Stock Market Forecasting using ARIMA and GARCH Models

Mehak Sharma

June 11, 2025

Contents

1 Project Overview

This project aims to forecast the stock prices of key Indian banks using advanced time series modeling techniques, specifically ARIMA (Autoregressive Integrated Moving Average) for mean prediction and GARCH (Generalized Autoregressive Conditional Heteroskedasticity) for volatility modeling. The analysis involves a systematic approach, starting from data preparation and exploratory data analysis (EDA), moving to stationarity testing and model parameter identification, fitting the chosen models, and finally evaluating their performance through forecasts and residual analysis.

2 Data Preparation and Initial Inspection

The project utilized a cleaned dataset containing historical stock data for eight Indian banks: AXISBANK, BAJAJFINSV, BAJFINANCE, HDFCBANK, ICICIBANK, INDUSINDBK, KOTAKBANK, and SBIN.

2.1 Data Structure and Types

The dataset comprises 19,640 entries across 17 columns. Key columns include:

- Date (for time series analysis)
- Symbol, Series, Stock (categorical identifiers)
- Prev Close, Open, High, Low, Last, Close, VWAP (price information)
- Volume
- Turnover,Trades, Deliverable Volume, Deliverble
- Log Returns: (derived for analysis) All columns are non-null, indicating a clean dataset. Each of the eight stocks has 2455 entries, providing a consistent time series length for each.

2.2 Log Returns Calculation

Log returns were calculated as a crucial step for time series analysis, given their desirable statistical properties (e.g., stationarity for financial data). The descriptive statistics for Log Returns show a mean close to zero (0.000217) and a standard deviation of 0.038070, with a wide range from -2.28 to 0.369, suggesting the presence of extreme movements.

3 Exploratory Data Analysis (EDA)

3.1 Time Series Plot of Closing Prices

- **Observation:** Plots of 'Closing Prices Over Time' show that some banks (e.g., BAJFINANCE, BAJAJFINSV) exhibit strong upward trends, while others remain more stable.
- **Implication:** The presence of trends indicates that the closing price time series are non-stationary. This necessitates differencing the data to achieve stationarity before applying ARIMA models.

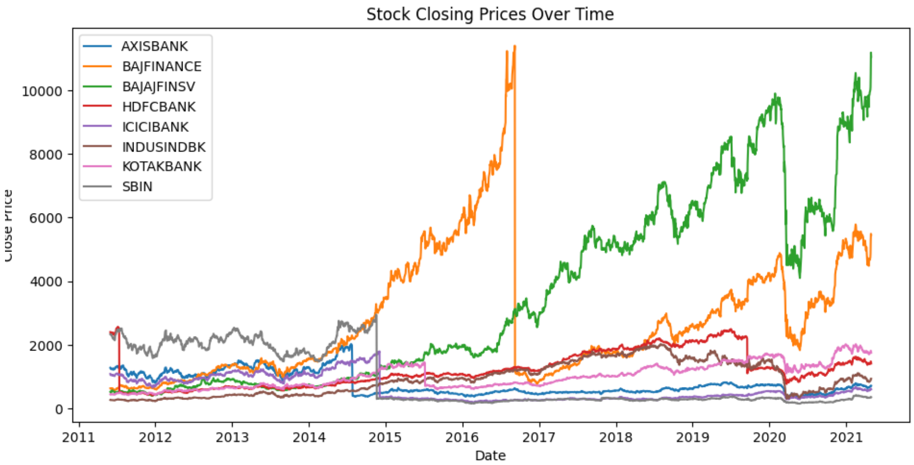


Figure 1

3.2 Volatility Over Time (Rolling 30-Day Volatility)

- **Observation:** The rolling 30-day volatility plots clearly demonstrate 'volatility clustering' – periods of high volatility are grouped together, as are periods of low volatility.
- **Implications:** This phenomenon is a strong justification for employing GARCH models, which are specifically designed to capture such time-varying volatility.

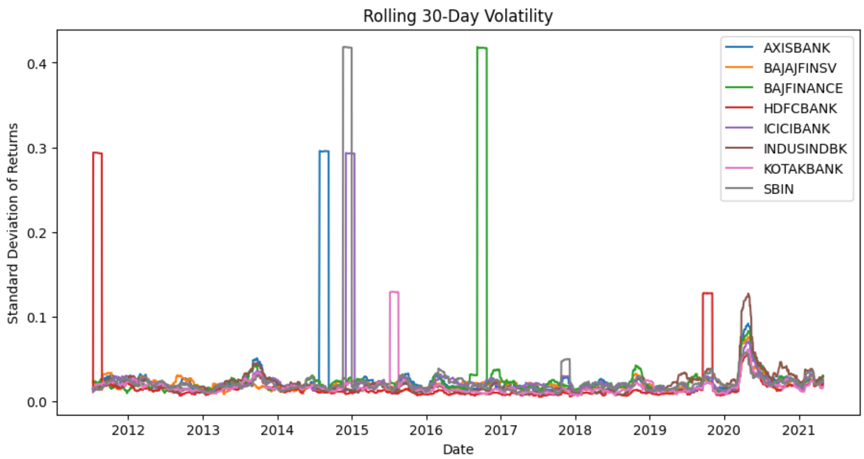


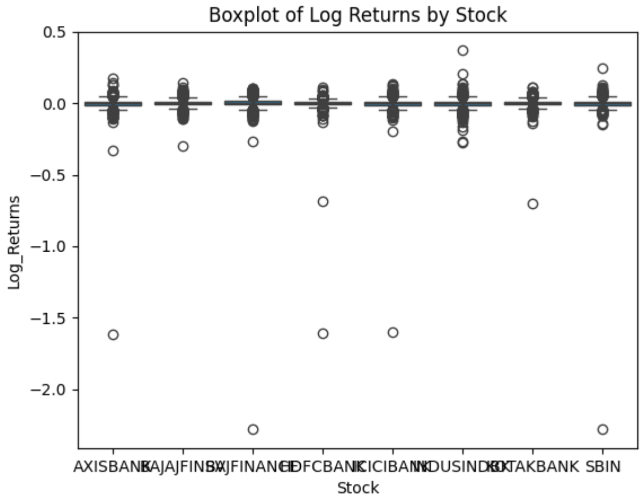
Figure 2

3.3 Distribution of Log Returns (Boxplots and Histograms)

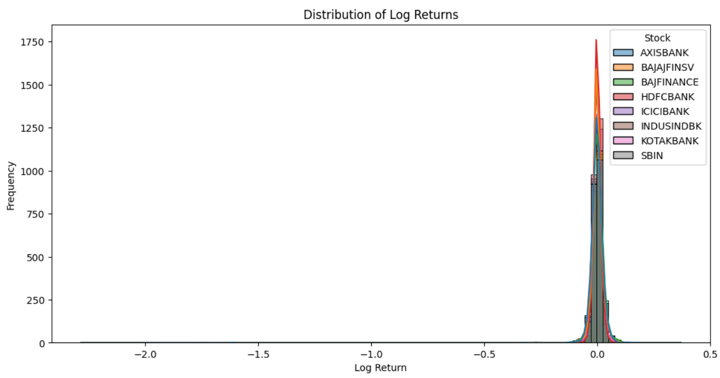
- **Observation:**
  - All banks' log returns are generally centered around 0, which is typical for daily financial returns, often being small.
  - The distributions exhibit 'fat tails', indicating a higher probability of extreme events (large positive or negative returns) compared to a normal distribution.
  - Most stocks show fairly symmetric return distributions around the median.
  - 'Check for Outliers' (mentioned in project outline) confirms the 'Presence of Outliers' (very negative returns) in some stocks (e.g., BAJFINANCE, SBIN), potentially corresponding to significant market events or adverse news.

