International Journal of Statistics and Applied Mathematics 2019; 4(6): 24-31

International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2019; 4(6): 24-31

www.mathsjournal.com
Received: 10-09-2019
Accepted: 12-10-2019

KK Saxena
Professor of Statistics in
The Department of Statistics,
University of Dodoma, Dodoma,
Tanzania

Guntram Martin Wagofya
Assistant Lecturer in
The Department of Mathematics
and Statistics, Mbeya University
of Science and Technology,
Mbeya, Tanzania

Corresponding Author:
KK Saxena
Professor of Statistics in
The Department of Statistics,
University of Dodoma, Dodoma,
Tanzania

Performance analysis of the loan repayment by poor borrowers in rural areas of saccos in Njombe, Tanzania by using time series arima modelling

KK Saxena and Guntram Martin Wagofya

Abstract
This paper describes the application of the Box-Jenkins methodology in time series analysis in measuring the performance of SACCOS in respect of loan repayment. An autoregressive integrated moving average (ARIMA) model is fitted and then forecasts are made to see the viability of the loan repayment. The study focuses on time series modelling and forecasting of loan repayment at Kifanya SACCOS Limited in Njombe in Tanzania. The findings show that the best time series model for monthly loan repayment is ARIMA (1, 1, 2). This model fits well the historical data and hence can be used in short-term forecasting of loan repayment. The results shows that the loan repayment at Kifanya SACCOS Limited in Njombe in Tanzania follows an increasing trend, which is an indicator of its good governance.

Keywords: Time series, loan repayment, forecasting and ARIMA model

1. Introduction

Savings and Credit Cooperatives (SACCOS) are one of the Micro Finance Institutions (MFIs) in Tanzania. They provide a wide range of savings and loan products to those individual poor borrowers in rural areas, who cannot afford to go the banks. The SACCOS are working as a bridge between these small borrowers and the formal financial institutions. These cooperative societies offer small sized loans to members to whom the formal financial institutions will not offer loans or credit in view of the various reasons most importantly the high interest rate and other formalities in processing the applications of poor people. So in other way, SACCOS are providing ways to help small and poor creditors in rural areas particularly those who, otherwise are under the poverty line. These under privileged rural persons take loans from SACCOS and return the loan payments with small interest rate.

This study aims at fitting a time series model that can be used to forecast the trend of loan repayment at Kifanya SACCOS Ltd by using the Box-Jenkins procedure. ARIMA model for forecasting is recommended in this study because the model assumes that the observations of time series are statistically dependent on each other, meaning that the autocorrelation between successive values of the time series data under study is not zero (Kumar & Anand, 2015) [7].

A time series is a sequence of observations on a particular variable recorded at discrete intervals of time (Adhikari, 2013) [2]. Time series analysis involves the techniques of studying a given time series by analyzing its patterns, fitting models for those patterns and using the fits to predict the future values of that time series. Time series analysis is a very important aspect of statistics which is applicable in many fields where forecasting is highly needed. Forecasting is very important in areas of economics, business, finance, medicine, environment, industry and so on.

The Box-Jenkins (or ARIMA) procedure is among the most widely used forecasting methods in time series analysis. ARIMA Models are excellent short-term forecasting linear models for a wide variety of time series. These models have shown efficiency and have outperformed complex structural models in short-term forecasting (Adebiyi, Adewumi, & Ayo, 2014) [1]. A time series model is considered to be linear when the current value of the series is a linear function of past observations (M. Zhang, 2018) [12].

In time series analysis the interest is to obtain a suitable time series model that can be used to predict the future values of the series. The fitted model becomes useful when it is physically meaningful and observes the principle of parsimony. The principle of parsimony recommends that when we have more than one model the one with fewer parameters is preferred to the one with more parameters.

