chauk and Shwebo townships are obtained from Department of Meteorology and Hydrology in Mandalay. The present study is carried out for Central Dry Zone in Myanmar for a period of 13 years. This data is used for the Yearly, Monthly and Seasonal Rainfall Probability Analysis. These data series will analyzed by using the Box-Jenkins Method (1970) and this method consists of four steps: model identification, parameter estimation, diagnostic checking and forecasting.

The autoregressive, moving average and autoregressive moving average models are used in stationary time series analysis. A time series is supposed to be stationary, if the mean of the series and the covariance among its observation do not change over time and do not follow any trend. In practice, most time series are non-stationary, so to fit stationary models, it is necessary to get rid of the non-stationary source of variation. To solve out this, introduced by ARIMA model which generally overcomes this limitation by the non-stationary data into a stationary. Box and Jenkins have generalized this model to deal with seasonality. Their proposed model is identified as the Seasonal ARIMA (SARIMA) model.

Firstly, the monthly rainfall data are described and analyzed. The SACF and SPACF functions are used to identify the order p and q of the model in the identification stage.

4.2 Descriptive Statistics of Volumes of Rainfall Data in Chauk Townships (2016-2018)

The monthly rainfall data of Minimum, Maximum and Average was shown in Table (4.2).

Table (4.2) Volumes of Rainfall data in Chauk (2006-2018)

Year / Month	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Max	Min	Average
Jan	0	0	10	0	0	0	0	0	0	25	0	0	41	41	0	6
Feb	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Mar	0	0	0	0	5	8	8	0	0	0	0	8	0	8	0	2
Apr	8	0	0	3	0	12	0	5	0	13	0	85	2	85	0	10
May	55	249	96	79	55	83	5	46	32	11	79	113	137	249	5	80
Jun	128	81	50	42	102	54	25	84	52	63	135	53	192	192	25	82
Jul	99	12	26	16	3	12	28	15	34	111	71	47	35	111	3	39
Aug	134	48	38	35	92	213	70	33	34	122	260	77	71	260	33	94
Sep	176	125	110	63	125	73	154	226	78	85	189	132	63	226	63	123
Oct	172	89	115	95	387	290	68	218	70	168	248	132	70	387	68	163
Nov	8	27	0	0	0	0	7	0	37	10	50	17	0	50	0	12
Dec	0	0	0	0	14	5	0	0	0	1	0	9	13	14	0	3
Max	176	249	115	95	387	290	154	226	78	168	260	132	192	387	78	194
Min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Average	65	53	37	28	65	63	30	52	28	51	86	56	52	387	0	51

Sources : Monthly Rainfall data: Department of Meteorology and Hydrology in Mandalay

4.3 Descriptive Statistics of Volumes of Rainfall Data in Shwebo Townships (2016-2018)

The monthly rainfall data of Minimum, Maximum and Average was shown in Table (4.2).

Table (4.3) Volumes of Rainfall data in Shwebo Township (2006-2018)

Year / Month	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	Min	Max	Average
Jan	0	0	32	0	0	0	0	0	0	1	0	0	23	0	32	4
Feb	0	9	0	0	0	0	0	0	15	0	3	0	0	0	15	2
Mar	0	0	0	0	4	51	4	0	0	14	0	11	0	0	51	6
Apr	89	0	2	28	12	138	8	12	11	17	23	8	17	0	138	28
May	76	188	113	181	55	159	16	60	44	94	87	201	120	16	201	107
Jun	218	133	91	192	165	87	48	60	61	14	161	83	229	14	229	119
Jul	158	21	53	16	27	69	78	47	20	368	106	143	45	16	368	89
Aug	195	176	105	137	157	279	88	302	238	116	427	243	48	48	427	193
Sep	182	82	62	139	152	88	279	221	259	211	124	196	79	62	279	160
Oct	104	146	137	60	362	284	115	191	68	203	74	140	219	60	362	162
Nov	98	47	6	0	0	0	20	0	3	7	17	48	0	0	98	19
Dec	3	0	0	0	34	5	0	2	0	0	0	11	0	0	34	4
Min	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	2
Max	218	188	137	192	362	284	279	302	259	368	427	243	229	62	427	193
Average	94	67	50	63	81	97	55	75	60	87	85	90	65	18	186	74

Sources : Monthly Rainfall data: Department of Meteorology and Hydrology in Mandalay

Chauk's the average rainfall is peak on May to October from 2006 to 2018. The lowest contribution to monthly rainfall is less in January, February, March, April, November and December and Figure (4.2) show that 2010 has the maximum rainfall contribution is at 387mm. This is followed by years 2011 (290mm), 2013 (218mm) and 2016 (248mm), all of which have significant rainfall.

Shwebo's the average rainfall is peak on May to October from 2006 to 2018. The lowest contribution to monthly rainfall is less in January, February, March, April, November and December and Figure (4.3) show that 2016 has the maximum rainfall contribution is at 427mm. This is followed by years 2015 (368mm), 2010 (362mm) and 2018 (229mm), all of which have significant rainfall.

4.4 Plot of original monthly rainfall data in Chauk Township

The original time series was plotted with the monthly rainfall data recorded from 2006 to 2018 employing the Box-Jenkin ARIMA model-building procedure.

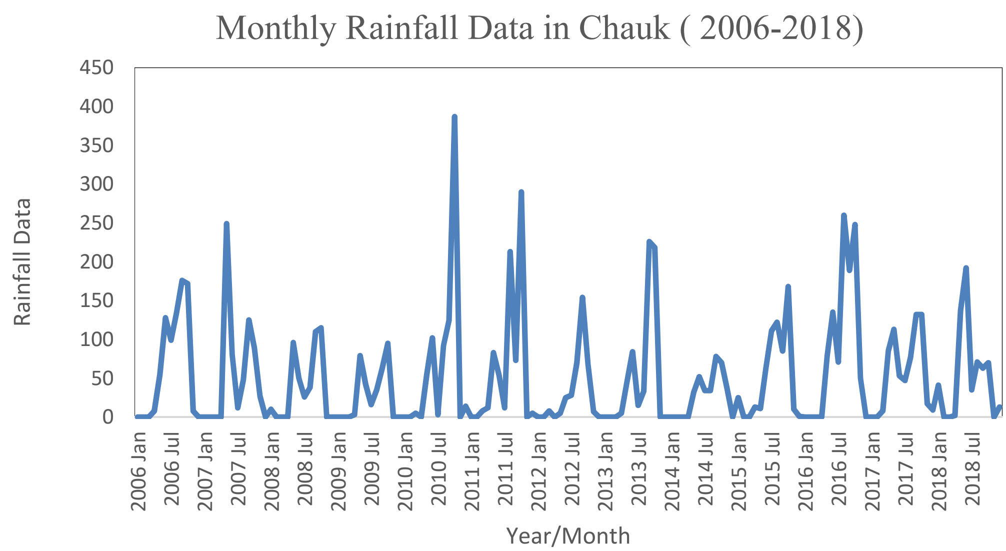


Figure (4.4) Plot of original monthly rainfall data in Chauk Township (2006-2018)

The plot of data is given in Figure (4.4), it indicates that the series is stationary in the mean but may not be stationary in variance. It suggests that square root transformation for variance stabilization. This series indicate a quite obvious that there exists seasonality in the data has a seasonal pattern with peaks and valleys in the same months of the year. So, this series need to check whether the rainfall series is seasonality or not.

4.5 Test of Seasonality

The results for testing the seasonality in Chauk Township (2006-2018) are shown in Table (4.5).

Table (4.5) ANOVA Table for rainfall series in Chauk Township

Source of variation	Sum of Square	Degree of Freedom	Mean Square Error	F-ratio
Due to Month	SSM= 443233.86	11	MSM= 40293.99	19.77
Due to Year	SSY= 41637.44	12	MSY= 3469.79	
Error (Residual)	SSE= 268992.72	132	MSE=2037.82	
Total	SST= 753864.02	155		

At 5% level of significance, the critical value K = F (0.05, 11, 132) is 1.75. Meanwhile the computed F-value = 19.77 is greater than K=1.75, it can be identified that the monthly rainfall series exists seasonality.

4.6 Seasonal Variation

The seasonal variation of monthly rainfall data series in Chauk from 2006 to 2018 are computed by the ratio to moving averages method.

Table (4.6) Seasonal Indexes for rainfall data in Chauk (2006-2018)

Month	Seasonal Index
January	0.1642
February	0.0376
March	0.0737
April	0.2074
May	1.5726
June	1.7615
July	0.8056
August	1.7025
September	2.4005
October	2.9522
November	0.2326
December	0.0903

The seasonal variation of monthly rainfall data series from 2006 to 2018 is calculated by the ratio to moving average method under multiplicative decomposition of time series which consists of 156 observations and it was shown in Appendix. The result of Table (4.6), the seasonal index are shown that the lowest value of seasonal index is in February, March and December and the month of May, June, August, September and October have the highest value of seasonal index. The peak period is in September and the lowest period is February with the seasonal index of rainfall data series in Chauk Township.

4.7 Box-Jenkins ARIMA Model for Rainfall data series in Chauk Township

The monthly rainfall series in Chauk Township from 2006 to 2018, applied Box-Jenkins Methodology which is Model Identification, Parameter Estimation, Diagnostic Checking and Forecasting. The data series consist of 156 observations.

4.7 Identification

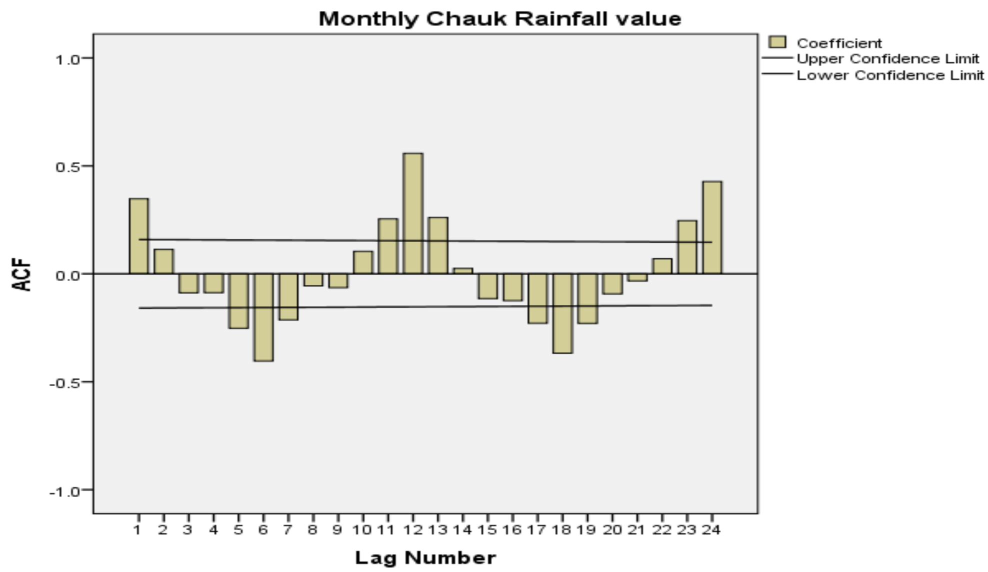
The SACF and SPACF function of rainfall values are calculated and shown in Table (4.7.1) and Figure (4.7.1). These results pointed out that the seasonality trend exists and it should be eliminated to achieve stationary of the time series data.

Table (4.7.1) Sample ACF and Sample PACF for the original series Z_t of Rainfall in Chauk Township

	$\hat{\rho}_k$ for $\{Z_t\}$												
Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	1-12	0.347	0.113	-0.088	-0.088	-0.252	-0.404	-0.213	-0.056	-0.064	0.104	0.254	0.558
Z_t	St-E	0.079	0.079	0.079	0.079	0.078	0.078	0.078	0.077	0.077	0.077	0.077	0.076
	13-24	0.26	0.025	-0.115	-0.124	-0.229	-0.367	-0.229	-0.093	-0.033	0.07	0.246	0.428
	St-E	0.076	0.076	0.076	0.075	0.075	0.075	0.075	0.074	0.074	0.074	0.073	0.073

	$\hat{\phi}_{kk}$ for {Zt}												
Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	1-12	0.347	-0.009	-0.142	-0.014	-0.23	-0.313	0.025	-0.003	-0.18	0.137	0.129	0.383
Z_t	St-E	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	13-24	-0.016	-0.118	-0.058	0.016	-0.022	-0.104	-0.05	-0.094	-0.02	-0.018	0.058	0.066
	St-E	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08

In Figure (4.7.1) the values of SACF are statistically significant at lag 1,6,12 and 24 and shows a damping sine wave. The SPACF has relatively large spikes at 1, 6 and 12.



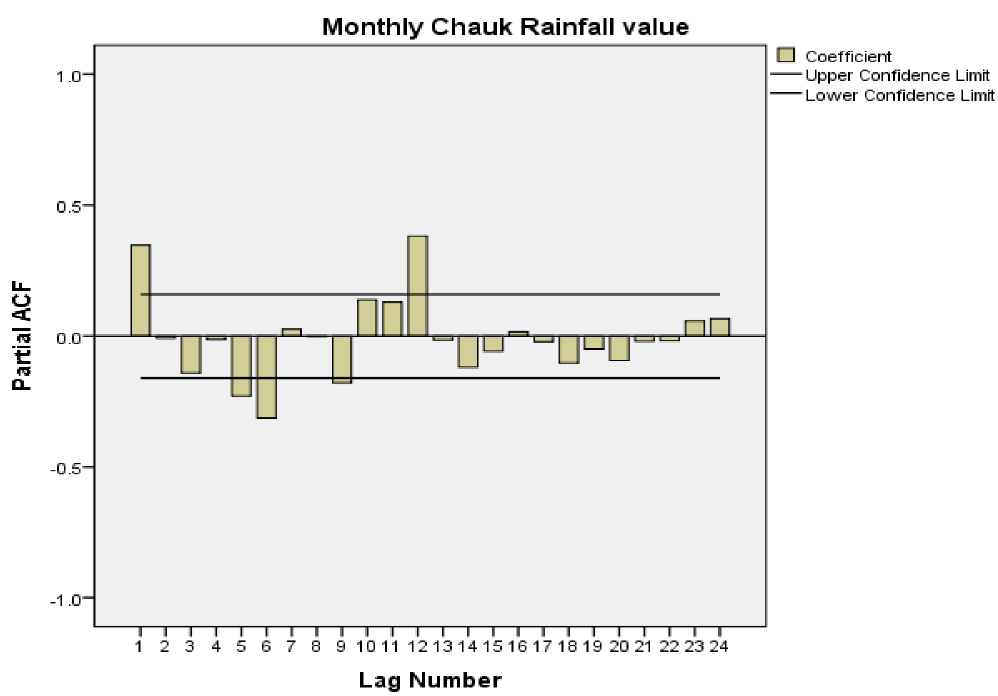


Figure (4.7.1) Sample Correlogram for Original Series of Rainfall value in Chauk Township

With reference to Figure (4.4), the plot is clearly shown the series may not be stationary in variance. So, this series required transformation for variance stabilization. It suggests that a square root transformation applied to the data. For power transformation, Residual mean square errors calculated with the two transformations which are shown in Table (4.7.2).

Table (4.7.2) Residual mean square errors in the power transformation

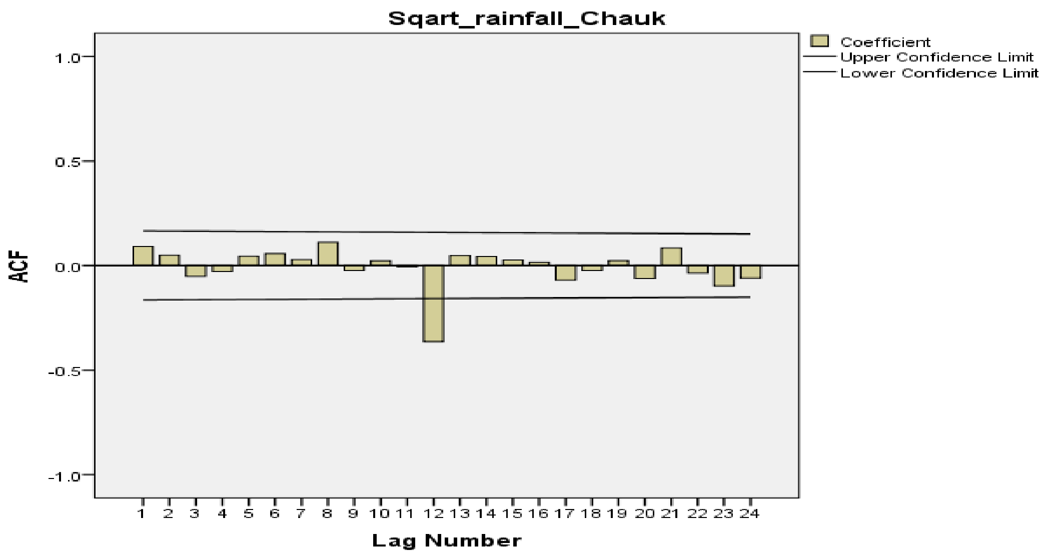
Transformation	Residual mean square errors (RMSE)
Square root	79.59
Logarithmic	83.54

According to the result, Square root transformation of RMSE is lower than Logarithmic transformation value. So, it suggests that a square root transformation is suitable to apply it to the original series to obtain the variance stationary. The sample ACF and sample PACF of square root transformation series $\sqrt{Z_t}$ were computed as shown in Table (4.7.3) and Figure (4.7.3).

