Advances and Applications in Statistics
© 2024 Pushpa Publishing House, Prayagraj, India
https://pphmjopenaccess.com
https://doi.org/10.17654/0972361724050
Volume 91, Number 7, 2024, Pages 921-942 **P-ISSN: 0972-3617**

PREDICTING NON-PERFORMING LOAN IN GHANA USING SARIMA MODEL

Albert Ayi Ashiagbor*, Kojo Amonkwandoh Essel-Mensah, Christian Akrong Hesse and Godwin Gidisu

Department of Business Management
University of Pretoria
Pretoria, South Africa
e-mail: albert.ashiagbor@up.ac.za

Department of Mathematical Statistics and Actuarial Science
University of the Free State
Bloemfontein, South Africa

Department of Economics and Actuarial Science
University of Professional Studies
Accra, Ghana

Department of Data Management and Analytics
GCB Bank
Accra, Ghana

Received: March 13, 2024; Accepted: May 6, 2024

2020 Mathematics Subject Classification: 62P20, 62P25, 62M10, 91B02, 91B70, 91G80, 91B84.

Keywords and phrases: non-performing loans, financial stability, time series analyses, SARIMA.

*Corresponding author

How to cite this article: Albert Ayi Ashiagbor, Kojo Amonkwandoh Essel-Mensah, Christian Akrong Hesse and Godwin Gidisu, Predicting non-performing loan in Ghana using SARIMA model, Advances and Applications in Statistics 91(7) (2024), 921-942. https://doi.org/10.17654/0972361724050

Published Online: May 30, 2024

Abstract

Banks' non-performing loans are an important indicator of financial stability, which assumes critical importance since they reflect asset quality, credit risk, and efficiency in the allocation of resources to productive sectors in an economy. This study used the seasonal autoregressive integrated moving average technique, which plays an important role in many practical fields, to examine the monthly non-performing loan data in Ghana from 2006 to 2022. The autocorrelation function, partial autocorrelation function, Akaike information criterion, and Bayesian information criterion analytical tools were used to assess the reliability of the models. The results of the study suggest that the SARIMA $(0, 1, 2)\times(1, 0, 0)_{12}$ model was appropriate for modeling the non-performing loan in Ghana. This model had the lowest benchmark value compared with other competitive models. It was found that the forecasted values indicate relative stability of the monthly non-performing loan in the banking industry for the next few years, though at a high level. We recommend that policies to target minimizing monthly non-performing loans, hence financial stability. We further recommend that policymakers as well as banking sector regulators need to ensure good corporate governance and risk management practices to avoid loan defaults in the banking sector.

1. Introduction

Financial stability is considered as an essential spectrum of sustained and rapid economic progress (Khan et al. [19], Adebayo [1] and Ozili [25]). The regulatory definition of the non-performing loans (NPLs) varies across jurisdictions (Kalfaoglou [15], and Pirgaip and Uysal [26]). However, according to international standards, banks are often advised to consider NPLs when principal and interest payments are 90 days or more past due or when future repayments are not anticipated to be fully made (Ari et al. [6], and Kartal et al. [17]). Thus, NPL is considered as a loan in which the borrower has defaulted in repayments of principal and interest for a specified period.

NPLs can be used as a serious signal of banking crises as it affects the economic growth of a nation by decreasing credit development (Handley [14], Alnabulsi et al. [4], and Kern et al. [18]). A low level of NPLs shows a strong monetary system of the country while high NPLs indicate a weak financial position.

Existing studies about the Ghanaian financial sector indicate the rising levels of NPLs and the consequences associated with such patterns to an emerging economy like Ghana (Amoah et al. [5], Cantah et al. [9], and Kamasa et al. [16]).

The Ghanaian banking industry's NPL ratio, as indicated by the central bank, has deteriorated to 18.0 percent in April 2023 from 14.3 percent in April 2022, reflecting higher loan impairments and elevated credit risks.

Hence, the main objective of current study is to develop a forecasting model to predict NPLs in Ghana using the seasonal autoregressive integrated moving average (SARIMA) time series approach.

Forecasting is important because it helps investors and the government in the financial sector to make informed decisions. Furthermore, this study's findings will greatly assist the central bank of Ghana in making appropriate financial policies in the nearest future in relation to NPLs. The detrimental effect that NPLs may have on banking sector and the economy if not analyze properly makes it necessary to examine the NPLs in the banking industry in Ghana. Again, the motivation of the current study is to demonstrate the role that accurate forecasting of NPLs can play in formulating financial stability strategies and plans in developing countries, using Ghana as a case study.

There are many studies on NPLs in different countries both developed and developing economies (Erdas and Ezanoglu [11], Foglia [12], and Gjeçi et al. [13]). Also, the following studies (Osei-Assibey and Asenso [24], and Liu et al. [21]) have delved into NPLs in Ghana with insightful works. However, their approaches are significantly different from the current study. None of these studies used the Box-Jenkin's SARIMA methodology to empirically examine the NPLs in Ghana.

2. Material and Methods

This section focuses on the various techniques that are used to achieve the objectives of the study. Here, we explain the basic concept and algorithmic processes of ARIMA and SARIMA models.

