Time Series Analysis and Volatility Modeling of SBIN.NS and INFY.NS

Srikrishna Das
MA24M025

Guid: Prof. Neelesh S Upadhye

September 16, 2025

Department of Mathematics
Indian Institute of Technology Madras

Contents

1 Introduction

This report details the process of constructing and validating marginal distribution models for the financial returns of two prominent Indian stocks: State Bank of India (SBIN.NS) and Infosys (INFY.NS). The primary objective is to create a robust model that accurately captures the stylized facts of financial returns, including volatility clustering, leverage effects, and heavy-tailed distributions.

The modeling framework integrates three key components:

1. An Autoregressive Moving Average (ARMA) model for the conditional mean
2. A Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model with Student's t-distribution for conditional volatility
3. Extreme Value Theory (EVT) for modeling the tail distributions

The final composite model is validated using the Probability Integral Transform (PIT) and Kolmogorov-Smirnov (KS) tests to ensure adequate description of the entire return distribution.

2 Data Acquisition and Preparation

2.1 Data Loading

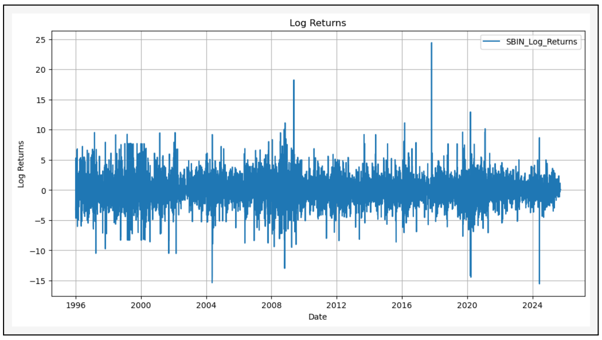
Historical daily stock price data for SBIN.NS and INFY.NS was obtained from Yahoo Finance covering the maximum available period. Daily log returns were calculated using the standard formula:

$$R_t = 100 \times \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1}$$

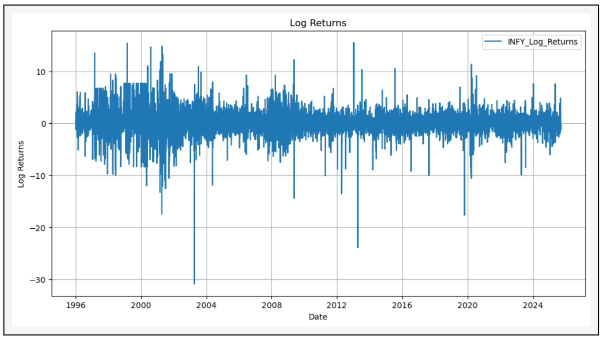
where P_t represents the closing price at time t. These log returns form the foundation for all subsequent modeling procedures. The initial log returns recorded on 1996-01-02 were -0.405307% for INFY.NS and -3.234742% for SBIN.NS.

2.2 Log Returns Visualization

Both stock return series exhibit clear evidence of **volatility clustering**, where periods of high volatility tend to be followed by similar high-volatility periods, and vice versa for low volatility periods.



(a) **Daily Log Returns (%) for SBIN.NS**



(b) **Daily Log Returns (%) for INFY.NS**

Figure 1: **Daily Log Returns for SBIN.NS and INFY.NS**

3 ARMA Model Analysis

The optimal ARIMA specification for each stock's log returns was determined using a systematic approach. The dataset was partitioned with 95% allocated for training purposes. The `pm.auto_arima` function performed an exhaustive search to identify the model minimizing the Akaike Information Criterion (AIC), with parameters: `start_p=1`, `start_q=1`, `max_p=7`, `max_q=7`, `start_P=0`, `m=3`, `test='adf'`, `seasonal=False`, and `D=1`.

3.1 Best ARIMA Model for SBIN.NS

The optimal model identified for SBIN.NS was ARIMA(0,0,1)(0,0,0) with intercept.

Table 1: **SARIMAX Results for Best ARIMA Model – SBIN.NS**

Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Model Summary					
Dep. Variable	y	**No. Obs.**	7,086	**Model**	SARIMAX(0,0,1)
Log Likelihood	-16,132.767	**AIC**	32,271.535	**BIC**	32,292.132
Date	Tue, 16 Sep 2025	**Time**	10:46:41	**HQIC**	32,278.628
Sample	0 – 7,086	**Cov. Type**	opg		
Coefficients					
	coef	**std err**	**z**	**P>|z|**	**[0.025, 0.975]**
intercept	0.0577	0.030	1.937	0.053	[-0.001, 0.116]
ma.L1	0.0603	0.008	7.554	0.000	[0.045, 0.076]
sigma2	5.5597	0.052	107.154	0.000	[5.458, 5.661]
Diagnostics					

Continued on next page

Table 1: **SARIMAX Results for Best ARIMA Model – SBIN.NS (Continued)**

Parameter	Value	Parameter	Value	Parameter	Value
Ljung-Box (Q)	0.01	**Prob(Q)**	0.94	**Jarque-Bera (JB)**	6,399.71
Prob(JB)	0.00	**Hetero. (H)**	0.70	**Prob(H)**	0.00
Skew	0.18	**Kurtosis**	7.64		

Model Performance Metrics:

- **RMSE:** 1.6094
- **AIC:** 32,271.5347
- **BIC:** 32,292.1323

3.2 Best ARIMA Model for INFY.NS

The optimal model identified for INFY.NS was ARIMA(2,0,1)(0,0,0) with intercept.