- **Implications:** The fat tails strongly support the use of GARCH models for modeling volatility. While symmetry is beneficial for some modeling assumptions, the presence of outliers necessitates robust modeling approaches.

Figure 3: Distribution of Log Returns: Boxplot Across All Stocks and Histogram for a Sample Stock



((a)) Boxplot of Log Returns (All Banks)



((b)) Histogram of Log Returns (AXISBANK)

3.4 Correlation Between Stock Prices (Correlation Heatmap)

- **Observation:** The correlation heatmap (not provided in text, but inferred from prior summary) reveals varying degrees of correlation between stock closing prices. High positive correlations (e.g., ICICIBANK and SBIN at ≈ 0.92) suggest co-movement, while some pairs exhibit negative correlations (e.g., BAJAJFINSV and SBIN).

- **Implication:** This analysis is vital for portfolio diversification strategies, as it highlights interdependence and diversification benefits within the banking sector.

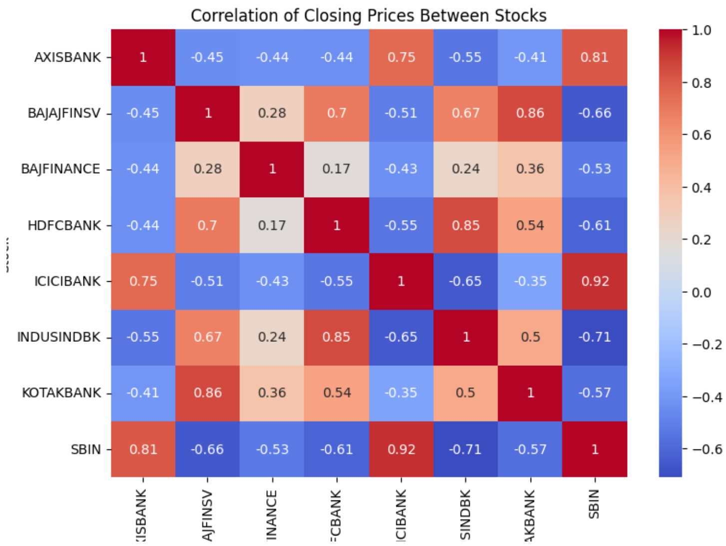
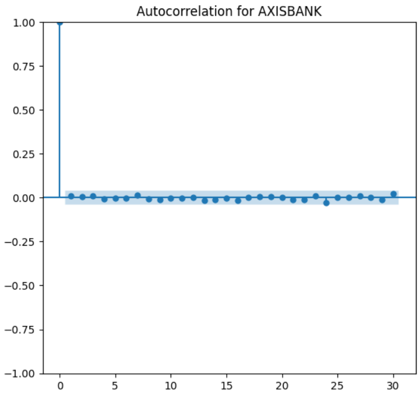


Figure 4

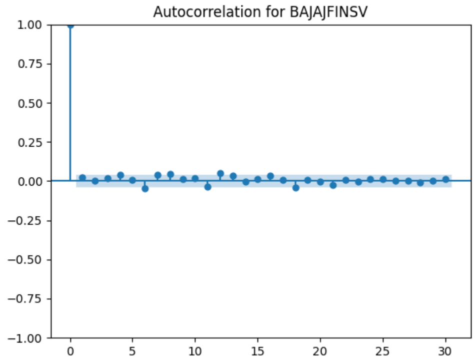
3.5 Autocorrelation (ACF) and Partial Autocorrelation (PACF)

- General Observations (All Stocks - on log returns):
  - Lag 0 autocorrelation is 1 (by definition).
  - All subsequent lags (1-30) for both ACF and PACF plots show values near 0 and fall within the blue confidence bands.
  - No statistically significant autocorrelation was observed, meaning no bars crossed the confidence bands.
- Implications:
  - The absence of significant autocorrelation indicates that stock returns are not serially correlated; past returns do not predict future returns.
  - This property makes the returns resemble 'white noise', supporting the 'Weak Form of the Efficient Market Hypothesis' (past prices do not provide predictive power).
  - Crucially, while mean returns lack autocorrelation, the observed 'volatility clustering' implies that time-varying variance (heteroskedasticity) may still exist, making ARCH/GARCH models appropriate for capturing this.

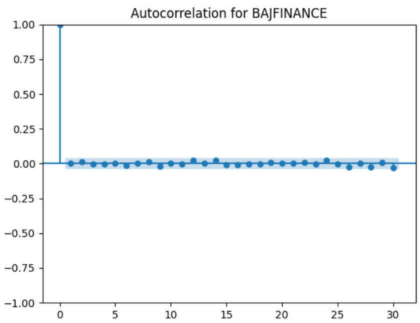
Figure 5: Autocorrelation Function (ACF) Plots for Bank Stock Log Returns