2. Literature Review

Time series modelling is very important in economic forecasting which plays a very crucial role in decision making process in both public and private sector. These sectors base many of their decisions on future expected economic conditions or on the predictions of specific indicators of interest such as income growth, exchange rates, inflation, interest rates, unemployment and many others. But the realization of economic outcomes is a vast, dynamic and stochastic process which makes forecasting very difficult and forecast errors unavoidable. However, forecast accuracy and reliability can be improved by using appropriate models and methods such as ARIMA and VAR models (Ghysel & Marcellino, 2018) [5]. Otooet *et al*. (2015) analyzed loan default at Minescho Cooperative Credit Union in Ghana by using Regression and ARIMA models. The fitted ARIMA (3,2,2) model indicated an increasing trend on the amount of loan default. Iwueze *et al*. (2013) [6] fitted a time Series model and showed that ARIMA (2,1,0) was the appropriate model to describe the pattern of the financial series. Ayo (2014) [1] applied ARIMA model in forecasting stock price and showed that ARIMA model is strong in short-term forecasting. Zhang (2013) [11] applied Autoregressive Integrated Moving Average (ARIMA) model, the Vector Autoregressive (VAR) model and the First-order Autoregressive (AR(1)) model in modelling and forecasting the regional GDP. The findings showed that all the three models were valid in generating short-term forecasts. Nochai & Nochai (2006) applied ARIMA Model in forecasting three types of oil palm price; farm price, wholesale price and pure oil price and showed that the proper ARIMA Models for forecasting farm price, wholesale price and pure oil price of oil palm were ARIMA (2,1,0), ARIMA (1,0,1) and ARIMA (3,0,0) respectively.

In general, there is no sufficient literature on ARIMA modelling and forecasting of loan repayment particularly in savings and credit cooperative societies or credit unions. Therefore, this study adds literature on time series modelling and forecasting in the context of financial data.

3. Materials and Methodology

3.1 ARIMA Processes and their models

The ARIMA procedure developed by Box and Jenkins for modelling time series data is applicable when the modelling is done on a stationary time series. This procedure involves the following models.

- **Autoregressive model:** In autoregressive model, the forecast of a given variable is a linear combination of past values of that variable. An autoregressive model of order p which is denoted by AR(p) is written as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t \qquad (3.1)$$

Where Y_t is a time series and ε_t is white noise.

- **Moving average model**: In moving average model, the forecast of a given variable is a weighted average of past forecast errors. A moving average model of order p which is denoted by MA(q) is written as

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad (3.2)$$

Where ε_t is white noise.

- **Autoregressive moving average model**: This model is a combination of an autoregressive model and a moving average model. The model is denoted by ARMA (p,q) where p and q are the orders of autoregressive component and moving average component respectively and it is written as

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q} \qquad (3.3)$$

Where ε_t is white noise.

- **Autoregressive integrated moving average (ARIMA) model:** An autoregressive integrated moving average (ARIMA) model is an ARMA model which is obtained when a time series is made stationary by differencing it say d times. This model is denoted by ARIMA(p,d,q) and is written as

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\varepsilon_t \qquad (3.4)$$

Where,

p, is the order of the autoregressive component
d, is the number of differences needed to obtain a stationary ARMA(*p*, *q*) process
q, is the order of the moving average component
ε_t, is white noise

- **Seasonal Autoregressive integrated moving average (SARIMA) model:** The model is developed by including seasonal parameters in non-seasonal ARIMA models. A seasonal ARIMA model with seasonal period *s* is written as

$$\phi(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Y_t = \theta(B)\Theta_Q(B^s)\varepsilon_t \quad (3.5)$$

Where,
$\phi(B) = (1-\phi_1 B - \ldots - \phi_p B^p)$ and $\theta(B) = (1-\theta_1 B - \ldots - \theta_q B^q)$ are the non-seasonal AR and MA characteristic operators respectively.

$\Phi_P(B^s) = (1-\Phi_1 B^s - \ldots - \Phi_p B^{Ps})$ and $\Theta_Q(B^s) = (1-\Theta_1 B^s - \ldots - \Theta_Q B^{sQ})$ are the seasonal AR and MA characteristic operators respectively.

D is the number of seasonal differences
P is the order of the seasonal autoregressive process
Q is the order of seasonal moving average process
p, d, q and ε_t have already been defined in (3.4)

3.2 The Box - Jenkins Methodology
The Box-Jenkins methodology is an iterative procedure used to build an ARIMA model in three steps. The three steps involved in this procedure are;

Model identification
In this step a tentative model is suggested after studying the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of a stationary time series. Usually, the analysis starts with preliminary assessment for stationarity of the time series by plotting the data. This graph is helpful in identifying important features such as trend, seasonality and cyclic fluctuations in a given time series. Time plots are helpful in deciding whether the given time series needs to be transformed before doing any further analysis or not.