Table (4.7.3) Sample ACF and Sample PACF Function for square root transformation of Rainfall values in Chauk Townships

(a)	$\hat{\rho}_k$ for $\{\sqrt{Z_t}\}$												
Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	1-12	0.347	0.113	-0.088	-0.088	-0.252	-0.404	-0.213	-0.056	-0.064	0.104	0.254	0.558
$\sqrt{Z_t}$	St-E	0.079	0.079	0.079	0.079	0.078	0.078	0.078	0.077	0.077	0.077	0.077	0.076
	13-24	0.260	0.025	-0.115	-0.124	-0.229	-0.367	-0.229	-0.093	-0.033	0.070	0.246	0.428
	St-E	0.076	0.076	0.076	0.075	0.075	0.075	0.075	0.074	0.074	0.074	0.073	0.073

(b)	$\hat{\emptyset}_{kk}$ for $\{\sqrt{Z_t}\}$												
Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	1-12	0.347	-0.009	-0.142	-0.014	-0.230	-0.313	0.025	-0.003	-0.180	0.137	0.129	0.383
$\sqrt{Z_t}$	St-E	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080
	13-24	-0.016	-0.118	-0.058	0.016	-0.022	-0.104	-0.050	-0.094	-0.020	-0.018	0.058	0.066
	St-E	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080	0.080



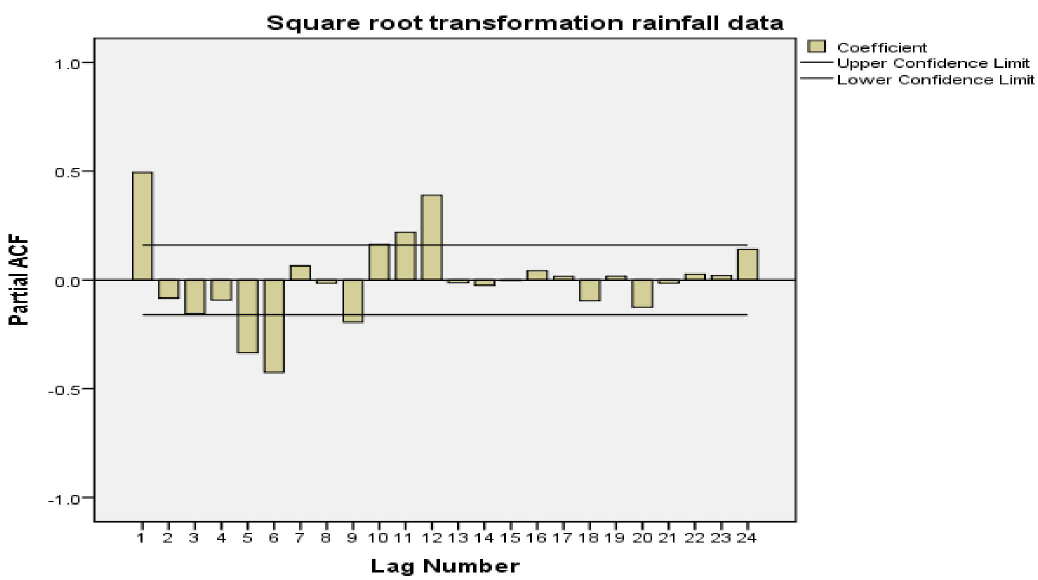


Figure (4.7.3) Sample Correlogram for square root transformation of Rainfall values in Chauk Township

In Figure (4.7.3) the sample ACF illustrate a single spike at lag 12. The sample PACF has relatively large spike at lags 1,5,6 and12. This series have strong seasonal variation that the first seasonal differencing is called for. To eliminate seasonality, the sample ACF and sample PACF of first seasonal differenced series $(1 - B^{12})\sqrt{Z_t}$ were computed as shown in Table (4.7.4) and Figure (4.7.4).

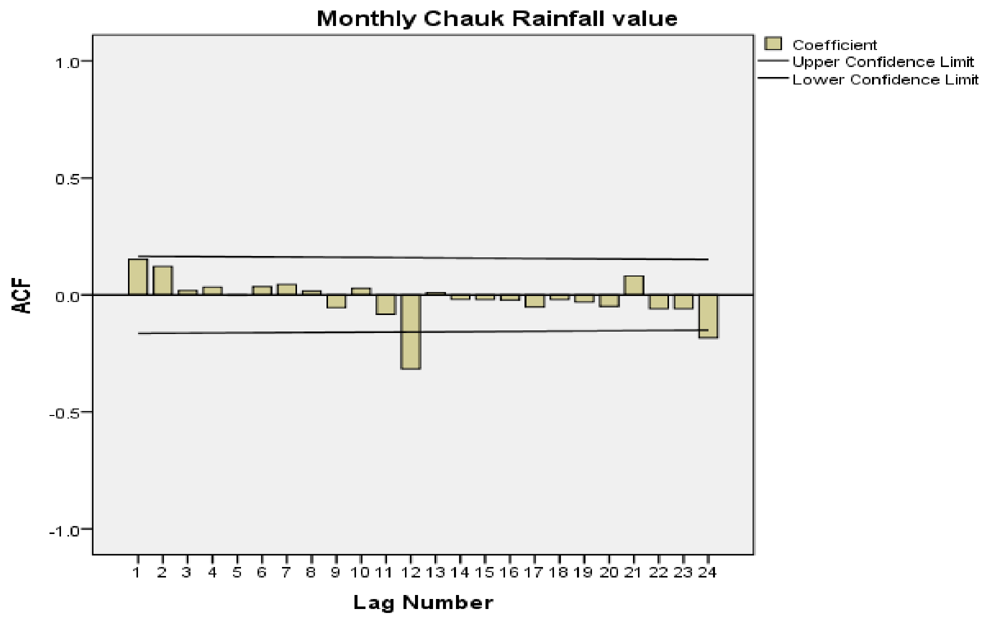
Table (4.7.4) Sample ACF and Sample PACF Function for First Seasonal Difference transformation Series (W_t) of Rainfall values in Chauk Townships

(a) $\hat{\rho}_k$ for $\{W_t = (1 - B^{12})\sqrt{Z_t}\}$

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	1-12	0.151	0.121	0.018	0.033	-0.002	0.035	0.045	0.016	-0.054	0.029	-0.083	-0.316
W_t	St-E	0.082	0.082	0.082	0.082	0.081	0.081	0.081	0.08	0.08	0.08	0.08	0.079
	13-24	0.009	-0.019	-0.02	-0.022	-0.052	-0.02	-0.031	-0.049	0.081	-0.059	-0.059	-0.184
	St-E	0.079	0.079	0.078	0.078	0.078	0.077	0.077	0.077	0.076	0.076	0.076	0.076

(b) $\hat{\emptyset}_{kk}$ for $\{W_t = (1 - B^{12})\sqrt{Z_t}\}$

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	1-12	0.151	0.101	-0.014	0.021	-0.010	0.032	0.038	-0.003	-0.067	0.044	-0.084	-0.317
W_t	St-E	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
	13-24	0.126	0.035	-0.046	0.006	-0.055	0.018	0.020	-0.072	0.082	-0.029	-0.139	-0.280
	St-E	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083



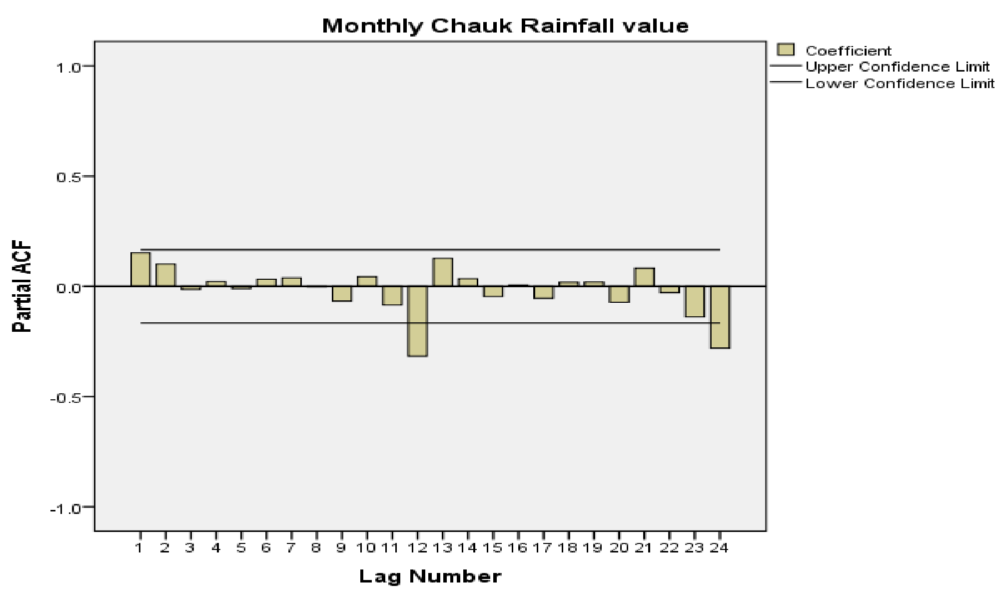


Figure (4.7.4) Sample Correlogram for First Seasonal Difference transformation Series of Rainfall values in Chauk Township

According to the Figure (4.7.4), both of the values of SACF and SPACF are same spike after lags 12 lie inside of the confidence limits respectively. It can be suggested that the series $(1-B^{12})\sqrt{Z_t}$ might be considered SAR (1) model as tentative model for this series. Thus the tentative model for the series is SAR (1) process and checked parameter estimation as shown in Table (4.8.1).

$$(1 - \Phi B^{12})\sqrt{Z_t} = \theta_0 + a_t$$

4.8 Parameter Estimation

Using SAR (1) model, the estimated parameters with their statistics were shown in Table (4.8.1) and Figure (4.8.1).

Table (4.8.1) Estimated Parameters and Model Statistics for SAR (1) Model of Rainfall values in Chauk Township

	Estimate	SE	t	P-value
Constant	52.540	9.944	5.284	.000
Φ	.591	.065	9.085	.000

The following estimated model was obtained

$$(1 + 0.591B^{12})\, Z_t = 52.540 + a_t$$
$$(0.065) \qquad\qquad (9.944)$$

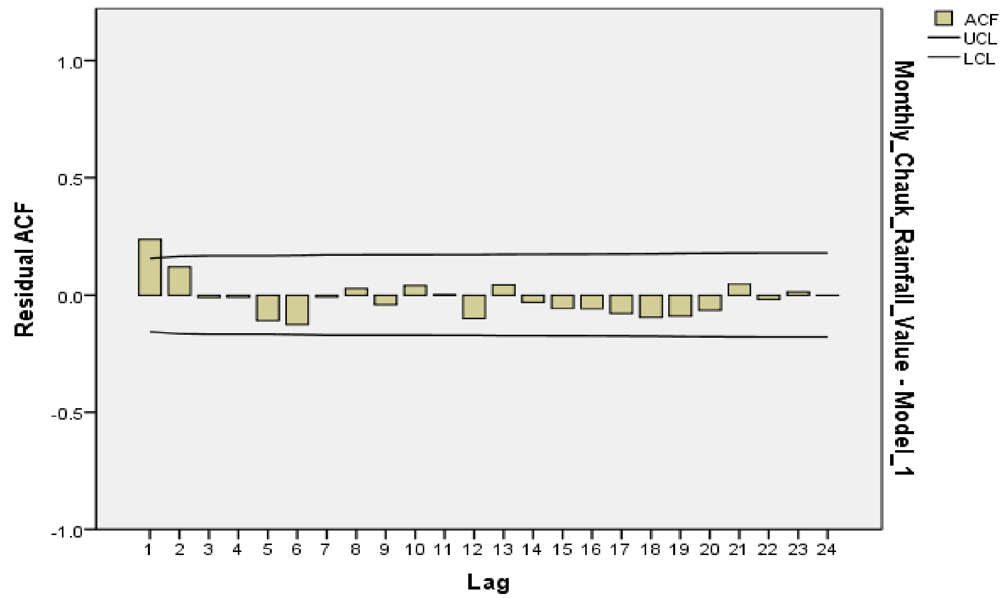
The estimation of SAR (1) Model of rainfall values in Chauk Township give θ_0 = 52.540 with the estimated standard error 9.944. The estimated parameter Φ = 0.591, the estimated standard error 0.065, the test statistics t is 9.085 with P-value is 0.000. So, there is no evidence to reject the null hypothesis $\Phi = 0$.

Furthermore, the sample ACFs and the sample PCFs of residual for the above tentative model were shown in Table (4.8.2) and Figure (4.8.2).

Table (4.8.2) Estimated Autocorrelation and Partial Autocorrelation Function of Residual for SAR (1) Model of Rainfall values in Chauk Township

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	.236	.120	-.011	-.010	-.109	-.127	-.010	.028	-.041	.039	.002	-.099
St-E	.080	.084	.086	.086	.086	.086	.088	.088	.088	.088	.088	.088
13-24	.043	-.031	-.057	-.059	-.078	-.095	-.090	-.065	.047	-.019	.014	-.001
St-E	.089	.089	.089	.089	.089	.090	.090	.091	.091	.091	.091	.091

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	.236	.068	-.057	-.004	-.105	-.086	.061	.034	-.072	.054	-.028	-.127
St-E	.080	.080	.080	.080	.080	.080	.080	.080	.080	.080	.080	.080
13-24	.124	-.053	-.084	.005	-.088	-.093	-.001	-.056	.038	-.018	-.039	-.034
St-E	.080	.080	.080	.080	.080	.080	.080	.080	.080	.080	.080	.080



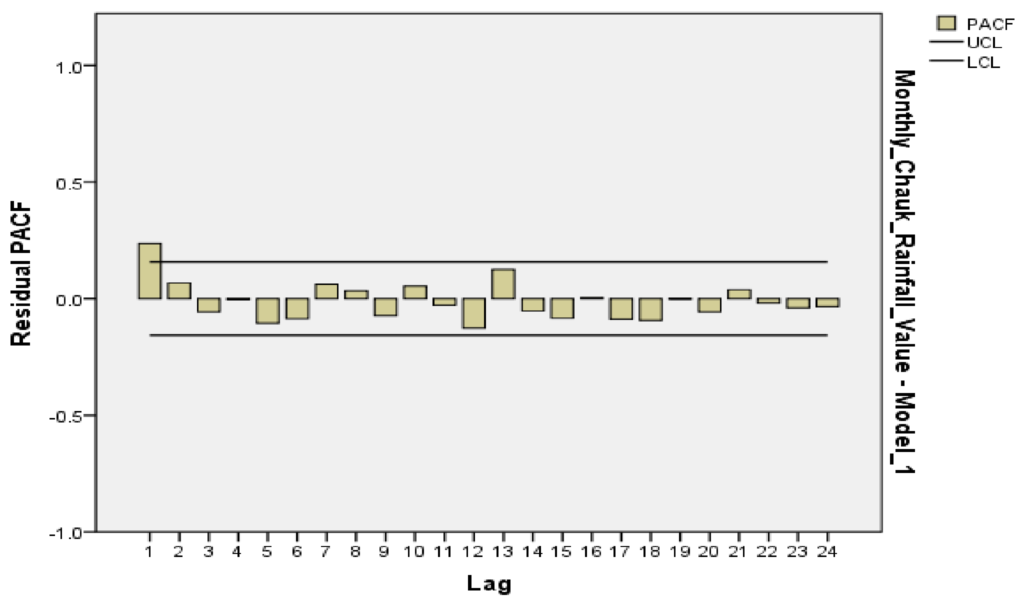


Figure (4.8.2) Sample ACF and Sample PACF of Residual values for SAR (1) Model of Rainfall Series in Chauk Township

In Figure (4.8.2), the sample ACF and Sample PACF of residual is cut off and significant spike at lag 1 and this model exhibit a pattern that the residual series are not white noise process. That the model to an ARIMA (1,0,0) x $(1,1,0)_{12}$ model considered as another tentative model that is

$$(1-\emptyset B)(1-\Phi B^{12})\sqrt{Z_t} = \theta_0$$

Seasonal ARIMA (1,0,0) x $(1,1,0)_{12}$ model, the estimated parameter with their statistics were shown in Table (4.8.3).

Table (4.8.3) Estimated Parameters and Model Statistics for Seasonal ARIMA (1,0,0) x $(1,1,0)_{12}$ Model of Rainfall values in Chauk Township

	Estimate	SE	t	P-value
Constant	.012	.214	.055	.956
$\emptyset$	.143	.083	1.707	.090
Φ	-.437	.080	-5.481	.000

The following estimated model was obtained

$$(1 - 0.143B)(1 + 0.437B^{12})\sqrt{Z_t} = 0.012$$

(0. 083) (0.080)

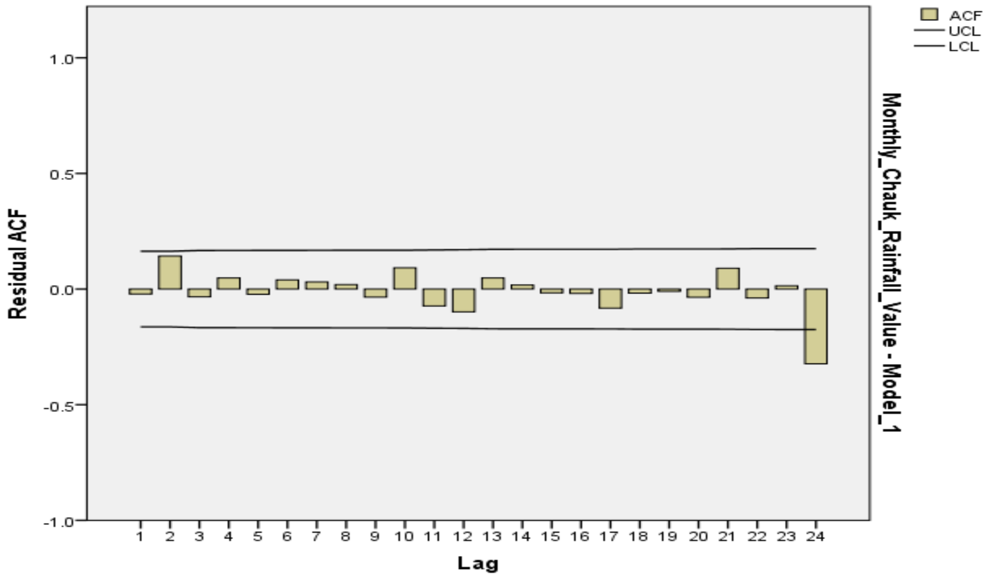
The estimation (1,0,0) x $(1,1,0)_{12}$ Model of rainfall values in Chauk Township give θ_0 = 0.012, the estimated standard error is 0.214 with the test statistics is 0.055. The estimate parameter $\emptyset = 0.143$, the estimated standard error is 0.083, the test statistics is 1.707 with p-value is 0.090 and Φ = -0.437, the estimated standard error 0.080 with p-value is 0.000. So, there is no evidence to reject the null hypothesis.