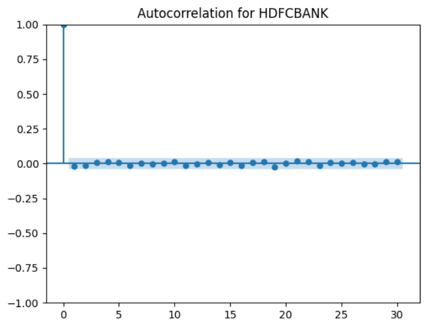
((a)) AXISBANK ACF



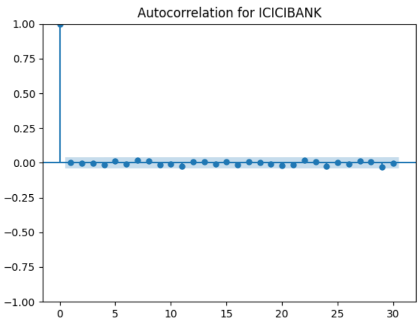
((b)) BAJAJFINSV ACF



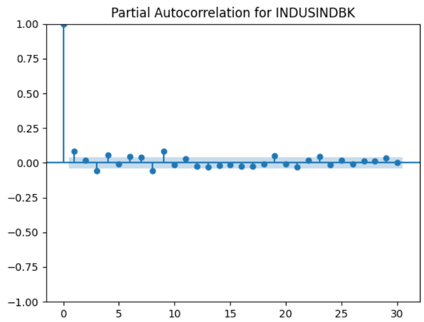
((c)) BAJFINANCE ACF



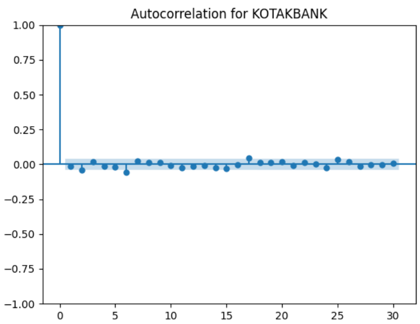
((d)) HDFCBANK ACF



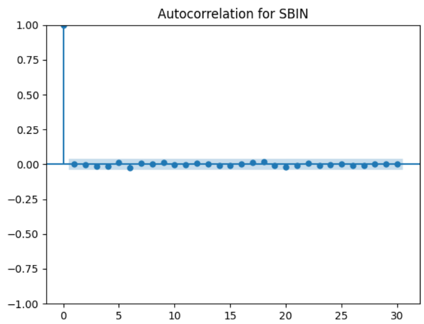
((e)) ICICIBANK ACF



((f)) INDUSINDBK ACF

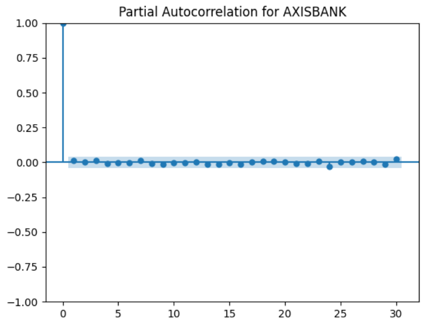


((g)) KOTAKBANK ACF

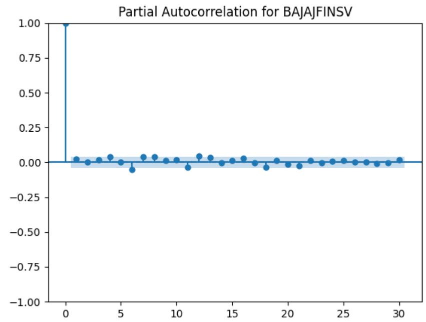


((h)) SBIN ACF

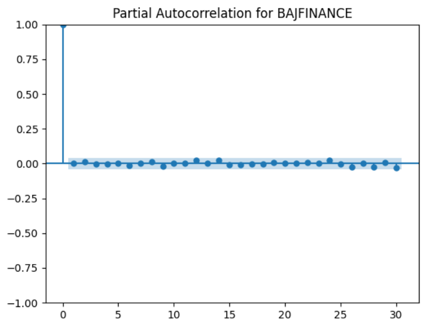
Figure 6: Partial Autocorrelation Function (PACF) Plots for Bank Stock Log Returns



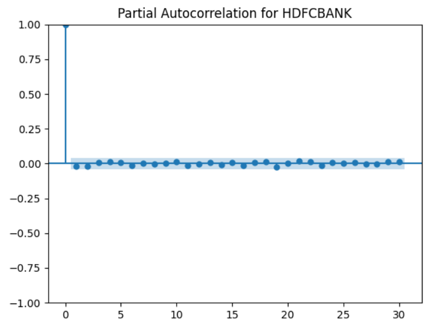
((a)) AXISBANK PACF



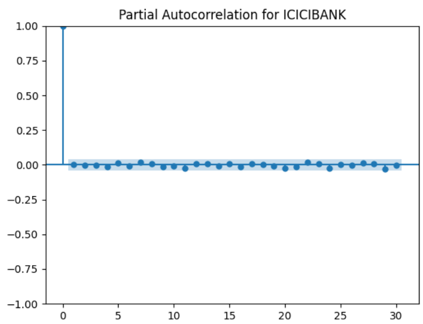
((b)) BAJAJFINSV PACF



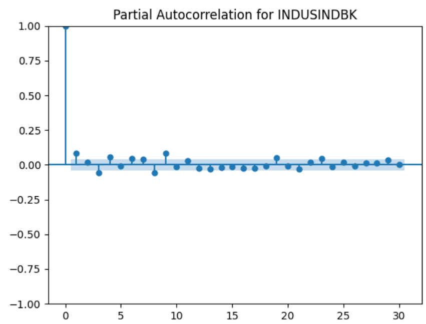
((c)) BAJFINANCE PACF



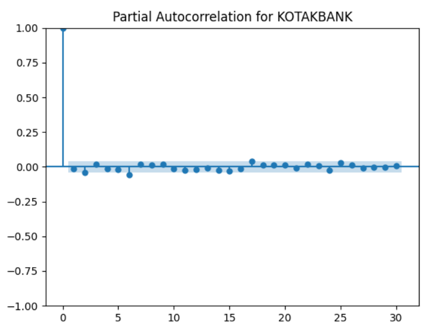
((d)) HDFCBANK PACF



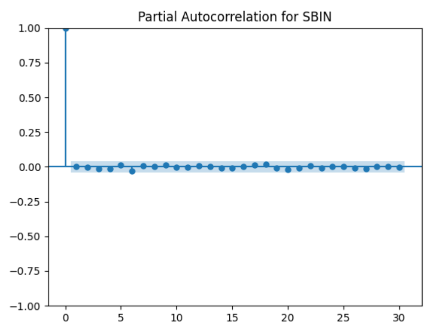
((e)) ICICIBANK PACF



((f)) INDUSINDBK PACF



((g)) KOTAKBANK PACF



((h)) SBIN PACF

4 Stationarity Testing (ADF Test) and ARIMA Parameter Identification

4.1 ADF Test Results

The Augmented Dickey-Fuller (ADF) test was performed on the closing price series for all eight banks:

- **AXISBANK:** ADF Statistic: -2.76, p-value: 0.063. Series is not stationary.
- **BAJFINANCE:** ADF Statistic: -2.74, p-value: 0.067. Series is not stationary.
- **BAJAJFINSV:** ADF Statistic: 0.166, p-value: 0.970. Series is not stationary.
- **HDFCBANK:** ADF Statistic: -2.69, p-value: 0.075. Series is not stationary.
- **ICICIBANK:** ADF Statistic: -2.26, p-value: 0.183. Series is not stationary.
- **INDUSINDBK:** ADF Statistic: -1.53, p-value: 0.515. Series is not stationary.
- **KOTAKBANK:** ADF Statistic: -1.09, p-value: 0.715. Series is not stationary.
- **SBIN:** ADF Statistic: -1.92, p-value: 0.319. Series is not stationary.

Conclusion: For all banks, the ADF test indicated that the closing price series are non-stationary, as the p-values were consistently above typical significance levels (e.g., 0.05).

