

## FORECASTING QUARTERLY PERFORMANCE OF FOREIGN BANKS IN INDIA :AN ARIMA MODEL APPROCH

**K.C.Bhanu<sup>1</sup>, Dr.P.Uma Maheswari Devi<sup>2</sup>**

<sup>1</sup>Research Scholar, Department of Commerce and Management Studies  
Adikavi Nannayya University ,Rajahmundry

<sup>2</sup>Associate Professor , Department of Commerce and Management Studies  
Adikavi Nannayya University ,Rajahmundry  
[bhanu1605@gmail.com](mailto:bhanu1605@gmail.com),[umadevi\\_4@yahoo.com](mailto:umadevi_4@yahoo.com)

### ABSTRACT

From April of 2023 to March of 2024, this study will forecast the quarterly performance of international banks in India. The study utilizes quarterly data from March 2019 through March 2023. The Autoregressive Integrated Moving Average (ARIMA) model, a common machine learning approach for time series forecasting, is used to accomplish this goal. At first, the dataset is investigated and preprocessed to make sure it can be used for modeling. Next, we apply the ARIMA model to the historical data in order to project outcomes for the subsequent four quarterly time intervals. Due to its ability to minimize mistakes and maximize prediction accuracy, the ARIMA model's optimal parameters are found to be (0, 1, 0) after extensive testing. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are computed to verify the ARIMA model of choice. Useful information about the model's goodness-of-fit and whether or not it reflects the underlying patterns and dynamics of the data may be gleaned from these metrics. The research checks the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to make sure the time series is stationary. The ARIMA model may be improved with the use of these diagnostic tools by locating any residual non-stationary components in the data. Researchers hope that these findings would help stakeholders and decision-makers in India better understand the likely future performance of international banks in the country. In addition, this demonstration of the usefulness of machine learning techniques for forecasting financial time series data is exemplified by the use of ARIMA modeling techniques in this scenario.

**Keywords:** Foregin Bank, ARIMA, Forecasting, India

### INTRODUCTION:

The financial system is essential to the development and stability of any economy. Foreign banks' presence in India has helped broaden the country's banking sector, increase access to a wider range of financial services, and attract foreign investment and knowledge workers. The ability of politicians, investors, and regulators to understand and predict the performance of these international banks is crucial.

Using the Autoregressive Integrated Moving Average (ARIMA) model, this study digs into time series forecasting to project the quarterly performance of international banks in India for the next calendar year, 2023-2024. This research is based on data collected quarterly from March of 2019 .By employing state-of-the-art machine learning methods, we want to provide accurate forecasts for the development of these financial institutions.

The ARIMA model is a popular statistical technique for analyzing time series data because it utilizes the strengths of autoregression, differencing, and moving averages. We start making projections for the following four quarters, paying special attention to being as precise as possible.

The financial industry benefits greatly from accurate predictions of the performance of international banks operating in India. It gives policymakers and regulators the information they need to make educated choices that will help maintain a healthy financial system. As an added bonus, these predictions may be used by financiers to weigh the odds of various returns and losses, so informing their investment decisions.

The Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are compared to determine the quality of the selected ARIMA model. These metrics provide invaluable feedback on the model's efficacy, allowing us to fine-tune it so that it more accurately represents the dynamics and intricacies of the financial time series data.

Important considerations include model validation and the stationarity of the time series data. With the use of the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF), we hope to spot any lingering non-stationary components in the data. Taking into account such factors that may affect the model's performance makes for a more solid and trustworthy forecasting procedure.

The results of this research have important ramifications for a wide range of players in the financial industry, helping them better predict the long-term success of foreign banks in India. The importance of sophisticated analytical approaches in financial time series data forecasting is further highlighted by demonstrating the efficacy of machine learning techniques like ARIMA in this setting. Following this introduction, we provide a thorough discussion of our study's design, findings, and implications, illuminating opportunities for expansion as well as potential obstacles for international banks doing business in India.

