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.

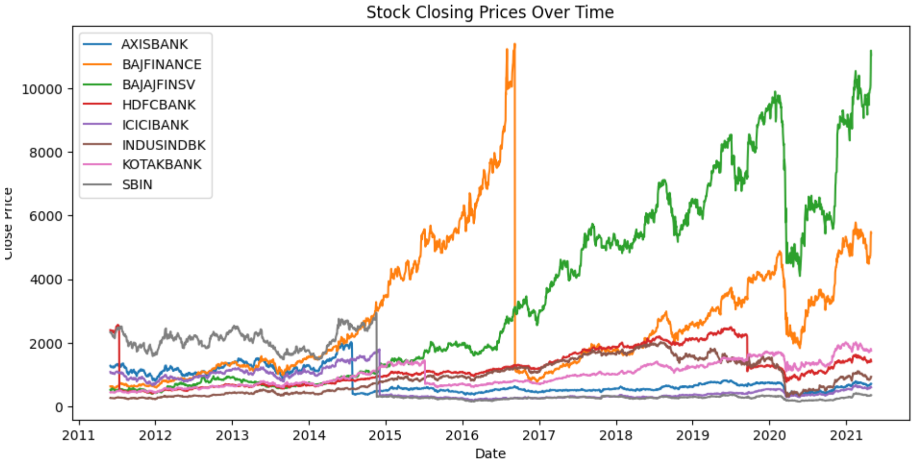


Figure 1

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.

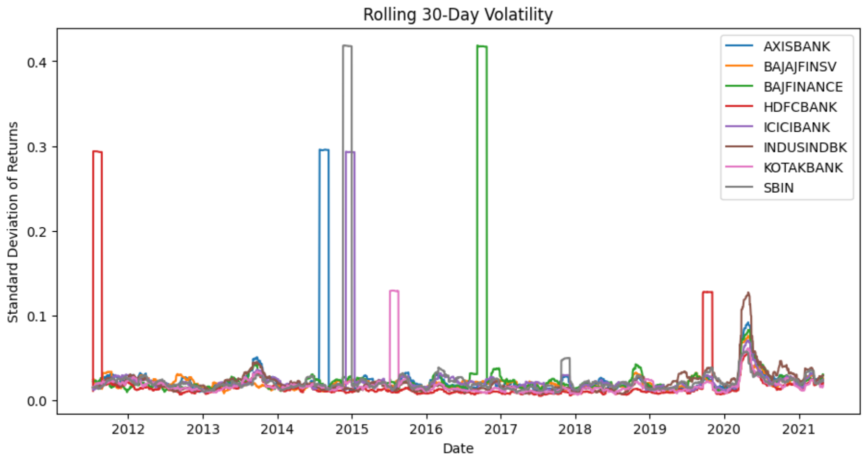


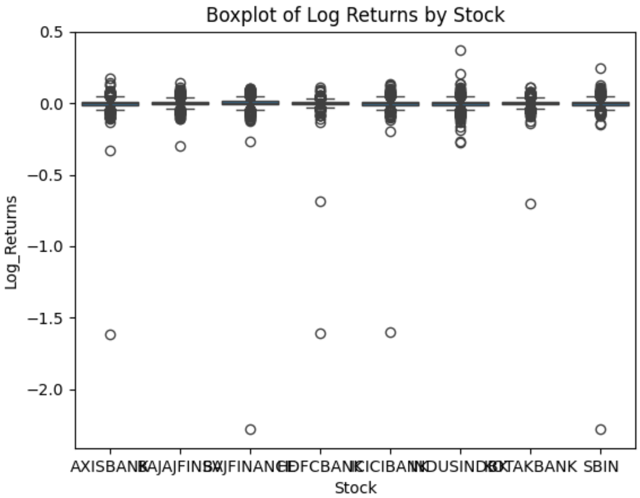
Figure 2

3.3 Distribution of Log Returns (Boxplots and Histograms)

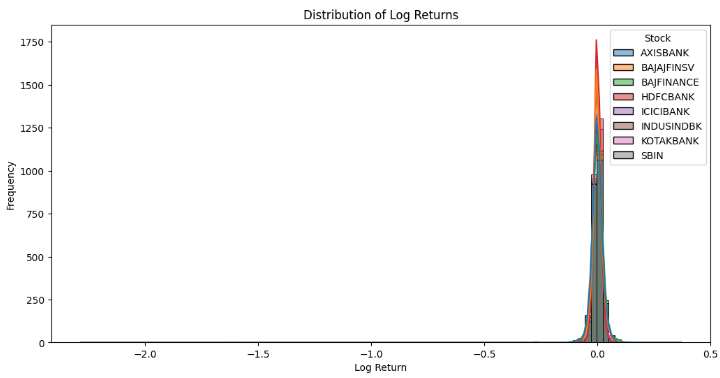
- **Observation:**
  - All banks' log returns are generally centered around 0, which is typical for daily financial returns, often being small.
  - The distributions exhibit 'fat tails', indicating a higher probability of extreme events (large positive or negative returns) compared to a normal distribution.
  - Most stocks show fairly symmetric return distributions around the median.
  - 'Check for Outliers' 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 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.

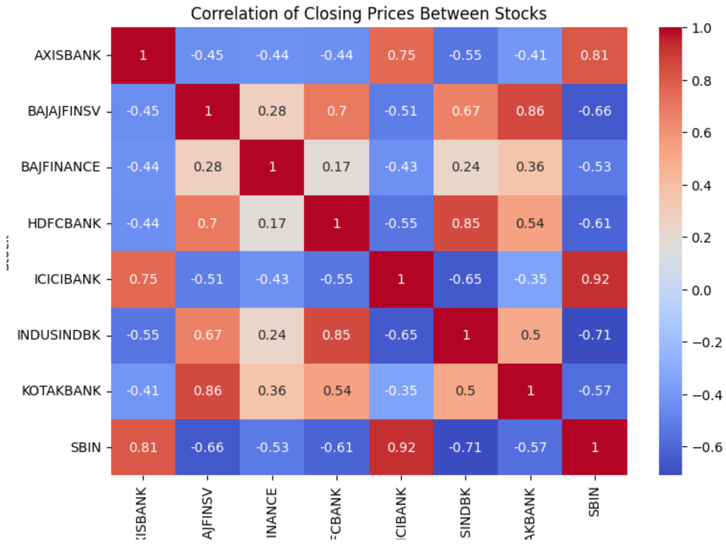
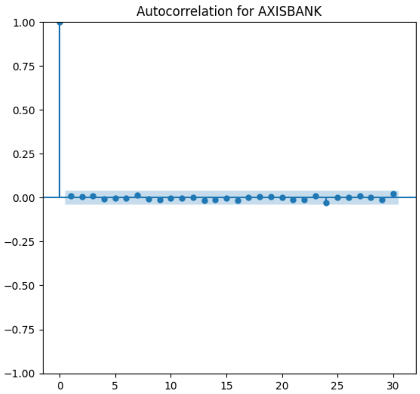


Figure 4

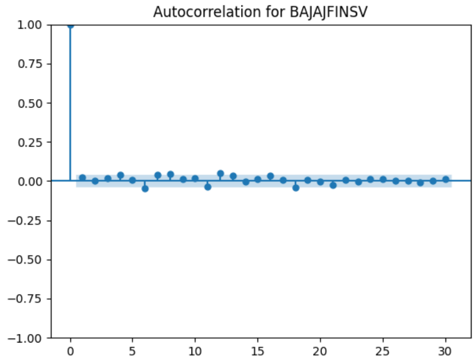
3.5 Autocorrelation (ACF) and Partial Autocorrelation (PACF)

- 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.