Moreover, the sample ACFs and the sample PCFs of residual for the tentative model were shown in Table (4.8.4) and Figure (4.8.4).

Table (4.8.4) Estimated Autocorrelation and Partial Autocorrelation Function of Residual for (1,0,0) x $(1,1,0)_{12}$ Model of Rainfall values in Chauk Township

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	-.016	.112	-.074	.039	.007	.069	.035	.103	-.020	.064	-.017	-.087
St-E	.083	.083	.084	.085	.085	.085	.085	.085	.086	.086	.087	.087
13-24	.056	.084	.018	.006	-.076	-.045	.006	-.044	.096	-.005	-.062	-.237
St-E	.087	.088	.088	.088	.088	.089	.089	.089	.089	.090	.090	.090

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	-.016	.112	-.072	.026	.024	.058	.038	.093	-.018	.046	-.003	-.114
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083
13-24	.061	.093	-.016	-.012	-.066	-.055	.026	-.044	.073	.015	-.093	-.254
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083



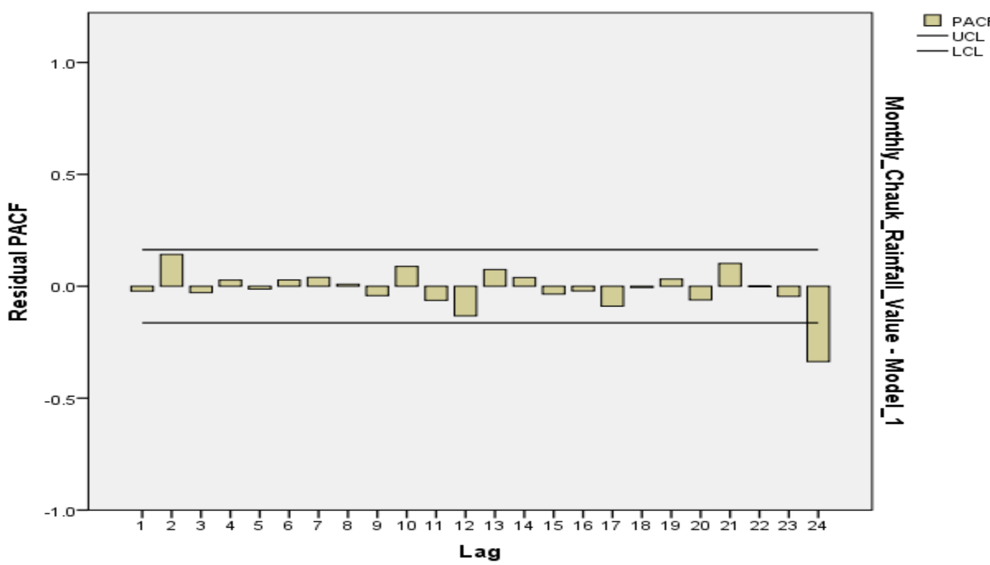


Figure (4.8.4) Sample ACF and Sample PACF of Residual values for (1,0,0) x $(1,1,0)_{12}$ Model of Rainfall values in Chauk Township

In Figure (4.8.4), the sample ACF and Sample PACF of residual have a significant spike at lag 24. It indicates that the residual series are not white noise process and the series exhibit the patterns. So, it can be concluded that the tentative SAR (1) model is inadequate. Then the another SAR (2) model was considered to a seasonal ARIMA (0,0,0) x $(2,1,0)_{12}$ model, that is

$$(1 - \Phi_1 B^{12} - \Phi_2 B^{24})\sqrt{Z_t} = \theta_0$$

Using Seasonal ARIMA (0,0,0) x $(2,1,0)_{12}$ model, the estimated parameter of their statistics were shown in Table (4.8.5).

Table (4.8.5) Estimated Parameters and Model Statistics for Seasonal ARIMA (0,0,0) x $(2,1,0)_{12}$ Model of Rainfall values in Chauk Township

	Estimate	SE	T	P-value
Constant	.052	.141	.367	.714
Φ_1	-.549	.088	-6.224	.000
Φ_2	-.308	.093	-3.326	.001

The estimated model was obtained:

$$\underset{(0.088)}{(1 + 0.549B^{12}} + \underset{(0.093)}{0.308B^{24})}\sqrt{Z_t} = \underset{(0.141)}{0.052}$$

According to Table (4.8.5), it can be found that this table gives $\theta_0 = 0.052$ and the estimated standard error is 0.141. The estimated parameter of Φ_1= -0.549, the estimated standard error is 0.088, the test statistics t is -6.224 with their p-value are 0.000. Hence there are no evidence to reject the null hypothesis Φ_1 =0. And the estimated parameter is Φ_2 = -0.308, the estimated standard error is 0.093, and the test statistics t is -3.326 with P-value is 0.001 at 1% level of significant. Hence there are no evidence to reject the null hypothesis Φ_2 = 0.

4.9 Diagnostic Checking

Checking the model adequacy for the residual ACF and PACF of the modified model as shown in Table (4.9.1) and Figure (4.9.1).

$$\gamma_k(\hat{a}_t) \pm 2.\widehat{S.E}\,[\gamma_k\,(\hat{a}_t)]$$

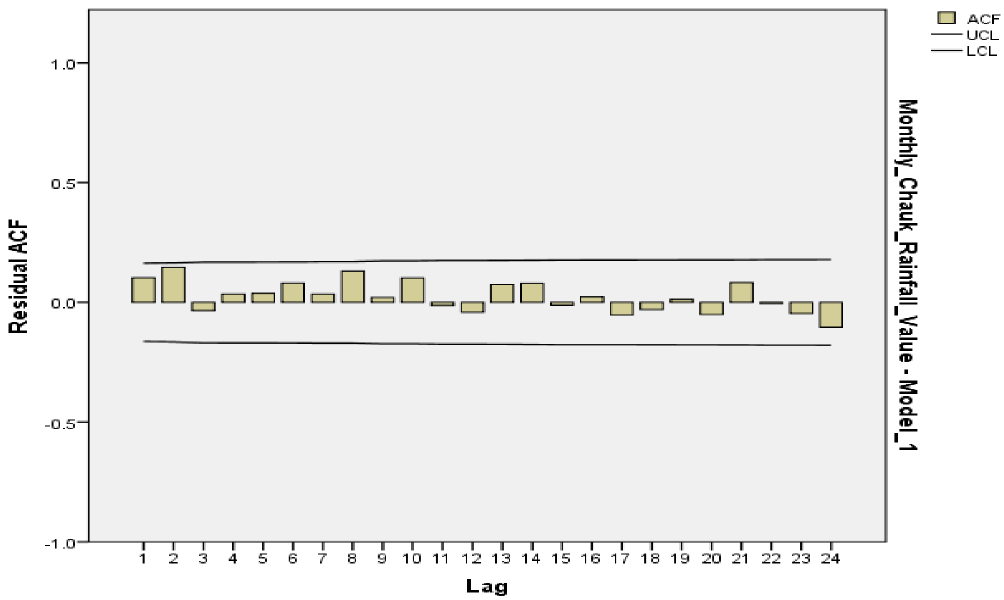
Where,

$$\widehat{S.E}\,[\gamma_k\,(\hat{a}_t)] = \frac{1}{\sqrt{n}}$$

Table (4.9.1) Estimated Autocorrelation and Partial Autocorrelation Function of Residual for (0,0,0) x $(2,1,0)_{12}$ Model of Rainfall values in Chauk Township

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	.102	.146	-.035	.035	.038	.081	.034	.131	.020	.101	-.014	-.041
St-E	.083	.084	.086	.086	.086	.086	.087	.087	.088	.088	.089	.089
13-24	.074	.080	-.012	.023	-.053	-.030	.012	-.051	.082	-.005	-.046	-.104
St-E	.089	.090	.090	.090	.090	.090	.090	.090	.091	.091	.091	.091

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	.102	.137	-.063	.024	.048	.064	.012	.114	-.003	.068	-.025	-.073
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083
13-24	.093	.063	-.070	.000	-.038	-.053	.031	-.057	.076	-.002	-.086	-.098
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083



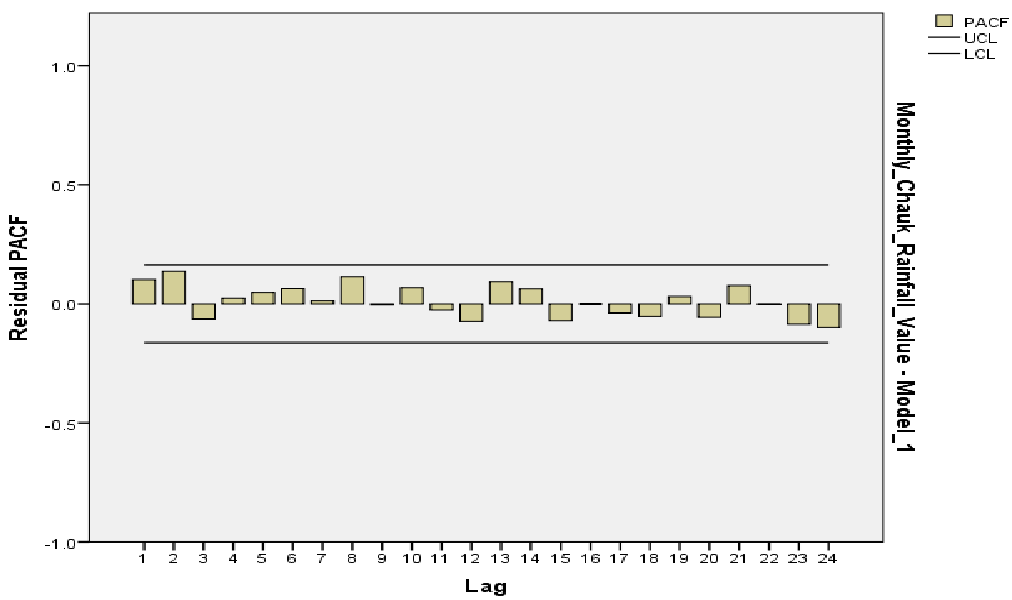


Figure (4.9.1) Sample ACF and Sample PACF Function of Residual Values for ARIMA (0,0,0)x (2,1,0)$_{12}$ Model of Rainfall value in Chauk Township

In checking model adequacy, Table (4.9.1) and Figure (4.9.1) give the value of residual ACF and PACF of considered model lie inside the confidence limits and the residuals are white noise process. So, this suggested that this model adequate. Additionally, the autocorrelation of $\hat{a}_t$ can be taken as not significant different from zero.

Furthermore, the autocorrelation of residuals are analyzed by using the test statistics Q.

$H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0$ (There is no correlation between the residuals)

Table (4.9.2) Model Statistics of SARIMA (0,0,0) x $(2,1,0)_{12}$ model for Rainfall value in Chauk Township

The value of residual for SARIMA (0,0,0) x $(2,1,0)_{12}$ are shown in Table (4.9.2).

Model Statistics							
Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
Stationary R-squared	R-squared	RMSE	Normalized BIC	Statistics	DF	Sig.	
0.213	0.355	56.369	8.167	13.674	16	0.623	0

According to Table (4.9.2), an overall check is performed by using the test statistics Q, the observed value of Q is 13.674 and it is not significant since p-value is 0.623 which is greater than $\alpha = 0.5$, there is no autocorrelation between the residuals. Thus, it can be considered that the SARIMA (0,0,0) x $(2,1,0)_{12}$ model the fitted is adequate for this data and used to forecast rainfall data in Chauk Township.

4.10 Forecasting

Since the model SARIMA (0,0,0) x $(2,1,0)_{12}$ adequate, this model can be used to forecast the future rainfall in Chauk Township. Below Table (4.10.1) and Figure (4.10.1) shows the forecast values of rainfall data in Chauk township, Myanmar. According to these forecasts Rainfall data series which will decrease in the future; this result indicates that the model provides an acceptable and fitted to predict the rainfall data series.

Table (4.10.1) Forecast Values with 95% Limits for Rainfall data in Chauk Township

Year	Forecast Value	95% Limits		Actual Value
		UCL	LCL	
Jan-19	18	79	8	2
Feb-19	9	36	34	0
Mar-19	9	45	26	0
Apr-19	18	78	9	0
May-19	123	274	23	24
Jun-19	145	309	33	81
Jul-19	58	167	1	26
Aug-19	129	285	26	149
Sep-19	123	275	23	12
Oct-19	141	302	31	24
Nov-19	20	84	7	30
Dec-19	15	70	12	0

According to Table (4.10.1) and Figure (4.10.1), the result indicates that the forecast values of SARIMA (0,0,0) x $(2,1,0)_{12}$ model is not very close to the observed value but the forecast values are between the upper confidence limit (UCL) and the lower limit (LCL). According to the forecast values, monthly rainfall in Chauk Township is expected to decrease in the future year.

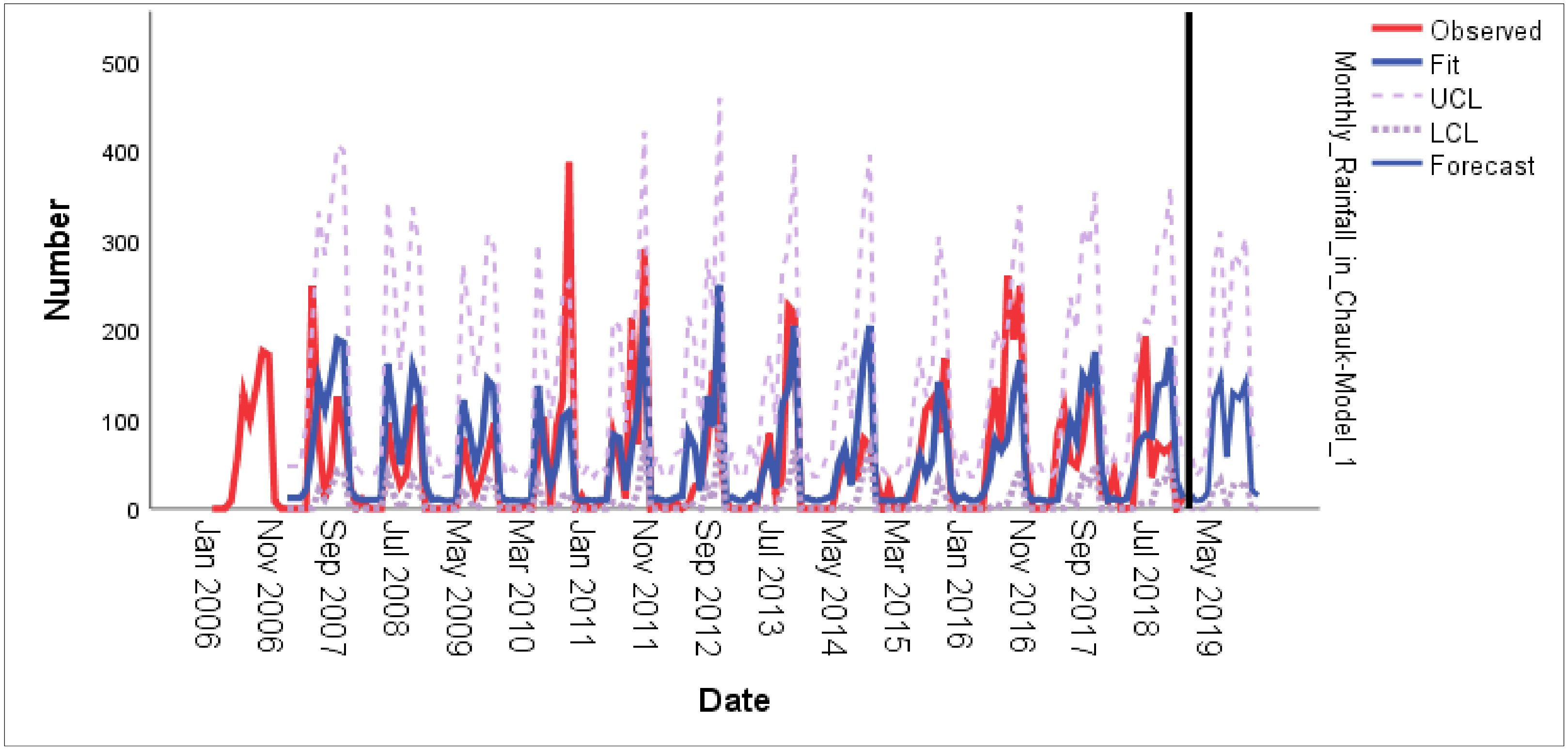


Figure (4.10.1) Forecast Values with 95% Limits for Rainfall value in Chauk Township for SARIMA (0,0,0) x $(2,1,0)_{12}$ Model

4.11 Plot of original monthly rainfall data in Shwebo Township

The number of rainfall data in Shwebo Township are shown in Table (4.11) and figure (4.11).