Table 2: **SARIMAX Results for Best ARIMA Model – INFY.NS**

Parameter	Value	Parameter	Value	Parameter	Value
Model Summary					
Dep. Variable	y	**No. Obs.**	7,086	**Model**	SARIMAX(2,0,1)
Log Likelihood	-16,260.519	**AIC**	32,531.038	**BIC**	32,565.367
Date	Tue, 16 Sep 2025	**Time**	10:47:27	**HQIC**	32,542.860
Sample	0 – 7,086	**Cov. Type**	opg		
Coefficients					
	coef	**std err**	**z**	**P>|z|**	**[0.025, 0.975]**
intercept	0.0322	0.014	2.343	0.019	[0.005, 0.059]
ar.L1	0.7885	0.096	8.187	0.000	[0.600, 0.977]
ar.L2	-0.0719	0.008	-9.203	0.000	[-0.087, -0.057]
ma.L1	-0.7234	0.096	-7.501	0.000	[-0.912, -0.534]
sigma2	5.7638	0.039	146.730	0.000	[5.687, 5.841]
Diagnostics					
Ljung-Box (Q)	0.00	**Prob(Q)**	0.97	**Jarque-Bera (JB)**	32,296.81

Continued on next page

Table 2: **SARIMAX Results for Best ARIMA Model – INFY.NS (Continued)**

Parameter	Value	Parameter	Value	Parameter	Value
Prob(JB)	0.00	**Hetero. (H)**	0.28	**Prob(H)**	0.00
Skew	-0.44	**Kurtosis**	13.42		

Model Performance Metrics:

- **RMSE:** 1.5297
- **AIC:** 32,531.0379
- **BIC:** 32,565.3673

4 GJR-GARCH-StudentT Model Analysis

Following the ARIMA model fitting, residuals were extracted and used to estimate a GJR-GARCH model with Student's t-distribution for modeling conditional volatility. The `arch_model` function was configured with parameters: `vol='Garch'`, `p=1`, `o=1`, `q=1`, `dist='studentst'`, and `mean='Constant'`.

4.1 GJR-GARCH-StudentT Model for SBIN.NS

Table 3: **Constant Mean – GJR-GARCH Model Results for SBIN.NS**

Parameter	Value	Parameter	Value	Parameter	Value
Model Summary					
Dep. Variable	None	**R-squared**	0.000	**Adj. R-squared**	0.000
Mean Model	Constant Mean	**Vol Model**	GJR-GARCH	**Distribution**	Standardized Student's t
Log-Likelihood	-15,383.9	**AIC**	30,779.9	**BIC**	30,821.1
Method	Maximum Likelihood	**No. Obs.**	7,086	**Df Residuals**	7,085
Date	Tue, Sep 16 2025	**Time**	10:47:27	**Df Model**	1
Mean Model Parameters					
	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**

Continued on next page

Table 3: **Constant Mean – GJR-GARCH Model Results for SBIN.NS (Continued)**

Parameter	Value	Parameter	Value	Parameter	Value
μ	-8.4595e-03	2.206e-02	-0.383	0.701	[-5.170e-02, 3.478e-02]
Volatility Model Parameters					
ω	0.1378	3.602e-02	3.827	1.299e-04	[0.0672, 0.208]
α	0.0745	1.232e-02	6.044	1.501e-09	[0.0503, 0.0986]
γ	0.0358	1.280e-02	2.799	5.130e-03	[0.0107, 0.0609]
β	0.8869	1.669e-02	53.146	0.000	[0.854, 0.920]
Distribution Parameters					
ν	5.8373	0.415	14.063	6.379e-45	[5.024, 6.651]

Model Performance: AIC = 30,779.8577, BIC = 30,821.0530

4.1.1 Residual Diagnostics for SBIN.NS

Comprehensive residual analysis was conducted on the standardized residuals from the GJR-GARCH model.

Ljung-Box Test Results:

Lags	LB Statistic	p-value
	Residuals	
10	29.213	0.0012
20	43.201	0.0019
	Squared Residuals	
10	7.562	0.6715
20	14.046	0.8281