2.1. The ARIMA model

Autoregressive integrated moving average, or ARIMA(p, d, q), is a forecasting method for univariate time series data. As its name suggests, it supports both autoregressive and moving average elements. The integrated element refers to differencing allowing the method to support time series data with a trend. A problem with ARIMA is that it does not support seasonal data. That is a time series with a repeating cycle. ARIMA expects data that is either not seasonal or has the seasonal component removed, e.g., seasonally adjusted via methods such as seasonal differencing. An ARMA(p, q) model can be expressed as

$$y_i = \phi_0 + \sum_{j=1}^{p} \phi_j y_{i-j} + \varepsilon_i + \sum_{j=1}^{q} \theta_j \varepsilon_{i-j}. \tag{1}$$

If $\phi_0 = 0$ (i.e., the mean of the stochastic process is zero), then equation (1) can be expressed using the lag operator as

$$\phi(L) y_i = \theta(L)\varepsilon_i, \tag{2}$$

where

$$\phi(L) y_i = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p) = \left(1 - \sum_{j=1}^{p} \phi_j L^j\right) y_i, \tag{3}$$

$$\theta(L)\varepsilon_i = (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q) = \left(1 + \sum_{j=1}^{q} \theta_j L^j\right)\varepsilon_i. \tag{4}$$

Alternatively, we can express an ARIMA(p, d, q) process y_i without constant as

$$\phi(L)(1 - L)^d y_i = \theta(L)\varepsilon_i. \tag{5}$$

We can add the constant term back in

$$\phi(L)(1 - L)^d y_i = c + \theta(L)\varepsilon_i. \tag{6}$$

The number of times the series is differenced determines the order of d. The ARIMA(p, d, q) components, that is autoregressive (AR) and the moving average (MA) signatures are determined using non-seasonal and seasonal autocorrelation functions (ACF) and partial autocorrelation function (PACF) plots. A theoretical AR model of order p has an ACF that decays and a PACF that cuts off at lag p while a theoretical MA model of order q consists of a PACF that decays and an ACF that cuts off at lag q. The model with the minimum AIC and BIC values is selected as the model that fits the data best (Liu et al. [21]).

2.2. Seasonal ARIMA (SARIMA)

The SARIMA model extends an ARIMA model by taking seasonality into account. Such models are expressed as $(p, d, q) \times (P, D, Q)_m$, where (p, d, q) are as for an ARIMA model, while $(P, D, Q)_m$ express the seasonal autoregressive, integration and moving average components where the seasonality period is m. The SARIMA model is one of the most effective linear models for seasonal time series forecasting. As indicated earlier, the SARIMA model takes the same form as the ARIMA model, but now there are additional terms that reflect the seasonality part of the model. That is, the orders p and q are for the non-seasonal AR and MA components, respectively. The orders of the seasonal AR and MA components are P and Q, respectively. Also, the orders of differencing for the non-seasonal and seasonal are d and D, respectively.

Specifically, a SARIMA$(p, d, q) \times (P, D, Q)_m$ model without constant can be expressed as

$$\phi(L)\Phi(L^m)(1 - L)^d (1 - L^m)^D y_i = \theta(L)\Theta(L^m)\varepsilon_i, \tag{7}$$

where

$$\Phi(L)\, y_i = (1 - \Phi_1 L - \Phi_2 L^2 - \cdots - \Phi_p L^p)\, y_i = \left(1 - \sum_{j=1}^{P} \Phi_j L^j\right) y_i \qquad (8)$$

and

$$\Theta(L)\varepsilon_i = (1 + \Theta_1 L + \Theta_2 L^2 + \cdots + \Theta_q L^Q)\varepsilon_i = \left(1 + \sum_{j=1}^{Q} \Theta_j L^j\right)\varepsilon_i. \qquad (9)$$

Here, we have P seasonal autoregressive terms (with coefficients $\Phi_1, ..., \Phi_P$), Q seasonal moving average terms (with coefficients $\Theta_1, ..., \Theta_Q$) and D seasonal differencing based on m seasonal periods.

For purpose of clarity, let us assume that the two types of differencing (corresponding to d and D) have already been done. Then a SARIMA $(1, 0, 1) \times (1, 0, 1)_{12}$ model takes the following form:

$$\phi(L)\Phi(L^{12})\, y_i = \theta(L)\Theta(L^{12})\varepsilon_i, \qquad (10)$$

$$(1 - \phi_1 L)(1 - \phi_1 L^{12})\, y_i = (1 + \theta_1 L^1)(1 + \theta_1 L^{12})\varepsilon_i, \qquad (11)$$

$$(1 - \phi_1 L^1 - \Phi_1 L^{12} + \phi_1 \Phi_1 L^{13})\, y_i = (1 + \theta_1 L^1 + \Theta_1 L^1 + \theta_1 \Theta_1 L^{12} + \theta_1 \Theta_1 L^{13})\varepsilon_i, \qquad (12)$$

$$y_i - \phi_1 y_{i-1} - \Phi y_{i-12} + \phi_1 \Phi_1 y_{i-13} = \varepsilon_i + \theta_1 \varepsilon_{i-1} + \theta_1 \varepsilon_{i-12} + \theta_1 \Phi_1 \varepsilon_{i-13}. \qquad (13)$$

The residuals can therefore be expressed as:

$$\varepsilon_i = y_i - \phi_1 y_{i-1} - \Phi y_{i-12} + \phi_1 \Phi_1 y_{i-13} - \theta_1 \varepsilon_{i-1} - \theta_1 \varepsilon_{i-12} - \theta_1 \Phi_1 \varepsilon_{i-13}. \qquad (14)$$

Likewise, a SARMA $(p, q) \times (P, Q)_m$ model without constant can be expressed as

$$\phi(L)\Phi(L^m)\, y_i = \theta(L)\Theta(L^m)\varepsilon_i \qquad (15)$$

or, equivalently, as

$$\left(1-\sum_{j=1}^{p}\phi_j L^j\right)\left(1-\sum_{j=1}^{p}\Phi_j L^{jm}\right)y_i = \left(1+\sum_{j=1}^{q}\theta_j L^j\right)\left(1+\sum_{j=1}^{q}\Theta_j L^{jm}\right)\varepsilon_i, \quad (16)$$

$$y_i = \left(\sum_{j=1}^{p}\phi_j L^j\right)y_i + \left(\sum_{j=1}^{p}\Phi_j L^{jm}\right)y_i - \left(\sum_{j=1}^{p}\phi_j L^j\right)\left(\sum_{j=1}^{p}\Phi_j L^{jm}\right)y_i$$