Parameter estimation
In this step methods such as the method of moments, method of maximum likelihood and the method of least squares are applied to estimate the parameters of the suggested model. Usually, this is achieved by using statistical software packages such as SPSS, STATA, Minitab, SAS and R.

Diagnosis of the model
To be useful, the tentatively identified model is examined for its adequacy. This is done by analyzing the residuals. If the fitted model is adequate then the residuals should behave like white noise and the fit can be used to generate forecasts. When the fit is not adequate, we have to go to step one where another model has to be considered.

3.3 Summary of the Box-Jenkins Methodology
The three steps involved in the Box – Jenkins procedure are summarized in the following diagram.

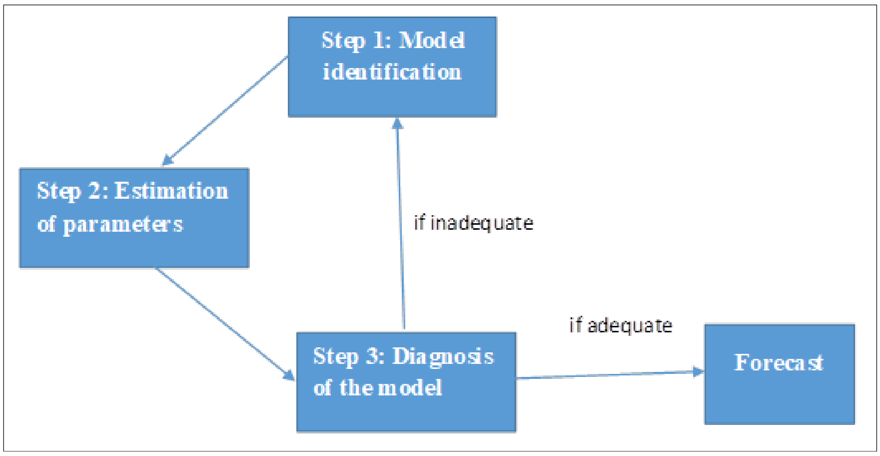

Fig 1: Outline of the Box-Jenkins Procedure

4. Data Analysis and Discussion of Results

Step 1: Model identification

Before getting a tentative model, the loan repayment data are plotted as shown in the time plot displayed in figure 2. In addition, seasonal decomposition has been done to analyze the components of this series, and the results are displayed in figure 3. From these figures it is clear that the loan repayment time series is not stationary in mean as well as in variance. For that reason, the series has to be transformed.