ARCH LM Test: p-value = 0.8366

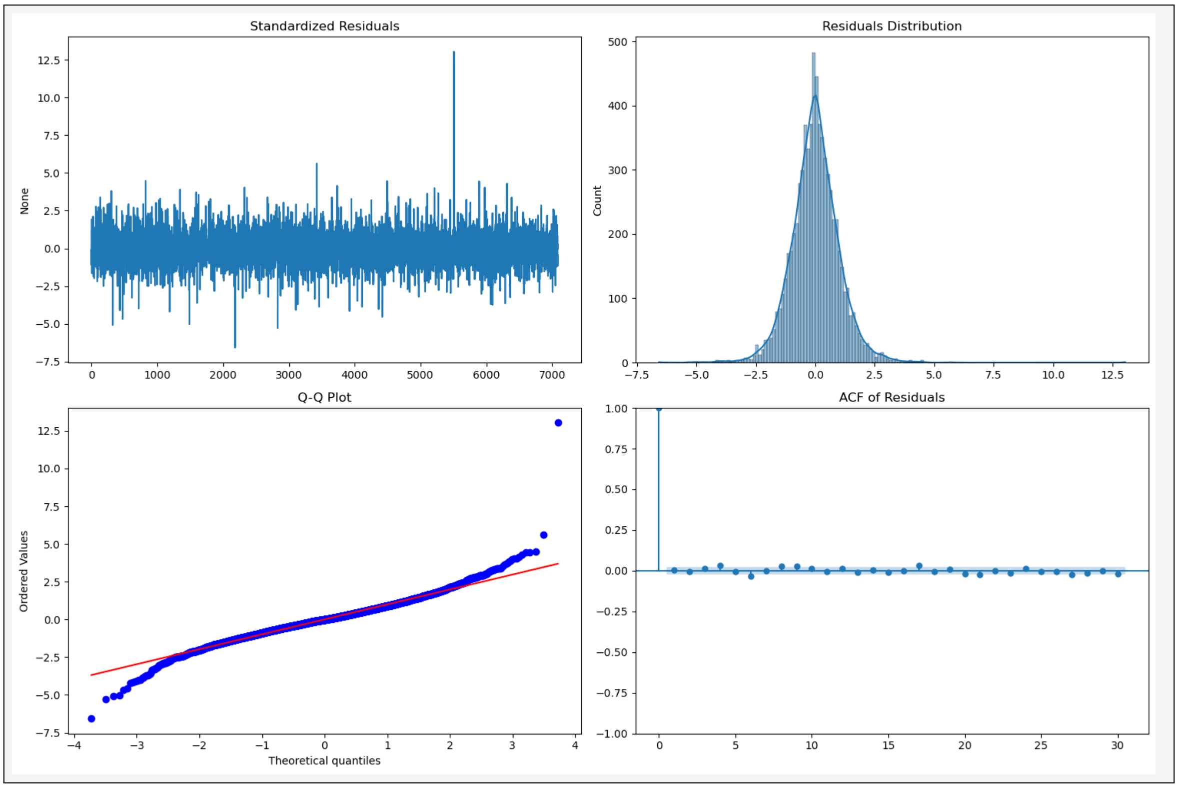


Figure 2: Residual Diagnostics for SBIN.NS GJR-GARCH Model

4.2 GJR-GARCH-StudentT Model for INFY.NS

Table 4: **Constant Mean – GJR-GARCH Model Results for INFY.NS**

Parameter	Value	Parameter	Value	Parameter	Value
Model Summary					
Dep. Variable	None	**R-squared**	0.000	**Adj. R-squared**	0.000
Mean Model	Constant Mean	**Vol Model**	GJR-GARCH	**Distribution**	Standardized Student's t
Log-Likelihood	-14,747.6	**AIC**	29,507.3	**BIC**	29,548.5
Method	Maximum Likelihood	**No. Obs.**	7,086	**Df Residuals**	7,085
Date	Tue, Sep 16 2025	**Time**	10:47:29	**Df Model**	1
Mean Model Parameters					
	coef	**std err**	**t**	**P>|t|**	**95.0% Conf. Int.**

Continued on next page

Table 4: **Constant Mean – GJR-GARCH Model Results for INFY.NS (Continued)**

Parameter	Value	Parameter	Value	Parameter	Value
μ	-0.0197	1.905e-02	-1.037	0.300	[-5.709e-02, 1.759e-02]
Volatility Model Parameters					
ω	0.2404	7.272e-02	3.306	9.466e-04	[0.0979, 0.383]
α	0.1397	2.934e-02	4.762	1.921e-06	[0.0822, 0.197]
γ	0.0313	2.520e-02	1.243	0.214	[-1.808e-02, 8.069e-02]
β	0.8145	4.008e-02	20.322	8.214e-92	[0.736, 0.893]
Distribution Parameters					
ν	4.1291	0.228	18.105	2.924e-73	[3.682, 4.576]

Model Performance: AIC = 29,507.2999, BIC = 29,548.4951

4.2.1 Residual Diagnostics for INFY.NS

Ljung-Box Test Results:

Lags	LB Statistic	p-value
Residuals		
10	7.689	0.6592
20	20.503	0.4269
Squared Residuals		
10	7.578	0.6700
20	11.702	0.9260