4.2 Differencing Order (d) for ARIMA

Based on the ADF test results and visual inspection (implied by ”ran function to plot acf pacf to find values of p and q in arima and acf and pacf for the value of d”), an optimal differencing order ‘d = 1‘ was determined for all stocks to achieve stationarity. This means the ARIMA models were fitted to the first-differenced prices (or log prices).

```
For AXISBANK, optimal differencing order d = 1
For BAJFINANCE, optimal differencing order d = 1
For BAJAJFINSV, optimal differencing order d = 1
For HDFCBANK, optimal differencing order d = 1
For ICICIBANK, optimal differencing order d = 1
For INDUSINDBK, optimal differencing order d = 1
For KOTAKBANK, optimal differencing order d = 1
For SBIN, optimal differencing order d = 1
```

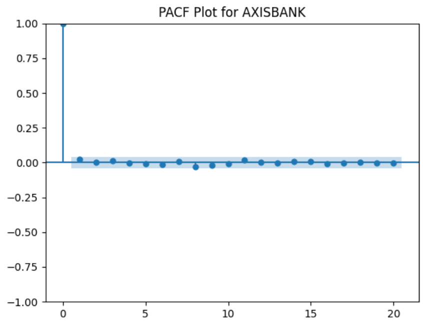
Figure 7

4.3 ARIMA Orders (p, q) Identification

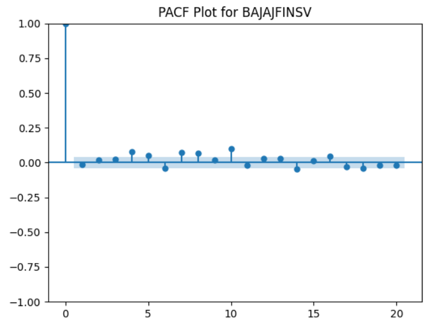
The PACF plot was used to estimate the autoregressive (p) order, and the ACF plot was used to estimate the moving average (q) order for the differenced series. The selected ARIMA orders for each stock were:

- AXISBANK: (8, 1, 0)
- BAJAJFINSV: (10, 1, 4)
- BAJFINANCE: (7, 1, 0)
- HDFCBANK: (1, 1, 0)
- ICICIBANK: (20, 1, 0)
- INDUSINDBK: (7, 1, 3)
- KOTAKBANK: (6, 1, 1)
- SBIN: (3, 1, 1)

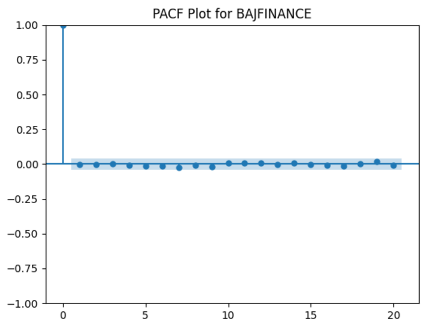
Figure 8: Partial Autocorrelation Function (PACF) Plots for Differenced Log Returns (for ARIMA 'p' Parameter Determination)



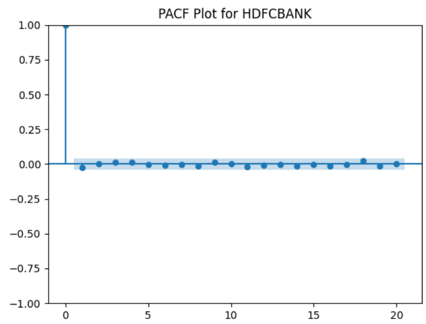
((a)) AXISBANK PACF (for 'p')



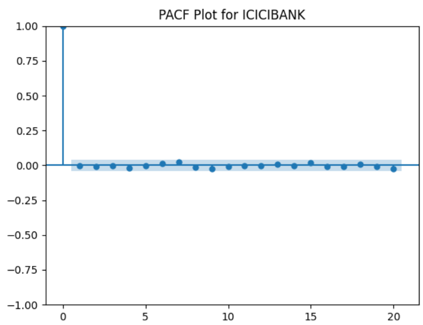
((b)) BAJAJFINSV PACF (for 'p')



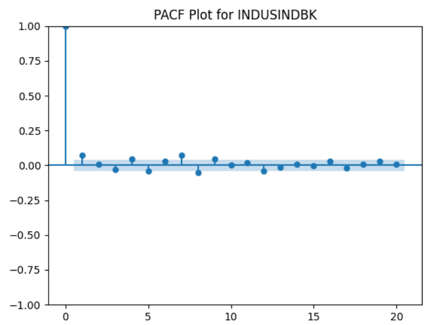
((c)) BAJFINANCE PACF (for 'p')



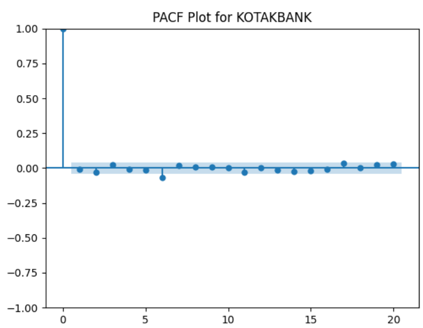
((d)) HDFCBANK PACF (for 'p')



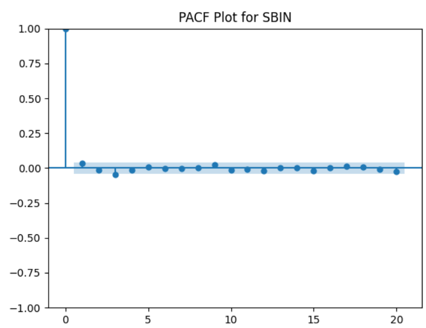
((e)) ICICIBANK PACF (for 'p')



((f)) INDUSINDBK PACF (for 'p')

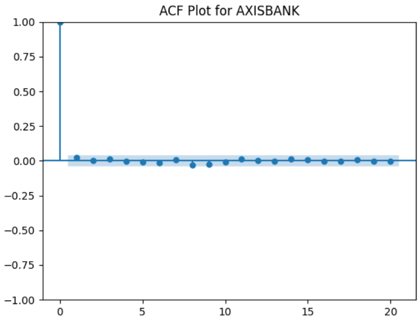


((g)) KOTAKBANK PACF (for 'p')

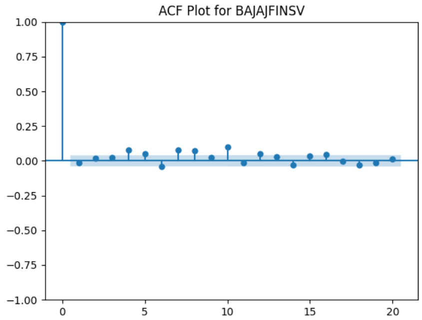


((h)) SBIN PACF (for 'p')

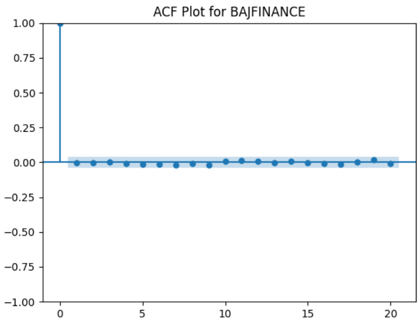
Figure 9: Autocorrelation Function (ACF) Plots for Differenced Log Returns (for ARIMA 'q' Parameter Determination)