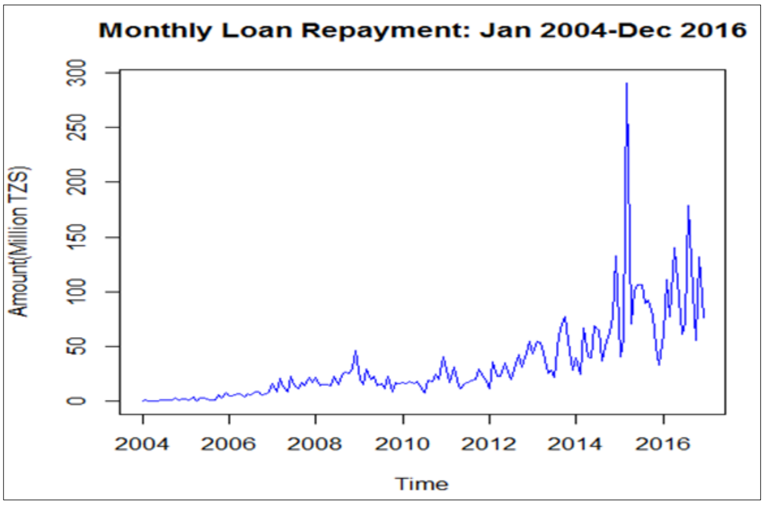

Fig 2: Time series plot for the monthly loan repayment data

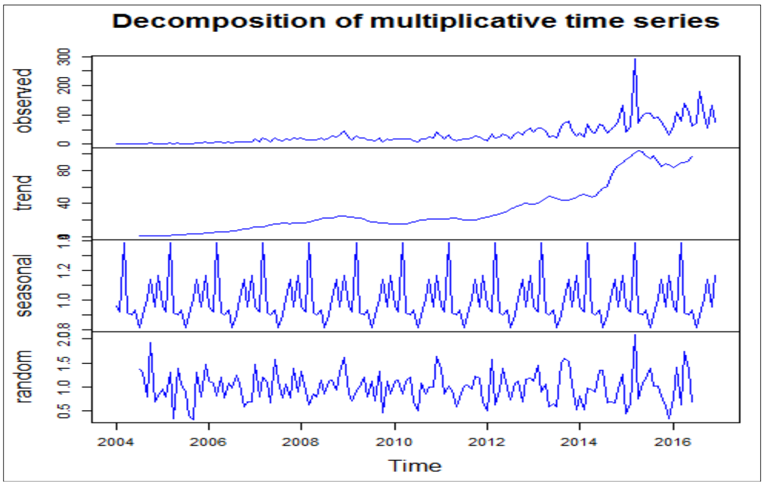

Fig 3: Components of the monthly loan repayment time series

The first transformation is conducted to stabilize the changing variance of the series. This was done by applying natural logarithms to the loan repayment data, and the transformed series is displayed in figure 4. Now the transformed series appears to have stable variance. Let us denote this time series with stable variance as X_t.

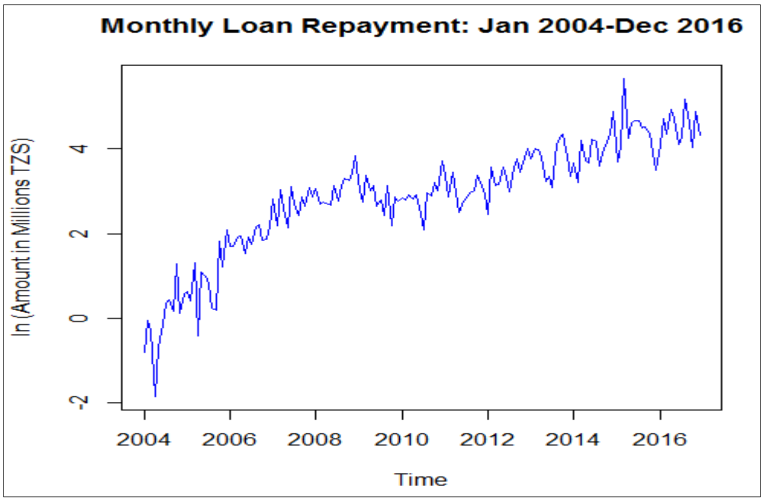

Fig 4: Time Series Plot of the log transformed loan repayment data

The second transformation is conducted to remove the trend of the series. The trend is removed by differencing the series and the differenced series W_t, is displayed in figure 5. By visual inspection, this series appears to be stationary but its stationarity is confirmed by conducting a formal test usually a unit root test. The unit root test that was employed here is the Augmented Dickey-Fuller (ADF) test.

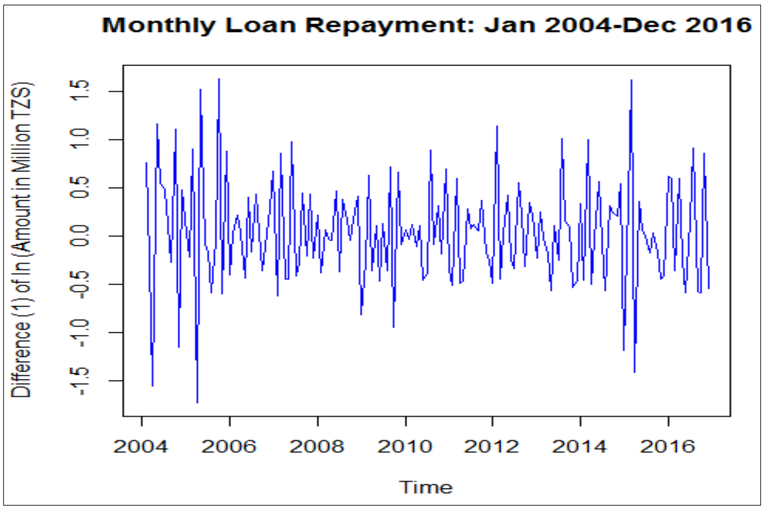

Fig 5: Time Series Plot of the first difference of the log transformed data

Unit Root Test for Stationarity of the Time Series

Augmented Dickey-Fuller (ADF) Test: In this test the hypothesis is;