ARCH LM Test: p-value = 0.9102

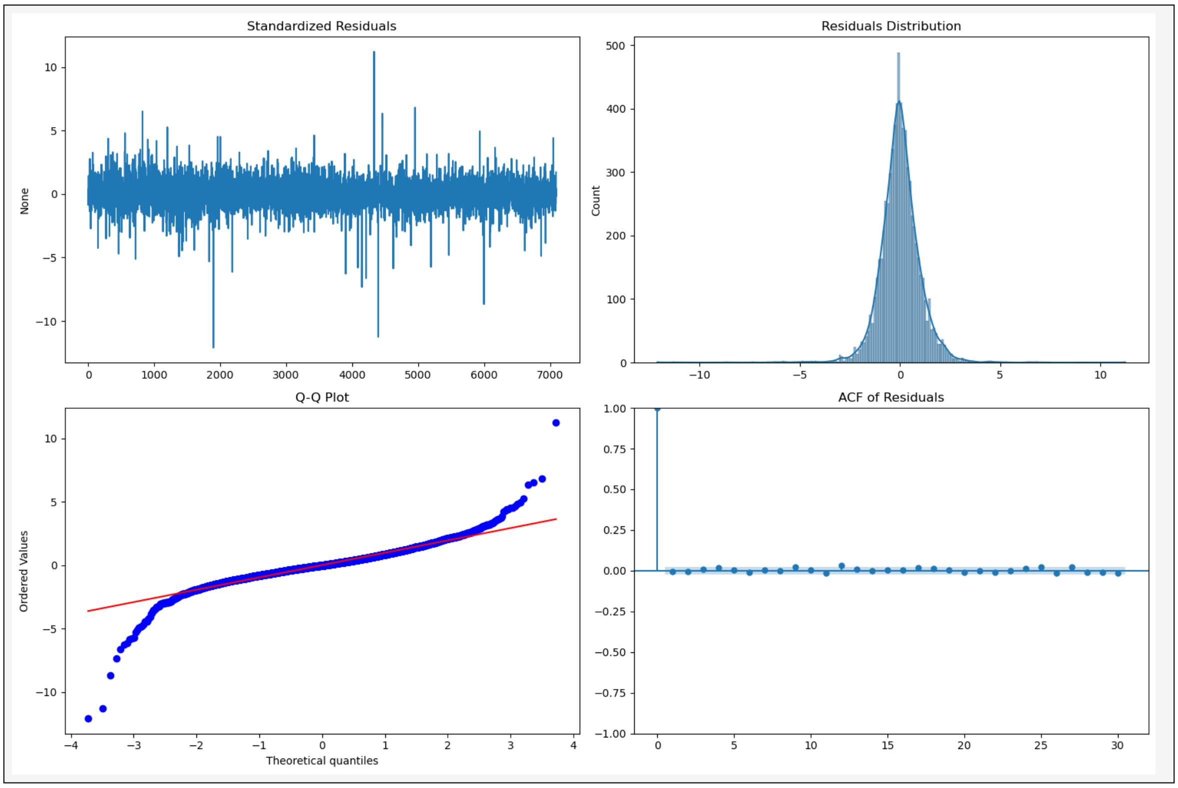


Figure 3: Residual Diagnostics for INFY.NS GJR-GARCH Model

5 Rolling Forecasting

Rolling forecasting procedures were implemented for both stocks using the fitted GJR-GARCH models on ARIMA residuals. The models were recursively re-estimated at each step with updated residual histories, generating one-step-ahead volatility forecasts. Actual volatility was approximated using a 40-day rolling standard deviation window.