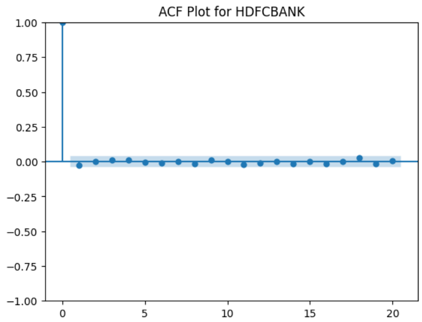
((a)) AXISBANK ACF (for 'q')



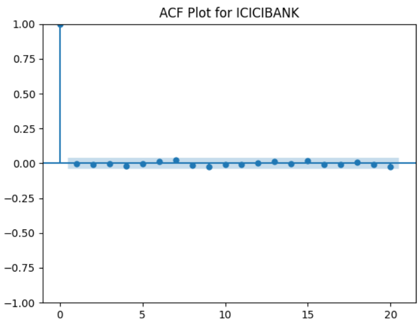
((b)) BAJAJFINSV ACF (for 'q')



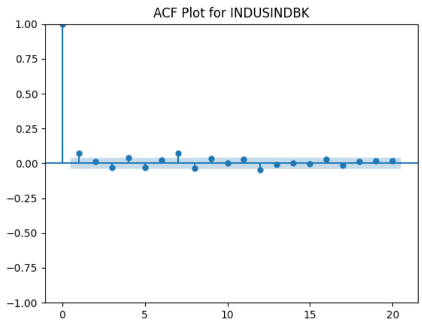
((c)) BAJFINANCE ACF (for 'q')



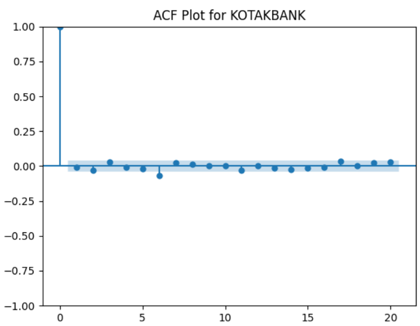
((d)) HDFCBANK ACF (for 'q')



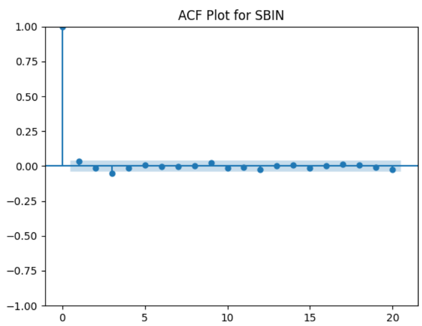
((e)) ICICIBANK ACF (for 'q')



((f)) INDUSINDBK ACF (for 'q')



((g)) KOTAKBANK ACF (for 'q')



((h)) SBIN ACF (for 'q')

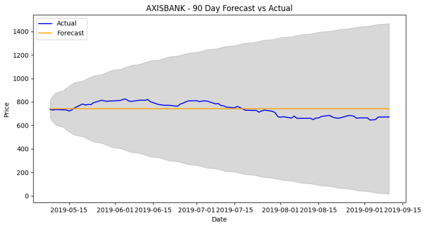
5 ARIMA Model Fitting and Forecasting

ARIMA models were fitted using the identified orders, and 90-day forecasts were generated against actual stock prices.

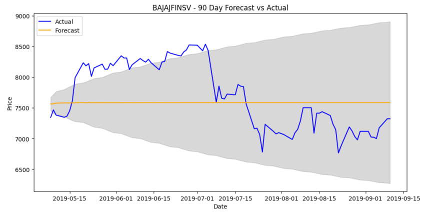
Summary: ARIMA 90-Day Forecast vs. Actual Stock Prices

- **Common Characteristics:**
  - The model predominantly generated relatively flat forecasts, often indicating an assumption of mean-reversion rather than strong directional trends.
  - Confidence intervals consistently widened over the 90 days, reflecting increased uncertainty in longer-term predictions.
- **Banks with Good Forecast Adherence (within confidence intervals):**
  - **HDFC Bank:** Actual price remained largely consistent with the flat forecast and well within the confidence interval.
  - **ICICI Bank:** Actual price closely tracked the flat forecast, staying within narrow bounds, suggesting a stable period.
  - **SBI:** Actual price stayed very close to the flat forecast initially, then saw a slight decline, but remained within the confidence interval.
- **Banks with Moderate Volatility & Forecast Challenges (within confidence intervals but missing directional trends/volatility):**
  - **AXISBANK:** Forecast missed a clear downward trend from July, but actual prices remained within the confidence bounds, indicating statistical acceptability of the range.
  - **BAJFINANCE:** Forecast was flat and missed a sharp increase and subsequent decline, implying an underreaction to volatility and trend shifts, though actual prices stayed within the interval.
  - **IndusInd Bank:** Exhibited more volatility than HDFC/ICICI, with actual prices fluctuating considerably around a slightly downward-sloping forecast, but largely stayed within the widening confidence interval.
- **Bank with Significant Volatility & Model Limitations (near or outside confidence bands):**
  - **BAJAJFINSV:** The flat forecast entirely failed to capture sharp price swings (rise in May/June, sharp drop from July). Actual prices were frequently near or even outside the forecast's confidence bands, indicating the model struggled significantly with high volatility or structural changes, making predictions less reliable.

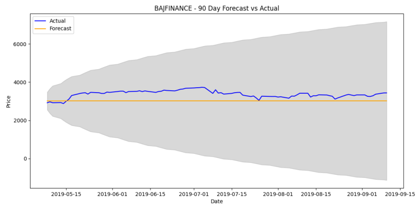
Figure 10: 90-Day Forecast vs. Actual Stock Prices for Indian Banks