## LITERATURE REVIEW

Fang et al.,(2001), Fuzzy ARIMA model for forecasting the forging exchange market. This study creates a fuzzy ARIMA (FARIMA) model by combining the time-series ARIMA(p,d,q) model and the fuzzy regression model to predict the NT dollar/US dollar exchange rate. Future value distributions are given by FARIMA, and the model supports interval models with interval

parameters. When compared to the ARIMA model, this one requires less data points to accurately predict both the best and worst case scenarios.

Adewole et al.,(2011). Artificial Neural Network Model for forecasting Foreign Exchange Rate . Forecasting models cannot manage foreign exchange data uncertainty and volatility. This paper created an artificial neural network foreign exchange rate forecasting model (AFERFM) to address these issues. Training and forecasting comprised the design. Back propagation was used to train foreign exchange rates and estimate input. Sigmoid Activation Function (SAF) converted the input to [0, 1]. For training-consistent output, learning weights were randomly assigned in the range [-0.1, 0.1]. SAF's hyperbolic tangent increased learning speed and efficiency. Feed-forward Network improved back propagation. Multilayer Perceptron Network forecasts. Back propagation for foreign currency rate assessment and forecasting utilized Oanda datasets. Matlab7.6 and Visual Studio were used to implement the idea since they enable forecasting systems. The system was evaluated using mean square error and standard deviation using a 0.10 learning rate, an input layer, 3 hidden layers, and an output layer. HFERFM, the best-known related study, with 69.9% accuracy compared to AFERFM's 81.2%. The new method enhanced foreign exchange rate forecasts.

Ngan (2016), Forecasting Foreign Exchange Rate by using ARIMA Model: A Case of VND/USD Exchange Rate. Forecasting foreign currency rates helps Vietnamese commercial joint stock banks manage foreign exchange risk. This research offers a four-step Arima model to anticipate VND/USD foreign exchange rates in 2016 using real data from 2013 to 2015. After forecasting foreign exchange data, we compare them with real foreign exchange rate data to check the suitability of Arima model for forecasting foreign exchange rate in Vietnam. The results show that Arima model is suitable for short-term forecasting.

Jere et al.,(2017). Forecasting Foreign Direct Investment To Zambia: A Time Series Analysis. This study examines SES, HWES, and ARIMA. Zambia's 1970–2014 net FDI inflows were anticipated using the best fit model. Zambia receives foreign direct investment. Zambia's yearly Net FDI inflows highlight the methodology throughout the study. ARIMA (1, 1, 5) has the lowest error, making it the best fit model. Net FDI inflows are expected to rise 44.36% by 2024. Policymakers need forecasts. Accurate projections enable sound policymaking and strategic planning. This report can help Zambian FDI policymakers develop FDI-promoting strategies.

## **METHOD AND APPROACH**

**Time series :-** A time series is a sequence of data points or observations collected and recorded over a specific period of time, where the data points are ordered based on their corresponding time index. In other words, it is a collection of data points that are indexed or labeled by time.

Time series data is commonly encountered in various fields, including finance, economics, engineering, environmental science, and many others. It provides valuable insights into the behavior and patterns of phenomena that evolve over time.

The characteristics of time series data include:

1. **Time Index:** Each data point in a time series is associated with a specific time index or timestamp, indicating when the observation was made.
2. **Sequential Order:** The data points in a time series are arranged in chronological order, with each subsequent observation occurring after the previous one.
3. **Temporal Dependence:** Time series data often exhibits a certain level of dependence or correlation between observations. The value at a given time point can be influenced by its previous values or exhibit patterns over time.
4. **Seasonality:** Many time series exhibit regular patterns or variations that repeat at fixed intervals, known as seasonality. For example, retail sales may have higher values during holiday seasons.
5. **Trend:** Time series data often shows a long-term trend or systematic changes over time. Trends can be increasing (upward trend), decreasing (downward trend), or exhibit more complex patterns.
6. **Randomness:** Time series data can also contain random or unpredictable fluctuations, known as noise or random variations. These random components make it challenging to accurately forecast future values.

Analyzing time series data involves various techniques and methods, including trend analysis, seasonality detection, forecasting, and modeling. Time series analysis aims to understand and extract useful information from the data, uncover underlying patterns, and make predictions about future behavior based on historical observations.