5.1 Rolling Forecasting for SBIN.NS

Forecast Accuracy: RMSE = 0.5047

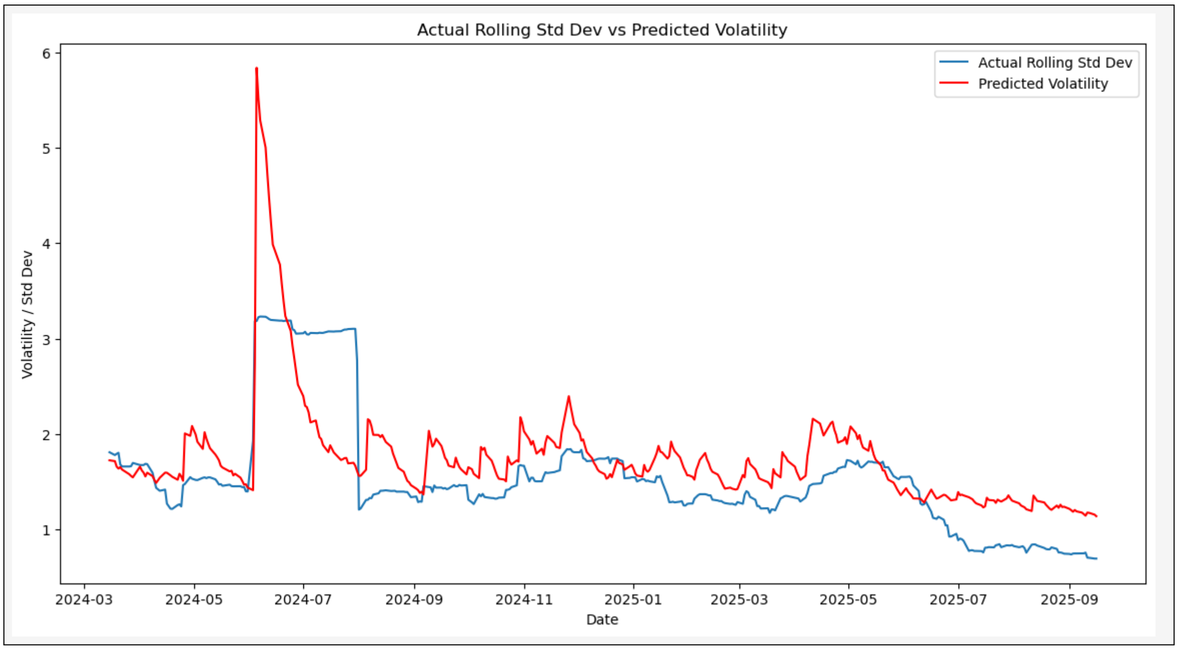


Figure 4: Actual Rolling Standard Deviation vs. Predicted Volatility – SBIN.NS

5.2 Rolling Forecasting for INFY.NS

Forecast Accuracy: RMSE = 0.4664

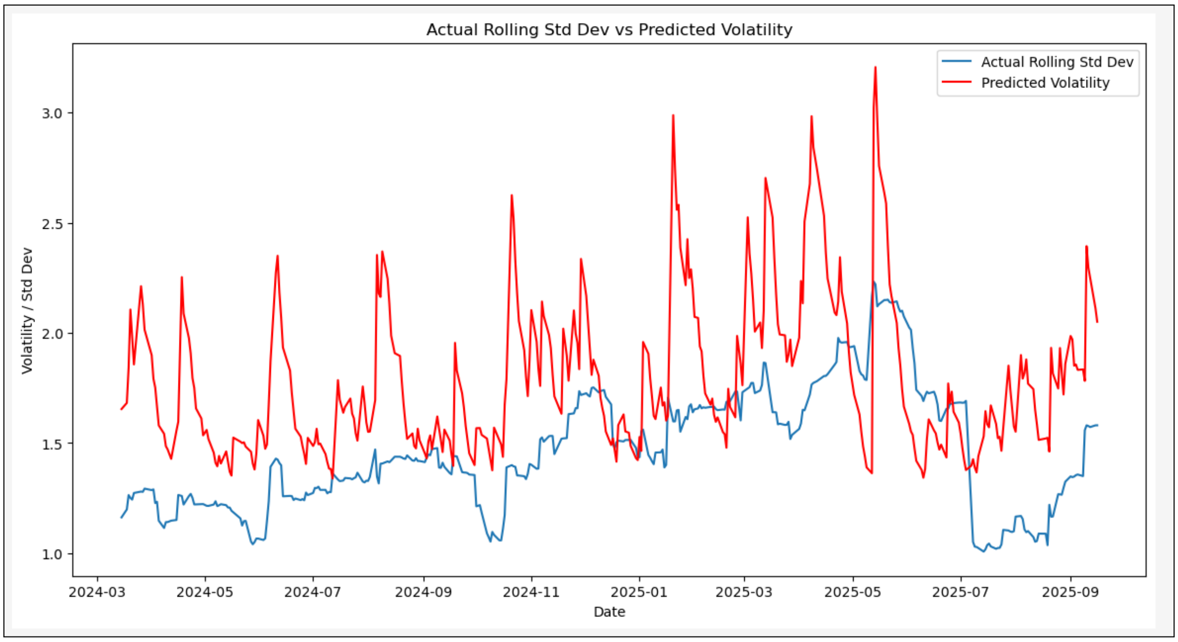


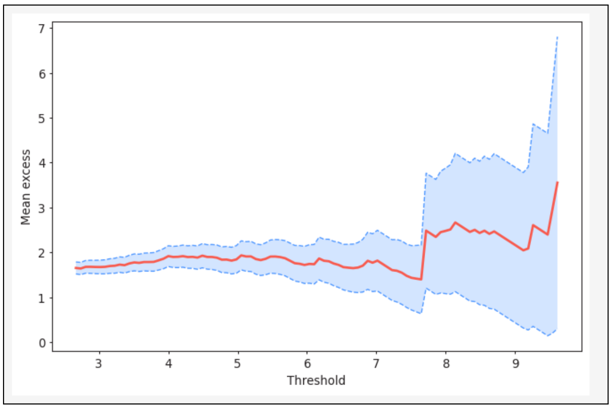
Figure 5: Actual Rolling Standard Deviation vs. Predicted Volatility – INFY.NS

6 Extreme Value Theory (EVT) Analysis

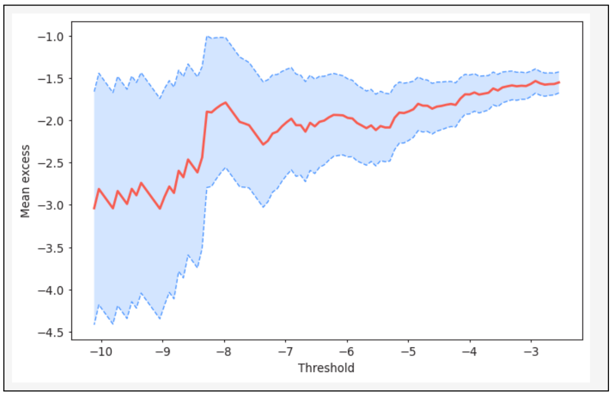
The Peak Over Threshold (POT) methodology was employed to fit Generalized Pareto Distributions (GPD) to the tail regions of the log return distributions. Threshold selection involved systematic evaluation across quantile ranges, computing Kolmogorov-Smirnov p-values and Anderson-Darling statistics for exceedances, with scoring mechanisms to identify optimal fits. A smooth cumulative distribution function was constructed combining Student's t-distribution for the central body with GPD tails, connected via logistic splicing. Model validation utilized PIT histograms and KS tests.

6.1 EVT Analysis for SBIN.NS

6.1.1 Mean Residual Life and Parameter Stability

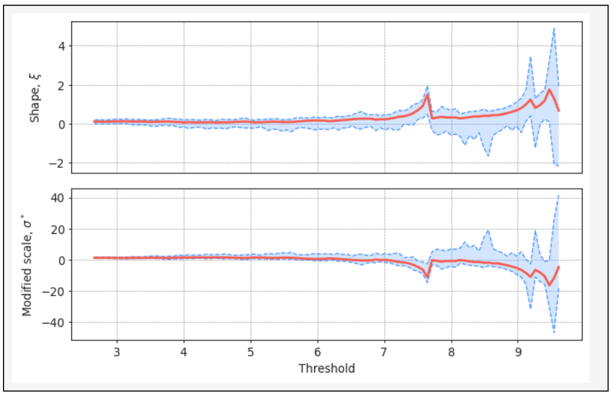


(a) MRL Plot (Upper Extremes)

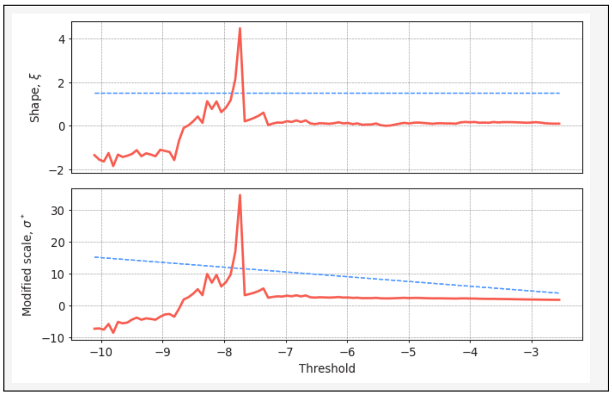


(b) MRL Plot (Lower Extremes)

Figure 6: Mean Residual Life Plots – SBIN.NS



(a) Parameter Stability (Upper Extremes)



(b) Parameter Stability (Lower Extremes)

Figure 7: Parameter Stability Plots – SBIN.NS

6.1.2 Optimal Threshold Selection for SBIN.NS

Upper Tail Threshold:

- **Quantile:** 0.950
- **Threshold:** 3.7167
- **Exceedances:** 373
- **Shape Parameter:** 0.0756

- **Scale Parameter:** 1.6505
- **KS p-value:** 0.4344
- **AD Statistic:** 1.1088
- **Combined Score:** 0.3235

Lower Tail Threshold:

- **Quantile:** 0.045
- **Threshold:** -3.6935
- **Exceedances:** 336
- **Shape Parameter:** 0.1344
- **Scale Parameter:** 1.4497
- **KS p-value:** 0.9168
- **AD Statistic:** 0.3173
- **Combined Score:** 0.8851

6.1.3 Smooth CDF and PIT Validation for SBIN.NS

The composite distribution combines Student's t-distribution for the central region with GPD tails, seamlessly connected through logistic splicing.