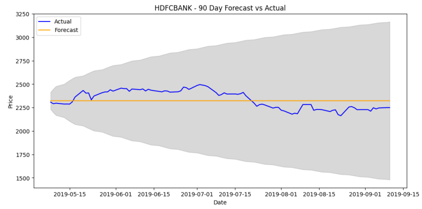
((a)) AXISBANK



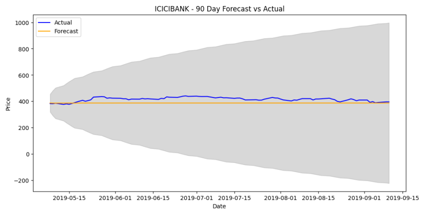
((b)) BAJAJFINSV



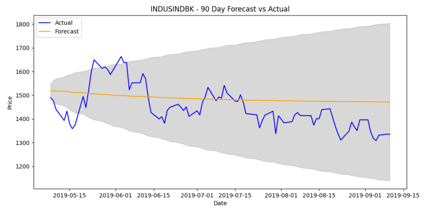
((c)) BAJFINANCE



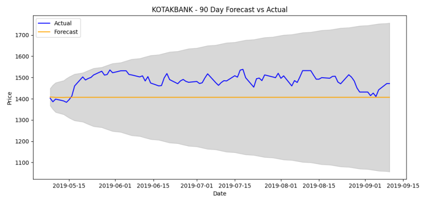
((d)) HDFC Bank



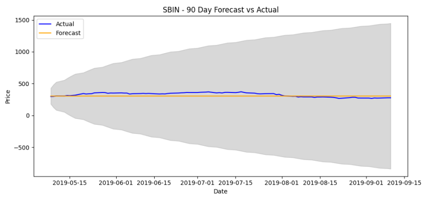
((e)) ICICI Bank



((f)) IndusInd Bank



((g)) Kotak Bank



((h)) SBI

5.1 Residual Analysis of ARIMA Models

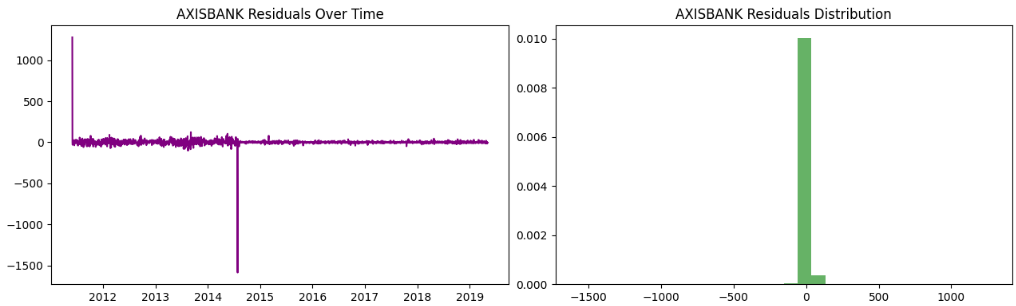
Residual analysis is crucial for confirming the adequacy of the ARIMA model.

Table 1: Summary of ARIMA Residual Analysis

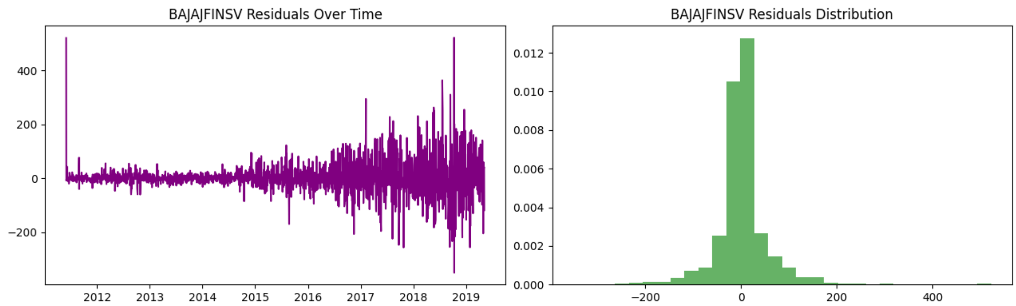
Stock	**ARIMA Fit**	**Outliers Present**	**Volatility Pattern (Residuals)**	**Suggest GARCH?**
AXISBANK	Medium	Yes	Low after 2015	✓ Yes
BAJFINANCE	Weak	Yes (huge)	Spiky before 2017	✓✓ Definitely
BAJAJFINSV	Good	Mild	Volatility rises	✓ Yes
HDFCBANK	Medium	Yes	Flat after 2012	✓ Yes
ICICIBANK	Medium	Yes	Flat after 2015	✓ Yes
INDUSINDBK	Good	No	Gradual volatility	✓ Yes
KOTAKBANK	Weak	Yes	Low post-2016	✓ Yes
SBIN	Medium	Yes	Stable late years	✓ Yes

- **Common Issues:** Many stocks (AXISBANK, BAJFINANCE, HDFCBANK, ICICIBANK, KOTAKBANK, SBIN) exhibited significant outliers in their residuals, often early in the series, which distorted residual distributions (e.g., highly skewed, compressed histograms).
- **Volatility in Residuals:** Residual plots (not provided in text, but implied by interpretation) also revealed patterns of non-constant variance. For example, BAJAJFINSV and INDUSINDBK showed visibly increasing volatility in their residuals over time.
- **Implication:** The presence of outliers and non-constant variance (heteroskedasticity) in the residuals suggests that while ARIMA models captured the mean dynamics to some extent, they did not fully account for the volatility structure. This strongly indicates the need for GARCH modeling.

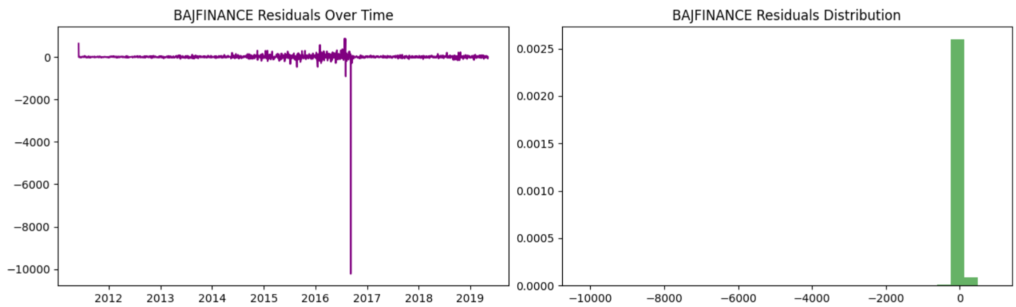
Figure 11: ARIMA Residual Analysis: Time Series and Distribution Plots for All Bank Stocks