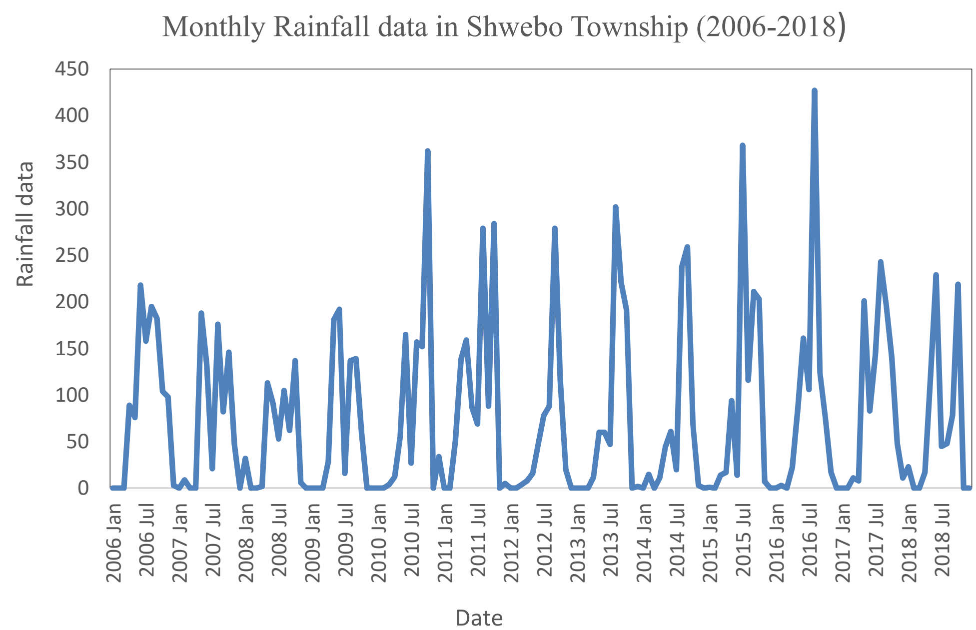


Figure (4.11) Plot of original monthly rainfall data in Shwebo Township (2006-2018)

In Figure (4.11) presents information about the monthly rainfall data in Shwebo from 2006 to 2018. It can be seen that the pattern of rainfall data has similar pattern throughout the years 2006 to 2018 where the series seen to be mean stationary but may be nonstationary in variance that this series required transformation for variance stabilization. In this figure, the series is also repetitive in nature due to seasonal variation and there exists seasonality in the data has a seasonal pattern with peaks and valleys in the same months of the year.

4.12 Test of Seasonality

The results for testing the seasonality in Shwebo Township (2006-2018) are shown in Table (4.12).

Table (4.12) ANOVA Table for rainfall series in Shwebo Township

Source of variation	Sum of Square	Degree of Freedom	Mean Square Error	F-ratio
Due to Month	SSM= 443233.86	11	MSM= 40293.99	19.77
Due to Year	SSY= 41637.44	12	MSY= 3469.79	
Error (Residual)	SSE= 268992.72	132	MSE=2037.82	
Total	SST= 753864.02	155		

At 5% level of significance, the critical value K = F (0.05, 11, 132) is 1.75. Meanwhile the computed F-value = 19.77 is greater than K=1.75, it can be concluded that the monthly rainfall data exists seasonality.

4.13 Seasonal Variation

The seasonal variation of monthly rainfall data series in Shwebo from 2006 to 2018 are computed by the ratio to moving averages method.

Table (4.13) Seasonal Indexes for rainfall data in Shwebo Township (2006-2018)

Month	Seasonal Index
January	0.1332
February	0.0451
March	0.0539
April	0.1659
May	1.5916
June	1.9246
July	1.1628
August	2.2755
September	2.1526
October	2.1329
November	0.3153
December	0.0486

The seasonal variation of monthly rainfall data series from 2006 to 2018 is calculated by the ratio to moving average method under multiplicative decomposition of time series which consists of 156 observations and it was shown in Appendix. The result of Table (4.13), the seasonal index are shown that the lowest value of seasonal index is in February, March, November, December and the month of May, June, August, September and October have the highest value of seasonal index. The peak period is in August and the lowest period is February with the seasonal index of rainfall data series in Shwebo Township.

4.14 Box-Jenkins ARIMA Model for Rainfall data series in Shwebo Township

The monthly rainfall series in Chauk Township from 2006 to 2018 analyzed as per Box-Jenkins Method which is Model Identification, Parameter Estimation, Diagnostic Checking and Forecasting. The data series consists of 156 observations.

4.14 Identification

The SACF and SPACF function of rainfall values are calculated and shown in Table (4.14.1) and Figure (4.14.1). These results pointed out that the seasonality exist and it should be eliminated to achieve stationary of the time series data.

Table (4.14.1) Sample Autocorrelation Function and Sample Partial Autocorrelation Function for the original series Z_t of Rainfall in Shwebo Township

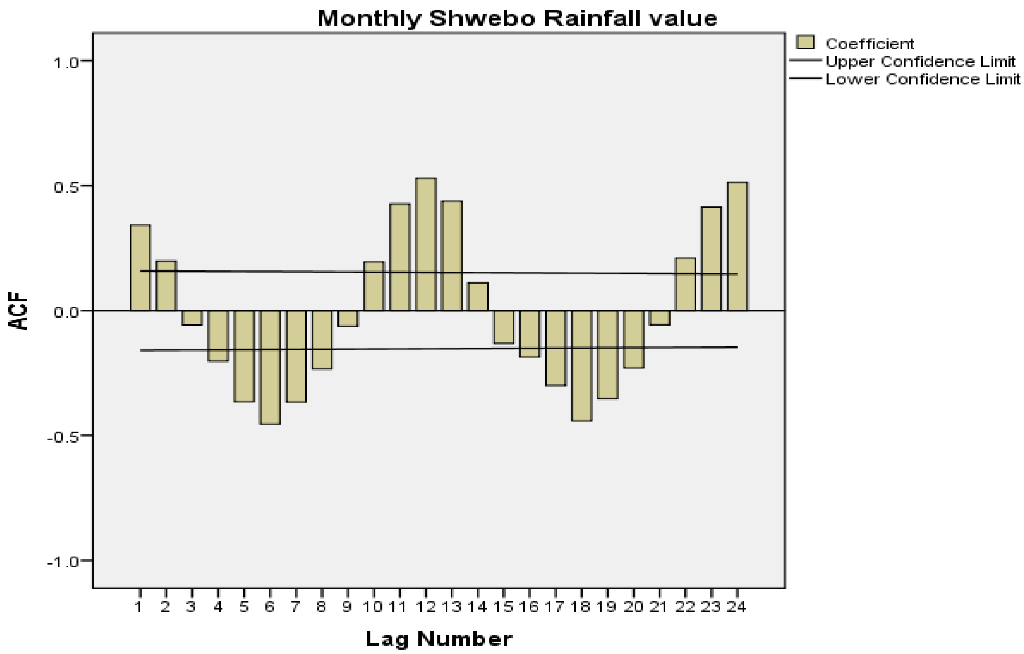
(a) $\hat{\rho}_k$ for {Zt}

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
Zt	01-12	0.342	0.198	-0.058	-0.202	-0.364	-0.454	-0.366	-0.234	-0.063	0.194	0.426	0.53
	St-E	0.079	0.079	0.079	0.079	0.078	0.078	0.078	0.077	0.077	0.077	0.077	0.076
	13-24	0.438	0.111	-0.131	-0.186	-0.3	-0.441	-0.352	-0.23	-0.058	0.21	0.413	0.512
	St-E	0.076	0.076	0.076	0.075	0.075	0.075	0.075	0.074	0.074	0.074	0.073	0.073

(b) $\hat{\phi}_{kk}$ for {Zt}

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
Zt	01-12	0.342	0.092	-0.172	-0.179	-0.257	-0.3	-0.187	-0.151	-0.129	0.049	0.205	0.263
	St-E	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	13-24	0.207	-0.119	-0.239	0.01	0.074	-0.095	-0.014	-0.063	-0.137	0.032	0.078	0.109
	St-E	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08

According to Table (4.14.1), the values of SACF are high at lags 1,6,12, 18 and 24 then slowly declines and the value of SPACF is cutoff at lag 1.



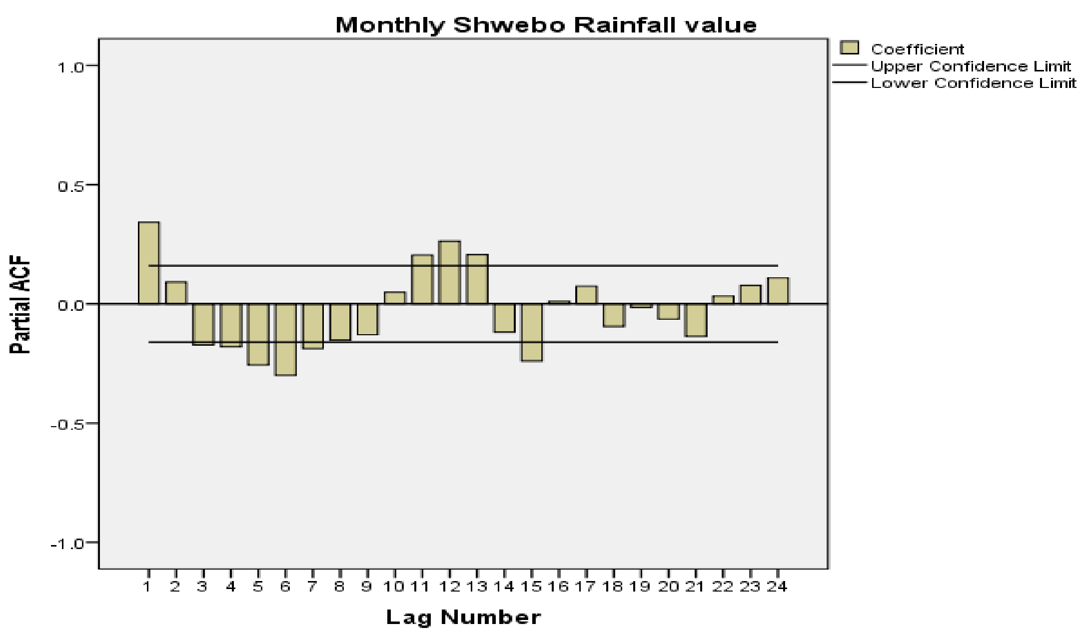


Figure (4.14.1) Sample Correlogram for Original Series of Rainfall data in Shwebo Township

With reference to Figure (4.14.1), the sample autocorrelation function ACF indicates a strong variation with seasonal period 12 and the sample ACFS of the original series slowly decay and there are significant seasonal spike of sample PACFS at lag 1.

The series seen to be mean stationary, but the variance may not be is stationary and is not stabilization. So, this series need to transformation for variance stabilization. For power transformation, Residual mean square errors calculated with the two transformations which are shown in Table (4.14.2).

Table (4.14.2) Residual mean square errors in the power transformation of Shwebo Township

Transformation	Residual Mean Square Error (RMSE)
Square root	108.75
Logarithmic	113.81

According to the result, Square root transformation of RMSE is lower than Logarithmic transformation value. So, it suggests that a square root transformation is suitable to apply it to the original series to obtain the variance stationary. The sample ACF and sample PACF of square root transformation series $\sqrt{Z_t}$ were computed as shown in Table (4.14.3) and Figure (4.14.3).

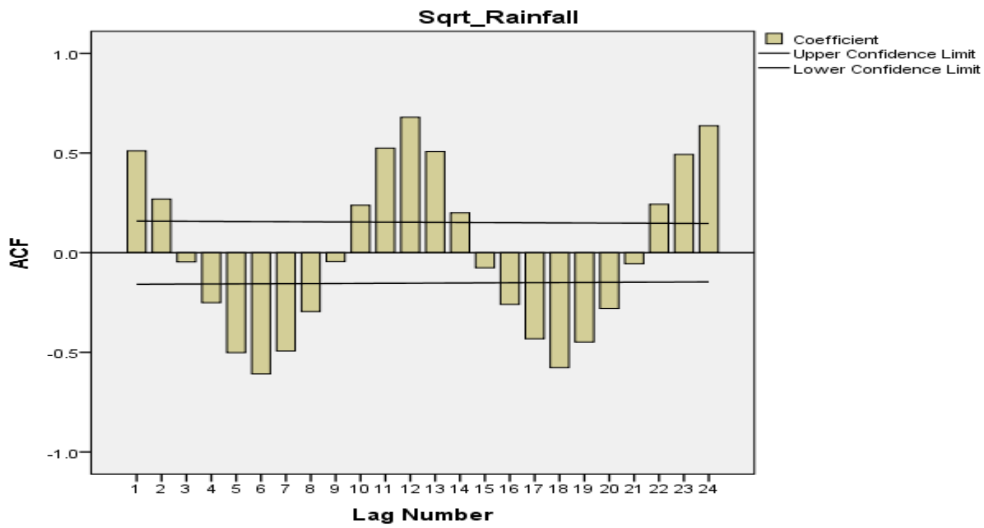
Table (4.14.3) Sample ACF and Sample PACF Function for square root transformation of Rainfall values in Shwebo Townships

(a) $\hat{\rho}_k$ for $\sqrt{Z_t}$

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
$\sqrt{Z_t}$	01-12	0.511	0.269	-0.046	-0.25	-0.501	-0.608	-0.494	-0.295	-0.045	0.238	0.525	0.679
	St-E	0.079	0.079	0.079	0.079	0.078	0.078	0.078	0.077	0.077	0.077	0.077	0.076
	13-24	0.508	0.2	-0.076	-0.259	-0.432	-0.577	-0.448	-0.28	-0.055	0.243	0.493	0.637
	St-E	0.076	0.076	0.076	0.075	0.075	0.075	0.075	0.074	0.074	0.074	0.073	0.073

(b) $\hat{\phi}_{kk}$ for $\sqrt{Z_t}$

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
$\sqrt{Z_t}$	01-12	0.511	0.011	-0.254	-0.194	-0.357	-0.343	-0.152	-0.133	-0.091	0.043	0.214	0.322
	St-E	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
	13-24	0.059	-0.175	-0.167	-0.004	0.079	-0.084	0.03	-0.067	-0.15	0.025	0.043	0.155
	St-E	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08



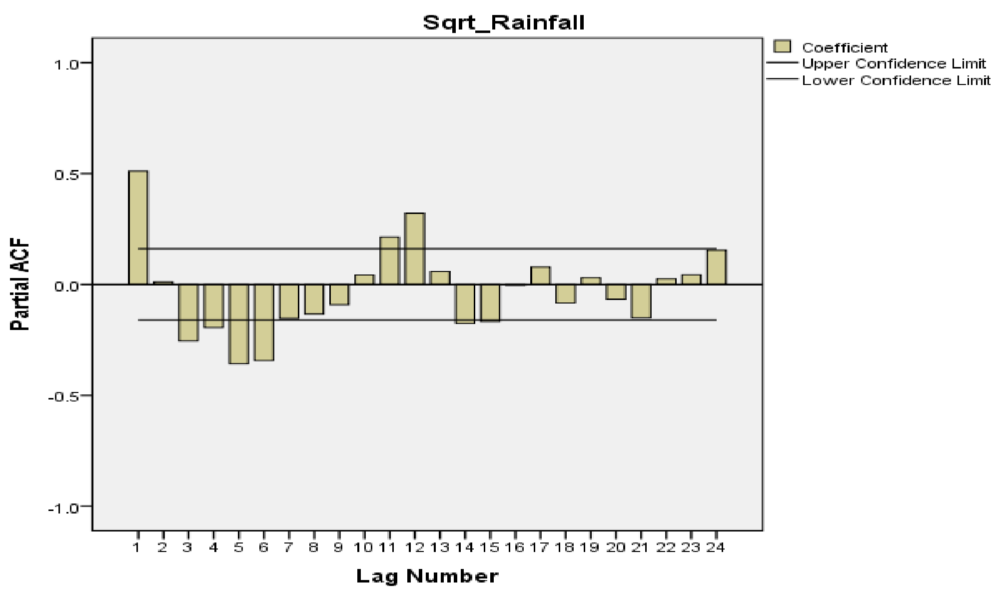


Figure (4.14.3) Sample Correlogram for square root transformation of Rainfall values in Shwebo Township

In Figure (4.14.3) the sample ACF illustrate a damping sine-cosine wave and the value of SACF at lag 1, 6, 12, 14 lie outside the confidence limits and the value of SPACF are significant seasonal spike at lags 1 and 12 which is a 12 period of seasonality. Hence, it should be transformed by taking the first seasonal differences to eliminate seasonality. The sample ACF and sample PACF of first seasonal differenced series $(1 - B^{12})\sqrt{Z_t}$ were computed as shown in Table (4.14.4) and Figure (4.14.4).