### Box-Jenkins Methodology

Box-Jenkins, also known as the Box-Jenkins methodology or Box-Jenkins approach, is a widely used and powerful technique for time series analysis and forecasting. It was developed by George Box and Gwilym Jenkins in the 1970s and has become a standard approach in the field of time series modeling. The Box-Jenkins methodology consists of three main steps: model identification, model estimation, and model diagnostic checking.

#### 1. Model Identification:

The first step in the Box-Jenkins approach is to identify an appropriate model that best represents the underlying structure of the time series. This involves determining the orders of autoregressive (AR), differencing (I), and moving average (MA) components, denoted as (p, d, q), respectively. Model identification is typically done through the analysis of autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, which help identify the significant lags and the presence of autoregressive and moving average patterns.

## 2. Model Estimation:

Once the model has been identified, the next step is to estimate the model parameters using a suitable estimation method. Maximum likelihood estimation (MLE) is commonly used for estimating the parameters of the ARMA model. The estimation process involves finding the values of the AR and MA coefficients that maximize the likelihood function based on the available data.

## 3. Model Diagnostic Checking:

After estimating the model, it is important to assess the adequacy of the model by conducting diagnostic checks. This involves analyzing the residuals (the differences between the observed values and the values predicted by the model) to ensure they meet certain assumptions, such as being normally distributed and exhibiting no significant autocorrelation. Various statistical tests and plots, such as ACF of residuals and Ljung-Box test, are employed to assess the model's goodness of fit.

If the diagnostic checks reveal that the model assumptions are not adequately met, further iterations of model identification, estimation, and diagnostic checking may be performed to refine the model until a satisfactory fit is achieved.

The Box-Jenkins methodology is known for its flexibility and versatility, allowing for the modeling of complex time series patterns. It has been successfully applied in various fields for forecasting, analyzing, and modeling time series data.

## ARIMA

The ARIMA (Autoregressive Integrated Moving Average) model is a widely used time series forecasting method. It combines autoregressive (AR), differencing (I), and moving average (MA) components to capture the underlying patterns and dependencies in a time series. Here's the mathematical description of an ARIMA (p, d, q) model:

Let  $Y_t$  be the observed value of the time series at time  $t$ .

### 1. Autoregressive (AR) component:

The autoregressive part models the relationship between the current observation and a linear combination of past observations.

AR(p) component:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t$$

Here,  $c$  is a constant term,  $\phi_1, \phi_2, \dots, \phi_p$  are the autoregressive coefficients,  $\varepsilon_t$  is the error term assumed to be white noise with zero mean and constant variance.

### 2. Differencing (I) component:

The differencing part helps in removing trend or seasonality from the time series data. It calculates the difference between consecutive observations to achieve stationarity.

I(d) component:

$$Y'_t = Y_t - Y_{t-d}$$

Here,  $Y'_t$  is the differenced series,  $Y_t$  is the original series, and  $d$  represents the order of differencing.

3. Moving Average (MA) component:

The moving average part models the dependency between the current observation and a linear combination of past error terms.

MA(q) component:

$$Y_t = \mu + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots + \phi_q \varepsilon_{t-p}$$

Here,  $\mu$  is the mean of the time series,  $\theta_1, \theta_2, \dots, \theta_q$  are the moving average coefficients, and  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p}$  are the lagged error terms.

Combining all three components, the ARIMA(p, d, q) model can be written as:

$$Y'_t = c + \phi_1 Y'_{t-1} + \phi_2 Y'_{t-2} + \dots + \phi_p Y'_{t-p} + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \dots + \phi_q \varepsilon_{t-p} + \varepsilon_t$$

Here,  $Y'_t$  represents the differenced series after applying the autoregressive and moving average components,  $\varepsilon_t$  and is the error term in the model.

The parameters  $p$ ,  $d$ , and  $q$  are determined based on the characteristics of the time series data and are typically estimated using methods like autocorrelation function (ACF) and partial autocorrelation function (PACF) analysis.

## ACF and PACF

ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) are statistical tools used to analyze and identify the correlation structure within a time series. Both ACF and PACF provide insights into the relationship between an observation and its lagged values. Here's the mathematical description of ACF and PACF:

### 1. Autocorrelation Function (ACF):

The ACF measures the correlation between an observation and its lagged values at different time lags.

