Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 46 No. 9
September 2025

Analysis of Private Sector Bank Index Prices Using Arima Model

[1]Mousumi Banerjee , [2]Priyanka, [3]Nisha Bhadana, [4]Sudhir Singh,
[5]Pradeep Kumar Singh

[1,2,4] Department of Applied Science,KCC Institute of Technology and Management,Greater Noida-201306, U.P

[3] Department of Computer Science and Engineering Vishveshwarya Group Of Institutions , Greater Noida, U.P

[5] Department of Basic Sciences and Humanities, Pranveer Singh Institute of Technology, Kanpur, U.P

Abstract: One of the most important financial assumptions that investors face is and will always be stock price forecasting. There are many approaches to accurately predict a company's share price, most of which depend on different variables that affect the share price in the market. Time series data analysis is one of the main models used in data analysis. A significant class of models in machine learning, econometrics, and statistics is time series forecasting. Predictions from a time series model are typically predicated on the idea that historical trends will recur in the future. This paper establish comparison of four Arima (1,0,1), (5,1,1) (1,1,1) & (4,1,2) modal for private sector bank index from 10/06/24 to 20/06/24. The result shows that (5,1,1) gives the best result.

Key words: Arima Model , Stock Prices, E-view, Time Series. Garch Model.

1. Introduction

It is very difficult to forecast the stock prices though stock prices is part of time series analysis. Stock prices are non-linear and non-stationary due to a variety of factors that influence them. Due to the large amount of noise and constantly shifting conditions, forecasting the stock price index requires a number of processes. Time series data related to stock prices. The data used in the financial business, namely stock price indices, are frequently highly volatile, stochastic, and heteroscedastic. Volatility as a measure of uncertainty is applied to time series data that exhibit conditional heteroscedasticity, that is, non-uniform variance. Heteroscedasticity serves as the variance to be modeled in the ARCH and GARCH models, allowing us to know Hajizadeh et al. (2012) made an effort to improve GARCH-type models capacity to predict return volatility. They put forth two hybrid models that combine the EGARCH model and neural networks.

The Glosten et al. (1993) study improved the model and suggested GJR GARCH. The premise behind their model is that volatility reacts asymmetrically to both positive and negative shocks. This research investigates the existence of long memory in the mean and volatility of the Naira per Dollar exchange rate series using models of autoregressive fractionally integrated moving average (ARFIMA), generalised autoregressive conditional heteroscedastic (GARCH), and fractionally integrated generalised autoregressive conditional heteroscedastic (FIGARCH) origins.Floros &Christos (2008) investigated the application of GARCH-type models for modelling volatility and explaining financial market risk using daily data from Israel (TASE-100 index) and Egypt (CMA General Index). The investigation came to the conclusion that there is substantial evidence to support the above models' ability to describe daily returns. Fereshteh and Hossein (2013) utilised GARCH (1-1) and GARCH (2-2) to examine the volatility of a chosen pharmaceutical group, car group, and oil business utilizing daily index data from 2006 to 2010. The outcome demonstrated volatility feedback in the oil and pharmaceutical industries. Volatilities have a positive impact on output in the pharmaceutical group, but a negative impact on the oil group. Furthermore, it wasn't verified in the car group. Zhang et al. (2018) used eighteen macroeconomic and eighteen technical factors to study oil price forecasting. For a mean-variance investor, the outcomes produced certainty comparable return gains and demonstrated accurate forecasting.A time series is simply a collection of observations arranged chronologically. Bozarth (2016) Time series forecastingmodels forecast demand using mathematical methods derived from historical data. Zhang(2017) used this method to present a hybrid ARIMA and ANN approach for time series forecasting. To capture various types of relationships in the time series data, Zhang's hybrid model combines the linear ARIMA and nonlinear multilayer perceptrons models.To achieve accurate predictions in the foreign exchange market, Yu et al. (2005) proposed a novel onlinear ensemble forecasting model that integrates neural networks and generalised linear auto regression (GLAR).Box and Jenkins(1976)suggested using the sample data's autocorrelation function (ACF) and partial autocorrelation function (PACF) as the fundamental tools to determine the order of the ARIMA model.