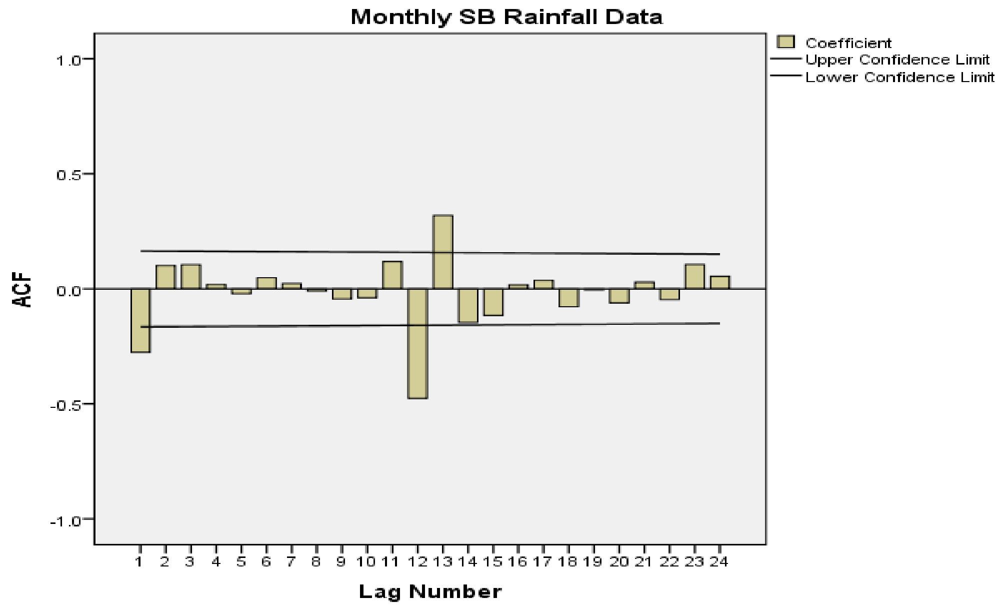
Table (4.14.4) Sample ACF and Sample PACF Function for First Seasonal Difference Series (W_t) of Rainfall values in Shwebo Townships

(a) $\hat{\rho}_k$ for $\{W_t = (1 - B^{12}) \sqrt{Z_t}\}$

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	01-Dec	-0.28	0.10	0.10	0.02	-0.02	0.05	0.02	-0.01	-0.04	-0.04	0.12	-0.48
W_t	St-E	.082	.082	.082	.082	.081	.081	.081	.080	.080	.080	.080	.079
	13-24	0.32	-0.15	-0.12	0.02	0.04	-0.08	0.00	-0.06	0.03	-0.05	0.11	0.05
	St-E	.079	.079	.078	.078	.078	.077	.077	.077	.076	.076	.076	.076

(b) $\hat{\phi}_{kk}$ for $\{W_t = (1 - B^{12}) \sqrt{Z_t}\}$

Lag	K	1	2	3	4	5	6	7	8	9	10	11	12
	01-Dec	-0.28	0.03	0.15	0.09	-0.02	0.01	0.03	0.01	-0.06	-0.09	0.11	-0.44
W_t	St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083
	13-24	0.13	0.00	-0.10	-0.03	0.05	0.01	-0.02	-0.11	0.00	-0.05	0.22	-0.17
	St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083



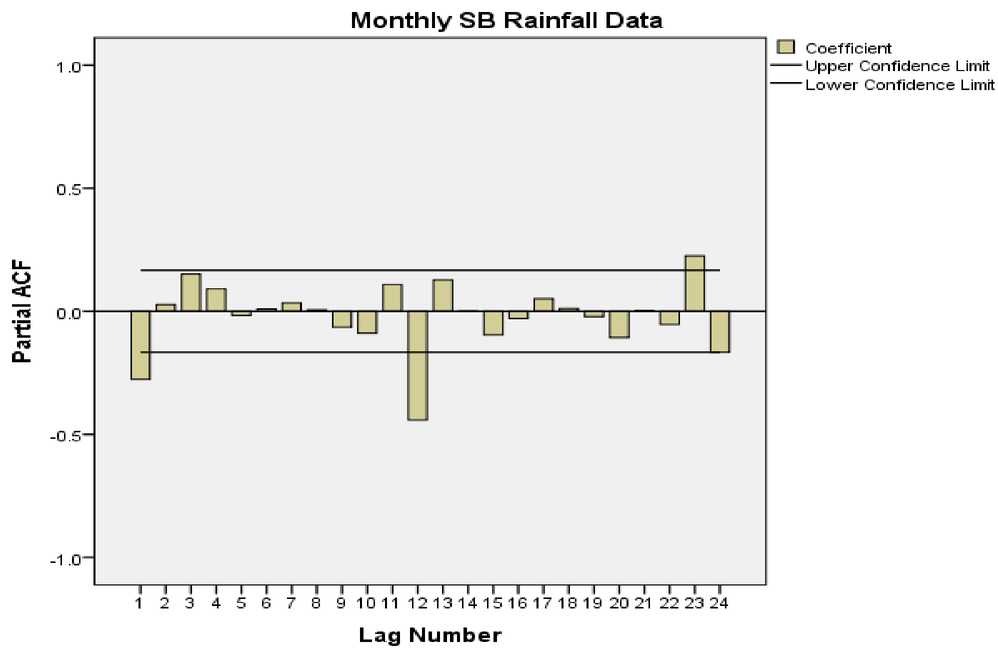


Figure (4.14.4) Sample Correlograms for First Seasonal Difference Series of Shwebo Township

According to the Figure (4.14.4), the values of SACF and SPACF are spike at lags 1, 2 lie inside of the confidence limits accordingly. The possible model might be considered SAR (1) model as tentative model to this series $(1 - B^{12})\sqrt{Z_t}$.

Hence, the tentative model for the series is SAR (1) process.

$$(1 - \Phi B^{12})\sqrt{Z_t} = \theta_0 + a_t$$

4.15 Parameter Estimation

Using SAR (1) model, the estimated parameters with their statistics were shown in Table (4.15.1) and Figure (4.15.1).

Table (4.15.1) Estimated Parameters and Model Statistics for SAR (1) Model of Rainfall values in Shwebo Township

	Estimate	SE	t	P-value
Constant	-.966	4.166	-.232	.817
Φ	-.512	.075	-6.837	.000

The following estimated model was obtained

$$(1 + 0.512B^{12})\sqrt{Z_t} = -0.966 + a_t$$

(0.075) (4.166)

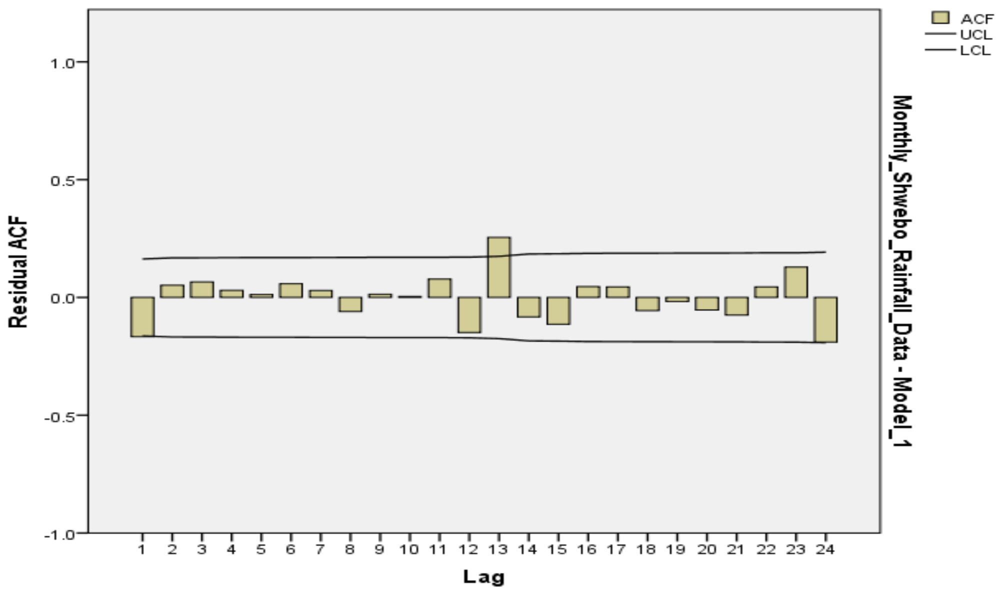
According to Table (4.15.1), the estimation of SAR (1) Model of rainfall values in Shwebo Township give θ_0 = -0.966, the estimated standard error 4.166 and Φ = -0.512, the estimated standard error 0.075 with P-value is 0.000. So, there is no evidence to reject the null hypothesis $H_0 : \Phi = 0$.

In addition, the sample ACFs and the sample PCFs of residual for the tentative model were shown in Table (4.15.2) and Figure (4.15.2).

Table (4.15.2) Estimated Autocorrelation and Partial Autocorrelation Function of Residual for SAR (1) Model of Rainfall values in Shwebo Township

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	-.167	.052	.066	.030	.013	.058	.029	-.060	.014	.004	.078	-.148
St-E	.083	.086	.086	.086	.086	.086	.087	.087	.087	.087	.087	.087
13-24	.255	-.082	-.114	.046	.045	-.056	-.017	-.053	-.075	.045	.129	-.190
St-E	.089	.094	.095	.095	.096	.096	.096	.096	.096	.097	.097	.098

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	-.167	.025	.081	.054	.021	.056	.042	-.059	-.021	-.002	.087	-.127
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083
13-24	.215	-.009	-.143	-.020	.056	-.027	-.043	-.084	-.049	.037	.177	-.215
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083



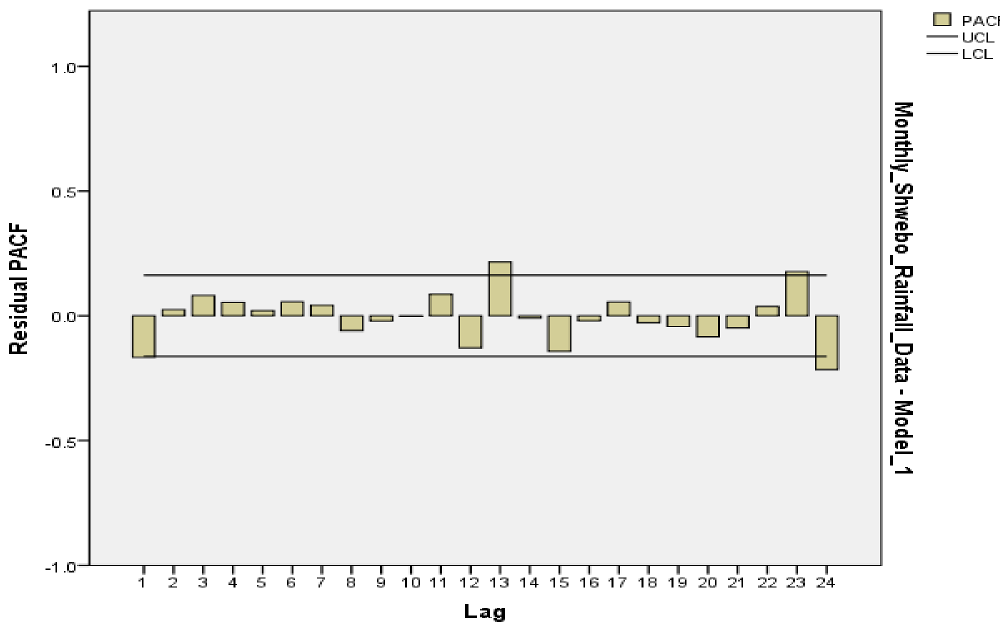


Figure (4.15.2) Sample ACF and Sample PACF of Residual values for SAR (1) Model of Rainfall Series in Shwebo Township

In Figure (4.15.2), the sample ACF and Sample PACF of residual is cut off after lag 1 lie inside confidence limits and this model exhibit a pattern that the residual series are not white noise process. So, (0,0,0) x $(2,1,0)_{12}$ model considered as another tentative model which is

$$(1-\Phi_1 B^{12}-\ \Phi_2 B^{24})\sqrt{Z_t}\ =\theta_0$$

Using Seasonal ARIMA (0,0,0) x $(2,1,0)_{12}$ model, the estimated parameter of their statistics were shown in Table (4.15.3).

Table (4.15.3) Estimated Parameters and Model Statistics for Seasonal ARIMA (0,0,0) x $(2,1,0)_{12}$ Model of Rainfall values in Shwebo Township

	Estimate	SE	T	P-value
Constant	-.022	.157	-.138	.891
Φ_1	-.672	.090	-7.481	.000
Φ_2	-.259	.090	-2.872	.005

The estimated model was obtained:

$$(1 + 0.672B^{12} + 0.259B^{24})\sqrt{Z_t} \ = -0.022$$

(0.090) (0.090) (0.157)

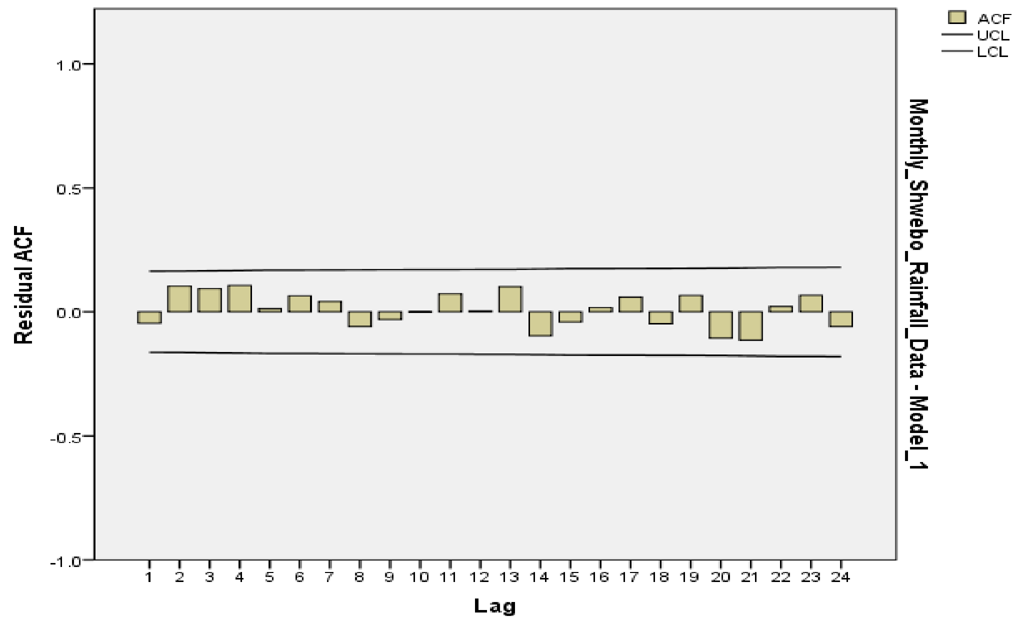
According to Table (4.15.3), it can be found that the table give $\theta_0 = 0.052$, the estimated standard error is 0.141. The estimated parameter of Φ_1= 0.672 and Φ_2 = -0.259, both of the estimated standard error are 0.090, the test statistics t is -7.481 and -2.872. Since their p-value are 0.000 and 0.005 at 1% level of significant. Hence there are no evidence to reject the null hypothesis Φ_1 =0 and Φ_2 = 0.

Furthermore, the sample ACFs and the sample PCFs of residual for the tentative model were shown in Table (4.15.4) and Figure (4.15.4).

Table (4.15.4) Estimated Autocorrelation and Partial Autocorrelation Function of Residual for (0,0,0) x $(2,1,0)_{12}$ Model of Rainfall values in Shwebo Township

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	-.046	.104	.094	.107	.014	.065	.042	-.060	-.031	.000	.072	.004
St-E	.083	.084	.084	.085	.086	.086	.086	.087	.087	.087	.087	.087
13-24	.004	.102	-.096	-.041	.018	.059	-.048	.066	-.106	-.114	.022	.066
St-E	.087	.087	.088	.089	.089	.089	.089	.089	.090	.091	.092	.092

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	-.046	.102	.104	.108	.005	.035	.025	-.081	-.061	-.011	.090	.038
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083
13-24	.105	-.099	-.091	-.004	.060	-.020	.071	-.094	-.129	.008	.084	-.029
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083



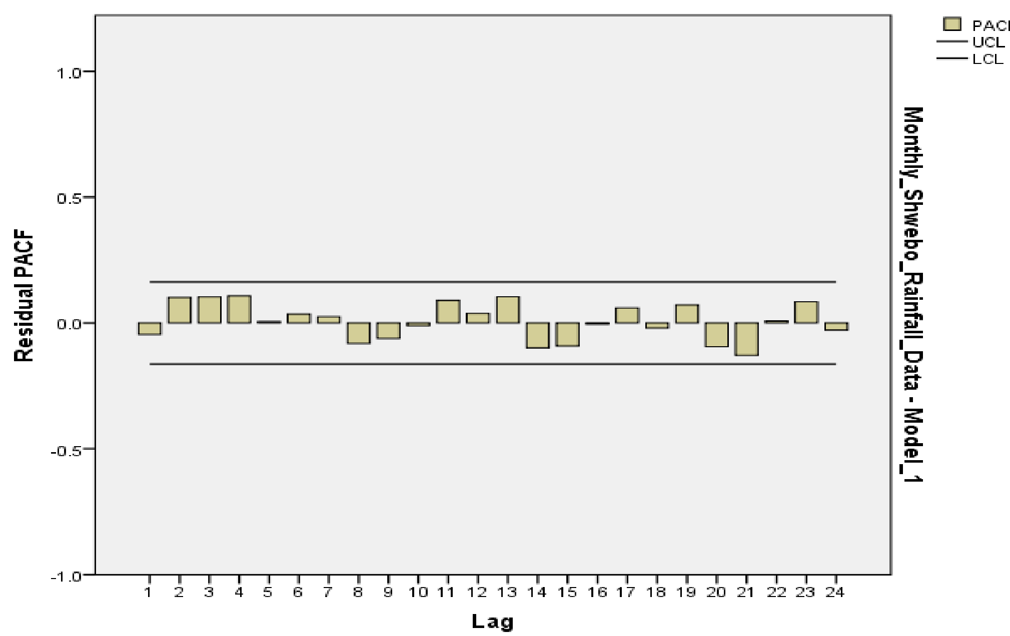


Figure (4.15.4) Sample ACF and Sample PACF Function of Residual Values for ARIMA (0,0,0) x $(2,1,0)_{12}$ Model of Rainfall value in Shwebo Township

In Figure (4.15.4), the sample ACF and Sample PACF of residual for the above tentative model which indicate that the residual series are white noise process and the series exhibit no patterns. But the tentative (1,0,0) x $(1,1,0)_{12}$ model modified as another tentative model which is

$$(1 - \emptyset B)(1 - \Phi B^{12})\sqrt{Z_t} = \theta_0$$

Seasonal ARIMA (1,0,0) x $(1,1,0)_{12}$ model of the estimated parameter with their statistics were shown in Table (4.15.5).