ACF at lag  $k$ , denoted as  $\rho(k)$ , is defined as:

$$\rho(k) = \text{Corr}(Y_t, Y_{t-k})$$

Here,  $Y_t$  and  $Y_{t-k}$  are the values of the time series at time  $t$  and  $t-k$ , respectively.  $\text{Corr}$  represents the correlation coefficient between the two variables.

The ACF provides a measure of the linear relationship between the current observation and its lagged values. It quantifies the extent to which the values in the time series are correlated with each other over different lags. The ACF plot displays the correlation coefficients for various lags, which helps in understanding the presence of autocorrelation in the data.

## **2. Partial Autocorrelation Function (PACF):**

The PACF measures the correlation between an observation and its lagged values, accounting for the intermediate correlations through the removal of the effects of shorter lags.

PACF at lag  $k$ , denoted as  $\phi(k,k)$ , is defined as:

$$\phi(k,k) = \text{Corr}(Y_t, Y_{t-k} \mid Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$$

Here,  $Y_t$ ,  $Y_{t-k}$ , and  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$  represent the values of the time series at the respective time points.

The PACF helps in identifying the direct relationship between the current observation and its lagged values, after accounting for the intermediate correlations. It provides insights into the unique contribution of each lag to the current observation, which is useful for determining the appropriate lag order in autoregressive models.

Both ACF and PACF are commonly used in time series analysis to identify the order of autoregressive (AR) and moving average (MA) components in models like ARIMA and to understand the correlation structure of the time series data.

## **AIC, AICC, BIC**

AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and AICC (Corrected Akaike Information Criterion) are statistical measures used for model selection and comparison in the context of time series analysis and other statistical modeling. They provide a quantitative assessment of the goodness of fit of different models and help in selecting the most appropriate model based on their respective criteria. Here's the mathematical description of AIC, BIC, and AICC:

### **1. Akaike Information Criterion (AIC):**

AIC is a measure of the relative quality of a statistical model. It balances the trade-off between goodness of fit and model complexity, penalizing models with a larger number of parameters.

AIC for a model with parameter  $p$  and log-likelihood value  $L$  is calculated as:

$$\text{AIC} = -2 * L + 2 * p$$

Here, a lower AIC value indicates a better model fit. The term  $-2 * L$  represents the maximized log-likelihood of the model, and  $2 * p$  is the penalty term that accounts for the number of parameters in the model.

## 2. Bayesian Information Criterion (BIC):

BIC is a criterion similar to AIC but places a stronger penalty on the number of parameters. It incorporates a prior belief that simpler models are more likely to be true.

BIC for a model with parameter  $p$  and log-likelihood value  $L$  is calculated as:

$$BIC = -2 * L + p * \log(n)$$

Here,  $n$  represents the sample size. BIC penalizes models with a larger number of parameters more strongly than AIC, making it more suitable for model selection when the sample size is relatively small.

## 3. Corrected Akaike Information Criterion (AICC):

AICC is a modification of AIC that adjusts for small sample sizes. It takes into account both the model complexity and the sample size to provide a more accurate measure of model fit.

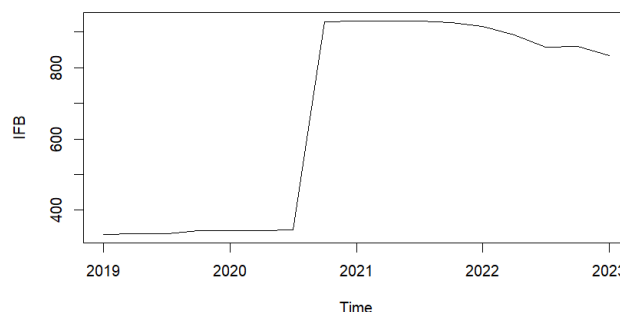
AICC for a model with parameter  $p$ , log-likelihood value  $L$ , and sample size  $n$  is calculated as:

$$AICC = AIC + (2 * p * (p + 1)) / (n - p - 1)$$

The additional term in AICC accounts for the correction based on the number of parameters and the sample size. AICC is particularly useful when the sample size is small and provides a more reliable measure of model fit compared to AIC.