2. Arima Model: Time series analysis primarily relies on four key models:moving average (MA), auto-regressive (AR), auto-regressive moving average (ARMA), and auto-regressive integrated moving average (ARIMA). The ARIMA model is utilized extensively in time forecasting. The ARIMA model is widely recognized forits accuracy and adaptability in predicting a broad range of time series data sources.The purpose of the ARIMA model is to forecast and analyze time series data by figuring out the best values for the p, d and q parameters. The smoothness of the series is used to calculate the number of differences, or d, for a particular time series data set. The ACF and PCAF plots are then used to derive the values of p and q for the AR(p) and MA(q) components. The ARIMA model's generic form is

$$Y_t = c + \emptyset_1 Y_{t-1} + \emptyset_2 Y_{t-2} + \ldots . \emptyset_p Y_{t-p} + \emptyset_2 \delta_{t-1} + \emptyset_1 \delta_{t-2} + \emptyset_q \delta_{t-q} + \varepsilon$$

The predicted outcome of the variable is denoted by Y_t, while the preceding values of the dependent variable or autoregressive terms are represented by $Y_{t-1}, Y_{t-2}, ---- Y_{t-p}$. The error term is represented by ε_t while the moving average terms or error lagged values are denoted by α and θ.

3. Methodology: To develop an ARIMA model, you need to do two main steps: model identification, and parameter estimation. Data on private sector index stock prices during 10 days from June 10 to June 20 was obtained from NSE. Because it provides a fair and reliable indicator of equity value, supplementingthe simple market price, we decided to utilise the adjusted closing price for our research. E-Views software, version 10, was the instrument utilizes to perform the time series analysis.

3.1 Model Identification: In order to develop an ARIMA model, also known as the Box & Jenkins's model, we first plot the raw data graph in order to gain a general understanding of the data and analyse its patterns. This enables us to ascertain whether the data series is stationary. The principal assumption is that the stationary. The mean and variance of the process would not change over time. Nonetheless, in the domains of finance and economics, time-varying conditional variance is commonly observed in series. Additionally, non-stationary in variance is often seen in high frequency data, such as hourly, daily, and weekly data. The Auto correlation Function, the ACF plot, and the Augmented Dicker Fuller Test (Unit Root Test) are instruments used to verify the stationary of the data series. The auto correlation of the time series is plotted versus lags using the ACF. In a dataset, lag is the interval of time between one observation and its predecessor. To achieve stationary, data transformation will be performed if the data series is not stationary. We use differencing in this work in order to achieve stationaryby computing the difference between successive observations, We can eliminatetrend patterns fromdata by using differentiating. Usually, $d = 1$ or first differentiating is enough to stabilize the mean and make the time series data st ationary. Excessive differentiating often results in superfluous correlation within the model. A prerequisite for using the ARIMA model is stationery. To determine the quantity of AR or MA terms in the ARIMA model, we look at the correlogram pattern, which is represented by plots of the autocorrelation function (ACF) and partial autocorrelation function (PACF).

3.2 Parameter setting: In the ARIMA (p, 1, q) model, we use the trial-and-error method to find the order of p and q. Our "simplest to best" approach is used until we get the best parsimonious model. The best-fitting model in this instance is based on the statistically significant coefficients. The Akaike-Information Criterion (AIC) and the Bayesian or Schwarz Information Criterion (SIC) scores are used to identify the best model. Model with lowest AIC and SIC values is the suggested model.

4. Data Preprocessing

4.1 Stationary test: As intended, Table 1 depicts the non-stationary behaviour of share prices. The ADF test statistic, which is negligible at the 5% level, attests to the outcome. Therefore, natural logarithms have been used for stock prices in order to induce stationarity. The change in the log of stock prices, or the first difference, has then been considered. Table 2 shows that the outcome is a stationary series.