Table (4.15.5) Estimated Parameters and Model Statistics for Seasonal ARIMA (1,0,0) x $(1,1,0)_{12}$ Model of Rainfall values in Shwebo Township

	Estimate	SE	T	P-value
Constant	-1.033	3.589	-.288	.774
$\emptyset$	-.171	.083	-2.064	.041
Φ	-.484	.077	-6.269	.000

The following estimated model was obtained

$$\underset{(0.083)}{(1 + 0.171B)}\underset{(0.077)}{(1 + 0.484B^{12})}\sqrt{Z_t} = -1.033$$

The estimation (1,0,0) x $(1,1,0)_{12}$ Model of rainfall values in Shwebo Township give θ_0 = -1.022, the estimated standard error is 3.589. The estimated parameter $\emptyset$ = -0.171, the estimated standard error 0.083, the test statistics t is -2.064 with p-value is 0.041. Hence, there is no evidence to reject the null hypothesis $\emptyset = 0$ and The estimated parameter Φ = -0.484, the estimated standard error is 0.077, the test statistics t is -6.269 with p-value is 0.000, there is no evidence to reject the null hypothesis Φ = 0.

4.16 Diagnostic Checking

In checking the model adequacy for the residual ACF and PACF of the modified model as shown in Table (4.16.1) and Figure (4.16.1).

$$\gamma_k(\hat{a}_t) \pm 2.\widehat{S.E}\,[\gamma_k\,(\hat{a}_t)]$$

Where,

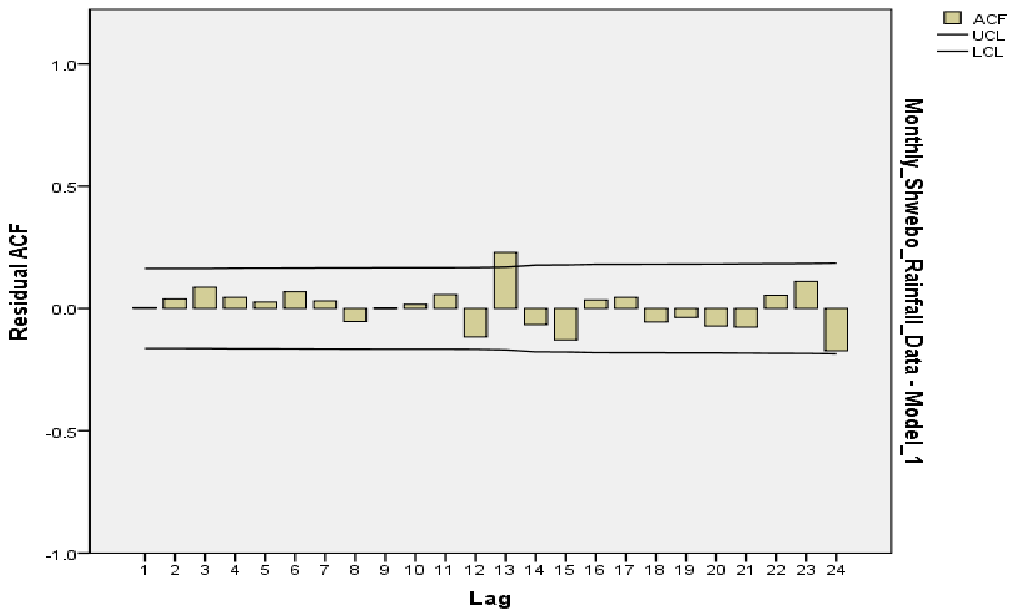
$$\widehat{S.E}\,[\gamma_k\,(\hat{a}_t)] = \frac{1}{\sqrt{n}}$$

Table (4.16.1) Estimated Autocorrelation and Partial Autocorrelation Function of Residual for (1,0,0) x $(1,1,0)_{12}$ Model of Rainfall values in Shwebo Township

The values of residual ACF and PACF of seasonal ARIMA (1,0,0) x $(1,1,0)_{12}$ model are shown in Table (4.16.1).

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	.003	.039	.087	.045	.026	.068	.031	-.053	.002	.017	.056	-.116
St-E	.083	.083	.083	.084	.084	.084	.085	.085	.085	.085	.085	.085
13-24	.230	-.066	-.128	.035	.045	-.055	-.037	-.072	-.076	.054	.110	-.173
St-E	.086	.091	.091	.092	.092	.092	.093	.093	.093	.093	.094	.095

Lag K	1	2	3	4	5	6	7	8	9	10	11	12
01-12	.003	.039	.087	.044	.020	.058	.023	-.064	-.014	.011	.062	-.118
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083
13-24	.231	-.073	-.132	.011	.050	-.041	-.050	-.086	-.024	.071	.133	-.235
St-E	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083	.083



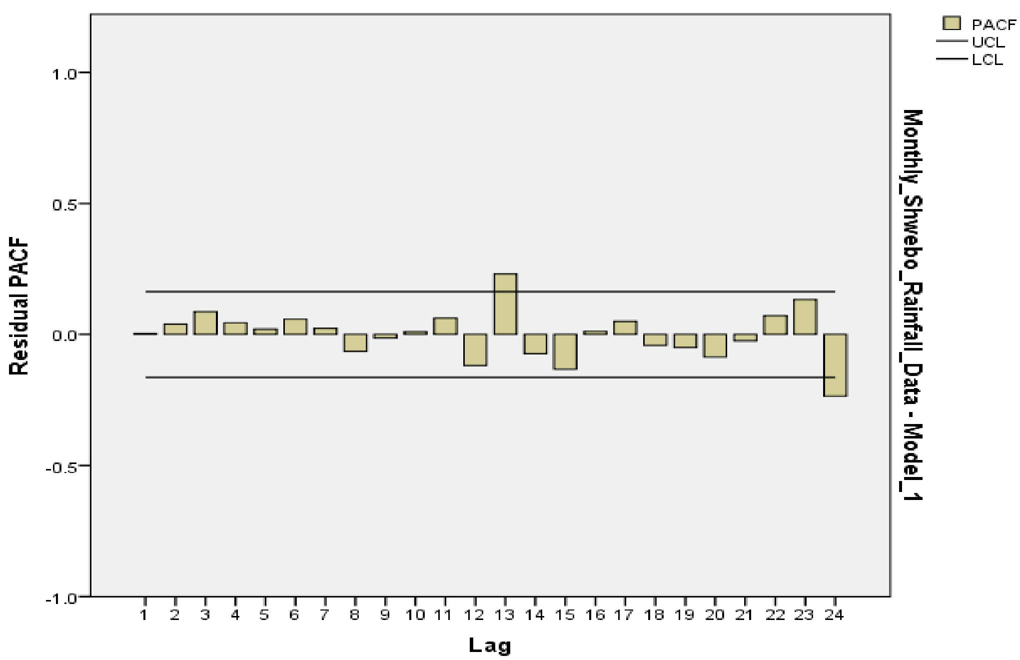


Figure (4.16.1) Sample ACF and Sample PACF Function of Residual values for ARIMA (1,0,0) x (1,1,0)$_{12}$ Model of Rainfall value in Shwebo Township

In checking model adequacy, Figure (4.16.1), the values of residual ACF is spike at lag 12 and the value of residual PACF is spike at lags 12 and 24 for seasonal ARIMA (1,0,0) x (0,1,1)$_{12}$ which exhibit a pattern and the residuals are not white noise process. But the p-value is significant at the level of 1%. Hence, this suggested that this model is inadequate and the autocorrelation of $\widehat{a_t}$ can be taken as not significant different from zero.

So, the above tentative SARIMA (0,0,0) x $(2,1,0)_{12}$ model was considered as the fitted model for this data and used to forecast rainfall data in Shwebo township because the value of residual ACFs and PACFs are parsimonious and exhibit no patterns then they lie inside the confidence limits and white noise process.

Furthermore, the autocorrelation of residuals are analyzed by using the test statistics Q.

H_0 : $\rho_1 = \rho_2 = \cdots = \rho_k = 0$ (There is no correlation between the residuals)

Table (4.16.2) Model Statistics of SARIMA (0,0,0) x $(2,1,0)_{12}$ model for Rainfall value in Shwebo Township

The value of residual for SARIMA (0,0,0) x $(2,1,0)_{12}$ are shown in Table (4.16.2).

Model Statistics							
Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
Stationary R-squared	R-squared	RMSE	Normalized BIC	Statistics	DF	Sig.	
.275	.406	70.925	8.627	11.867	16	.753	0

According to Table (4.16.2), an overall check is performed by using the test statistics, the observed value of Q is 11.867 and it is not significant since p-value is 0. 753 which is greater than $\alpha = 0.05$, there is no autocorrelation between the residuals. Thus, it can be identified that the SARIMA (0,0,0) x $(2,1,0)_{12}$ model is adequate for this data and used to forecast rainfall values in Shwebo Township.

4.17 Forecasting

The models have been identified, estimated and analyzed to the rainfall values in Shwebo Township where SARIMA (0,0,0) x $(2,1,0)_{12}$ is adequate, this model can be used to forecast the future rainfall value in Shwebo Township from January 2019 to December 2019. Below Table (4.17.1) and Figure (4.17.1) shows the forecast values of rainfall data in Shwebo Township. According to these forecasts Rainfall data series which will decrease in the future; this results indicates that the model provides an acceptable fitted to predict the rainfall data series.

Table (4.17.1) Forecast values with 95% Limits for Rainfall data in Shwebo Township

Year	Forecast Value	95% Limits		Actual Value
		UCL	LCL	
Jan-19	14	70	28	2
Feb-19	12	52	41	0
Mar-19	14	67	30	0
Apr-19	26	111	10	0
May-19	152	348	25	24
Jun-19	155	353	26	81
Jul-19	107	275	9	26
Aug-19	209	435	52	149
Sep-19	145	338	22	12
Oct-19	154	352	26	24
Nov-19	27	115	9	30
Dec-19	14	67	30	0

According to table (4.17.1), the result indicates that the forecast values of SARIMA (0,0,0) x $(2,1,0)_{12}$ model is not very close to the observed value but the values that are between the upper confidence limit (UCL) and the lower limit (LCL). According to the forecast values, monthly rainfall in Shwebo Township is expected to decrease in the future year.

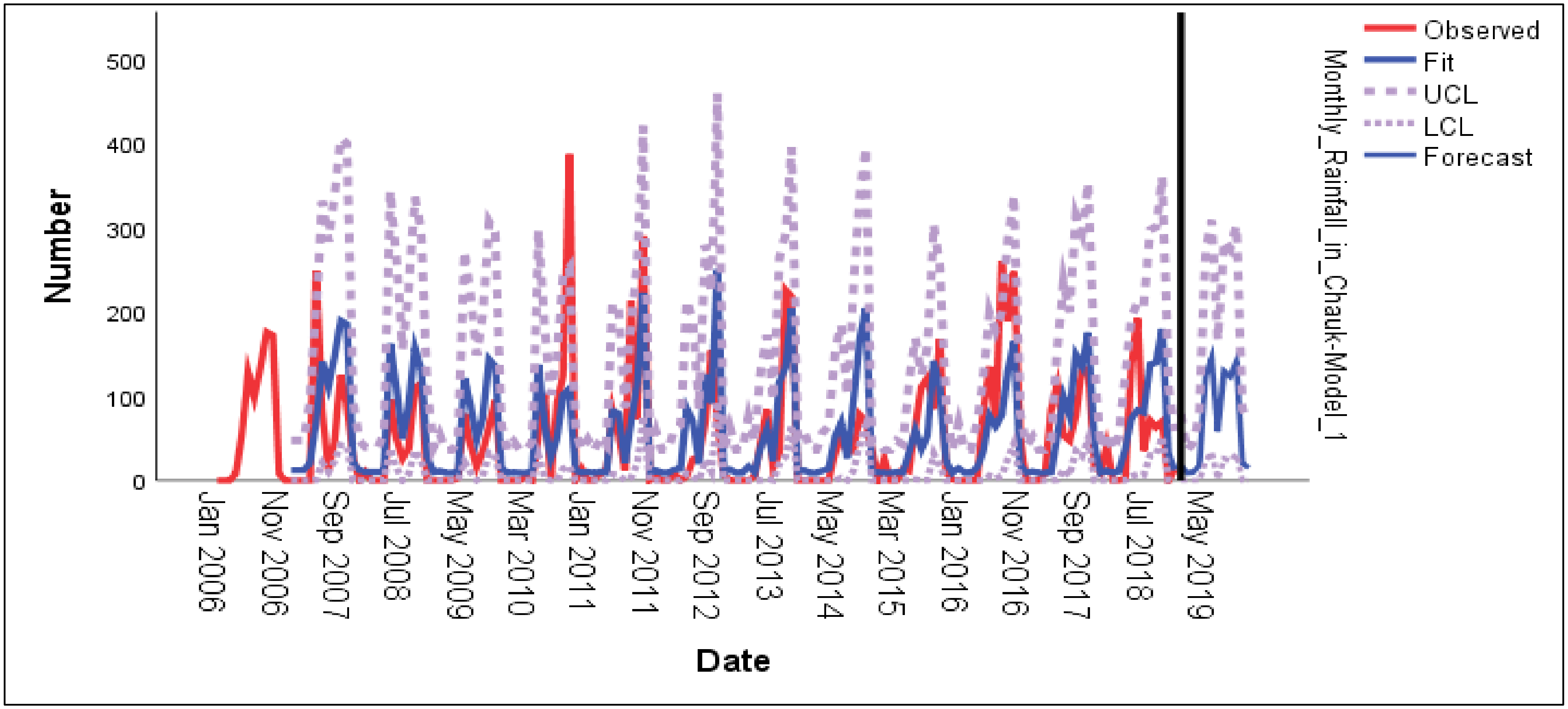


Figure (4.17.1) Forecast Values with 95% Limits for Rainfall values in Shwebo Township for SARIMA (0,0,0) x $(2,1,0)_{12}$ Model

CHAPTER V
CONCLUSION

This study focuses on the modeling and forecasting of the recorded monthly Rainfall in Chauk and Shwebo township in Myanmar by modeling monthly data from January 2006 to December 2018 using time series SARIMA model which can be obtained by using four iteratively Box-Jenkins steps and provided the prediction of the monthly Rainfall series in Chauk and Shwebo township in Myanmar. Following the box and Jenkins methodology, the time series modeling involves transformation of the data in order to achieve township, followed by identification of the fitted models, estimation of model parameters, diagnostic checking of the assumption model and finally forecasting of future values.

The model estimation for the volume of Rainfall flow was carried out by using Statistical Package for Social Science (SPSS) and partially of EViews software. Theoretical and estimated autocorrelation function (ACF) and partial autocorrelation (PACF) play important role in the construction of SARIMA as in the building of non-seasonal ARIMA models. The coefficients of the model are estimated using technique such as maximum Likelihood estimation (MLE) and the ordinary lest square (OLS) and the estimated residuals are then analyzed using the ACF and PACF to diagnose if the residuals are consistent with the hypothesis that the residuals are white noise. The results indicate that SARIMA $(0,0,0) \times (2,1,0)_{12}$ for Chauk Township and $(0,0,0) \times (2,1,0)_{12}$ for Shwebo Township are identified as the fitted model. This appropriate models for these two townships were used to find forecasts for 2019. The forecasts clearly show rainfall pattern for each stow

Evaluating and forecasting the volume of monthly rainfall are very essential. This model is considered appropriate to predict the monthly rainfall for the upcoming years to assist decision makers establish priorities for water demand, storage and distribution. (JECS, 2016). The normal rainfall pattern has also shifted. It has decreased in the months January, February, March and December in Chauk and Shwebo Township as well. More studies are needed to conclude on the long-term rainfall trends in Myanmar. In summary, analysis of observational data from the past 13 years in combination with model-based climate scenarios shows that the climatic conditions in

Myanmar's central Dry Zone will be decreasing number of rainy days. In the dry land, the livelihood is determined by favour of climatic condition. Thus, climate change effects become rooted for farmer vulnerability.

In conclusion, farmers' vulnerability emerges from the complex nature of the environmental and socio- economic interaction. However, this paper focuses on the further rainfall data which will help the vulnerability of the farmers for the specific small area, these issues are crucial for the consideration of dry zone agricultural development. This analysis has shown based on the consideration of dry zone regional development, the primacy is basic software and hardware infrastructure development such as implementation of water tube well, supporting agricultural tools and micro finance activity, etc. These kinds of development undermine the impacts of natural hazards and so the government should take it into great account for the future economic development of our country through agriculture.

REFERENCE

Al-Ansari NA (2013) Management of Water Resources in Iraq: Perspectives and Prognoses. J Engineering 5: 667-684.