$$+\left(\sum_{j=1}^{q}\theta_j L^j\right)\varepsilon_i + \left(\sum_{j=1}^{Q}\Theta_j L^{jm}\right)\varepsilon_i + \left(\sum_{j=1}^{q}\theta_j L^j\right)\left(\sum_{j=1}^{Q}\Theta_j L^{jm}\right)\varepsilon_i + \varepsilon_i \quad (17)$$

and so

$$y_i = \left(\sum_{j=1}^{p}\phi_j L^j + \sum_{j=1}^{p}\Phi_j L^{jm} - \sum_{j=1}^{p}\sum_{k=1}^{p}\phi_j \Phi_k L^{km+j}\right)$$

$$+\left(\sum_{j=1}^{q}\theta_j L^j + \sum_{j=1}^{Q}\Theta_j L^{jm} + \sum_{j=1}^{q}\sum_{k=1}^{Q}\theta_j \Theta_k L^{mk+j}\right)\varepsilon_i. \quad (18)$$

In the case where there is a constant term φ_0, this expression takes the form

$$y_i = \phi_0 + \left(\sum_{j=1}^{p}\phi_j y_{i-1} + \sum_{j=1}^{p}\Phi_j y_{i-jm} - \sum_{j=1}^{p}\sum_{k=1}^{p}\phi_j \Phi_k y_{i-km-j}\right)$$

$$+\left(\sum_{j=1}^{q}\theta_j \varepsilon_{i-j} + \sum_{j=1}^{Q}\Theta_j \varepsilon_{i-jm} + \sum_{j=1}^{q}\sum_{k=1}^{Q}\theta_j \Theta_k \varepsilon_{i-km-j}\right) + \varepsilon_i. \quad (19)$$

For the parameters estimation of SARIMA models, we use the maximum likelihood estimation technique.

2.3. Modeling approach

This study relied on the Box-Jenkins' concept of SARIMA to model the NPLs in Ghana. The algorithmic process of SARIMA model is depicted in Figure 1. The framework for modeling SARIMA basically consists of three steps, namely: identification, estimation of parameters and diagnostic checking.

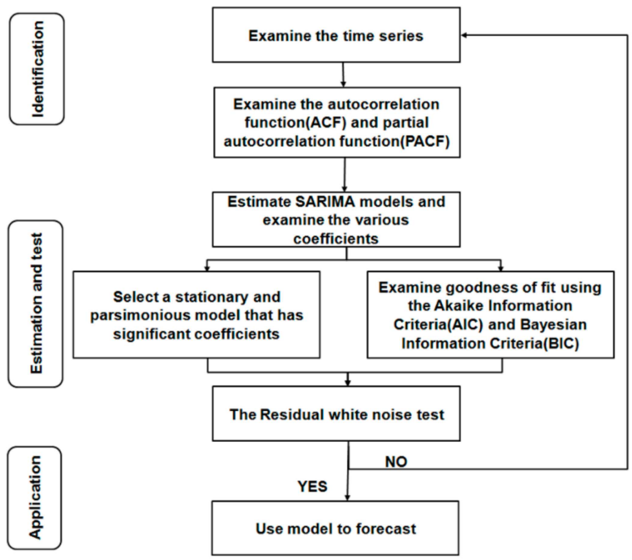

Figure 1. The framework for modeling SARIMA.

2.3.1. Model identification and selection

Given that commodity NPLs series are likely to exhibit a seasonal component, the current study hypothesizes that the SARIMA model best fits the data under discussion. The first and most crucial step in modeling is determining the best order for the SARIMA$(p, d, q)(P, D, Q)$ model (Koula et al. [20]). The model identification step involves the use of the techniques to determine the values of p, q, P, Q, D and d. In this study, the values are determined by using autocorrelation function (ACF), partial autocorrelation function (PACF) and the auto.arima function in Rstudio. The appropriate model selection involves selecting the model that best fits the available data taking into consideration the principle of parsimony.

2.3.2. Model parameters estimation and performance measurement

There is the need to estimate the parameters after the identification and selection of the tentative model. Akaike information criterion (AIC), modified Akaike information criterion (AICc), normalized Bayesian information criterion (BIC), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) are used. Model performance was measured to assess the regeneration capability based on these statistical measures. The minimum value of AIC, AICc, BIC, RMSE, MAPE and MAE among the chosen models is selected as the best-fitted model. Again, the maximum likelihood estimation method was used to estimate the coefficients $(\phi, \theta, \Phi, \text{ and } \Theta)$ of the appropriate model. Because the previous lagged observations of noise terms cannot be detected in a SARIMA framework, the MLE technique is adopted over ordinary least squares (OLS) regression analysis due to its efficiency (Boniface et al. [8]).

2.3.3. Model diagnostic/assessment of the model's adequacy

The diagnosis and the validation of an ARIMA family of models rest essentially on the characteristics of the residues (Koula et al. [20]). That is the diagnostic checks are performed on the residuals to see if they are randomly and normally distributed. In otherwise, the diagnostic check is based on the residuals of the model. We used the residual plot versus the fitted values to check if the residuals are randomly scattered. An overall check of the model adequacy was made using the Ljung-Box Q statistics.

Furthermore, the autocorrelation is another diagnostic test used in this study. A correlogram can be used to test the autocorrelation in the residuals. If there is no serial connection, autocorrelations and partial autocorrelations should be near to zero at all lags. Thus, these tests are performed to check for homoscedasticity and autocorrelation of the residuals of the model.

2.3.4. Test of stationarity

Testing data for stationarity is very important in research where the underlying variables are based on time (Mushtaq [22]). Non-stationary time series data has statistical properties, which change with time. So, it is required to change the data into stationary time series data before building the predictive model. We tested for stationary by making use of the augmented Dickey Fuller (ADF) and Phillips-Perron (PP) tests.