Modal	Test for Unit root	t-statistics	Probability
Test statistics	First difference level	-6.9069	0.0005

Table 1: Automated Dickey Fuller test statistics

Modal	Test for Unit root	t-statistics	Probability
Test statistics	First difference level	-6.2931	< 0.01

Table 2: Automated Dickey Fuller test statistics

4.2 Model Estimation

Dependent variable: ADJ Closing Price

Sample: 10/06/24 to 20/06/24

Including observation: 7

Variable	**Coefficient**	**Std Error**	**t-Statistic**	**Prob**
C	7413.321	23.407	316.7132	0.000
AR(5)	-0.972552	0.869780	-1.1181	0.3450
MA(1)	0.454809	0.840342	0.5412270	0.6260
R-Square	0.969647	Mean dependent var		7424.2249
Adj Rsquare	0.939294	S.D.dependent var		63.10413
S.E. of regression	15.547	Akaika info criterian		10.73070
Sum square	725.2193	Schwartz criterian		10.69979
Log likilihood	-33.56746	Hannan-Quinn criter		10.34868
F-Statistic	31.94560	Durbin -Watson Stat		1.667711

Table 3: Arima (5,1,1) for private sector index.

The ARIMA (5, 1, 1) model, which was chosen using the standard Box-Jenkins approach, is shown in Table 1 with its findings. At the 5% level, the model's two parameters are both significant. The model's corrected R2 value is .96. But in order to make sure this model fits the data the best, a few other likely models are also tested and compared with one another.

Dependent variable: ADJ Closing Price

Sample: 10/06/24 to 20/06/24

Including observation:7

Variable	**Coefficient**	**Std Error**	**t-Statistic**	**Prob**
C	7424.351	42.63699	174.1293	0.000
AR(1)	0.909379	2.675472	0.339895	0.7563
MA(1)	-0.999773	527.8964	-0.001894	0.9986
R-Square	0.059707	Mean dependent var		7424.2249
Adj Rsquare	-0.880587	S.D.dependent var		63.10413
S.E. of regression	86.53752	Akaika info criterian		12.09534
Sum square	22466.23	Schwartz criterian		12.06444
Log likilihood	-38.333	Hannan-Quinn criter		11.71332
F-Statistic	0.634	Durbin -Watson Stat		0.376263

Table 4: Arima (1,0,1) for private sector index.

The ARIMA (1, 0, 1) model, which is frequently employed in several ARIMA investigations by academics, is represented by the results in Table 3. Given that one of the three model parameters is negligible at the 5% level, this model appears to be a little bit worse than the ARIMA (5, 1, 1) model or the one that uses the Box-Jenkins technique. Nonetheless, the two models have the same modified R2 value. In comparison to ARIMA (5, 1, 1), ARIMA (1, 0, 1) has higher values for the Hannan-Quinn criterion (HQC), the Schwarz Bayesian criterion (SBC), and the Akaike information criterion (AIC). ARIMA (5, 1, 1) can be regarded as a better model since the lower the information criteria values, the better the model.

Dependent variable: ADJ Closing Price

Sample: 10/06/24 to 20/06/24

Including observation:7

Variable	**Coefficient**	**Std Error**	**t-Statistic**	**Prob**
C	7413.321	23.407	316.7132	0.000
AR(5)	-0.972552	0.869780	-1.1181	0.3450
MA(1)	0.454809	0.840342	0.5412270	0.6260
R-Square	0.969647	Mean dependent var		7424.2249
Adj Rsquare	0.939294	S.D.dependent var		63.10413
S.E. of regression	15.547	Akaika info criterian		10.73070
Sum square	725.2193	Schwartz criterian		10.69979
Log likilihood	-33.56746	Hannan-Quinn criter		10.34868
F-Statistic	31.94560	Durbin -Watson Stat		1.667711

Table 5:Arima (1,1,1) for private sector index:

We further examined at the most basic auto regressive process of order 1, often known as the ARIMA (1, 1, 1) process or the AR (1) and MA(1) process, in an effort to further simplify or abridge our recommended model. All of the information criteria values, including AIC, SBC, and HQC, were found to be lower than those for ARIMA (1, 0, 1), however the adjusted R 2 value was determined to be the same as ARIMA (5, 1, 1). Table 5 summarize the outcomes of the ARIMA (1, 1, 1) model.