## RESULT

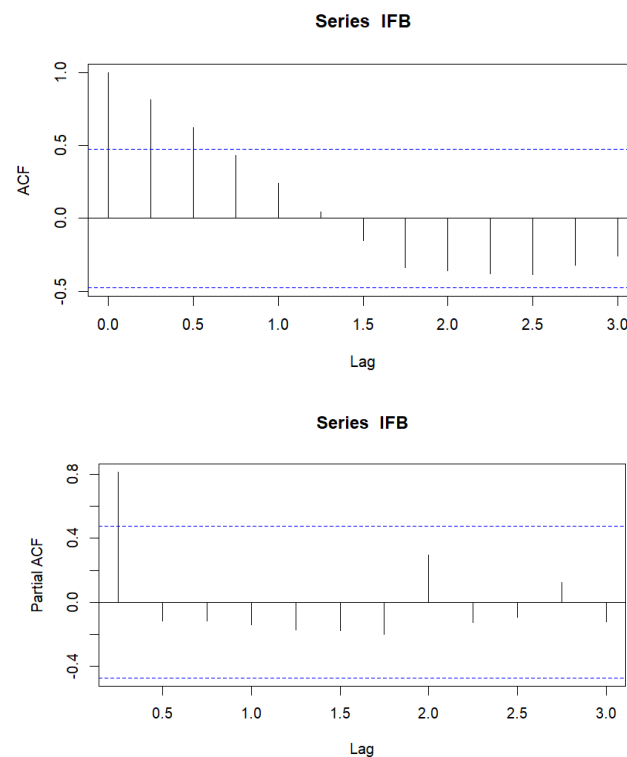
The following graph presents data that we have gathered quarterly on the number of foreign banks operating in India from 2019 all the way up to 2023 March. The years 2020 to 2021 marked the start of an increasing trend in that graph, which will be present all the way through 2022.



Using both stationary and non-stationary data to determine the AR and MA polynomial orders for India Foreign Bank By plotting the autocorrelation functions, we were able to determine the correct order of the autoregressive and moving average polynomials, as well as the values of  $p$



and  $q$ . This was made possible by the fact that we had access to the raw data. It is clear that the data are not stable since almost all of the autocorrelation lags, which are denoted by the notation  $(n/4th)$ , have values that are noticeably different from zero. Involving non-stationary data series that are related to India. It was discovered that the ACF plot depicted in the figure exhibited behavior that was not constant. On the other hand, PACF's figure shows that there was a significant spike at lag 1, which indicates that the series may include an autoregressive component of order one.



### Modelling Components

The procedure for determining the best fit for the foreign bank in India made use of a number of different ARIMA models, each of which had a different set of lags for the moving average and orders for the autoregressive series. The ARIMA model has been shown to be the most accurate (0,1,0) model. This is primarily attributable to the fact that the estimated value is lower than what was projected by the various other models. The task of handling the process of automatically include a constant falls under the purview of the Auto.arima() method. When  $d$  is more than one, the constant is never taken into account; on the other hand, when  $d$  is either less than zero or equal to one, the constant is always taken into account if the AIC value is enhanced as a result of its presence. After that, we used these models to make our forecasts for the Foreign Bank in India, and they covered the time period ranging from quarterly data 2019 to 2023 march.