In statistics, an ADF tests the null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. The model for ADF test is as follows:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p-1} + \varepsilon_t,$$

where α is a constant, β the coefficient on a time trend and p the lag order of the autoregressive process. It should be noted that $\alpha = 0$ and $\beta = 0$ correspond to modeling a random walk. Using the constant $\beta = 0$ corresponds to modeling random walk with a drift.

In statistics, the Phillips-Perron (PP) test is used in time series analysis to test the null hypothesis that a time series is integrated of order one (1). It builds on the Dickey-Fuller test of the null hypothesis $\rho = 1$ in $\Delta y_t = (\rho - 1) y_{t-1} + u_t$, where Δ is the first difference operator.

The other unit root test we used in addition is the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. Unlike the other tests, the null hypothesis for the KPSS test is that the time series is stationary, while the alternative hypothesis is that there is a unit root.

2.3.5. Forecasting

The advantages of SARIMA lie in its well-known statistical properties and its effective modeling process. One of our main objectives is to produce forecasts with minimum error as much as possible. In this section, we assess the forecasting performance of the SARIMA model. Following the work of Adubisi et al. [2], we assumed that the condition(s) under which the SARIMA model was constructed would persist in the periods for which the forecasts are made. The computation of the forecast values was carried out with forecast estimate function:

$$\hat{y}_{t+h/t} = \phi_1 \hat{y}_{t+h-1/t} + \phi_2 \hat{y}_{t+h-2/t} + \Theta_{12} \hat{\varepsilon}_{t+h-12/t}.$$

Thus, we used one-period-ahead forecasting using seasonal ARIMA modeling.

In this paper, SARIMA was applied using the statistical software, the R programming language and RStudio. The main advantage of SARIMA processes is its ability to model time series with trends, seasonal patterns and short-term correlation with a small data set.

3. Results and Discussion

In this section, the results of the modeling and prediction of the NPL with the Box-Jenkins SARIMA method are presented. The monthly dataset of Ghana's banking industry on non-performing loans which was obtained from the Central Bank of Ghana was used in this study.

3.1. Descriptive statistics

Descriptive statistics are an important part of this study which is used to describe the basic features of the data in the study. It helps to provide simple summaries about the sample and the measures. Measures of the central tendency and dispersion are used to describe the quantitative data. For the continuous data, test of the normality is an important step for deciding the measures of central tendency and statistical methods for data analysis.

The descriptive statistics as presented in Table 1 shows that the data is not normally distributed with a constant variance based on the coefficient of variation (CV), skewness (SKEW) and kurtosis (KURT) values. The values of the standard deviation (SD), median (MED), mean absolute deviation (MAD) are all presented in Table 1. More so, the mean (MIN) and the maximum (MAX) values of 14.53 and 23.45, respectively, suggest that NPLs in Ghana is relatively high.

Table 1. The descriptive statistics

OBS	MEAN	SD	MED	MAD	MIN	MAX	SKEW	KURT
202	14.53	4.07	14.21	4.09	6.14	23.45	–0.06	–0.55

The non-performing loan series plot as shown in Figure 2 depicts an increasing trend with non-constant variability including high peaks at specific periods in each year, suggesting the series requires some sort of decomposition and transformation to stabilize the variance.

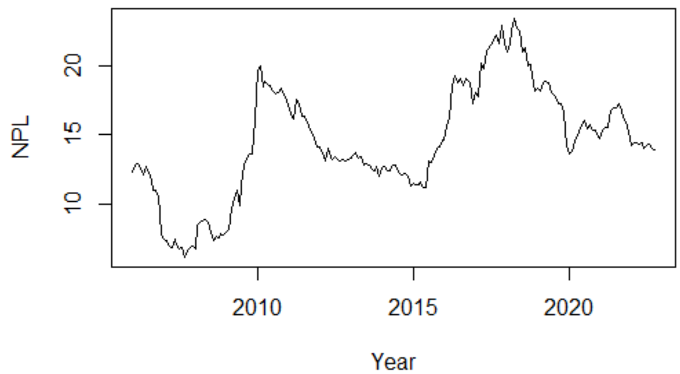

Figure 2. Plot of non-performing loans.

3.2. Decomposing time series

The theory of time series analysis gives opportunities of the idea of decomposing a times series into predictable and unpredictable components (Dodge [10]). The decomposition of time series is a statistical task that deconstructs a time series into several components, each representing one of the underlying categories of patterns. This is an important technique for all types of time series analysis, especially for seasonal adjustment. Decomposition of time series is applied to describe the trend and seasonal factors in a time series observation. From Figure 3, the time series decomposition reveals that the data series is separated into its constituent components. The decomposed plots of the NPLs data series were used to identify the time series components; seasonal, trend, cyclic, and random component in the data over the time (see Figure 3).

The up and down pattern of the observed time series is an indication of the seasonality of the NPLs data series and has a seasonal effect with the usual upward and downward pattern being experienced yearly over the study period. This indicates that the average monthly NPLs data series of each year was influenced by the seasonal components.

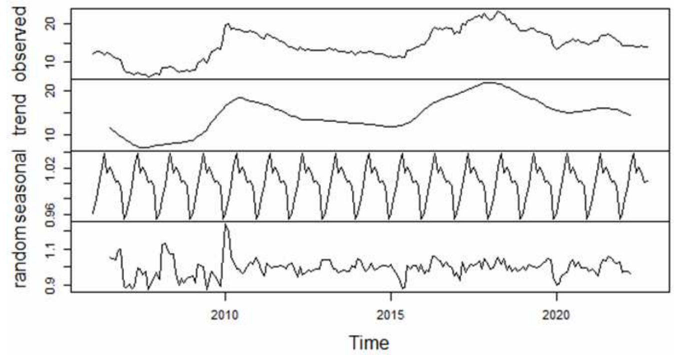

Figure 3. Decomposition of multiplicative non-performing loan time series data.