H_0: The time series is non-stationary or there is unit root.

H_1: The series is stationary or there is no unit root.

The STATA output gives the results as shown in table 1 below.

Table 1: Augmented Dickey-Fuller test for the loan repayment time series

	Test Statistic	1% Critical value	5% Critical value	10% Critical value
Z(t)	-7.809	-3.494	-2.887	-2.577

MacKinnon approximate p-value for Z(t) = 0.0000

The ADF test statistic is -7.809 for lag order 5 and a p-value of less than 0.0001 is recorded indicating that the series W_t is stationary.

From the stationary time series on loan repayment, the ACF and PACF plots were obtained. These plots are displayed in figure 6 and they indicate that there is one significant spike at lag 1 in the ACF indicating that MA (1) might be the non-seasonal component. The PACF plot shows two significant spikes suggesting that AR(1) and AR(2) are likely to be the non-seasonal components. Therefore, the ARIMA model for this time series is likely to contain both AR and MA processes.

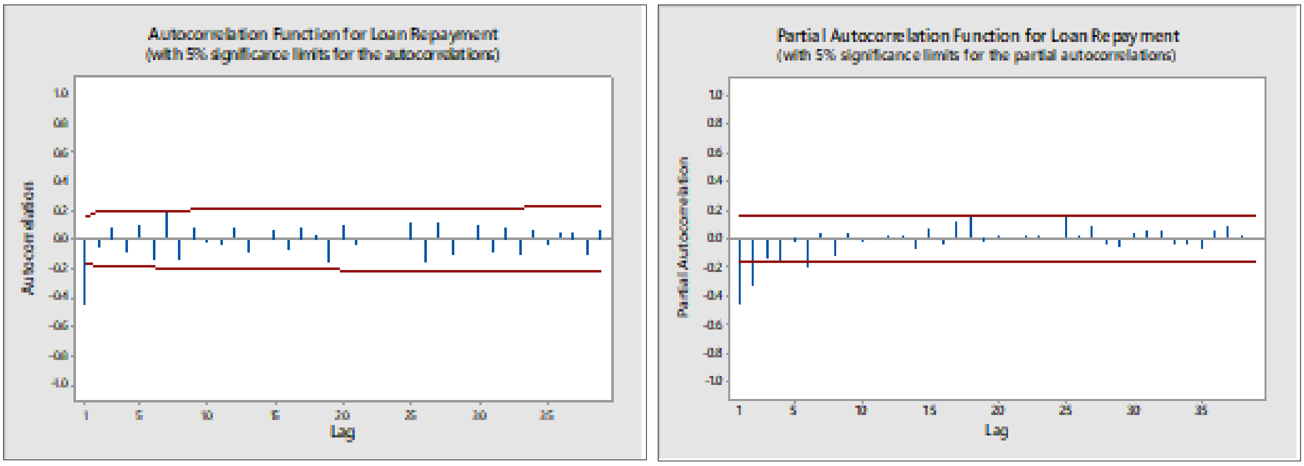

Fig 6: The ACF and PACF Plots for the transformed data

The ACF and PACF plots of the stationary series suggested the following models;

1. ARIMA (2, 1, 0)
2. ARIMA (1, 1, 2)
3. ARIMA (1, 1, 0)
4. ARIMA (0, 1, 1)

Step 2: Model Estimation

From suggested models parameter estimates were obtained by using Minitab. The best ARIMA model was selected on the basis of the following measures.