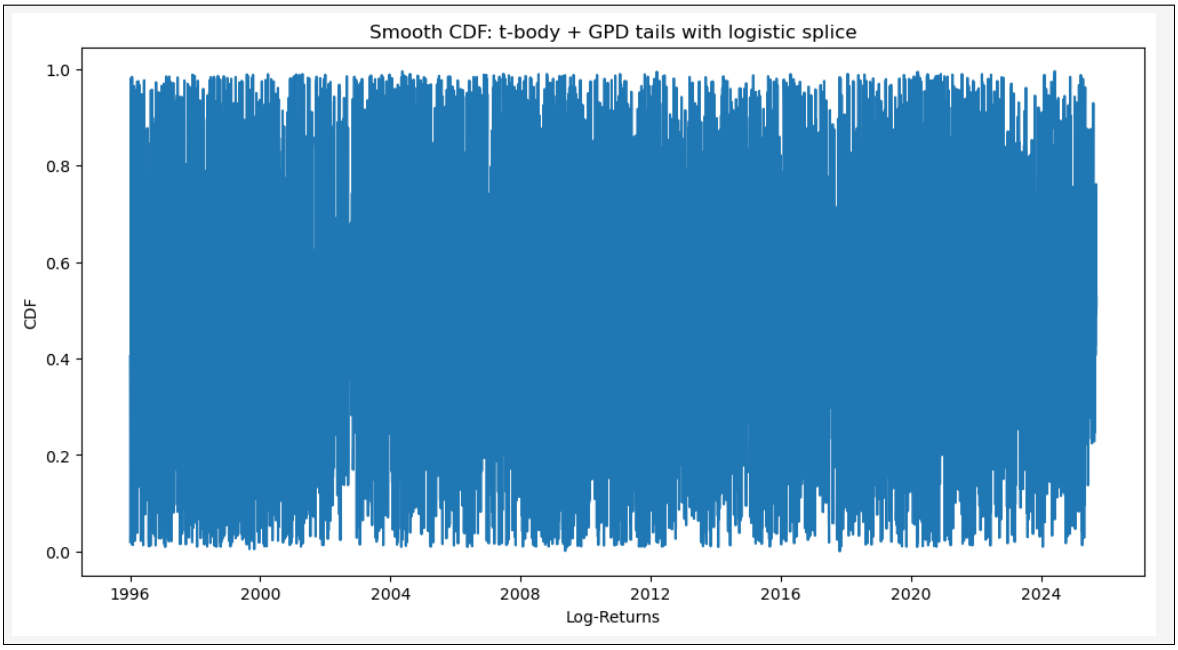


Figure 8: Smooth CDF: Student's t-body + GPD tails with logistic splice – SBIN.NS