3.3. Stationarity of the NPL data

Figure 1 displays the plot of historical NPLs data from January 2006 to December 2022. The series shows an overall upward trend, given the seasonal component of the series, the SARIMA model is appropriate because it captures seasonality that exists in the series.

The ADF test, PP test and KPSS test are applied for testing the stationarity for the non-performing loan dataset as presented in Table 1. From the table, based on the *p* values, it could be observed that one cannot accept that data series are stationary. Again the ACF and PACF in Figures 4 and 5 confirmed the non-stationarity of the NPL data.

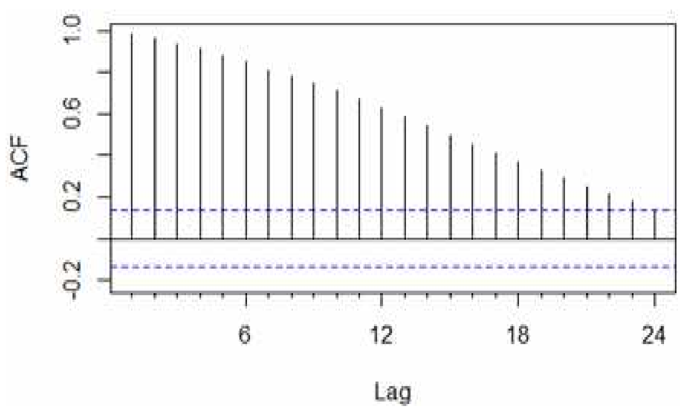

Figure 4. Autocorrelations of non-performing loans.

The series became stationary after the first differencing and seasonal differencing as shown in Table 2. Figure 4 shows the corresponding plot after differencing the NPL data.

Table 2. Unit root and stationary tests non-performing loans (not stationary)

Test type	Test statistics	Lag order	*p*-value
ADF	–2.2637	5	0.4656
PP	–6.4601	4	0.7459
KPSS	1.6522	4	0.01

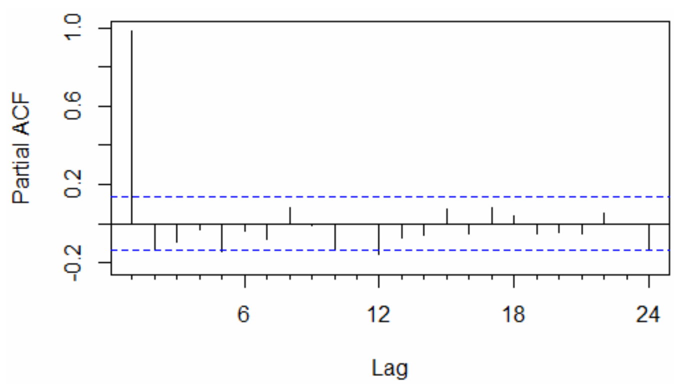

Figure 5. Partial autocorrelations of non-performing loans.

Table 3. Unit root and stationary tests non-performing loans model (stationary)

Test type	Test statistics	Lag order	*p*-value
ADF	–4.4635	5	0.01
PP	–207.6	4	0.01
KPSS	–0.095026	4	0.1

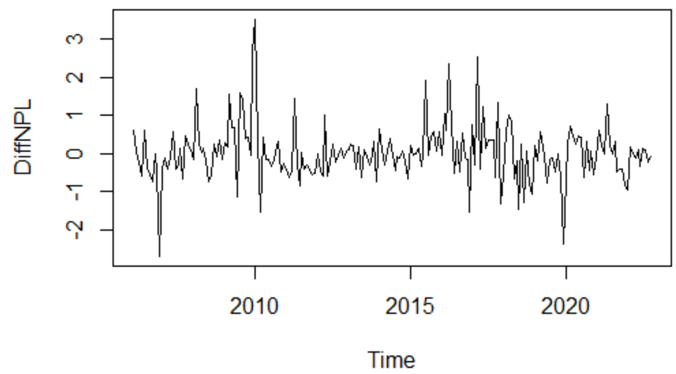

Figure 6. Plot of differencing non-performing loans.

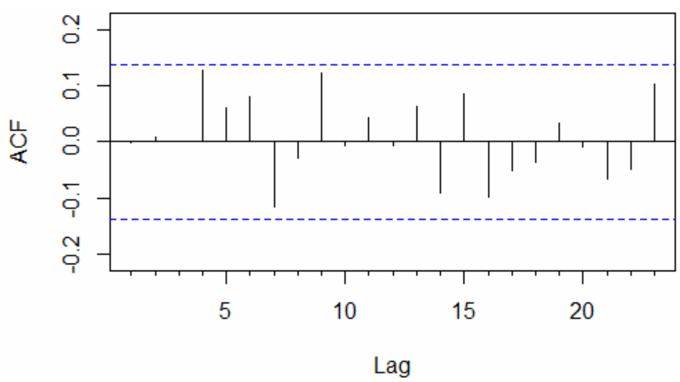

Figure 7. ACF of the first difference of the NPL data.

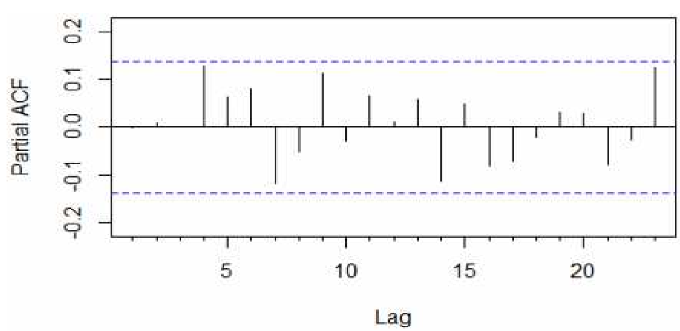

Figure 8. PACF of the first difference of the NPL data.