Table 2: Some statistics for the suggested loan repayment ARIMA models

Model	R-squared	RMSE	MAPE	MAE	Normalized BIC	AIC
ARIMA (2, 1, 0)	.646	20.776	40.131	11.448	6.133	208.43
ARIMA (1, 1, 2)	.678	19.864	38.652	10.999	6.075	200.83
ARIMA (1, 1, 0)	0.534	23.758	44.62	12.81	6.368	222.45
ARIMA (0, 1, 1)	.655	20.436	38.305	11.079	6.067	203.36

In selecting the best model from the list of suggested models, we consider the values of AIC and BIC. From table 2, AIC suggests that the best model is ARIMA (1, 1, 2) whilst BIC suggests the model ARIMA (0,1,1). The use of AIC as a selection criteria for the best model is recommended here because it agrees with the smallest value of RMSE which shows that there is minimum variation with the model suggested by AIC. Therefore, the best fit for the monthly loan repayment time series is the ARIMA (1,1,2). The parameters of the fitted ARIMA (1, 1, 2) were estimated in Minitab, and they appear to be significant as shown in table 3.

Table 3: Minitab Output for the Model ARIMA (1, 1, 2)

```
Final Estimates of Parameters
Type        Coef  SE Coef       T      P
AR   1  -0.8822   0.0523  -16.88  0.000
MA   1  -0.3291   0.0714   -4.61  0.000
MA   2   0.6430   0.0543   11.85  0.000
Differencing: 1 regular difference
Number of observations:  Original series 156, after differencing 155
Residuals:    SS =  31.2178 (backforecasts excluded)
              MS =  0.2054  DF = 152
Modified Box-Pierce (Ljung-Box) Chi-Square statistic
Lag             12      24      36      48
Chi-Square     5.5    13.8    22.6    40.2
DF               9      21      33      45
P-Value      0.787   0.877   0.914   0.676
```

Step 3: Diagnosis of the model

Diagnosis is done by studying the residuals to see if any pattern remains unaccounted for hence verifying that the fitted model is adequate. The fitted model is diagnosed by analysing autocorrelation in the residuals and some residual plots.

Portmanteau test: The modified portmanteau test was performed on residuals of the fitted model and results are shown in table 3. All p-values at different lags are greater than 0.05 indicating that the residuals are uncorrelated.

Residual plots: The ACF and PACF plots of the residuals in figure 7 clearly show that the residuals lie within the 95% confidence limits. This indicates that, the residuals left over after fitting the models are white noise. Further analysis on residual plots is done by considering the normal probability plot, residuals versus fitted value, histogram of the residuals and the time series plot of the residuals. These plots are displayed in figure 8. The normal probability plots and histograms indicate that the residuals are approximately normally distributed. The plots of residuals against the fitted value and against observation order indicate a stable variance in the series.

In general the analysis of residuals show that the fitted ARIMA model for the loan repayment time series is adequate and therefore the fit can be used in forecasting.