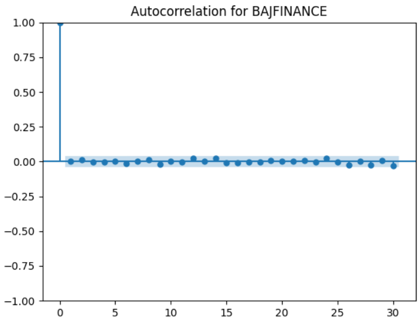
Figure 5: Autocorrelation Function (ACF) Plots for Bank Stock Log Returns



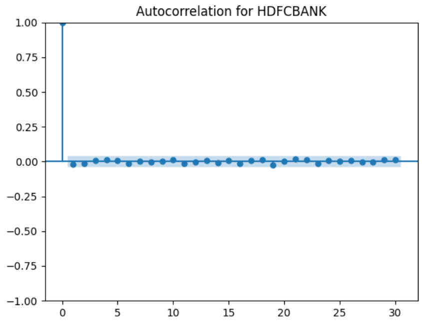
((a)) AXISBANK ACF



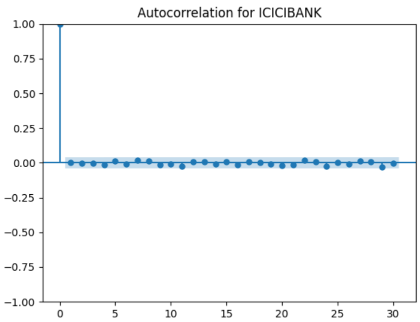
((b)) BAJAJFINSV ACF



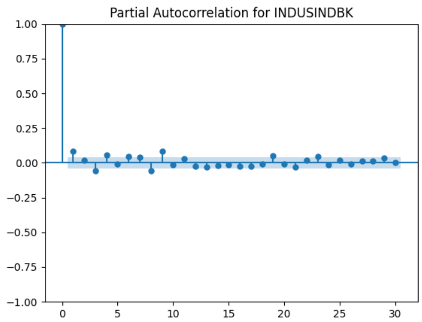
((c)) BAJFINANCE ACF



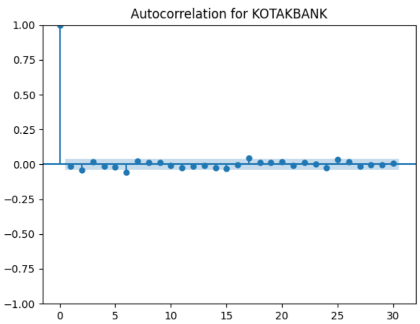
((d)) HDFCBANK ACF



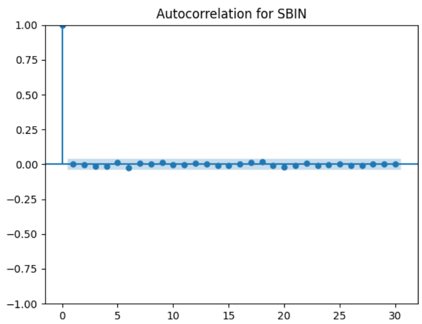
((e)) ICICIBANK ACF



((f)) INDUSINDBK ACF

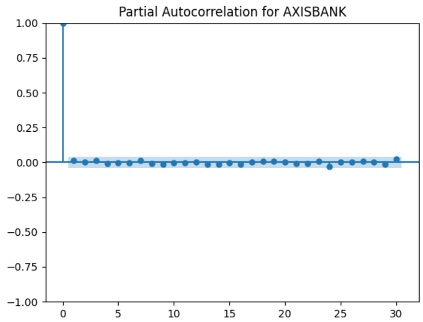


((g)) KOTAKBANK ACF

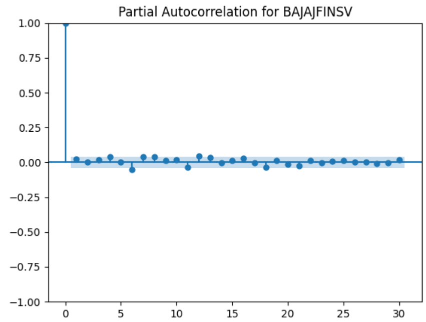


((h)) SBIN ACF

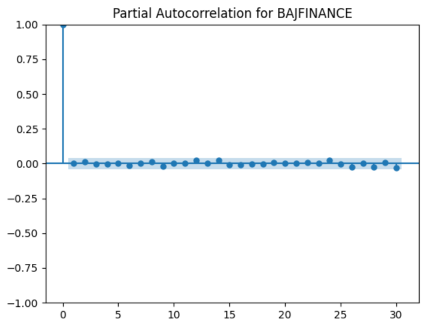
Figure 6: Partial Autocorrelation Function (PACF) Plots for Bank Stock Log Returns



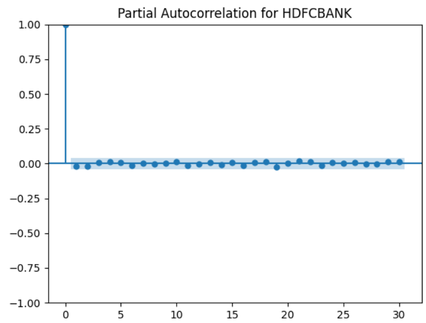
((a)) AXISBANK PACF



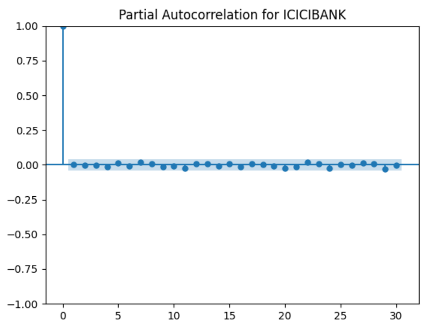
((b)) BAJAJFINSV PACF



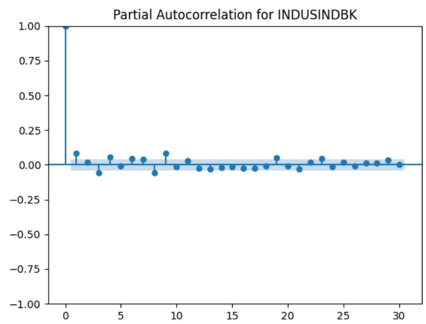
((c)) BAJFINANCE PACF



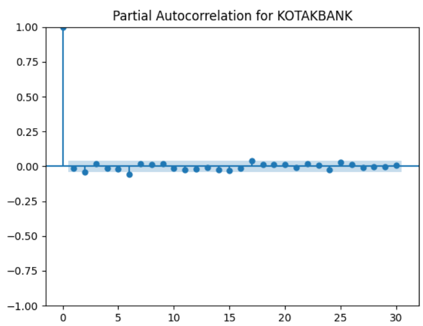
((d)) HDFCBANK PACF



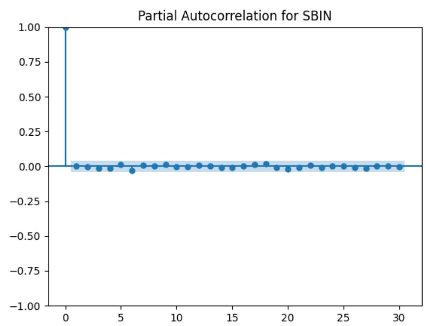
((e)) ICICIBANK PACF



((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).