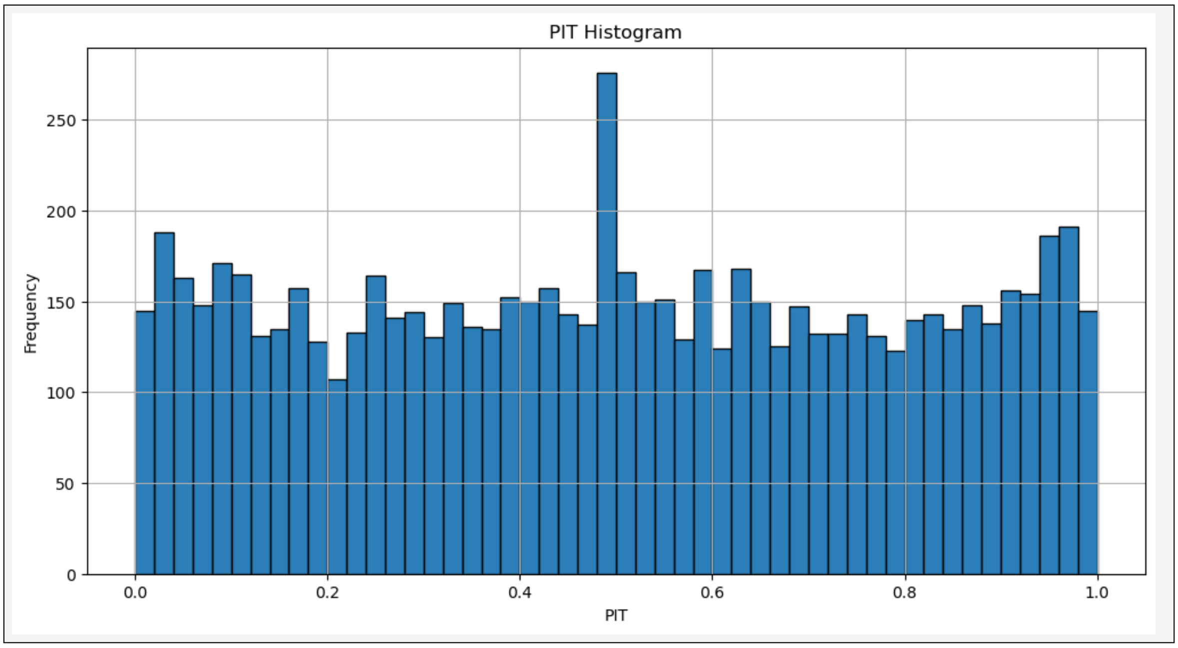


Figure 9: PIT Histogram – SBIN.NS

PIT Validation Results: KS Statistic = 0.0133, p-value = 0.1412

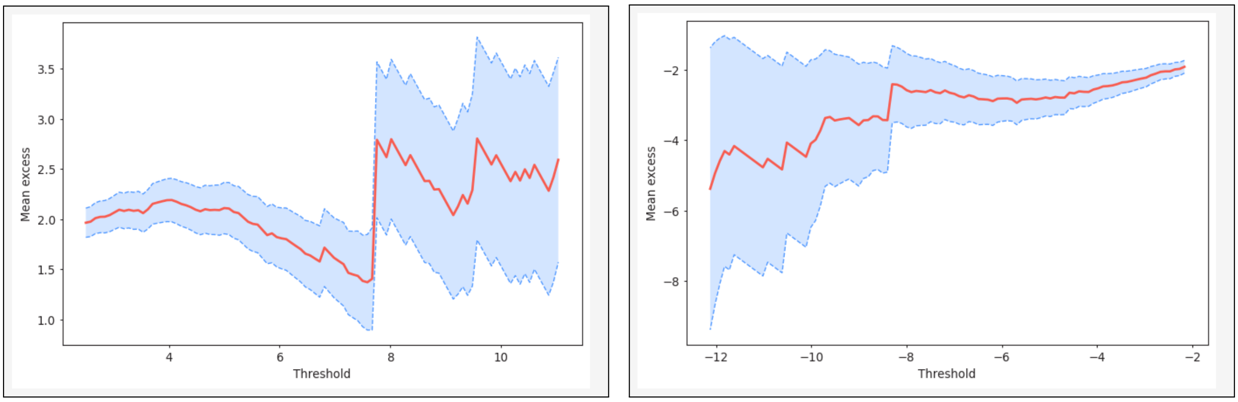
6.1.4 Global PIT Optimization for SBIN.NS

Systematic search across threshold combinations to maximize KS p-value for global PIT:

- **Optimal Combination:** (i=9, j=6)
- **KS Statistic:** 0.0133
- **p-value:** 0.1412
- **Upper Threshold:** Quantile = 0.950, Threshold = 3.7167, Exceedances = 373
- **Lower Threshold:** Quantile = 0.045, Threshold = −3.6935, Exceedances = 336

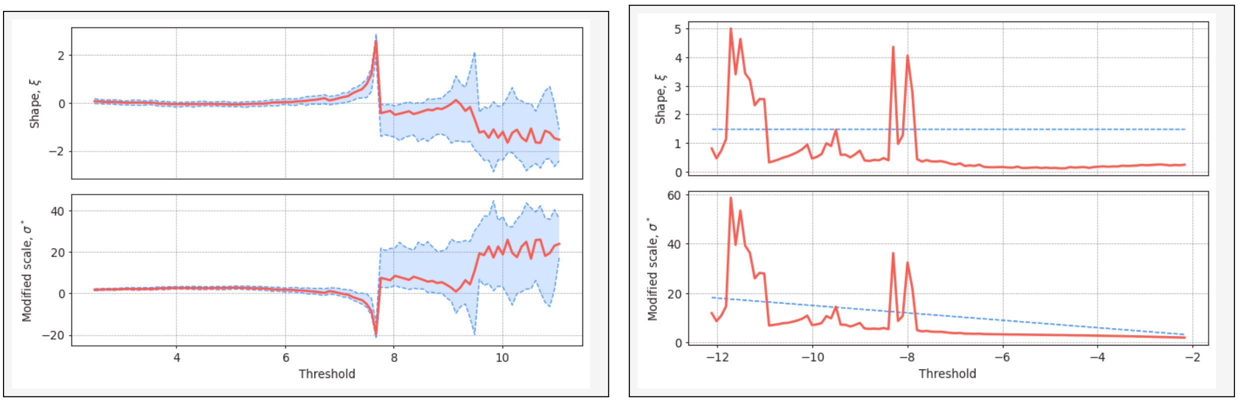
6.2 EVT Analysis for INFY.NS

6.2.1 Mean Residual Life and Parameter Stability



(a) MRL Plot (Upper Extremes) (b) MRL Plot (Lower Extremes)

Figure 10: Mean Residual Life Plots – INFY.NS



(a) Parameter Stability (Upper Extremes) (b) Parameter Stability (Lower Extremes)

Figure 11: Parameter Stability Plots – INFY.NS

6.2.2 Optimal Threshold Selection for INFY.NS

Upper Tail Threshold:

- **Quantile:** 0.925
- **Threshold:** 2.9710
- **Exceedances:** 560
- **Shape Parameter:** −0.0180
- **Scale Parameter:** 2.1022
- **KS p-value:** 0.2443
- **AD Statistic:** 1.3934
- **Combined Score:** 0.1050

Lower Tail Threshold:

- **Quantile:** 0.055
- **Threshold:** −3.0617
- **Exceedances:** 411
- **Shape Parameter:** 0.1899
- **Scale Parameter:** 1.8274
- **KS p-value:** 0.6968
- **AD Statistic:** 0.4515
- **Combined Score:** 0.6517