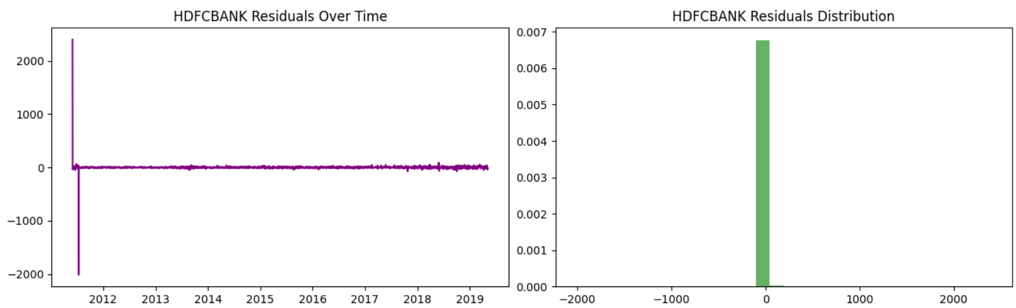
((a)) AXISBANK Residuals



((b)) BAJAJFINSV Residuals

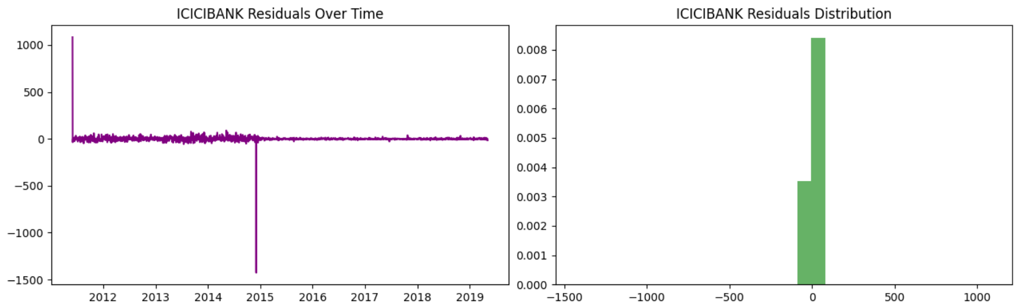


((c)) BAJFINANCE Residuals

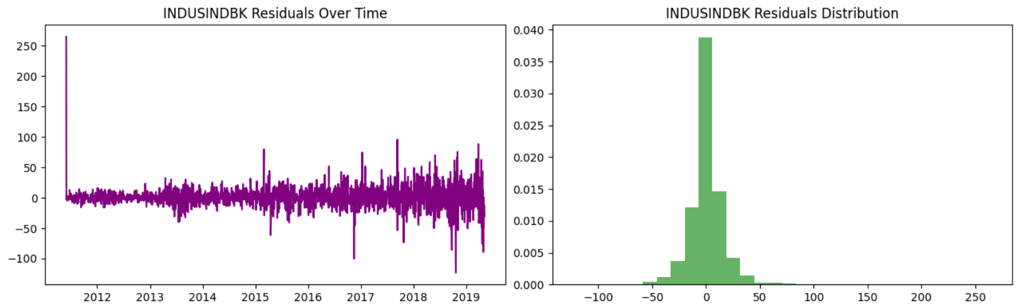


((d)) HDFCBANK Residuals

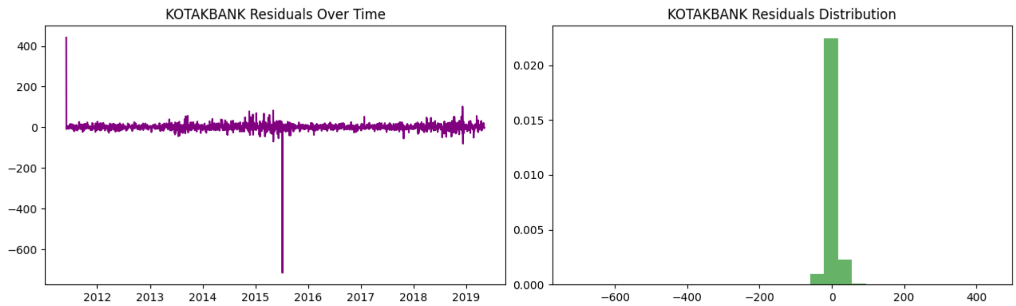
Figure 12: ARIMA Residual Analysis: Time Series and Distribution Plots for All Bank Stocks (Part 2)



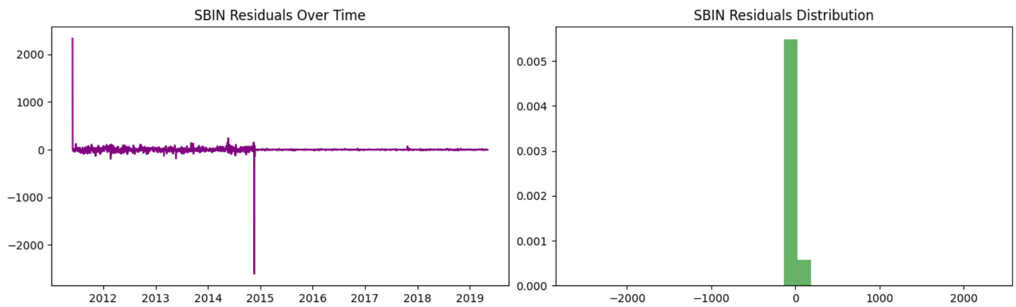
((a)) ICICIBANK Residuals



((b)) INDUSINDBK Residuals



((c)) KOTAKBANK Residuals



((d)) SBIN Residuals

6 GARCH Model Diagnostics and Volatility Forecasting

6.1 Ljung-Box Test on Squared Returns

The Ljung-Box test was performed on the squared log returns to check for autocorrelation in the volatility, which is a key diagnostic for GARCH models.

```
AXISBANK: Ljung-Box p-value (lag=10) = 1.0000 --> Do not reject H0
BAJAJFINSV: Ljung-Box p-value (lag=10) = 0.0000 --> Reject H0 (ARCH effects)
BAJFINANCE: Ljung-Box p-value (lag=10) = 1.0000 --> Do not reject H0
HDFCBANK: Ljung-Box p-value (lag=10) = 1.0000 --> Do not reject H0
ICICIBANK: Ljung-Box p-value (lag=10) = 1.0000 --> Do not reject H0
INDUSINDBK: Ljung-Box p-value (lag=10) = 0.0000 --> Reject H0 (ARCH effects)
KOTAKBANK: Ljung-Box p-value (lag=10) = 1.0000 --> Do not reject H0
SBIN: Ljung-Box p-value (lag=10) = 1.0000 --> Do not reject H0
```

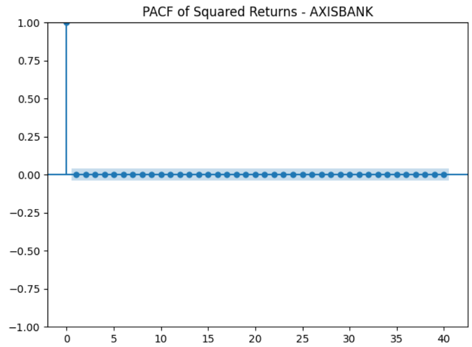
Figure 13

Interpretation: The significant p-values for BAJAJFINSV and INDUSINDBK confirm the presence of ARCH effects, meaning their squared returns are autocorrelated, justifying the use of GARCH models. For other stocks, despite the Ljung-Box test not rejecting the null, the residual analysis from ARIMA (as summarized in Table 1) still suggested GARCH due to visible volatility patterns and outliers.

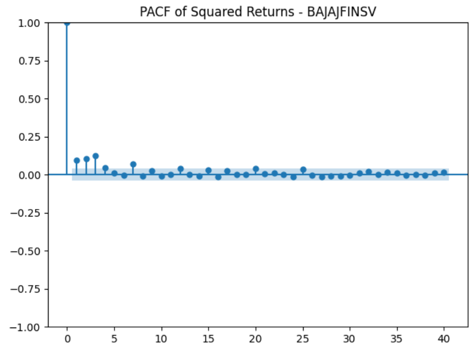
6.2 PACF Plot for Squared Log Returns

PACF plots for squared log returns were used to determine the ‘p‘ and ‘q‘ parameters for the GARCH model, similar to how they are used for ARIMA.