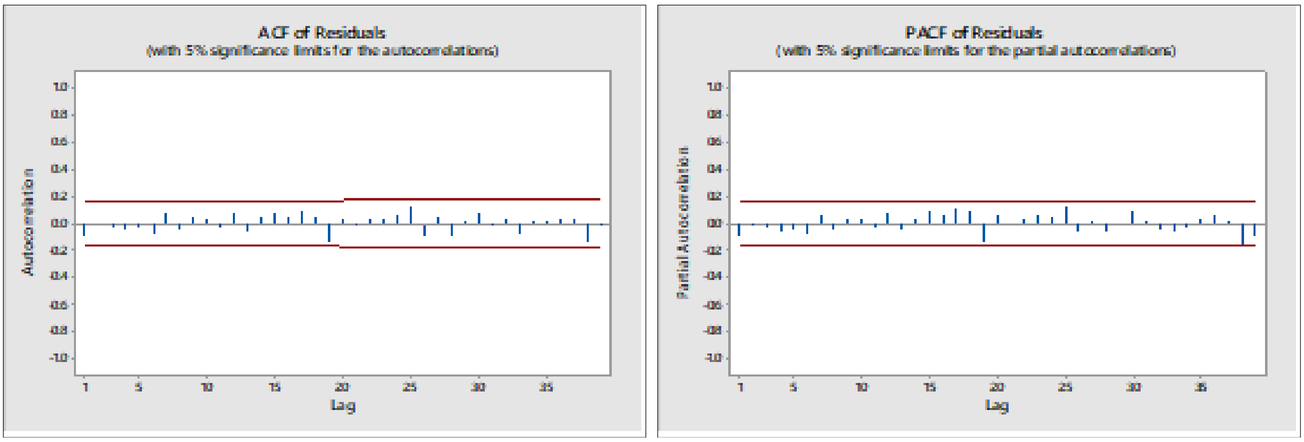

Fig 7: The ACF and PACF plots of the residuals for ARIMA(1,1,2)

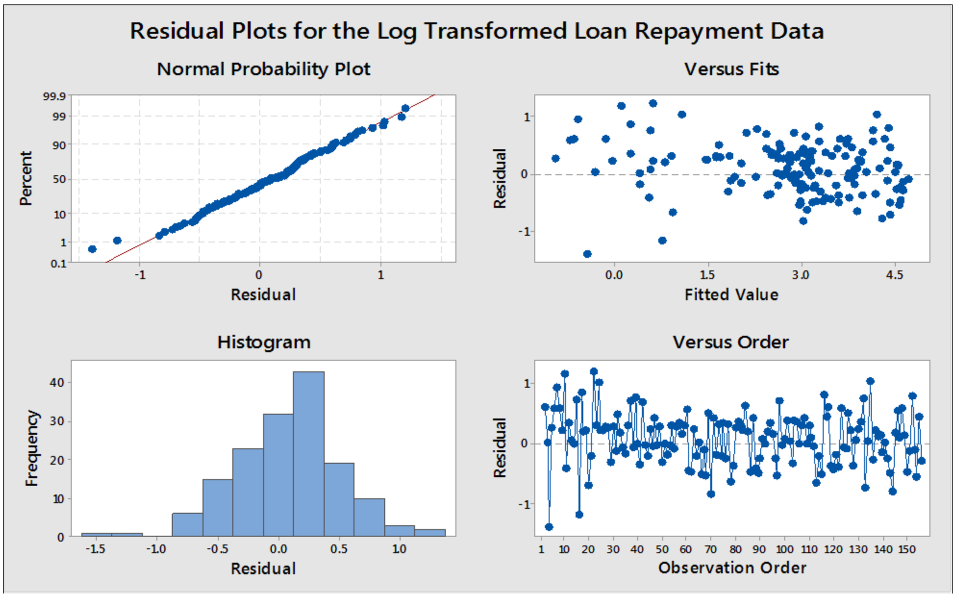

Fig 8: Residual plots for the Model ARIMA (1,1,2)

According to Tebbs (2013) [10], the fitted model ARIMA (1, 1, 2) can be written in its multiplicative form suggested by Box, Jenkins and Reinsel by using equation 3.4. Now the fit takes the form

$$(1-\phi_1 B)(1-B)X_t = (1-\theta_1 B - \theta_2 B^2)\varepsilon_t \tag{4.1}$$

Where $(1-B)X_t = X_t - X_{t-1}$ denoting the first difference of X_t, B is the backshift operator which changes the time period t to time period $t-1$ and e_t is the random error.

The estimated parameters for ARIMA (1, 1, 2) which are found in the Minitab output in table 3 are $\phi_1 = -0.8822, \theta_1 = -0.3291$ and $\theta_2 = 0.643.$

After substitution of these estimates in (4.1), the model becomes

$$(1+0.8822B)(1-B)X_t = (1+0.3291B - 0.643B^2)\varepsilon_t \tag{4.2}$$

The model in (4.2) can be expressed in terms of natural logarithms transformed data (X_t), and the fit becomes

$$\hat{X}_t = 0.1178X_{t-1} + 0.8822X_{t-2} + \varepsilon_t + 0.32918\varepsilon_{t-1} - 0.643\varepsilon_{t-2} \tag{4.3}$$

The fitted model (4.3) shows that the current value of the transformed loan repayment time series is a linear function of some of its past values together with its current and past shocks.