ARIMA Models(0,1,0) value of AIC=207.01, AICc=207.29, BICc=207.78

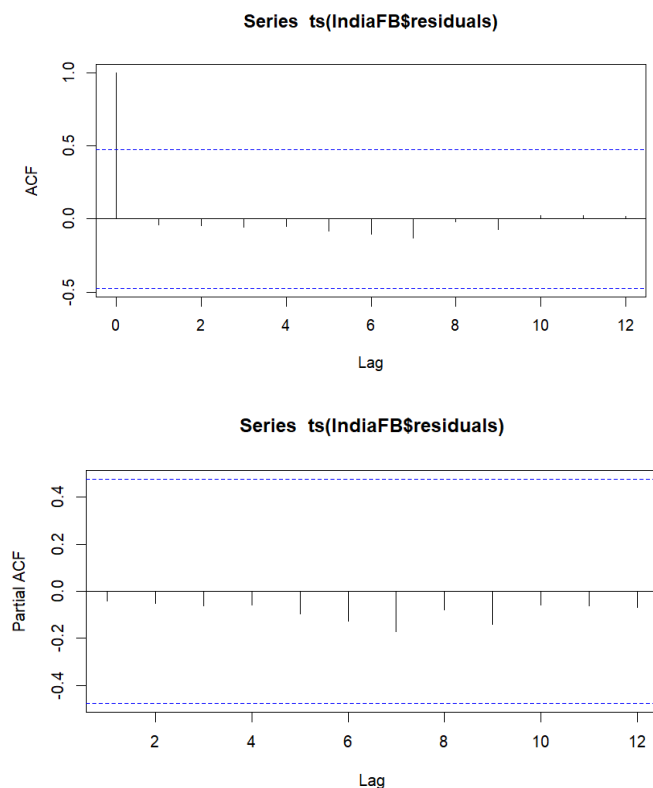
**Diagnostic Screening**

The model was studied by looking at the residuals to determine if there were any systematic patterns that could be eliminated to improve the selected ARIMA models. If there were, those patterns were discovered and taken into consideration during the analysis. In the event that there were, searches were conducted to locate and remove them. Following a series of experiments in which a variety of various combinations of delays were used for the autoregressive and moving average, it was discovered that the ARIMA (0,1,0) model was the one that provided the best accurate estimate of the total number of foreign banks operating in India.

The Fitted Model (0,1,0)

Ljung-Box Q statistic			
Model	X <sup>2</sup>	df	Sig
ARIMA(0,1,0)	0.41597	5	0.9949
Ljung-Box Q statistic			
Model	X <sup>2</sup>	df	Sig
ARIMA(0,1,0)	0.08148	2	0.9601

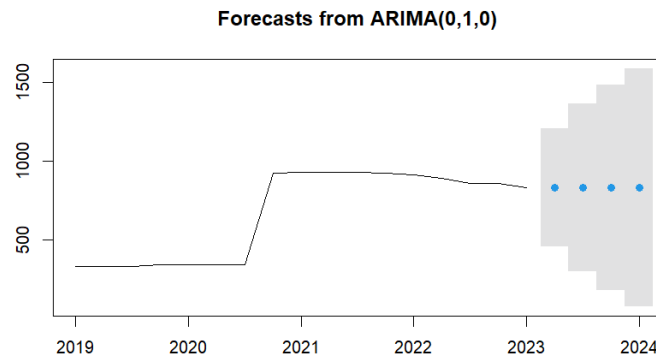
Residual ACFs and PACFs plot of India on ARIMA Model



From 2023 April to 2024 March , Four Quarterly forecast and value fir Foreign Bank in India

Year	Forecast Value	Lo(95)	Hi(95)
2023(Q2)	833	545	1120

2023(Q3)	834	426	1239
2023(Q4)	833	335	1330
2024(Q1)	832	258	1407



## CONCLUSION

In conclusion, this research study applied the ARIMA model in order to provide projections regarding the quarterly performance of international financial institutions operating in India for the following calendar year, which will begin in April 2023 and end in March 2024. We were able to get significant insights about the future trajectory of these financial institutions by conducting an analysis of quarterly data beginning in 2019 and continuing through March 2023.

The use of the ARIMA model produced encouraging findings, with the parameter combination (0, 1, 0) being determined to be the model that fits the data the best. This option was made due to its capacity to reduce the number of forecasting errors while simultaneously increasing accuracy, thereby providing a solid foundation for accurately anticipating the performance of international banks in India. The model was validated using the Akaike Information Criterion (AIC), as well as the Bayesian Information Criterion (BIC), which provided further confirmation of its applicability and effectiveness in capturing the underlying patterns and dynamics of the financial time series data.

In addition, the research into the stationarity of the time series data that we conducted with the help of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) enabled us to address any non-stationary components that were still present in the dataset. We improved the robustness and dependability of the ARIMA model, which resulted in an increase in the accuracy of our forecasts, and we did this by ensuring that the data were stationary.