3.4. Tentative model, selection and estimation

In order to get the most efficient model from a combination of different orders having seasonal and non-seasonal parts, we relied on the auto.arima function in the Rstudio to generate tentative models. Table 3 represents these tentative models. A closer look at Table 4 revealed that $\text{SARIMA}(0, 1, 2) \times (1, 0, 0)_{12}$ shows the lowest AIC value with the corresponding standard deviation (Std) of the model. Thus, this model should be considered as the best forecasting model. Again, Table 4 shows the estimated parameters of each of the tentative models.

Table 4. Tentative seasonal ARIMA models

Rank	SARIMA structure	Parameter estimates (standard error)	Selection criteria	*t*-value	Std of model
1	ARIMA $(0, 1, 2)(1, 0, 0)_{12}$	MA(1) = 0.1315 (0.0713)	Log-likelihood = -220.49	1.844	0.525
		MA(2) = 0.0892 (0.0653)	AIC = 443.2465	1.365	
		SAR(1) = 0.0613 (0.0727)	BIC = 462.1935	0.843	
2	ARIMA $(1, 1, 1)(1, 0, 0)_{12}$	AR(1) = 0.7934 (0.1662)	Log-likelihood = -219.74.74	4.775	0.521
		MA(1) = -0.6822 (0.2001)	AIC = 447.48	-3.410	
		SAR(1) = 0.0571 (0.0720)	BIC = 460.6934	0.793	
3	ARIMA $(1, 1, 0)(1, 0, 1)_{12}$	AR(1) = 0.1412 (0.0698)	Log-likelihood = -219.9	2.023	0.5124
		SAR(1) = 0.9923 (0.0863)	AIC = 447.8	11.504	
		SMA(1) = -0.9694 (0.1868)	BIC = 461.0095	-5.190	
4	ARIMA $(1, 1, 0)(1, 0, 0)_{12}$	AR(1) = 0.1402 (0.0703)	Log-likelihood = -221.11	1.993	0.5283
		SAR(1) = 0.0511 (0.0726)	AIC = 448.23	0.703	
			BIC = 458.1354		
5	ARIMA $(1, 1, 0)(2, 0, 0)_{12}$	AR(1) = 0.1372 (0.0702)	Log-likelihood = -220.14	1.954	0.5227
		SAR(1) = 0.0463 (0.0717)	AIC = 448.29	0.647	
		SAR(2) = 0.0994 (0.0711)	BIC = 461.4987	1.398	
6	ARIMA $(2, 1, 0)(1, 0, 0)_{12}$	AR(1) = 0.1269 (0.0707)	Log-likelihood = -220.34	1.794	0.5243
		AR(2) = 0.0874 (0.0703)	AIC = 448.69	1.242	
		SAR(1) = 0.0595 (0.0728)	BIC = 461.9027	0.818	
7	ARIMA $(0, 1, 0)(1, 0, 0)_{12}$		Log-likelihood = -223.08		0.5388
		SAR(1) = 0.0694 (0.0721)	AIC = 450.16	0.962	
			BIC = 456.762		

The forecast accuracy of the selected model is validated by applying a diagnosis check as shown in Table 4. From Table 5, it could be observed that the error measure value of $\text{SARIMA}(0, 1, 2) \times (1, 0, 0)_{12}$ is the lowest in many instances.

Table 5. Training set error measures

Measure	ARIMA $(1, 1, 0)(1, 0, 0)_{12}$	ARIMA $(0, 1, 2)(1, 0, 0)_{12}$	ARIMA $(1, 1, 0)(2, 0, 0)_{12}$	ARIMA $(1, 1, 0)(1, 0, 1)_{12}$	ARIMA $(0, 1, 0)(1, 0, 0)_{12}$	ARIMA $(2, 1, 0)(1, 0, 0)_{12}$	ARIMA $(1, 1, 1)(1, 0, 0)_{12}$
ME	0.00632	0.00575	0.00533	0.00199	0.00711	0.00546	0.00410
RMSE	0.72507	0.71178	0.72115	0.71407	0.73218	0.72226	0.70004
MAE	0.49252	0.49292	0.49213	0.48378	0.49486	0.49071	0.48904
MPE	-0.0583	-0.05223	-0.0392	-0.07906	-0.0697	-0.0472	-0.0185
MAPE	3.53879	3.52991	3.53669	3.47108	3.57618	3.53297	3.52063
MASE	0.98917	0.97999	0.98839	0.97162	0.99388	0.98554	0.98220
ACF1	-0.01196	-0.0013	-0.0129	-0.0128	0.13827	0.00033	0.00704

3.5. Residuals diagnostic of the selected model

In order to ensure that the basic assumptions of normality and autocorrelation underlined the principles of the Box-Jenkins' methodology in relation to the model residuals, we performed the diagnostic checks on residuals of the fitted model. From Figure 8, it could be found that the model residuals are independent and identically distributed with a constant mean and variance. Again there are no autocorrelations between the residuals. Thus, a closer look at Figure 9 shows that the histogram of the residuals' distribution is almost normally distributed and mean of the distribution seems to be zero.

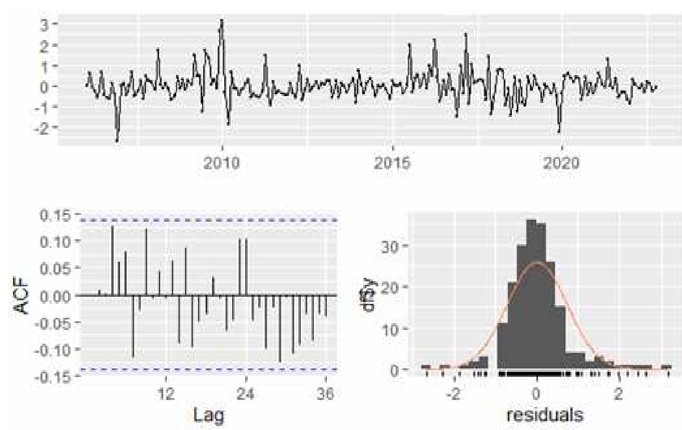

Figure 9. Residuals diagnostic from ARIMA $(0, 1, 2) \times (1, 0, 0)_{12}$.