Using the Fitted Model to Forecast

Having obtained an adequate model ARIMA(1,1,2), the fit has been used to generate out-of-sample forecasts of monthly loan repayment at Kifanya SACCOS Ltd for the period starting from January 2017 up to December 2020. The forecasts along with their corresponding lower and upper 95 per cent confidence limit are displayed in figure 9.

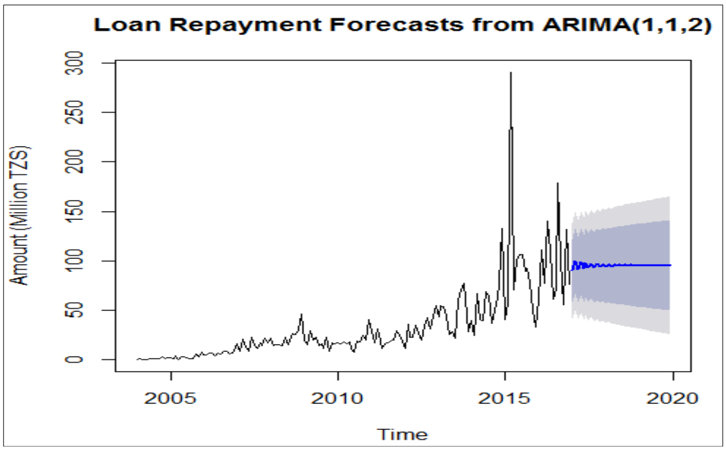

Fig 9: Out-of-sample forecasting of monthly loan repayment

5. Conclusion

The findings have shown that the fitted model for loan repayment fits well the historical data and can be used in short-term forecasting for the future values of this variable. However, in some months the forecasts may not agree strongly with the observed values as the series under study is very volatile and this increases unpredictability in its future values. But overall the loan repayment by borrowers is adequate enough and hence it is concluded that there is good governance of the SACCOS in Njombe, Tanzania.

6. References

1. Adebiyi AA, Adewumi AO, Ayo CK. Stock Price Prediction Using the ARIMA Model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 2014. https://doi.org/10.1109/UKSim.2014.67
2. Adhikari R, Agrawal RK. An Introductory Study on Time Series Modeling and Forecasting. Ratnadip Adhikari R. K. Agrawal, 2013.
3. Appiah T. Regression and Time Series Analysis of Loan Default At Minescho Cooperative Credit Union, Tarkwa. 2015; 4(08):188-195.
4. Dablemont S, Van Bellegem S, Verlezsen M. Modelling and Forecasting financial time series of Tick Data by functional analysis and neural networks, 2007, 18. Retrieved from http://hdl.handle.net/2078.1/23826
5. Ghysels E, Marcellino M. Applied Economic forecasting using Time Series Methods, Oxford University Press. Retrieved from arXiv preprint, 2018, arXiv:1302.6613, 2013 - arxiv.org
6. Iwueze IS, Nwogu EC, Nlebedim VU, Process A, Stability V, Reserves FE. Time Series Modeling of Nigeria External Reserves. 2013; 4(2):111-128.
7. Kumar M, Anand M. An Application of Time Series Arima Forecasting Model for Predicting Sugarcane Production in India. Studies in Business and Economics, 2015, 81-94. https://doi.org/10.1109/UKSim.2014.67
8. Lama A, Jha GK, Paul RK, Gurung B. Modelling and Forecasting of Price Volatility : An Application. Agricultural Economics Research Review. 2015; 28(1):73-82. https://doi.org/10.5958/0974-0279.2015.00005.1
9. Nochai R, Nochai T. Arima Model for Forecasting Oil Palm Price. Proceedings of the 2nd IMT-GT Regional Conference on Mathemathics, Statistics and Applications Universiti Sains Malaysia, Penang, 2005, 1-7.
10. Tebbs JM. Forecasting and Time, 2013. Retrieved from http://people.stat.sc.edu/Tebbs/stat520/f13notes.pdf
11. Zhang H, Rudholm N. Modeling and forecasting regional GDP in Sweden using autoregressive models, 2013.
12. Zhang M. Time Series: Autoregressive models R, MA, ARMA, ARIMA, 2018. Retrieved from http://people.cs.pitt.edu/~milos/courses/cs3750/lectures/class16.pdf