Dependent variable: ADJ Closing Price

Sample: 10/06/24 to 20/06/24

Including observation:7

Variable	**Coefficient**	**Std Error**	**t-Statistic**	**Prob**
C	7426.321	23.407	316.7132	0.000
AR(4)	-0.719898	0.869780	-1.1181	0.3450
MA(2)	-0.998847	0.840342	0.5412270	0.6260
R-Square	0.7517	Mean dependent var		7424.2249
Adj Rsquare	0.50342	S.D.dependent var		63.10413
S.E. of regression	44.456	Akaika info criterian		11.7410
Sum square	5932.282	Schwartz criterian		11.71037
Log likilihood	-37.09	Hannan-Quinn criter		11.71332
F-Statistic	3.023	Durbin -Watson Stat		0.65603

Table 6:Arima (4, 2, 1) for private sector index :

The prediction ability of ARIMA (4, 2, 1) was also examined in an effort to find the best possible model. Table 6 illustrates that every model parameter was highly significant. Compared to all the modalities, the model's adjusted R2 was marginally lower. However, the model's information criterion values were all marginally higher than those of ARIMA (1, 1, 1) and ARIMA (5, 1, 1), which further impacted the model's potential in comparison to ARIMA (1, 1, 1) and ARIMA (5, 1, 1).

4.3 Residual test : The line graph has three plotted lines with the following labels:

Residual(Blue) , Actual (orange),Fitted (green)

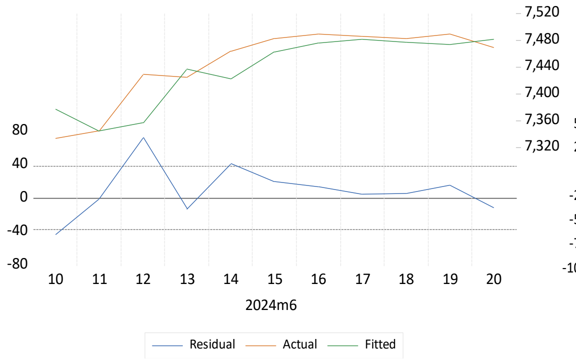

Fig 1:Residual series test plot for model(5,1,1)

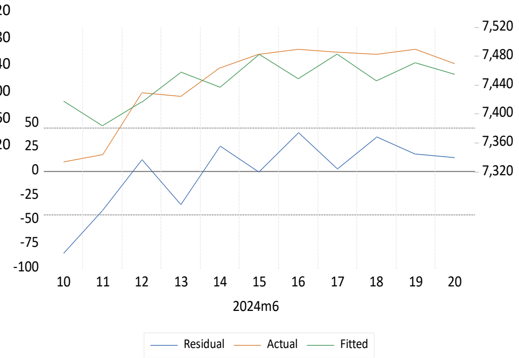

Fig 2:Residual series test plot for Model (1,0,1)

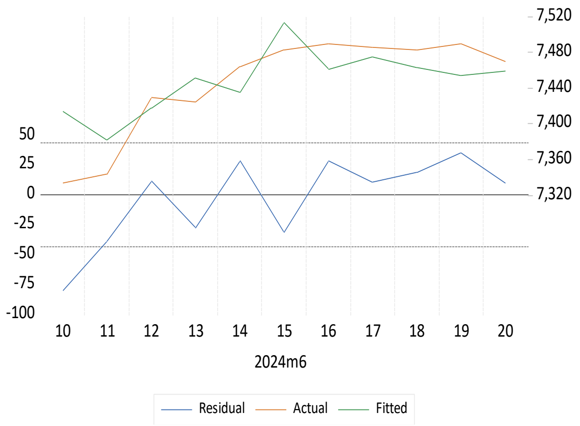

Fig 3: Residual series test plot for model(1,1,1)

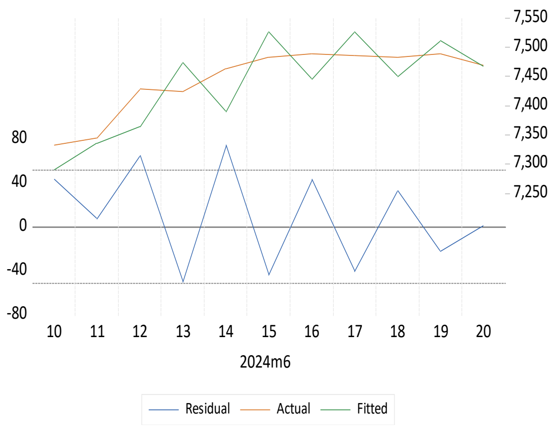

Fig 4: Residual series test plot for model (4,1,2)