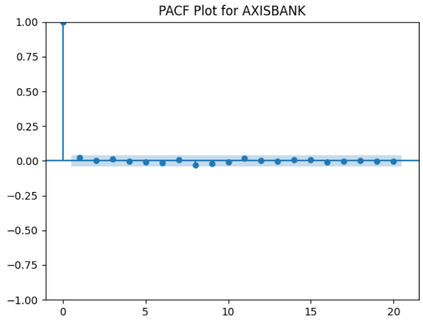
```
For AXISBANK, optimal differencing order d = 1
For BAJFINANCE, optimal differencing order d = 1
For BAJAJFINSV, optimal differencing order d = 1
For HDFCBANK, optimal differencing order d = 1
For ICICIBANK, optimal differencing order d = 1
For INDUSINDBK, optimal differencing order d = 1
For KOTAKBANK,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

Figure 7

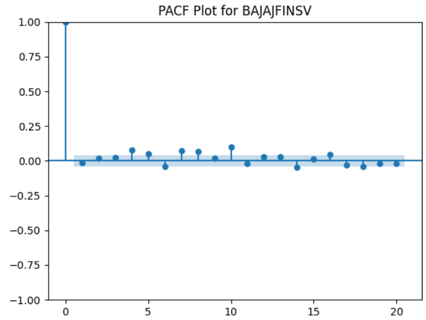
4.3 ARIMA Orders (p, q) Identification

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.

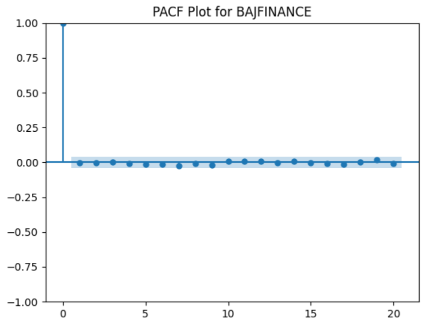
Figure 8: Partial Autocorrelation Function (PACF) Plots for Differenced Log Returns (for ARIMA 'p' Parameter Determination)



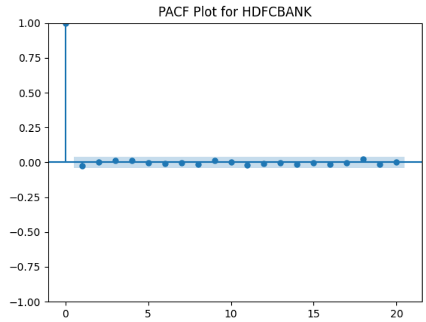
((a)) AXISBANK PACF (for 'p')



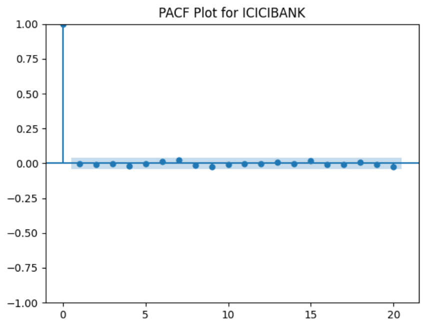
((b)) BAJAJFINSV PACF (for 'p')



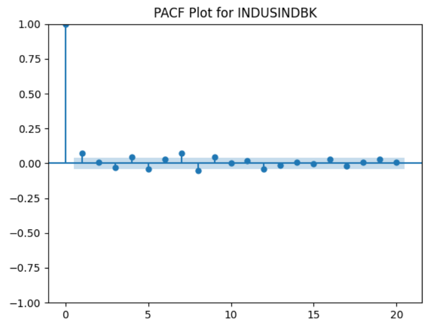
((c)) BAJFINANCE PACF (for 'p')



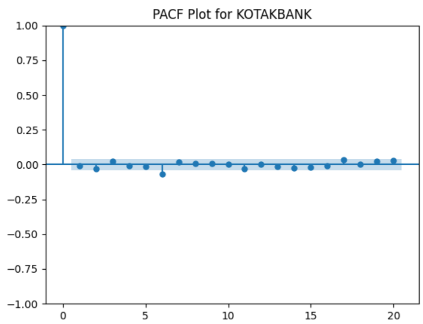
((d)) HDFCBANK PACF (for 'p')



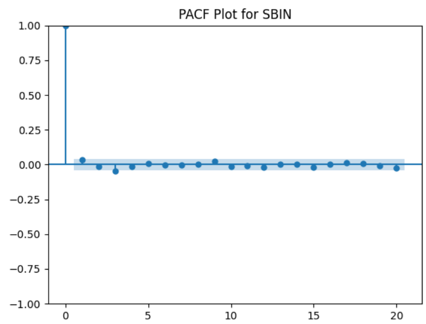
((e)) ICICIBANK PACF (for 'p')



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

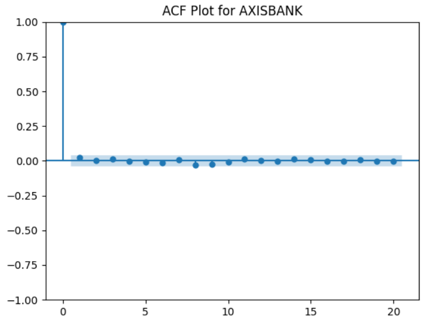


((g)) KOTAKBANK PACF (for 'p')

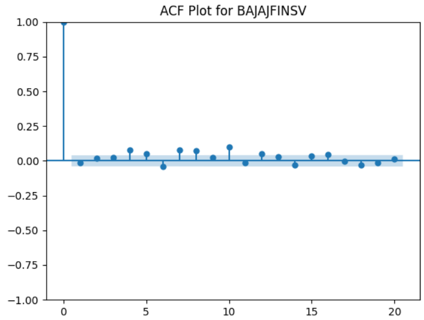


((h)) SBIN PACF (for 'p')

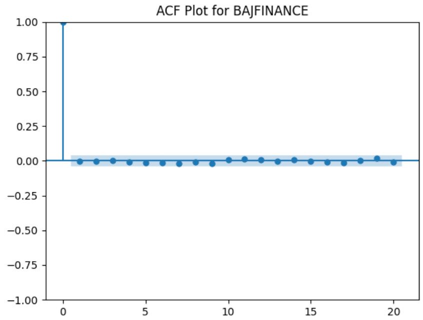
Figure 9: Autocorrelation Function (ACF) Plots for Differenced Log Returns (for ARIMA 'q' Parameter Determination)



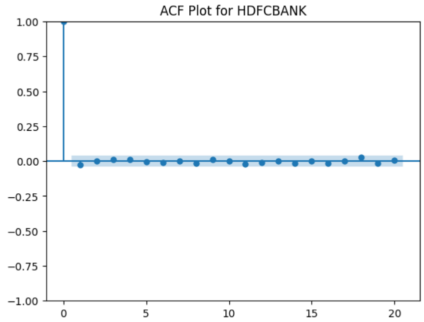
((a)) AXISBANK ACF (for 'q')



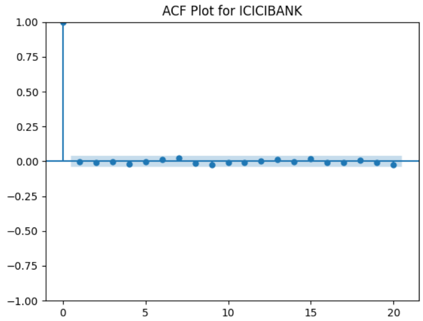
((b)) BAJAJFINSV ACF (for 'q')



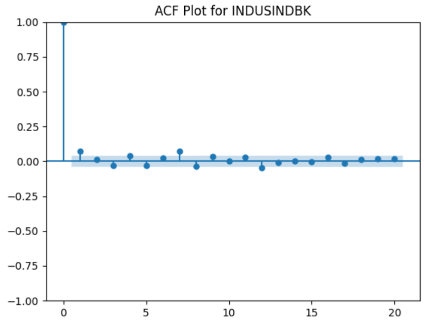
((c)) BAJFINANCE ACF (for 'q')



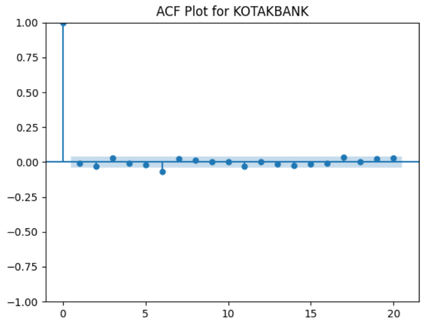
((d)) HDFCBANK ACF (for 'q')



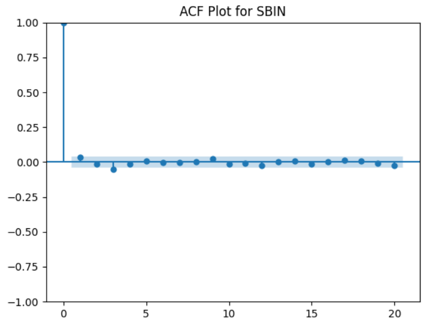
((e)) ICICIBANK ACF (for 'q')



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



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



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

5 ARIMA Model Fitting and Forecasting

5.1 Confirmation of ARIMA Orders using AIC Optimization

The high-order ARMA models initially suggested by visual inspection of the ACF and PACF plots (e.g., (10,1,10) for BAJAJFINSV) were tested against simpler candidates using the Akaike Information Criterion (AIC). The AIC penalizes model complexity, favoring simpler models that explain the data nearly as well (Principle of Parsimony).

The optimization confirmed that simpler models provide a better fit, minimizing the AIC.

5.2 Optimization Results for Ambiguous Stocks

The following results were obtained by running the pmdarima.autoarima function, fixing the differencing order at d=1 and searching for the optimal p and q up to 5:

```
Searching optimal ARIMA(p, 1, q) for BAJAJFINSV using AIC...
☑ Optimal Order for BAJAJFINSV: ARIMA(0, 1, 0) (AIC: 31305.09)

Searching optimal ARIMA(p, 1, q) for INDUSINDBK using AIC...
☑ Optimal Order for INDUSINDBK: ARIMA(1, 1, 0) (AIC: 22928.66)

Searching optimal ARIMA(p, 1, q) for KOTAKBANK using AIC...
☑ Optimal Order for KOTAKBANK: ARIMA(0, 1, 0) (AIC: 23752.14)

--- Final ARIMA Orders (Optimized) ---
{'AXISBANK': (0, 1, 0), 'BAJFINANCE': (0, 1, 0), 'HDFCBANK': (0, 1, 0), 'ICICIBANK': (0, 1, 0), 'SBIN': (0, 1, 0), 'BAJAJFINSV': (0, 1, 0), 'INDUSINDBK':
(1, 1, 0), 'KOTAKBANK': (0, 1, 0)}
```

Figure 10

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

Based on the ACF, PACF analysis, and AIC optimization, the ARIMA orders for the differenced stock price series are highly parsimonious. The differencing order is fixed at $d = 1$ for all series to ensure stationarity.

Table 1: Assigned ARIMA Orders for Bank Stocks

Stock	**ARIMA Order** (p, d, q)
AXISBANK	$(0, 1, 0)$
BAJFINANCE	$(0, 1, 0)$
BAJAJFINSV	$(0, 1, 0)$
HDFCBANK	$(0, 1, 0)$
ICICIBANK	$(0, 1, 0)$
KOTAKBANK	$(0, 1, 0)$
SBIN	$(0, 1, 0)$
INDUSINDBK	$(1, 1, 0)$

6 Interpretation of Forecast Plots

All forecast plots exhibit the characteristics of Random Walk models (ARIMA$(0, 1, 0)$ and ARIMA$(1, 1, 0)$), which are typical for efficient financial markets.

6.1 Key Pointers for ARIMA Forecast Visuals

- **The Forecast Line (Orange Line):**
  - **Pattern:** It is predominantly **flat (horizontal)** for all stocks across the 90-day horizon.
  - **Meaning:** This indicates the model predicts the future price will equal the **last observed price** ($Y_{t+k} \approx Y_t$). The differenced series (ΔY_t) is White Noise, meaning no linear pattern is found in the returns.

- **The Confidence Interval (Grey Shaded Area):**
  - **Pattern:** It **widens significantly** (diverges) over time.
  - **Meaning:** This represents the **accumulating uncertainty** in the forecast.
  - **Reason:** In Random Walk-type models, the forecast error variance grows linearly with the forecast step (k). This visually confirms that ARIMA models have no reliable long-term predictive power for price levels.

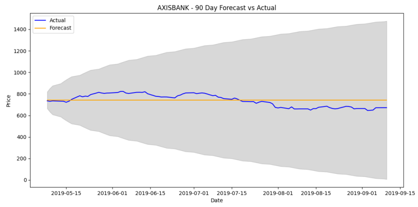
7 Stock-Specific Forecast Performance