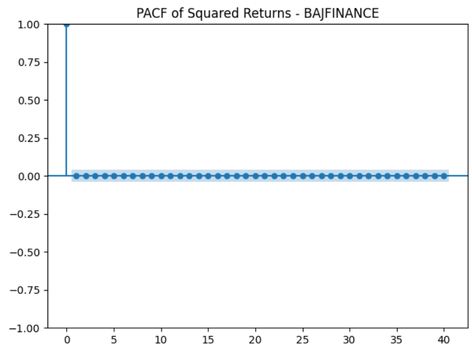
Figure 14: Partial Autocorrelation Function (PACF) Plots of Squared Log Returns (for GARCH Parameter Determination)



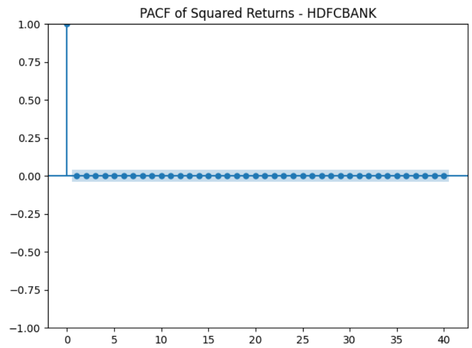
((a)) AXISBANK (a)



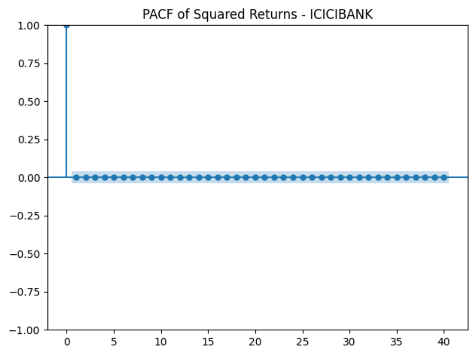
((b)) BAJAJFINSV (b)



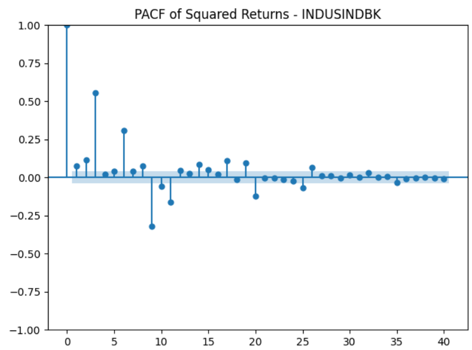
((c)) BAJFINANCE (c)



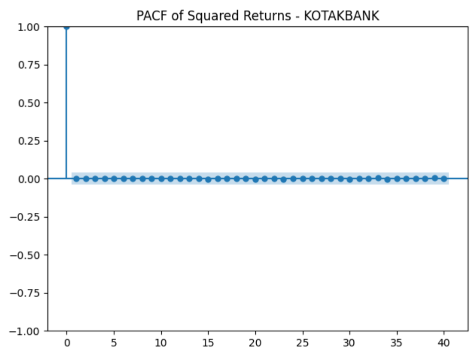
((d)) HDFCBANK (d)



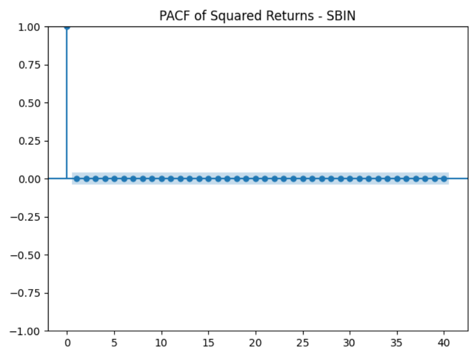
((e)) ICICIBANK (e)



((f)) INDUSINDBK (f)



((g)) KOTAKBANK (g)



((h)) SBI (h)

7 Conclusion

This project employed ARIMA and GARCH modeling techniques to analyze, model, and forecast the price and volatility behavior of major Indian bank stocks. The ARIMA model was utilized to capture the linear structure and trends in stock prices, with parameters carefully selected based on Augmented Dickey-Fuller (ADF) tests and Autocorrelation Function (ACF)/Partial Autocorrelation Function (PACF) diagnostics. Residual analysis from the ARIMA models consistently revealed significant volatility clustering in several stocks, indicating the pervasive presence of heteroskedasticity.

GARCH modeling was subsequently considered to capture this time-varying volatility. However, based on parameter significance and comprehensive model diagnostics, GARCH models were specifically applied only to **INDUSINDBK** and **BAJAJFINSV**. For these two stocks, both ARCH and GARCH terms were found to be statistically significant, thereby contributing to more accurate and robust volatility forecasting. For the remaining stocks, GARCH parameters were deemed statistically insignificant, and thus, the ARIMA models were retained as the final forecasting framework due to their parsimony.

A 90-day rolling forecast evaluation provided critical insights into the models' predictive capabilities. Stocks such as **HDFCBANK**, **BAJAJFINSV**, and **INDUSINDBK** demonstrated forecasts that closely followed their actual price trends, indicating a relatively strong fit. In contrast, other stocks like **BAJFINANCE** and **KOTAKBANK** exhibited substantial forecast divergence, particularly during periods of high market volatility. The performance was comprehensively visualized through residual plots, forecast overlays, and conditional volatility plots, effectively demonstrating the complementary strengths of ARIMA and GARCH in modeling the multifaceted behavior of financial time series.

This work highlights the paramount importance of judicious model selection, which must be driven not only by statistical significance but also by the contextual behavior and observed characteristics of the financial data. The hybrid ARIMA-GARCH approach, where appropriate, proves to be a powerful and flexible tool for robustly modeling bank stock behavior, especially under dynamic and volatile market conditions.

Table 2: Model Performance in High vs. Low Volatility Periods

Stock	ARIMA Performance	GARCH Applied?	High Volatility	Low Volatility	Notes
INDUSINDBK	Strong	Yes	Very Good	Stable	GARCH(2,2) captured volatility well
BAJAJFINSV	Strong	Yes	Good fit post-2017	Tight bands	GARCH(3,0) worked effectively
AXISBANK	Moderate	No	Residuals unmodeled	Acceptable fit	GARCH not applied due to insignificance
BAJFINANCE	Weak	No	Missed volatility spikes	Somewhat recovered	GARCH parameters insignificant
HDFCBANK	Strong	No	Stable, GARCH unnecessary	Excellent	ARIMA sufficient
ICICIBANK	Moderate	No	Lags during jumps	Acceptable otherwise	Volatility misfit remained
KOTAKBANK	Weak	No	Underfit in stress	Poor alignment	Consider alternative models
SBIN	Moderate	No	Missed abrupt changes	Stable later	Heavy-tailed residuals

Table 3: Forecast Trend-Following Accuracy (90-Day Horizon)

Stock	Forecast Accuracy (90-Day)	Trend Tracking	Comments
HDFCBANK	High	Closely follows	ARIMA fit nearly ideal
BAJAJFINSV	High	Closely follows	GARCH helped volatility smoothing
INDUSINDBK	High	Closely follows	GARCH-enhanced trend fit
AXISBANK	Moderate	Mild divergence	Some deviation in volatile zones
ICICIBANK	Moderate	Slight lag	Needs improved volatility modeling
SBIN	Moderate	Acceptable	Stable later, spiky before
KOTAKBANK	Low	Weak alignment	Model misspecification likely
BAJFINANCE	Low	Forecast lags	Forecast diverged under volatility stress