3.6. Forecast from the selected model

As we indicate earlier, forecasting provides an insight into the future variability, it aids in planning and decision-making. In this regard, we used the best fitted time series model SARIMA $(0, 1, 2) \times (1, 0, 0)_{12}$, to forecast the non-performing loan. The monthly data from January 2006 to December 2022 are considered for validation of the model which is as in-sample forecast and out-sample forecast data from January 2023 to December 2024 are used as to forecast, and it was satisfied that the residuals of all selected models are found to be approximately stationary and white noise. Figure 10 displays the forecast trend from the SARIMA $(0, 1, 2) \times (1, 0, 0)_{12}$.

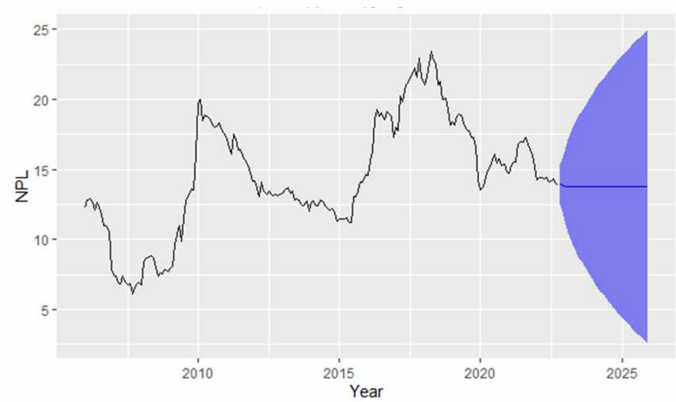

Figure 10. Forecast from SARIMA $(0, 1, 2) \times (1, 0, 0)_{12}$.

Table 6 displays specific data from predictive analysis. The forecast shows that the NPL has been relatively stable over time and is probably going to continue to do so. This study's findings are in line with those of earlier research done by other scholars such as Okyere and Mensah [23], Bacchiocchi et al. [7] and Amoah et al. [5], as well as the work of Ozili [25].

Table 6. Forecasted values with fitted model

Point	Month	Forecast	Low 95%	High 95%
2024	Jan	13.76927	7.032616	20.50593
2024	Feb	13.76995	6.782917	20.75698
2024	Mar	13.77002	6.541284	20.99876
2024	Apr	13.76957	6.306949	21.23220
2024	May	13.76995	6.080552	21.45935
2024	Jun	13.76867	5.859000	21.67834
2024	Jul	13.76920	5.645222	21.89317
2024	Aug	13.76957	5.436804	22.10234
2024	Sep	13.76863	5.232176	22.30509
2024	Oct	13.76830	5.032897	22.50369
2024	Nov	13.76810	4.837075	22.69913
2024	Dec	13.76790	4.645300	22.89051
2025	Jan	13.76768	4.457343	23.07801
2025	Feb	13.76772	4.273364	23.26207
2025	Mar	13.76772	4.092848	23.44260
2025	Apr	13.76770	3.915606	23.61978
2025	May	13.76772	3.741548	23.79389
2025	Jun	13.76764	3.570359	23.96492
2025	Jul	13.76767	3.402105	24.13324
2025	Aug	13.76770	3.236530	24.29886
2025	Sep	13.76764	3.073439	24.46184
2025	Oct	13.76762	2.912834	24.62240
2025	Nov	13.76761	2.754522	24.78069
2025	Dec	13.76759	2.598447	24.93674

Observing the prediction values, we can see that there is a relatively stable in pattern of the NPL series as such Ghana is likely to experience relative stable NPL for the year 2024 to 2025. This may be considered as a good sign for the Ghanaian in general, all things being equal, and for the banking sector in particular.

4. Conclusions and Recommendations

This study examined the historical and future patterns of NPLs in the banking industry of Ghana. The SARIMA model was used in this analysis. It was found that the SARIMA $(0, 1, 2) \times (1, 0, 0)_{12}$ showed the best results since the series contained a seasonal component. This model could serve as a useful tool for modeling and forecasting monthly NPLs in Ghana. Also, the forecasted values indicate relative stability in the NPLs in the banking industry for the next few years.

As indicated earlier, NPL analysis is a significant tool used by regulators to determine financial stability and bank asset quality. It is believed that this study contributes significantly in shading more light from the stochastic aspect of non-performing loan analysis in Ghana. From the results of the study, we recommend policies that would target towards minimizing NPLs, hence financial stability. The policy-makers as well as the banking sector regulators need to ensure good corporate governance and risk management techniques by the banks to avoid loan default in the banking sector.

References

[1] T. S. Adebayo, Towards unlocking the chain of sustainable development in the BRICS economies: analysing the role of economic complexity and financial risk, Geological Journal 58 (2023), 1810-1821. https://onlinelibrary.wiley.com/doi/pdf/10.1002/gj.4694.

[2] O. D. Adubisi, C. C. Eleke, T. T. Mom and C. E. Adubisi, Application of SARIMA to modelling and forecasting money circulation in Nigeria, Asian Research Journal of Mathematics 6(1) (2017), 1-10. https://doi.org/10.9734/ARJOM/2017/35555.

[3] C. Adusei and J. Bannerman, A retrospective study of non-performing loans of the Ghana banking sector between 1998 and 2019, Theoretical Economics Letters 12(2) (2022), 498-517. https://doi.org/10.4236/tel.2022.122028.