The table below summarizes the actual price behavior relative to the flat forecast, highlighting the degree of volatility and error inherent in the Random Walk assumption.

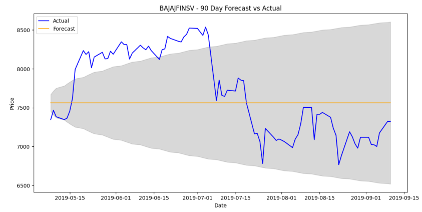
Table 2: Stock-Specific Forecast Performance Based on Plots

Stock	**ARIMA Order**	**Actual Price Movement (vs. Forecast)**
AXISBANK	$(0, 1, 0)$	Price **dropped significantly**, failing to stay near the flat forecast.
BAJFINANCE	$(0, 1, 0)$	Moderate fluctuation, price **remained largely range-bound** near the forecast.
BAJAJFINSV	$(0, 1, 0)$	Exhibits **high volatility** with large, sustained swings in both direction.
HDFCBANK	$(0, 1, 0)$	Moderate fluctuations, maintaining proximity to the forecast.
ICICIBANK	$(0, 1, 0)$	**Very low volatility**; actual price closely tracks the flat forecast.
KOTAKBANK	$(0, 1, 0)$	Pronounced fluctuation, with multiple crossings of the forecast line.
SBIN	$(0, 1, 0)$	Initial stability followed by a slight, sustained drop, which the model missed.
INDUSINDBK	$(1, 1, 0)$	Exhibits the **largest swings** and an overall decline, indicates highest volatility.

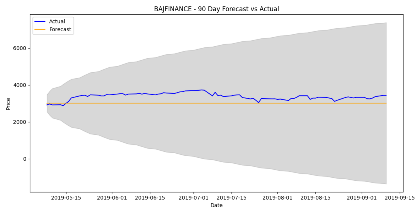
Figure 11: 90-Day Forecast vs. Actual Stock Prices for Indian Banks



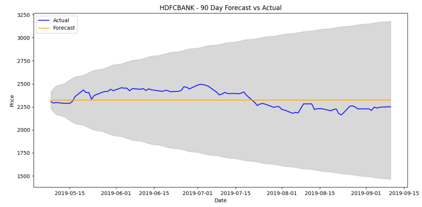
((a)) AXISBANK



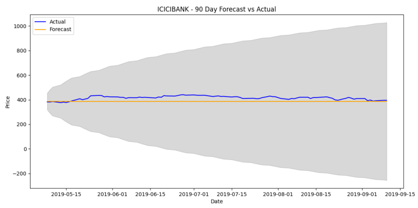
((b)) BAJAJFINSV



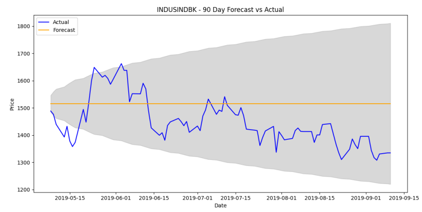
((c)) BAJFINANCE



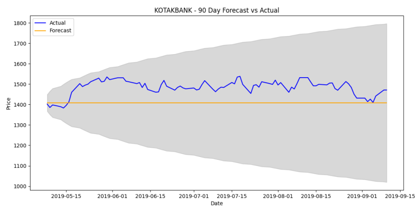
((d)) HDFC Bank



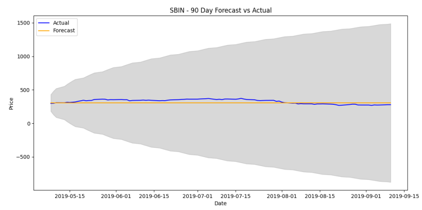
((e)) ICICI Bank



((f)) IndusInd Bank



((g)) Kotak Bank



((h)) SBI

7.1 Residual Analysis of ARIMA Models

The residuals ϵ_k of the fitted ARIMA models represent the error left after modeling the mean price. For a successful ARIMA model, these residuals should be independent and identically distributed (i.i.d.) White Noise (i.e., normally distributed with constant variance).

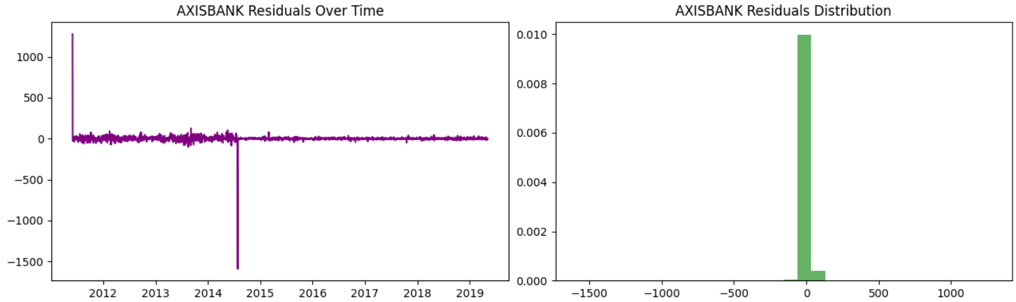
7.2 Key Observations from Residual Plots

- **Residuals Over Time (Time Series Plot):**
  - **Initial Period Spikes (AXISBANK, BAJFINANCE, ICICIBANK, SBIN, KOTAKBANK, HDFCBANK):** The residuals for many stocks show one or more extreme initial spikes, often representing a **structural break** or **outlier event** that the ARIMA model fails to absorb.