In Fig 1; The observation indicating the values of the residuals are not dispersed at random around zero may indicate that the model contains bias or non-linearity.Despite some deviations, the Actual and Fitted lines show a similar pattern, suggesting the model captures the overall trend.The Residuals are plotted separately and are represented by the difference between the Actual and Fitted lines.

In Fig 2: The residuals are not dispersed randomly along the horizontal zero line.There is a distinct pattern: a wave-like pattern of alternating ups and downs.A smooth upward trend is followed by a peak and a slight decline in the actual values.Although the fitted values are generally close, there is some discrepancy, particularl

In Fig 3: Model seems to be accurately capturing the data's trend because the fitted and actual lines are near to one another.Residual Pattern: The model's assumptions (such as linearity and homoscedasticity) may hold water if the residuals are dispersed randomly and don't exhibit any discernible pattern

In Fig 4: The residuals may have some cyclical or autocorrelated structure, as suggested by the alternating up-and-down pattern. It could mean that the model is lacking a seasonality or trend component.A more intricate model or additional refinement could enhance fit.

5. Conclusion: Analysis of the private sector bank index prices performance during ten days of NSE trading value gives us the ARIMA (5,1,1) model was selected from three alternative model parameters(1,0,1), (1,1,1) & (4,1,2)because it offers the best model and meets all fit statistics criteria, while the other three models(1,0,1), (5,1,1) (1,1,1) & (4,1,2) did not meet fit statistics requirements.ARIMA (5,1,1) model provides a strong and reliable framework for predicting the short-term price changes of the private sector bank index. Its selection emphasis how important it is to compare models rigorously using residual diagnostics and statistical fit measures. According to the model's excellent performance, recent price trends in the index show moving average and auto-regressive behaviour, which can be successfully captured for forecasting and investment decision-making. To improve predictive accuracy significantly, future studies might use a longer time period or include exogenous variables.

6. References:

1. Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. Journal of Finance, Vol. 48, pp.1779–1801.
2. Aliyu, M.A.,Dikko,H.G.,& Danbaba,U.A.,(2023). Statistical modelling for forecasting volatility in Naira per Dollar exchange rate using ARFIMA-GARCH and ARFIMA-FIGARCH models, World Scientific News, Vol. 176, pp. 27-42.
3. Floros, Christos (2008), Modelling volatility using GARCH models: evidence from Egypt and Israel.Middle Eastern Finance and Economics, Vol. 2, pp. 31-41.

4.Sha,N,Chandrasekar,K.S.,Dileap,M.R.,(2019):Stock Volatility in Muscat Securities Market,International Journal of Recent

Technology and Engineering (IJRTE) Vol 8, pp. 3942-396.2

5. Hajizadeh, E., Seifi, A., Zarandi, M.F., Turksen, I.B. (2012), A hybrid modelling approach for forecasting the volatility of S&P 500 index return. Expert Systems with Applications, Vol.39, pp.431-436.
6. Zhang. L et. al(2018):Trend analysis and forecast of PM 2.5 in Fuzhou. China using the ARIMA model, Vol. 95,pp.702-710.
7. Bozarth CB, Handfield RB (2016): Introduction to operations and supply chain management, 4thed. Raleigh: North Carolina State University,
8. G.P. Zhang(2003): Time series forecasting using a hybrid ARIMA and neural network model, Neurocomputing , Vol 50 , pp.159–175.
9 .L.Yu,S.Wang, & K.K. Lai (2005): A novel nonlinear ensemble forecasting model incor porating GLARandANNforforeignexchangerates,Computers and Operations Research , Vol 32 (10) pp. 2523–2541
10.P.Box,G.M.Jenkins(1976): Time Serie Analysis Forecastingand Control,Holden-day Inc., San Francisco, CA,