Alawaye A.I, Alao A.N (2017) Time Series Model and Analysis On Rainfall In Oshogbo Osun State, Nigeria

Amha G (2010) Modelling and forecasting monthly rainfall in Tigray region. Case of Mekellestation. (Unpublished M.Sc.Thesis), Department of Statistics, Addis Ababa University, Ethiopia.

D.A. Attach, G.M Bankole (2011) Time Series Analysis Model for Annual Rainfall Data in Lower Kaduna Catchment Kaduna, Nigeria

Daniel Eni, Fola J.Adeyeye (2015) Seasonal ARIMA Modeling and Forecasting of Rainfall in Warri Town, Nigeria

Harvey R, Andrew C, Souza RC (1987) Assessing and Modeling the Cyclical Behavior of Rainfall in North-East Brazil. Journal of Climate and Applied Meteorology 26: 1339-1344.

Mahsin MD, Yesmin A, Monira B (2012) Modeling Rainfall in Dhaka Division of Bangladesh Using Time Series Analysis. Journal of Mathematical Modelling and Application 1: 67-73.

MET Report, Myanmar Climate Report (2018)

MAS Programme, Time Series Analysis and Forecasting (2017)

Moet Moet Su Hlaing, M.Econ (Statistics), 2015 Seasonal ARIMA Modeling of Airways Transport and Railways Transport Series in Myanmar (2007-2013)

Pazvakawambwa G.T. and Ogunmokun A.A, A time-series forecasting model for Windhoek Rainfall, Namibia

P.E. Naill M. Momani (2009) Time Series Analysis Model for Rainfall Data in Jordan: Case Study for Using Time Series Analysis

Phyu Phyu Mon, M.Econ (Statistics), 2011 Application of Seasonal ARIMA Models

Ratnadip Adhikari R.K Agrawal (2017) An Introductory study on Time Series Modeling and Forecasting

Seyed A, Shamsnia M, Naeem S, Ali L (2011) Modeling weather parameter using stochastic methods (ARIMA Model)(Case Study: Abadeh station, Iran). International conference on Environment and Industrial innovation IPCBEE,12.

Soe Thu Zar Tint, M.Econ (Statistics), 2010 Time Series Forecasting Using Holt Winters Exponential Smoothing

Tariq and AbbasAbd (2014) Time Series Analysis of Nyala Rainfall Using ARIMA Method

Terence C.Mills (2019) Applied Time Series Analysis A Practical Guide to Modeling and Forecasting

Thomas Papalaskaris (2016) Stochastic monthly rainfall time series analysis, modeling and forecasting in Kavala City, Greece, North-Eastern Mediterranean basin

William W.S. Wei (2005) Second Edition Time Series Analysis – Univariate and Multivariate Methods

Yu War Hling (2018) Seasoanl Time Series Model of Production Transformers in Hitaci Soe Electronic And Machinery Co..LTD (2013-2017)

<u>Website</u>

https://www.itl.nist.gov

https://www.ststistics.org

https://www.un.org

https://www.researchgate.net

https://onlinelibrary.wiley.com

https://www.moezala.gov.mm

https://otests.com

https://people.duke.edu

https://help.xlstat.com

APPENDIX

Table (A-1.1)

Original monthly rainfall data in Chauk Township (2006-2018)

Year / Month	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Jan	0	0	10	0	0	0	0	0	0	25	0	0	41
Feb	0	0	0	0	0	0	0	0	0	0	0	0	0
Mar	0	0	0	0	5	8	8	0	0	0	0	8	0
Apr	8	0	0	3	0	12	0	5	0	13	0	85	2
May	55	249	96	79	55	83	5	46	32	11	79	113	137
Jun	128	81	50	42	102	54	25	84	52	63	135	53	192
Jul	99	12	26	16	3	12	28	15	34	111	71	47	35
Aug	134	48	38	35	92	213	70	33	34	122	260	77	71
Sep	176	125	110	63	125	73	154	226	78	85	189	132	63
Oct	172	89	115	95	387	290	68	218	70	168	248	132	70
Nov	8	27	0	0	0	0	7	0	37	10	50	17	0
Dec	0	0	0	0	14	5	0	0	0	1	0	9	13

Sources : Monthly Rainfall data: Department of Meteorology and Hydrology in Mandalay

Table (A-1.2)

Original monthly rainfall data in Shwebo Township (2006-2018)

Year / Month	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
Jan	0	0	32	0	0	0	0	0	0	1	0	0	23
Feb	0	9	0	0	0	0	0	0	15	0	3	0	0
Mar	0	0	0	0	4	51	4	0	0	14	0	11	0
Apr	89	0	2	28	12	138	8	12	11	17	23	8	17
May	76	188	113	181	55	159	16	60	44	94	87	201	120
Jun	218	133	91	192	165	87	48	60	61	14	161	83	229
Jul	158	21	53	16	27	69	78	47	20	368	106	143	45
Aug	195	176	105	137	157	279	88	302	238	116	427	243	48
Sep	182	82	62	139	152	88	279	221	259	211	124	196	79
Oct	104	146	137	60	362	284	115	191	68	203	74	140	219
Nov	98	47	6	0	0	0	20	0	3	7	17	48	0
Dec	3	0	0	0	34	5	0	2	0	0	0	11	0

Sources : Monthly Rainfall data: Department of Meteorology and Hydrology in Mandalay

Table (A-1.3)

Table (A-1.3) Seasonal Indexes for rainfall data in Chauk (2006-2018)

Time	Rainfall Data				Centered Moving Average (CMA)	Rainfall Data/CMA	Seasonal Index
2006 Jan	0						0.1642
2006 Feb	0						0.0376
2006 Mar	0						0.0737
2006 Apr	8						0.2074
2006 May	55						1.5726
2006 Jun	128	780	65.00	130.00	65.00	1.97	1.7615
2006 Jul	99	780	65.00	130.00	65.00	1.52	0.8056
2006 Aug	134	780	65.00	130.00	65.00	2.06	1.7025
2006 Sep	176	780	65.00	129.33	64.67	2.72	2.4005
2006 Oct	172	772	64.33	144.83	72.42	2.38	2.9522
2006 Nov	8	966	80.50	157.08	78.54	0.10	0.2326
2006 Dec	0	919	76.58	145.92	72.96	0.00	0.0903
2007 Jan	0	832	69.33	131.50	65.75	0.00	0.1642
2007 Feb	0	746	62.17	120.08	60.04	0.00	0.0376
2007 Mar	0	695	57.92	108.92	54.46	0.00	0.0737
2007 Apr	0	612	51.00	103.58	51.79	0.00	0.2074
2007 May	249	631	52.58	105.17	52.58	4.74	1.5726
2007 Jun	81	631	52.58	106.00	53.00	1.53	1.7615
2007 Jul	12	641	53.42	106.83	53.42	0.22	0.8056
2007 Aug	48	641	53.42	106.83	53.42	0.90	1.7025
2007 Sep	125	641	53.42	106.83	53.42	2.34	2.4005
2007 Oct	89	641	53.42	94.08	47.04	1.89	2.9522
2007 Nov	27	488	40.67	78.75	39.38	0.69	0.2326
2007 Dec	0	457	38.08	77.33	38.67	0.00	0.0903
2008 Jan	10	471	39.25	77.67	38.83	0.26	0.1642
2008 Feb	0	461	38.42	75.58	37.79	0.00	0.0376
2008 Mar	0	446	37.17	76.50	38.25	0.00	0.0737
2008 Apr	0	472	39.33	76.42	38.21	0.00	0.2074
2008 May	96	445	37.08	74.17	37.08	2.59	1.5726
2008 Jun	50	445	37.08	73.33	36.67	1.36	1.7615

Table (A-1.3) Seasonal Indexes for rainfall data in Chauk (2006-2018) (Continue)

Time	Rainfall Data				Centered Moving Average (CMA)	Rainfall Data/CMA	Seasonal Index
2008 Jul	26	435	36.25	72.5	36.25	0.717241	0.805591
2008 Aug	38	435	36.25	72.5	36.25	1.048276	1.702509
2008 Sep	110	435	36.25	72.75	36.375	3.024055	2.400451
2008 Oct	115	438	36.5	71.5833	35.7917	3.213038	2.952184
2008 Nov	0	421	35.0833	69.5	34.75	0	0.232585
2008 Dec	0	413	34.4167	68	34	0	0.090271
2009 Jan	0	403	33.5833	66.9167	33.4583	0	0.164244
2009 Feb	0	400	33.3333	62.75	31.375	0	0.037572
2009 Mar	0	353	29.4167	57.1667	28.5833	0	0.07366
2009 Apr	3	333	27.75	55.5	27.75	0.108108	0.207431
2009 May	79	333	27.75	55.5	27.75	2.846847	1.572596
2009 Jun	42	333	27.75	55.5	27.75	1.513514	1.761508
2009 Jul	16	333	27.75	55.5	27.75	0.576577	0.805591
2009 Aug	35	333	27.75	55.9167	27.9583	1.251863	1.702509
2009 Sep	63	338	28.1667	56.0833	28.0417	2.246657	2.400451
2009 Oct	95	335	27.9167	53.8333	26.9167	3.529412	2.952184
2009 Nov	0	311	25.9167	56.8333	28.4167	0	0.232585
2009 Dec	0	371	30.9167	60.75	30.375	0	0.090271
2010 Jan	0	358	29.8333	64.4167	32.2083	0	0.164244
2010 Feb	0	415	34.5833	74.3333	37.1667	0	0.037572
2010 Mar	5	477	39.75	103.833	51.9167	0.096308	0.07366
2010 Apr	0	769	64.0833	128.167	64.0833	0	0.207431
2010 May	55	769	64.0833	129.333	64.6667	0.850515	1.572596
2010 Jun	102	783	65.25	130.5	65.25	1.563218	1.761508
2010 Jul	3	783	65.25	130.5	65.25	0.045977	0.805591
2010 Aug	92	783	65.25	130.75	65.375	1.407266	1.702509
2010 Sep	125	786	65.5	132	66	1.893939	2.400451
2010 Oct	387	798	66.5	135.333	67.6667	5.719212	2.952184
2010 Nov	0	826	68.8333	133.667	66.8333	0	0.232585
2010 Dec	14	778	64.8333	130.417	65.2083	0.214696	0.090271
2011 Jan	0	787	65.5833	141.25	70.625	0	0.164244
2011 Feb	0	908	75.6667	147	73.5	0	0.037572

Table (A-1.3) Seasonal Indexes for rainfall data in Chauk (2006-2018) (Continue)

Time	Rainfall Data				Centered Moving Average (CMA)	Rainfall Data/CMA	Seasonal Index
2011 Mar	8	856	71.33333	134.5833	67.29167	0.118885449	0.07366
2011 Apr	12	759	63.25	126.5	63.25	0.18972332	0.207431
2011 May	83	759	63.25	125.75	62.875	1.320079523	1.572596
2011 Jun	54	750	62.5	125	62.5	0.864	1.761508
2011 Jul	12	750	62.5	125	62.5	0.192	0.805591
2011 Aug	213	750	62.5	125	62.5	3.408	1.702509
2011 Sep	73	750	62.5	124	62	1.177419355	2.400451
2011 Oct	290	738	61.5	116.5	58.25	4.978540773	2.952184
2011 Nov	0	660	55	107.5833	53.79167	0	0.232585
2011 Dec	5	631	52.58333	106.5	53.25	0.093896714	0.090271
2012 Jan	0	647	53.91667	95.91667	47.95833	0	0.164244
2012 Feb	0	504	42	90.75	45.375	0	0.037572
2012 Mar	8	585	48.75	79	39.5	0.202531646	0.07366
2012 Apr	0	363	30.25	61.08333	30.54167	0	0.207431
2012 May	5	370	30.83333	61.25	30.625	0.163265306	1.572596
2012 Jun	25	365	30.41667	60.83333	30.41667	0.821917808	1.761508
2012 Jul	28	365	30.41667	60.83333	30.41667	0.920547945	0.805591
2012 Aug	70	365	30.41667	60.16667	30.08333	2.326869806	1.702509
2012 Sep	154	357	29.75	59.91667	29.95833	5.140472879	2.400451
2012 Oct	68	362	30.16667	63.75	31.875	2.133333333	2.952184
2012 Nov	7	403	33.58333	72.08333	36.04167	0.194219653	0.232585
2012 Dec	0	462	38.5	75.91667	37.95833	0	0.090271
2013 Jan	0	449	37.41667	71.75	35.875	0	0.164244
2013 Feb	0	412	34.33333	74.66667	37.33333	0	0.037572
2013 Mar	0	484	40.33333	93.16667	46.58333	0	0.07366
2013 Apr	5	634	52.83333	105.0833	52.54167	0.095162569	0.207431
2013 May	46	627	52.25	104.5	52.25	0.880382775	1.572596
2013 Jun	84	627	52.25	104.5	52.25	1.607655502	1.761508
2013 Jul	15	627	52.25	104.5	52.25	0.28708134	0.805591
2013 Aug	33	627	52.25	104.5	52.25	0.631578947	1.702509
2013 Sep	226	627	52.25	104.0833	52.04167	4.342674139	2.400451

Table (A-1.3) Seasonal Indexes for rainfall data in Chauk (2006-2018) (Continue)

Time	Rainfall Data				Centered Moving Average (CMA)	Rainfall Data/CMA	Seasonal Index
2013 Oct	218	622	51.83333	102.5	51.25	4.2536585	2.952184
2013 Nov	0	608	50.66667	98.66667	49.33333	0	0.232585
2013 Dec	0	576	48	97.58333	48.79167	0	0.090271
2014 Jan	0	595	49.58333	99.25	49.625	0	0.164244
2014 Feb	0	596	49.66667	87	43.5	0	0.037572
2014 Mar	0	448	37.33333	62.33333	31.16667	0	0.07366
2014 Apr	0	300	25	53.08333	26.54167	0	0.207431
2014 May	32	337	28.08333	56.16667	28.08333	1.1394659	1.572596
2014 Jun	52	337	28.08333	58.25	29.125	1.7854077	1.761508
2014 Jul	34	362	30.16667	60.33333	30.16667	1.1270718	0.805591
2014 Aug	34	362	30.16667	60.33333	30.16667	1.1270718	1.702509
2014 Sep	78	362	30.16667	61.41667	30.70833	2.5400271	2.400451
2014 Oct	70	375	31.25	60.75	30.375	2.3045267	2.952184
2014 Nov	37	354	29.5	59.91667	29.95833	1.2350487	0.232585
2014 Dec	0	365	30.41667	67.25	33.625	0	0.090271
2015 Jan	25	442	36.83333	81	40.5	0.617284	0.164244
2015 Feb	0	530	44.16667	88.91667	44.45833	0	0.037572
2015 Mar	0	537	44.75	97.66667	48.83333	0	0.07366
2015 Apr	13	635	52.91667	103.5833	51.79167	0.2510056	0.207431
2015 May	11	608	50.66667	101.4167	50.70833	0.2169269	1.572596
2015 Jun	63	609	50.75	99.41667	49.70833	1.2673931	1.761508
2015 Jul	111	584	48.66667	97.33333	48.66667	2.2808219	0.805591
2015 Aug	122	584	48.66667	97.33333	48.66667	2.5068493	1.702509
2015 Sep	85	584	48.66667	96.25	48.125	1.7662338	2.400451
2015 Oct	168	571	47.58333	100.8333	50.41667	3.3322314	2.952184
2015 Nov	10	639	53.25	112.5	56.25	0.1777778	0.232585
2015 Dec	1	711	59.25	115.1667	57.58333	0.0173661	0.090271
2016 Jan	0	671	55.91667	123.3333	61.66667	0	0.164244
2016 Feb	0	809	67.41667	143.5	71.75	0	0.037572
2016 Mar	0	913	76.08333	158.8333	79.41667	0	0.07366
2016 Apr	0	993	82.75	168.8333	84.41667	0	0.207431
2016 May	79	1033	86.08333	172.0833	86.04167	0.91816	1.572596

Table (A-1.3) Seasonal Indexes for rainfall data in Chauk (2006-2018) (Continue)

Time	Rainfall Data				Centered Moving Average (CMA)	Rainfall Data/CMA	Seasonal Index
2016 Jun	135	1032	86	172	86	1.569767	1.761508
2016 Jul	71	1032	86	172	86	0.8255814	0.805591
2016 Aug	260	1032	86	172.6667	86.33333	3.01158301	1.702509
2016 Sep	189	1040	86.66667	180.4167	90.20833	2.09515012	2.400451
2016 Oct	248	1125	93.75	190.3333	95.16667	2.60595447	2.952184
2016 Nov	50	1159	96.58333	186.3333	93.16667	0.53667263	0.232585
2016 Dec	0	1077	89.75	177.5	88.75	0	0.090271
2017 Jan	0	1053	87.75	160.25	80.125	0	0.164244
2017 Feb	0	870	72.5	140.25	70.125	0	0.037572
2017 Mar	8	813	67.75	125.8333	62.91667	0.12715232	0.07366
2017 Apr	85	697	58.08333	113.4167	56.70833	1.49889787	0.207431
2017 May	113	664	55.33333	111.4167	55.70833	2.02842184	1.572596
2017 Jun	53	673	56.08333	115.5833	57.79167	0.91708724	1.761508
2017 Jul	47	714	59.5	119	59.5	0.78991597	0.805591
2017 Aug	77	714	59.5	118.3333	59.16667	1.30140845	1.702509
2017 Sep	132	706	58.83333	110.75	55.375	2.38374718	2.400451
2017 Oct	132	623	51.91667	105.8333	52.91667	2.49448819	2.952184
2017 Nov	17	647	53.91667	119.4167	59.70833	0.28471738	0.232585
2017 Dec	9	786	65.5	130	65	0.13846154	0.090271
2018 Jan	41	774	64.5	128.5	64.25	0.6381323	0.164244
2018 Feb	0	768	64	122.25	61.125	0	0.037572
2018 Mar	0	699	58.25	111.3333	55.66667	0	0.07366
2018 Apr	2	637	53.08333	104.75	52.375	0.03818616	0.207431
2018 May	137	620	51.66667	103.6667	51.83333	2.64308682	1.572596
2018 Jun	192	624	52	52	26	7.38461538	1.761508
2018 Jul	35						0.805591
2018 Aug	71						1.702509
2018 Sep	63						2.400451
2018 Oct	70						2.952184
2018 Nov	0						0.232585
2018 Dec	13						0.090271