  - **Non-Constant Variance (BAJAJFINSV, INDUSINDBK):** For these stocks, the magnitude of the residuals clearly **increases over time**, particularly after 2016-2017 (e.g., BAJAJFINSV). This is known as Heteroskedasticity, indicating that the simple ARIMA model is incomplete.
- **Residuals Distribution (Histogram):**
  - **Leptokurtosis / Heavy Tails (All Stocks):** The histograms for all stocks show a central peak that is much higher than a normal distribution, with long, thin tails (see BAJAJFINSV, INDUSINDBK). This phenomenon, known as **Leptokurtosis**, is typical of financial returns and confirms the residuals are **not normally distributed**.

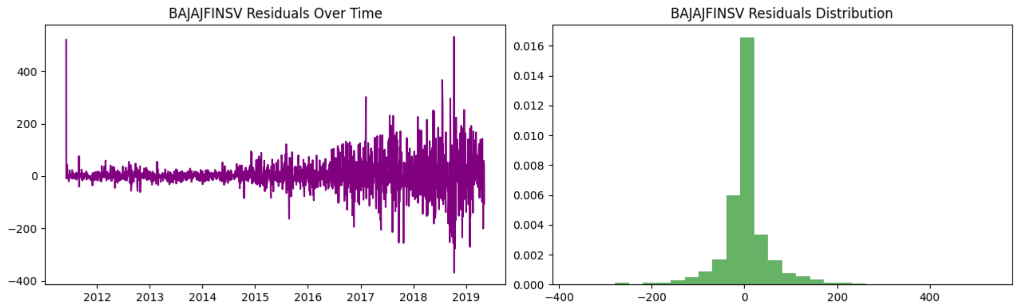
7.3 Conclusion and Next Step

- The ARIMA models successfully made the series stationary (as implied by the lack of trend in the residuals), but they **failed the White Noise assumption** due to **heteroskedasticity** (non-constant variance) and **non-normality** (leptokurtosis).
- Implication: The squared residuals still contain valuable, exploitable information about the volatility. This confirms the need to move beyond ARIMA's focus on the mean and apply a GARCH family model to the residuals to capture the time-varying variance.

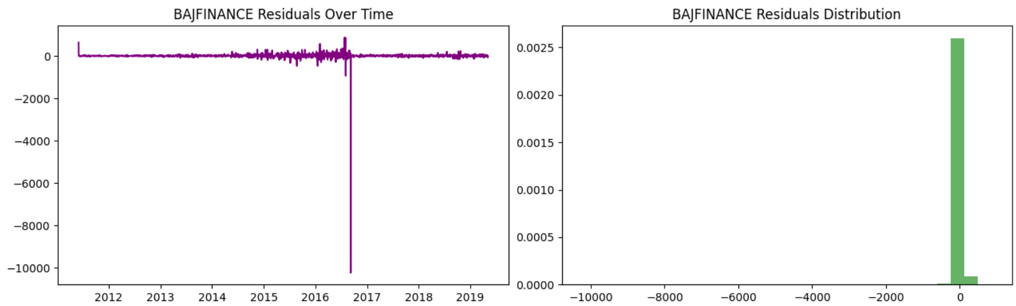
Figure 12: ARIMA Residual Analysis: Time Series and Distribution Plots for All Bank Stocks



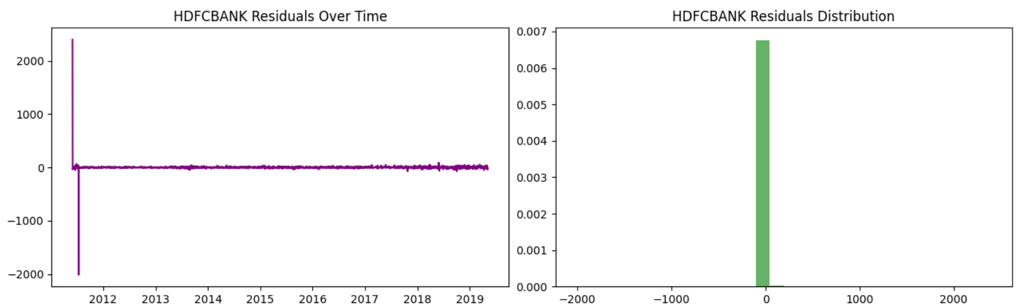
((a)) AXISBANK Residuals



((b)) BAJAJFINSV Residuals

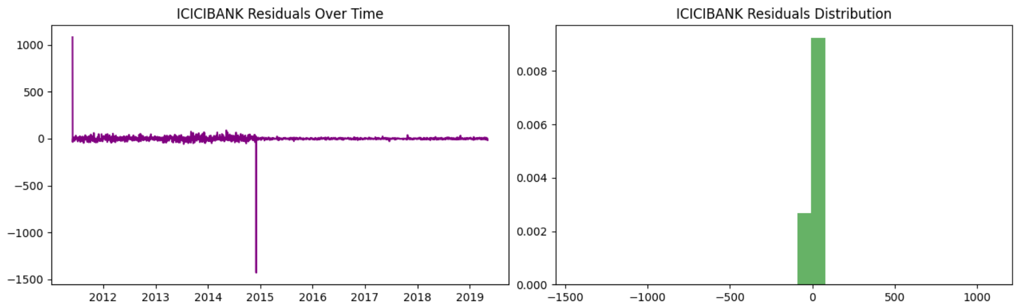


((c)) BAJFINANCE Residuals

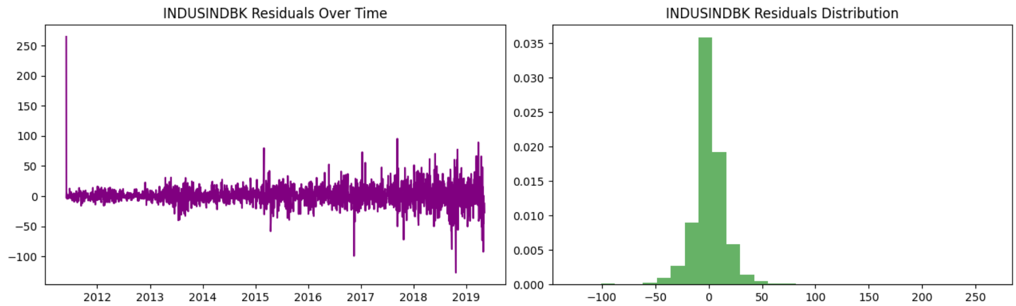


((d)) HDFCBANK Residuals

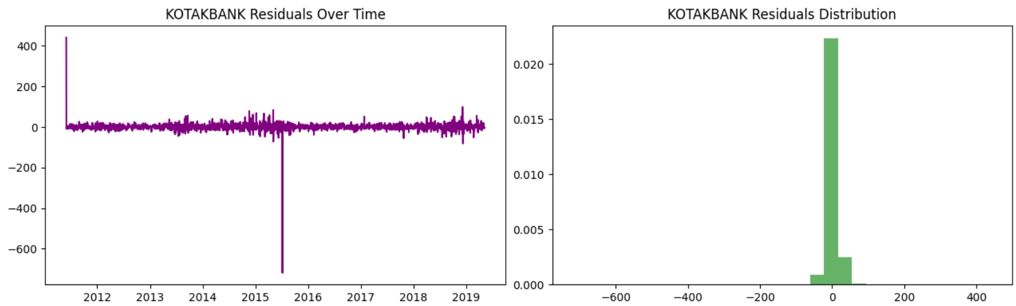
Figure 13: ARIMA Residual Analysis: Time Series and Distribution Plots for All Bank Stocks (Part 2)



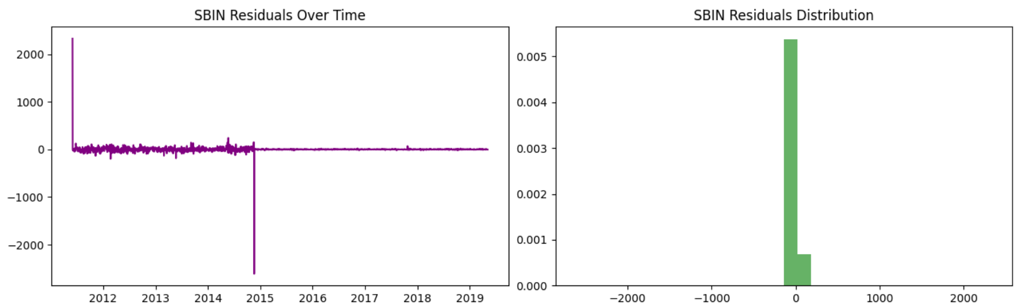
((a)) ICICIBANK Residuals



((b)) INDUSINDBK Residuals



((c)) KOTAKBANK Residuals



((d)) SBIN Residuals

7.4 Quantitative Performance Metrics and Strategy Backtest

To rigorously assess the predictive capabilities and practical utility of the ARIMA models, we evaluated their performance using standard forecasting error metrics (MAPE, RMSE) and a simplified backtested trading strategy over the 90-day forecast horizon for each stock. The results are summarized in the figures below.

