

Financial Time Series Analysis: Modeling SBI & INFY with ARIMA, GARCH, and EVT

In this project, I explore the financial time series of two stocks: State Bank of India (SBIN.NS) and Infosys (INFY.NS). My goal was to model their price movements, capture volatility, and analyze the extreme tail behavior of their returns. The analysis uses daily closing prices from January 1996 to September 2025.

The Workflow: From Raw Prices to Tail Risk

1. Data Prep and ARIMA Modeling

First, I pulled and cleaned the historical data. The price series were **non-stationary**, which I confirmed using **ADF** and **KPSS tests**. Then, I applied a first-order difference to make the series stationary before modeling.

Using the stationary data, I identified the best **ARIMA models** with `auto_arma` based on their AIC scores:

- **SBI:** ARIMA(0,1,1)
- **Infosys:** ARIMA(1,1,1)

The forecast performance was decent but highlighted the limitations of using ARIMA alone for volatile financial data.

2. Volatility Modeling with GARCH

To capture the **volatility clustering** present in the returns, I fitted **GARCH models**. An ARCH test confirmed this was a necessary step. The best models identified were:

- **SBI:** GARCH(1, 2)
- **Infosys:** GARCH(1, 5)

A **residual analysis** for both models showed they were well-specified. The residuals were nearly normal and had no significant autocorrelation, meaning the models successfully captured the volatility dynamics.

Extreme Value Theory (EVT) Analysis

The final step was to analyze the distribution's tails using **Extreme Value Theory (EVT)**, which is crucial for risk management. I took the standardized residuals from the GARCH models, focused on the losses, and identified an optimal threshold ($u = 1.2555$), which yielded 501 exceedances. I then fitted a **Generalized Pareto Distribution (GPD)** to these tail events.

A Critical Finding: The GPD Fit was Inadequate

To check if GPD model was actually a good fit for the data, by using **Kolmogorov-Smirnov (KS) goodness-of-fit test**:

- **KS Statistic:** 1.0000
- **KS p-value:** 0.0000

The resulting p-value was effectively zero. This forces the **rejection of the null hypothesis**, meaning the GPD model is a poor fit for the data's extreme tails. This is a critical finding, as it implies any risk measures (like VaR or ES) derived from this model would be unreliable.

Conclusion

While I successfully modeled the volatility of SBI and INFY returns with GARCH, the EVT application for tail risk was unsuccessful with this specific approach. The key takeaway is the **failure of the GPD fit**.