Table (A-1.4)

Table (A-1.4) Seasonal Indexes for rainfall data in Shwebo Township (2006-2018)

Time	Rainfall data				Centered Moving Average (CMA)	Rainfall Value/CMA	Seasonal Index
2006 Jan	0						0.1642
2006 Feb	0						0.0376
2006 Mar	0						0.0737
2006 Apr	89						0.2074
2006 May	76						1.5726
2006 Jun	218	1123	93.58	187.17	93.58	2.33	1.7615
2006 Jul	158	1123	93.58	187.92	93.96	1.68	0.8056
2006 Aug	195	1132	94.33	188.67	94.33	2.07	1.7025
2006 Sep	182	1132	94.33	181.25	90.63	2.01	2.4005
2006 Oct	104	1043	86.92	183.17	91.58	1.14	2.9522
2006 Nov	98	1155	96.25	185.42	92.71	1.06	0.2326
2006 Dec	3	1070	89.17	166.92	83.46	0.04	0.0903
2007 Jan	0	933	77.75	153.92	76.96	0.00	0.1642
2007 Feb	9	914	76.17	144.00	72.00	0.13	0.0376
2007 Mar	0	814	67.83	139.17	69.58	0.00	0.0737
2007 Apr	0	856	71.33	138.42	69.21	0.00	0.2074
2007 May	188	805	67.08	133.92	66.96	2.81	1.5726
2007 Jun	133	802	66.83	136.33	68.17	1.95	1.7615
2007 Jul	21	834	69.50	138.25	69.13	0.30	0.8056
2007 Aug	176	825	68.75	137.50	68.75	2.56	1.7025
2007 Sep	82	825	68.75	137.67	68.83	1.19	2.4005
2007 Oct	146	827	68.92	131.58	65.79	2.22	2.9522
2007 Nov	47	752	62.67	121.83	60.92	0.77	0.2326
2007 Dec	0	710	59.17	121.00	60.50	0.00	0.0903
2008 Jan	32	742	61.83	117.75	58.88	0.54	0.1642
2008 Feb	0	671	55.92	110.17	55.08	0.00	0.0376
2008 Mar	0	651	54.25	107.75	53.88	0.00	0.0737
2008 Apr	2	642	53.50	103.58	51.79	0.04	0.2074
2008 May	113	601	50.08	100.17	50.08	2.26	1.5726
2008 Jun	91	601	50.08	97.50	48.75	1.87	1.7615
2008 Jul	53	569	47.42	94.83	47.42	1.12	0.8056
2008 Aug	105	569	47.42	94.83	47.42	2.21	1.7025
2008 Sep	62	569	47.42	97.00	48.50	1.28	2.4005
2008 Oct	137	595	49.58	104.83	52.42	2.61	2.9522
2008 Nov	6	663	55.25	118.92	59.46	0.10	0.2326

Table (A-1.4) Seasonal Indexes for rainfall data in Shwebo Township (2006-2018) (Continue)

Time	Rainfall data				Centered Moving Average (CMA)	Rainfall Value/CMA	Seasonal Index
2008 Dec	0	764	63.6667	124.25	62.125	0	0.09027
2009 Jan	0	727	60.5833	123.833	61.9167	0	0.16424
2009 Feb	0	759	63.25	132.917	66.4583	0	0.03757
2009 Mar	0	836	69.6667	132.917	66.4583	0	0.07366
2009 Apr	28	759	63.25	126	63	0.44444	0.20743
2009 May	181	753	62.75	125.5	62.75	2.88446	1.5726
2009 Jun	192	753	62.75	125.5	62.75	3.05976	1.76151
2009 Jul	16	753	62.75	125.5	62.75	0.25498	0.80559
2009 Aug	137	753	62.75	125.833	62.9167	2.17748	1.70251
2009 Sep	139	757	63.0833	124.833	62.4167	2.22697	2.40045
2009 Oct	60	741	61.75	113	56.5	1.06195	2.95218
2009 Nov	0	615	51.25	100.25	50.125	0	0.23259
2009 Dec	0	588	49	98.9167	49.4583	0	0.09027
2010 Jan	0	599	49.9167	101.5	50.75	0	0.16424
2010 Feb	0	619	51.5833	104.25	52.125	0	0.03757
2010 Mar	4	632	52.6667	130.5	65.25	0.0613	0.07366
2010 Apr	12	934	77.8333	155.667	77.8333	0.15418	0.20743
2010 May	55	934	77.8333	158.5	79.25	0.69401	1.5726
2010 Jun	165	968	80.6667	161.333	80.6667	2.04545	1.76151
2010 Jul	27	968	80.6667	161.333	80.6667	0.33471	0.80559
2010 Aug	157	968	80.6667	165.25	82.625	1.90015	1.70251
2010 Sep	152	1015	84.5833	179.667	89.8333	1.69202	2.40045
2010 Oct	362	1141	95.0833	198.833	99.4167	3.64124	2.95218
2010 Nov	0	1245	103.75	201	100.5	0	0.23259
2010 Dec	34	1167	97.25	198	99	0.34343	0.09027
2011 Jan	0	1209	100.75	211.667	105.833	0	0.16424
2011 Feb	0	1331	110.917	216.5	108.25	0	0.03757
2011 Mar	51	1267	105.583	204.667	102.333	0.49837	0.07366
2011 Apr	138	1189	99.0833	198.167	99.0833	1.39277	0.20743
2011 May	159	1189	99.0833	195.75	97.875	1.62452	1.5726
2011 Jun	87	1160	96.6667	193.333	96.6667	0.9	1.76151
2011 Jul	69	1160	96.6667	193.333	96.6667	0.71379	0.80559
2011 Aug	279	1160	96.6667	189.417	94.7083	2.94589	1.70251
2011 Sep	88	1113	92.75	174.667	87.3333	1.00763	2.40045

Table (A-1.4) Seasonal Indexes for rainfall data in Shwebo Township (2006-2018) (Continue)

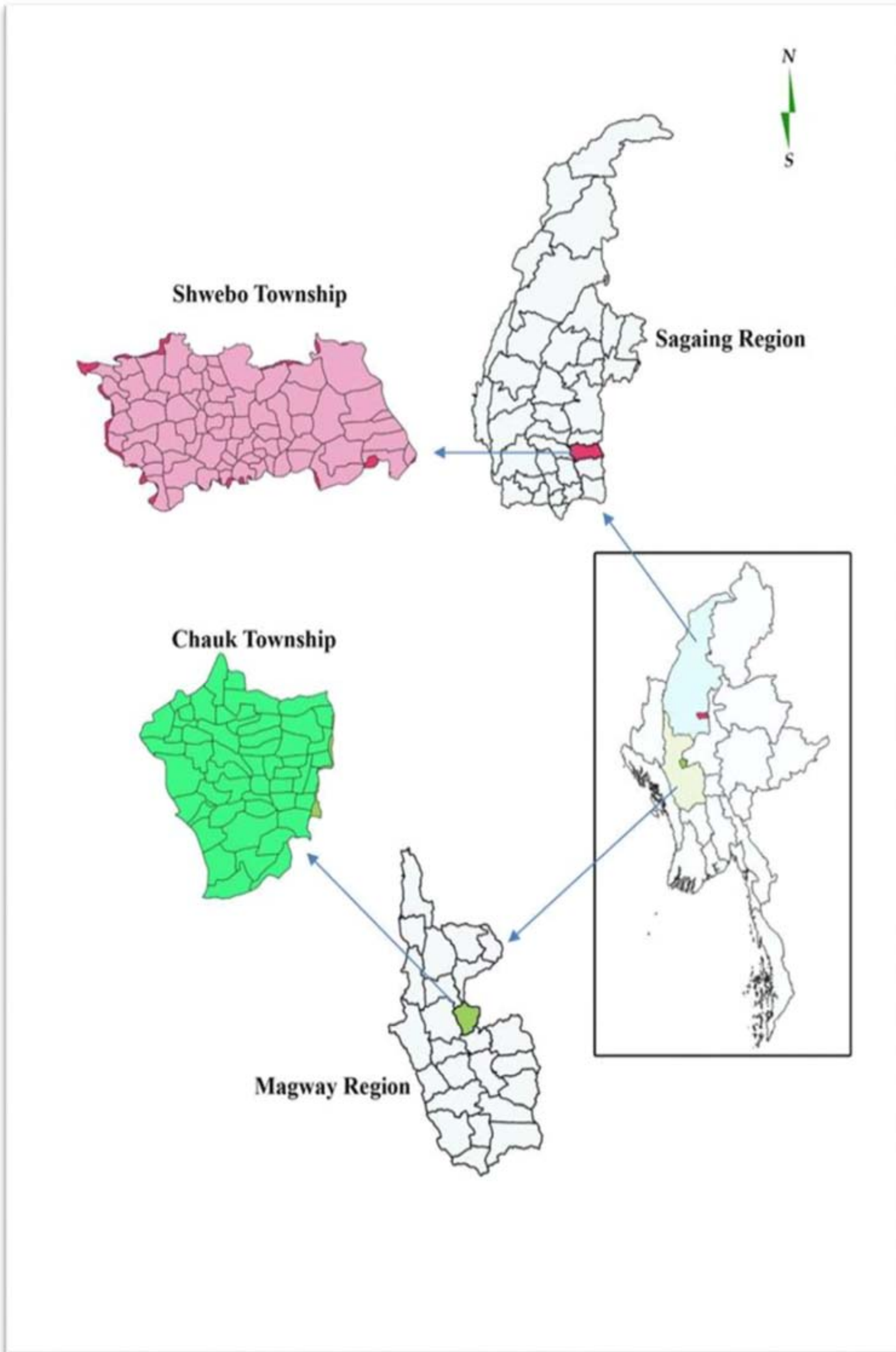
Time	Rainfall data				Centered Moving Average (CMA)	Rainfall Value/ CMA	Seasonal Index
2011 Oct	284	983	81.9167	151.917	75.9583	3.73889	2.95218
2011 Nov	0	840	70	136.75	68.375	0	0.23259
2011 Dec	5	801	66.75	134.25	67.125	0.07449	0.09027
2012 Jan	0	810	67.5	119.083	59.5417	0	0.16424
2012 Feb	0	619	51.5833	119.083	59.5417	0	0.03757
2012 Mar	4	810	67.5	120.917	60.4583	0.06616	0.07366
2012 Apr	8	641	53.4167	108.5	54.25	0.14747	0.20743
2012 May	16	661	55.0833	109.75	54.875	0.29157	1.5726
2012 Jun	48	656	54.6667	109.333	54.6667	0.87805	1.76151
2012 Jul	78	656	54.6667	109.333	54.6667	1.42683	0.80559
2012 Aug	88	656	54.6667	109	54.5	1.61468	1.70251
2012 Sep	279	652	54.3333	109	54.5	5.11927	2.40045
2012 Oct	115	656	54.6667	113	56.5	2.0354	2.95218
2012 Nov	20	700	58.3333	117.667	58.8333	0.33994	0.23259
2012 Dec	0	712	59.3333	116.083	58.0417	0	0.09027
2013 Jan	0	681	56.75	131.333	65.6667	0	0.16424
2013 Feb	0	895	74.5833	144.333	72.1667	0	0.03757
2013 Mar	0	837	69.75	145.833	72.9167	0	0.07366
2013 Apr	12	913	76.0833	150.5	75.25	0.15947	0.20743
2013 May	60	893	74.4167	149	74.5	0.80537	1.5726
2013 Jun	60	895	74.5833	149.167	74.5833	0.80447	1.76151
2013 Jul	47	895	74.5833	150.417	75.2083	0.62493	0.80559
2013 Aug	302	910	75.8333	151.667	75.8333	3.98242	1.70251
2013 Sep	221	910	75.8333	151.583	75.7917	2.91589	2.40045
2013 Oct	191	909	75.75	150.167	75.0833	2.54384	2.95218
2013 Nov	0	893	74.4167	148.917	74.4583	0	0.23259
2013 Dec	2	894	74.5	146.75	73.375	0.02726	0.09027
2014 Jan	0	867	72.25	139.167	69.5833	0	0.16424
2014 Feb	15	803	66.9167	137	68.5	0.21898	0.03757
2014 Mar	0	841	70.0833	129.917	64.9583	0	0.07366
2014 Apr	11	718	59.8333	119.917	59.9583	0.18346	0.20743
2014 May	44	721	60.0833	120	60	0.73333	1.5726
2014 Jun	61	719	59.9167	119.917	59.9583	1.01737	1.76151

Table (A-1.4) Seasonal Indexes for rainfall data in Shwebo Township (2006-2018) (Continue)

Time	Rainfall data				Centered Moving Average (CMA)	Rainfall Value/ CMA	Seasonal Index
2014 Jul	20	720	60	118.75	59.375	0.33684	0.80559
2014 Aug	238	705	58.75	118.667	59.3333	4.01124	1.70251
2014 Sep	259	719	59.9167	120.333	60.1667	4.30471	2.40045
2014 Oct	68	725	60.4167	125	62.5	1.088	2.95218
2014 Nov	3	775	64.5833	125.25	62.625	0.0479	0.23259
2014 Dec	0	728	60.6667	150.333	75.1667	0	0.09027
2015 Jan	1	1076	89.6667	169.167	84.5833	0.01182	0.16424
2015 Feb	0	954	79.5	155	77.5	0	0.03757
2015 Mar	14	906	75.5	162.25	81.125	0.17257	0.07366
2015 Apr	17	1041	86.75	173.833	86.9167	0.19559	0.20743
2015 May	94	1045	87.0833	174.167	87.0833	1.07943	1.5726
2015 Jun	14	1045	87.0833	174.083	87.0417	0.16084	1.76151
2015 Jul	368	1044	87	174.25	87.125	4.22382	0.80559
2015 Aug	116	1047	87.25	173.333	86.6667	1.33846	1.70251
2015 Sep	211	1033	86.0833	172.667	86.3333	2.44402	2.40045
2015 Oct	203	1039	86.5833	172.583	86.2917	2.35249	2.95218
2015 Nov	7	1032	86	184.25	92.125	0.07598	0.23259
2015 Dec	0	1179	98.25	174.667	87.3333	0	0.09027
2016 Jan	0	917	76.4167	178.75	89.375	0	0.16424
2016 Feb	3	1228	102.333	197.417	98.7083	0.03039	0.03757
2016 Mar	0	1141	95.0833	179.417	89.7083	0	0.07366
2016 Apr	23	1012	84.3333	169.5	84.75	0.27139	0.20743
2016 May	87	1022	85.1667	170.333	85.1667	1.02153	1.5726
2016 Jun	161	1022	85.1667	170.333	85.1667	1.89041	1.76151
2016 Jul	106	1022	85.1667	170.083	85.0417	1.24645	0.80559
2016 Aug	427	1019	84.9167	170.75	85.375	5.00146	1.70251
2016 Sep	124	1030	85.8333	170.417	85.2083	1.45526	2.40045
2016 Oct	74	1015	84.5833	178.667	89.3333	0.82836	2.95218
2016 Nov	17	1129	94.0833	181.667	90.8333	0.18716	0.23259
2016 Dec	0	1051	87.5833	178.25	89.125	0	0.09027
2017 Jan	0	1088	90.6667	166	83	0	0.16424
2017 Feb	0	904	75.3333	156.667	78.3333	0	0.03757
2017 Mar	11	976	81.3333	168.167	84.0833	0.13082	0.07366

Table (A-1.4) Seasonal Indexes for rainfall data in Shwebo Township (2006-2018) (Continue)

Time	Rainfall data				Centered Moving Average (CMA)	Rainfall Value/ CMA	Seasonal Index
2017 Apr	8	1042	86.8333	176.25	88.125	0.09078	0.20743
2017 May	201	1073	89.4167	179.75	89.875	2.23644	1.5726
2017 Jun	83	1084	90.3333	182.583	91.2917	0.90917	1.76151
2017 Jul	143	1107	92.25	184.5	92.25	1.55014	0.80559
2017 Aug	243	1107	92.25	183.583	91.7917	2.6473	1.70251
2017 Sep	196	1096	91.3333	183.417	91.7083	2.13721	2.40045
2017 Oct	140	1105	92.0833	177.417	88.7083	1.57821	2.95218
2017 Nov	48	1024	85.3333	182.833	91.4167	0.52507	0.23259
2017 Dec	11	1170	97.5	186.833	93.4167	0.11775	0.09027
2018 Jan	23	1072	89.3333	162.417	81.2083	0.28322	0.16424
2018 Feb	0	877	73.0833	136.417	68.2083	0	0.03757
2018 Mar	0	760	63.3333	133.25	66.625	0	0.07366
2018 Apr	17	839	69.9167	135.833	67.9167	0.25031	0.20743
2018 May	120	791	65.9167	130.917	65.4583	1.83323	1.5726
2018 Jun	229	780	65	65	32.5	7.04615	1.76151
2018 Jul	45						0.80559
2018 Aug	48						1.70251
2018 Sep	79						2.40045
2018 Oct	219						2.95218
2018 Nov	0						0.23259
2018 Dec	0						0.09027



Map 1. Location of Study Sites.