6.2.3 Smooth CDF and PIT Validation for INFY.NS

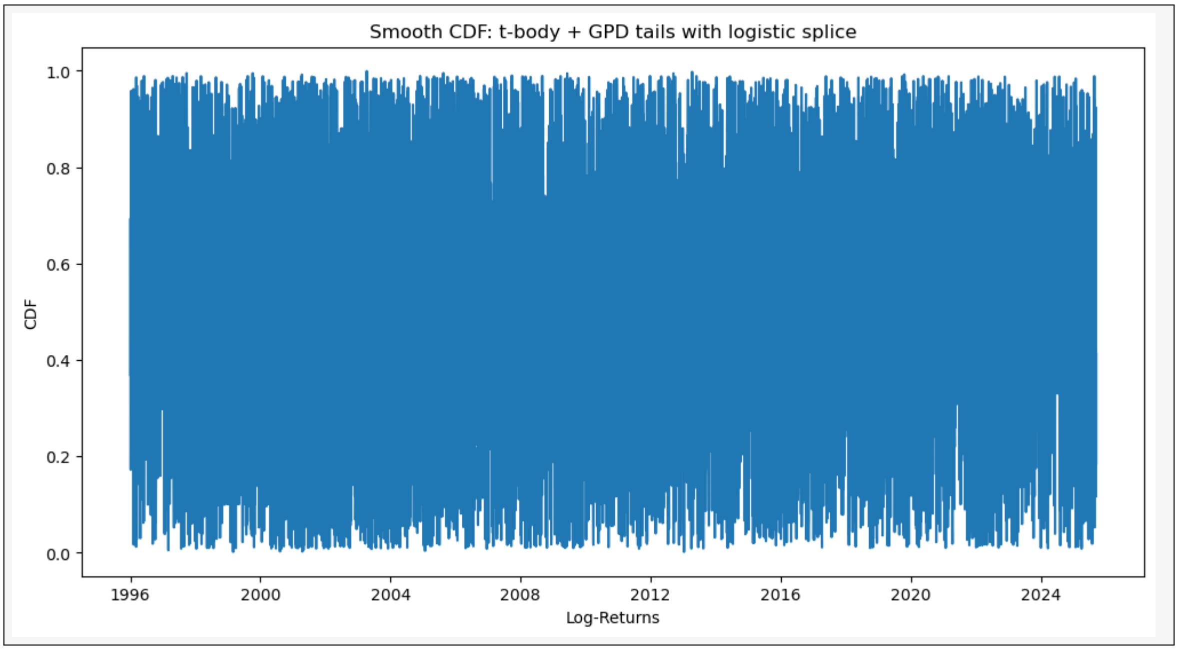


Figure 12: Smooth CDF: Student's t-body + GPD tails with logistic splice – INFY.NS

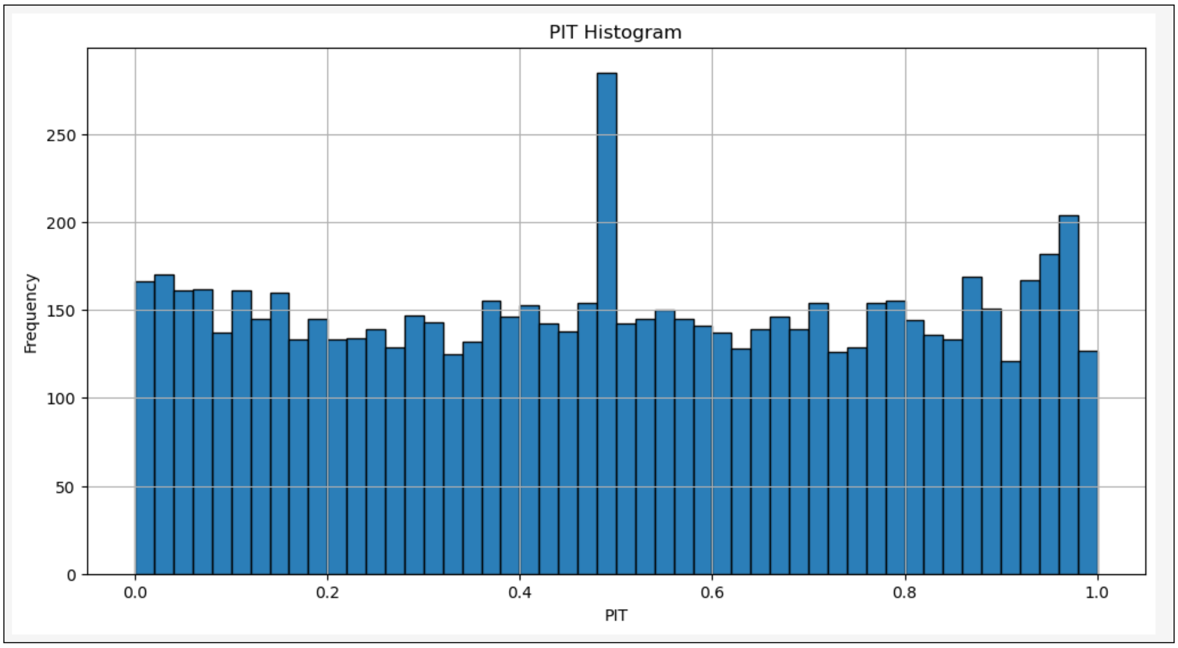


Figure 13: PIT Histogram – INFY.NS

PIT Validation Results: KS Statistic = 0.0116, p-value = 0.2668

6.2.4 Global PIT Optimization for INFY.NS

- **Optimal Combination:** (i=1, j=3)
- **KS Statistic:** 0.0116
- **p-value:** 0.2668
- **Upper Threshold:** Quantile = 0.925, Threshold = 2.9710, Exceedances = 560
- **Lower Threshold:** Quantile = 0.055, Threshold = −3.0617, Exceedances = 411