The findings of this study's forecasts and the insights it supplied contain significant implications for a variety of parties operating within the financial system. These projections can be utilized by policymakers and regulatory bodies in order to make informed decisions that contribute to the maintenance of a stable and favorable banking environment. In order to better direct their

investment plans and decisions, investors are able to evaluate the prospective risks and rewards of their investments based on the expected performance of international banks.

It is essential to recognize that the process of forecasting financial time series data is fraught with inherent uncertainties, and it is possible that the actual performance of foreign banks in India would diverge from the values that were expected. As a result, it is vital to maintain ongoing monitoring and improvement of the forecasting models in order to adapt to the shifting conditions and dynamics of the market.

The use of machine learning techniques such as ARIMA in financial forecasting illustrates the growing importance of sophisticated analytical tools in the process of decision-making. This study demonstrates the potential of machine learning algorithms in analyzing and predicting financial patterns, so paving the way for additional research and inquiry in this area of study.

In a nutshell, the forecasting research that was carried out in the course of this study sheds light on the future performance of foreign banks in India, hence giving significant information for stakeholders in the financial system. The ARIMA model and the accompanying procedures have been successfully implemented, which paves the way for future developments in forecasting techniques and contributes to the ongoing efforts to improve the financial sector's stability and efficiency in India.

## REFERENCE

1. Campa, J. M., & Chang, P. K. (1998). The forecasting ability of correlations implied in foreign exchange options. *Journal of International Money and Finance*, 17(6), 855-880.
2. Sermpinis, G., Theofilatos, K., Karathanasopoulos, A., Georgopoulos, E. F., & Dunis, C. (2013). Forecasting foreign exchange rates with adaptive neural networks using radial-basis functions and particle swarm optimization. *European Journal of Operational Research*, 225(3), 528-540.
3. Znaczkó, T. M. (2013). Forecasting Foreign Exchange Rates.
4. Jere, S., Kasense, B., & Chilyabanyama, O. (2017). Forecasting foreign direct investment to Zambia: a time series analysis. *Open Journal of Statistics*, 7(01), 122.
5. Ülengin, F., & Ülengin, B. (1994). Forecasting foreign exchange rates: A comparative evaluation of AHP. *Omega*, 22(5), 505-519.
6. Roy, N. C., Vamsi, N., & Rao, N. (2015). An appraisal and forecasting of NPAs in Indian banking industry. *Al-Barkaat Journal of Finance & Management*, 7(1), 31-40.
7. Naumenkova, S. V., & Domanetskyi, I. V. (2014). Forecasting of capitalization of banking institutions in Ukraine. *Financial and credit activity problems of theory and practice*, 2(17), 57-64.
8. Guo, H., & Savickas, R. (2008). Forecasting foreign exchange rates using idiosyncratic volatility. *Journal of Banking & Finance*, 32(7), 1322-1332.
9. Dominguez, K. M. (1986). Are foreign exchange forecasts rational?: New evidence from survey data. *Economics Letters*, 21(3), 277-281.

10. Raheem, F., & Iqbal, N. (2021, September). Forecasting foreign exchange rate: Use of FbProphet. In *2021 International Research Conference on Smart Computing and Systems Engineering (SCSE)* (Vol. 4, pp. 44-48). IEEE.
11. Tseng, F. M., Tzeng, G. H., Yu, H. C., & Yuan, B. J. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy sets and systems*, *118*(1), 9-19.
12. Adewole, A. P., Akinwale, A. T., & Akintomide, A. B. (2011). Artificial neural network model for forecasting foreign exchange rate.
13. Ngan, T. M. U. (2013). Forecasting foreign exchange rate by using ARIMA model: A case of VND/USD exchange rate. *Methodology*, *2014*, 2015.
14. Frankel, J. A., & Froot, K. A. (1990). Exchange rate forecasting techniques, survey data, and implications for the foreign exchange market.
15. Elliott, G. R. (1986). The changing competitive environment for the Australian banking/finance industry: review of a forecasting study. *International Journal of bank Marketing*, *4*(5), 31-40.
16. Al-rawashdeh, S. T., Nsour, J. H., & Salameh, R. S. (2011). Forecasting foreign direct investment in Jordan for the years (2011-2030). *International journal of business and Management*, *6*(10), 138.