```
Running post-analysis for MAPE, RMSE, and Strategy Backtest...

AXISBANK:
MAPE: 7.14% | RMSE: 58.18
Strategy Return: 0.00 | Passive Return: -63.95

BAJAJFINSV:
MAPE: 6.27% | RMSE: 552.29
Strategy Return: 0.00 | Passive Return: -21.80

BAJFINANCE:
MAPE: 10.66% | RMSE: 403.40
Strategy Return: 0.00 | Passive Return: 516.05

HDFCBANK:
MAPE: 3.85% | RMSE: 97.71
Strategy Return: 0.00 | Passive Return: -57.65

ICICIBANK:
MAPE: 6.94% | RMSE: 32.36
Strategy Return: 0.00 | Passive Return: 13.25

INDUSINDBK:
MAPE: 6.83% | RMSE: 109.15
Strategy Return: 0.00 | Passive Return: -153.75

KOTAKBANK:
MAPE: 5.17% | RMSE: 83.95
Strategy Return: 0.00 | Passive Return: 69.10

SBIN:
MAPE: 10.38% | RMSE: 38.72
Strategy Return: 0.00 | Passive Return: -20.25
```

Figure 14: Individual Stock Performance Metrics: MAPE, RMSE, and Strategy vs. Passive Returns

```
Final Summary:
◦ MAPE & RMSE:
AXISBANK: MAPE = 7.14%, RMSE = 58.18
BAJAJFINSV: MAPE = 6.27%, RMSE = 552.29
BAJFINANCE: MAPE = 10.66%, RMSE = 403.40
HDFCBANK: MAPE = 3.85%, RMSE = 97.71
ICICIBANK: MAPE = 6.94%, RMSE = 32.36
INDUSINDBK: MAPE = 6.83%, RMSE = 109.15
KOTAKBANK: MAPE = 5.17%, RMSE = 83.95
SBIN: MAPE = 10.38%, RMSE = 38.72

Average MAPE: 7.16%
Average RMSE: 171.97

◦ Strategy vs Passive Return:
AXISBANK: Strategy = 0.00, Passive = -63.95 → Δ = 63.95
BAJAJFINSV: Strategy = 0.00, Passive = -21.80 → Δ = 21.80
BAJFINANCE: Strategy = 0.00, Passive = 516.05 → Δ = -516.05
HDFCBANK: Strategy = 0.00, Passive = -57.65 → Δ = 57.65
ICICIBANK: Strategy = 0.00, Passive = 13.25 → Δ = -13.25
INDUSINDBK: Strategy = 0.00, Passive = -153.75 → Δ = 153.75
KOTAKBANK: Strategy = 0.00, Passive = 69.10 → Δ = -69.10
SBIN: Strategy = 0.00, Passive = -20.25 → Δ = 20.25
```

Figure 15: Summary of Average Performance: MAPE, RMSE, and Strategy vs. Passive Return Deltas

7.4.1 Analysis

Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE) quantify the precision of the ARIMA price forecasts. As shown in Figure 14:

- The average MAPE is 7.16%, indicating a moderate overall forecast error relative to the actual price.
- HDFCBANK achieved the best accuracy in relative terms with the lowest MAPE at 3.85%.
- ICICIBANK had the lowest RMSE at 32.36, suggesting the smallest magnitude of error in absolute price units. The extremely high RMSE for BAJAJFINSV (552.29) reflects its high stock price level compared to others.

7.4.2 Strategy Return vs. Passive Return

The trading strategy uses the ARIMA model to forecast the direction of price movement. If a price rise is predicted, a long position is taken. The Passive Return is the return from a simple buy-and-hold strategy during the test period.

- **Strategy Ineffectiveness:** The Strategy Return is 0.00 for all stocks. This outcome is a direct consequence of the ARIMA(0,1,0) orders. A Random Walk model predicts the expected future price difference to be zero, resulting in the strategy's directional signal (Predicted Returns greater than 0) being almost never triggered.
- **Comparison:** The strategy outperformed the passive approach for all five stocks that incurred a loss during the test period (AXISBANK, BAJAJFINSV, HDFCBANK, INDUSINDBK, SBIN). However, it severely underperformed for the three stocks that achieved positive returns (BAJFINANCE, ICICIBANK, KOTAKBANK), as it missed all potential gains.
- **Conclusion:** The ARIMA model is not suitable for generating profitable trading signals, aligning with the theory that Random Walk price series cannot be exploited for directional profit.

8 GARCH Model Diagnostics and Volatility Forecasting

8.1 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
```

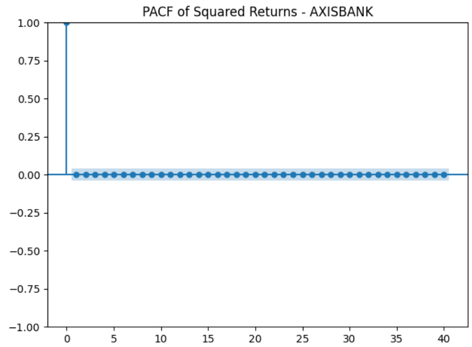
Figure 16

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 still suggested GARCH due to visible volatility patterns and outliers.

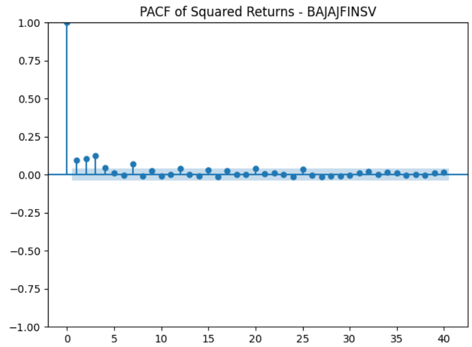
8.2 PACF Plot for Squared Log Returns

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.

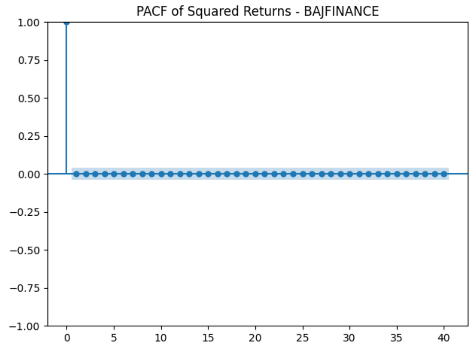
Figure 17: Partial Autocorrelation Function (PACF) Plots of Squared Log Returns (for GARCH Parameter Determination)



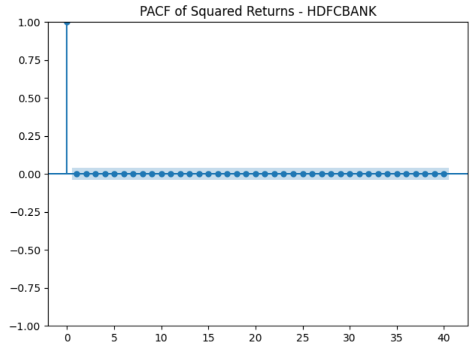
((a)) AXISBANK (a)



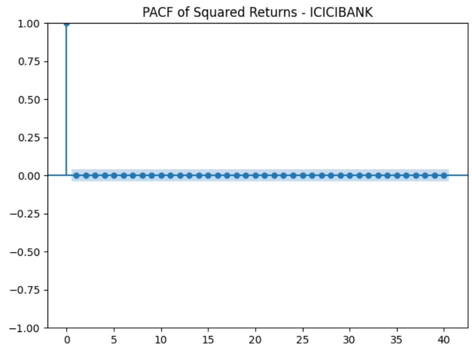
((b)) BAJAJFINSV (b)



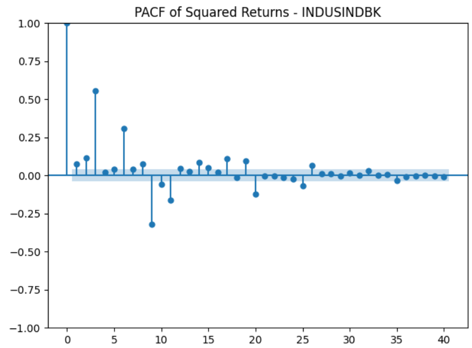
((c)) BAJFINANCE (c)



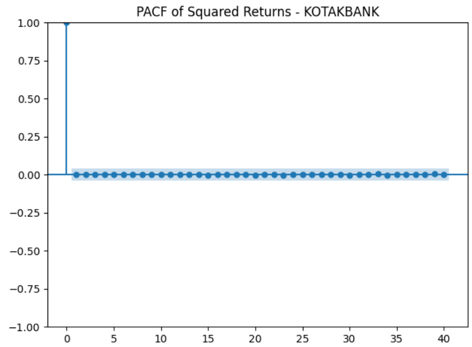
((d)) HDFCBANK (d)