[4] K. Alnabulsi, E. Kozarević and A. Hakimi, Non-performing loans as a driver of banking distress: a systematic literature review, Commodities 2(2) (2023), 111-130. https://doi.org/10.3390/commodities2020007.

[5] A. Amoah, R. K. Asiama and K. Korle, Applying the breaks to non-performing loans in Ghana, International Journal of Emerging Markets 18(8) (2023), 1978-1993. https://doi.org/10.1108/IJOEM-03-2020-0287.

[6] A. Ari, S. Chen and L. Ratnovski, The dynamics of nonperforming loans during banking crises: a new database, 2020. https://doi.org/10.2866/174853.

[7] A. Bacchiocchi, G. I. Bischi and G. Giombini, Non-performing loans, expectations and banking stability: a dynamic model, Chaos Solitons Fractals 157 (2022), 111906. https://doi.org/10.1016/j.chaos.2022.111906.

[8] O. I. Boniface, O. A. Chijioke and O. C. Justin, Effect of correlated measurement errors on estimation of population mean with modified ratio estimator, Biom. Biostat. Int. J. 11(2) (2022), 52-56. https://doi.org/10.15406/bbij.2022.11.00354.

[9] W. G. Cantah, K. N. Darfor, J. Nunoo, B. Afful Jr. and E. A. Wiafe, Bank competition and financial sector stability in Ghana, Financial Sector Development in Ghana: Exploring Bank Stability, Financing Models, and Development Challenges for Sustainable Financial Markets, Springer International Publishing, Cham, 2023, pp. 133-154. https://doi.org/10.1007/978-3-031-09345-6_6.

[10] Y. Dodge, Ed., The Oxford Dictionary of Statistical Terms, OUP Oxford, 2003.

[11] M. L. Erdas and Z. Ezanoglu, How do bank-specific factors impact non-performing loans: evidence from G20 countries, Journal of Central Banking Theory and Practice 11(2) (2022), 97-122. https://doi.org/10.2478/jcbtp-2022-0015.

[12] M. Foglia, Non-performing loans and macroeconomics factors: the Italian case, Risks 10(1) (2022), 21. https://doi.org/10.3390/risks10010021.

[13] A. Gjeçi, M. Marinč and V. Rant, Non-performing loans and bank lending behaviour, Risk Management 25(1) (2023), 7. https://doi.org/10.1057/s41283-022-00111-z.

[14] N. J. Handley, Growth in a time of debt, CFA Digest 40 (2010), 19-20. doi.org/10.2469/dig.v40.n3.19.

[15] F. Kalfaoglou, NPLs resolution regimes: challenges for regulatory authorities, Journal of Risk Management in Financial Institutions 11(2) (2018), 173-186. https://ideas.repec.org/a/aza/rmfi00/y2018v11i2p173-186.html.

[16] K. Kamasa, S. L. Afful and I. Bentum-Ennin, Impact of monetary policy rate on commercial banks' lending rate: empirical evidence from Ghana, Journal of Economic and Administrative Sciences (2023). http://dx.doi.org/10.1108/JEAS-07-2021-0141.

[17] M. T. Kartal, D. Kirikkaleli and F. Ayhan, Nexus between non-performing loans and economic growth in emerging countries: evidence from Turkey with wavelet coherence approach, International Journal of Finance and Economics 28(2) (2023), 1250-1260. https://doi.org/10.1002/ijfe.2474.

[18] A. Kern, E. Nosrati, B. Reinsberg and D. Sevinc, Crash for cash: offshore financial destinations and IMF programs, European Journal of Political Economy 78 (2023), 102359. https://doi.org/10.1016/j.ejpoleco.2023.102359.

[19] M. A. Khan, A. Siddique and Z. Sarwar, Determinants of non-performing loans in the banking sector in developing state, Asian Journal of Accounting Research 5(1) (2020), 135-145. https://doi.org/10.1108/AJAR-10-2019-0080.

[20] J. Koula, T. Tiho and A. C. Chiapo, On the analysis and modelling of the harmonized consumer price indices of West African economic and monetary union member states, Am. J. Theor. Appl. Stat. 9(6) (2020), 283-295. http://www.sciencepublishinggroup.com/j/ajtas.

[21] Q. Liu, M. A. Charleston, S. A. Richards and B. R. Holland, Performance of Akaike information criterion and Bayesian information criterion in selecting partition models and mixture models, Systematic Biology 72(1) (2023), 92-105. https://doi.org/10.1093/sysbio/syac081.

[22] R. Mushtaq, Augmented Dickey Fuller test, 2011. https://dx.doi.org/10.2139/ssrn.1911068.

[23] E. Okyere and A. C. Mensah, Macroeconomic and bank specific determinants of non-performing loans in Ghanaian banking sector, International Journal of Accounting, Finance and Risk Management 7(2) (2022), 40-48. http://www.sciencepublishinggroup.com/j/ijafrm.

[24] E. Osei-Assibey and J. K. Asenso, Regulatory capital and its effect on credit growth, non-performing loans and bank efficiency: evidence from Ghana, Journal of Financial Economic Policy 7(4) (2015), 401-420. http://dx.doi.org/10.1108/JFEP-03-2015-0018.

[25] P. K. Ozili, Sustainable development and bank nonperforming loans: are they correlated? Arab Gulf Journal of Scientific Research (2023). https://doi.org/10.1108/AGJSR-01-2023-0028.

[26] B. Pirgaip and A. Uysal, The impact of non-performing loan sales on the stock market: the role of corporate governance in an emerging market, Borsa Istanbul Review 23(3) (2023), 674-684. https://doi.org/10.1016/j.bir.2023.01.006.

[27] I. Zuhroh, M. Rofik and M. S. W. Suliswanto, Non-performing loans: endogenous constraint on the credit-growth nexus, Asian Economic and Financial Review 12(8) (2022). http://dx.doi.org/10.55493/5002.v12i8.4584.