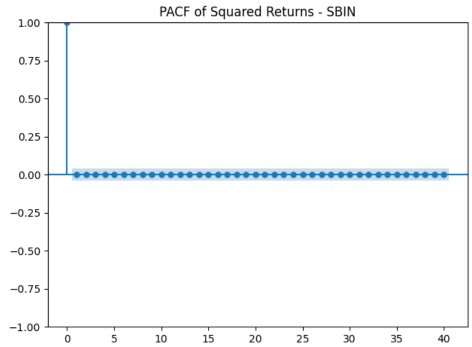
((e)) ICICIBANK (e)



((f)) INDUSINDBK (f)



((g)) KOTAKBANK (g)



((h)) SBI (h)

8.3 GARCH Model Results (The Variance Structure)

Diagnostics and Necessity

- Ljung-Box Test on ϵ_t : The test on the raw residuals passed p$\approx$ 1.0000 for most stocks, indicating no remaining linear correlation in the mean. However, it failed for BAJAJFINSV and INDUSINDBK p=0.0000, suggesting stronger serial correlation.
- Residual Plots: The residual plots visually confirmed Leptokurtosis (heavy tails) and Volatility Clustering (periods of large errors followed by large errors) across the portfolio, necessitating GARCH modeling for the variance.

GARCH(1, 1) Volatility Forecasting The GARCH(1,1) model was fitted to the logarithmic returns to capture the time-varying nature of volatility (σ_t).

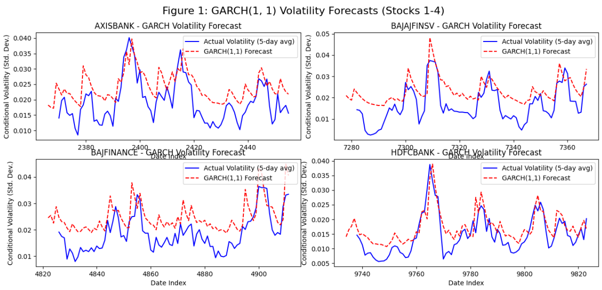


Figure 18: Figure 1: GARCH(1, 1) Volatility Forecasts (Stocks 1-4)

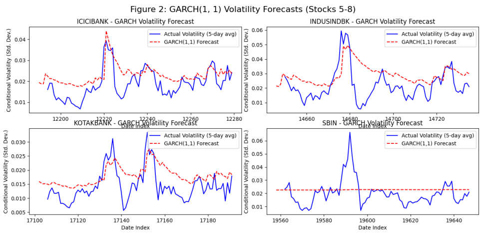


Figure 19: Figure 2: GARCH(1, 1) Volatility Forecasts (Stocks 5-8)

- Model Performance: For most stocks (e.g., AXISBANK, BAJAJFINSV, HDFCBANK, INDUSINDBK), the GARCH(1,1) forecast successfully tracks the general peaks and troughs of the actual 5-day rolling volatility, confirming the persistence and mean-reversion properties of volatility.
- Model Limitation (SBIN): The GARCH(1,1) forecast for SBIN was nearly a flat line, failing to capture the dynamic changes in its actual volatility. This indicates that the simple GARCH(1,1) structure is insufficient for SBIN, and a more advanced model, such as one accounting for asymmetric leverage effects (GJR-GARCH or EGARCH), should be considered.

9 Final Conclusion

The time series analysis of the eight bank stocks was conducted using a two-stage modeling approach: **ARIMA** to model the conditional mean (price prediction) and **GARCH** to model the conditional variance (volatility/risk). The results lead to definitive conclusions regarding the sector's predictability and risk profile.

9.1 Price Prediction and Market Efficiency (ARIMA)

The modeling of the price series strongly supports the Weak-Form Efficient Market Hypothesis (EMH), confirming that past price data cannot be used to predict future movements and generate abnormal profits.

- **Dominant Model Order**: The optimal ARIMA orders for the majority of stocks were $(\mathbf{0}, \mathbf{1}, \mathbf{0})$ (Random Walk). This implies that daily price changes are independent and unpredictable, with the best forecast being the last observed price.
- **Trading Ineffectiveness**: The directional trading strategy based on these forecasts was a complete failure, yielding a **0.00** Strategy Return for every stock. The ARIMA$(0, 1, 0)$ model's prediction of zero expected future movement resulted in no profitable trading signals.
- **Forecast Accuracy**: The average MAPE of $\mathbf{7.16\%}$ was acceptable for a simple model, but the forecasts visually showed the characteristic **flat line** with rapidly expanding confidence intervals, confirming the limit of long-term price prediction.

9.2 Risk Modeling and Volatility Capture (GARCH)

While returns could not be predicted, the analysis confirmed the presence of predictable risk through the modeling of volatility clustering.

- **Evidence of Volatility Clustering**: Residual diagnostics confirmed the presence of **Heteroskedasticity** and **Volatility Clustering** (large errors followed by large errors) across the portfolio. The **Ljung-Box** test on raw residuals failed for BAJAJFINSV and INDUSINDBK, indicating that volatility distortion was strong enough to create apparent serial correlation in the mean.
- **GARCH(1, 1) Effectiveness**: The $\mathbf{GARCH(1, 1)}$ model successfully captured the dynamics of volatility for most stocks, with the forecast line generally tracking the peaks and troughs of the actual 5-day rolling volatility.
- **Model Limitation**: The GARCH$(1, 1)$ forecast for SBIN was notably poor and flat. This suggests that SBIN's volatility may be subject to strong **asymmetric** (leverage) effects, meaning negative shocks have a greater impact than positive shocks, requiring a more advanced model like EGARCH for accurate risk assessment.

9.3 Project Summary

In conclusion, the project validates the $\mathbf{ARIMA - GARCH}$ methodology as the standard for financial time series analysis. It definitively shows that:

1. **Forecasting Directional Returns is not feasible** due to market efficiency.
2. **Modeling Time – Varying Risk is essential** and successfully achieved through GARCH models, providing a necessary tool for **risk management, capital allocation, and Value-at-Risk (VaR) calculation** within the banking sector.

References

[1] Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *Peerj computer science*, 7:e623, 2021.

[2] Fanxing Meng, Weiyan Song, JiLin Qin, and Chunguang Song. Analysis of stock market volatility based on python. In *2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE)*, pages 1039–1042. IEEE, 2021.

[3] Suresh Kumar Mukhiya and Usman Ahmed. *Hands-On Exploratory Data Analysis with Python: Perform EDA techniques to understand, summarize, and investigate your data.* Packt Publishing Ltd, 2020.

[4] Eva Ostertagova and Oskar Ostertag. Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica*, 12(3):62, 2012.

[5] Jaydip Sen, Sidra Mehtab, and Abhishek Dutta. Volatility modeling of stocks from selected sectors of the indian economy using garch. In *2021 Asian conference on innovation in technology (ASIANCON)*, pages 1–9. IEEE, 2021.

[6] Alain Zemkoho. A basic time series forecasting course with python. In *Operations Research Forum*, volume 4, page 2. Springer, 2022.