ARMA-GARCH and Copula Modeling for Indian Stocks

Vishal Nigam SM3800014
Data Science & Artificial Intelligence
University of Trieste

0.1 Introduction

This report is part of the **Data Science for Insurance** course project. We analyze the marginal and joint behavior of return series from three Indian stocks — **RELIANCE**, **INFY**, and **HDFCBANK** — using:

- ARMA-GARCH models for individual return dynamics
- Copula-based methods for dependence modeling

The dataset contains **daily adjusted closing prices** from **January 2015 to July 2025**, sourced from Yahoo Finance.

0.2 Project Objectives

- Model **univariate return volatility** using ARMA-GARCH
- Extract **i.i.d. standardized residuals**
- Convert residuals into **pseudo-observations**
- Study pairwise dependence with **copula families**
- Estimate **tail dependence** and **rank-based measures**
- Assess model fit via **GOF statistics** and **visual checks**

0.3 Methodology Overview

0.3.1 Univariate Modeling

We model log-returns of each stock using **ARMA-GARCH**: - ARMA for serial correlation - GARCH(1,1) to capture volatility clustering - Innovations follow a **Student-t** distribution - Residuals standardized and transformed to **pseudo-observations** via empirical CDF

0.3.2 Dependence Modeling (Copulas)

After filtering marginal effects, we model **pairwise dependence** via: - Rank-based measures: **Kendall's** , **Spearman's** - Tail dependence: **Upper ( )** and **Lower ( )** - Copula families: **Clayton**, **Gumbel**, **Gaussian**, **t-Copula** - Selection criteria: **log-likelihood**, **Cramér–von Mises (CvM)** test, and **graphical diagnostics**

0.4 R Packages Used

Task	Packages
Data retrieval	`quantmod`, `xts`
Time series modeling	`forecast`, `rugarch`,`tseries`
Copula modeling	`copula`, `VineCopula`
Visualization	`ggplot2`, `gridExtra`, `gtable`, `ggExtra`, base R

This workflow applies modern **risk modeling** tools in R, integrating **time series analysis** and **copula-based dependence**—vital in finance and insurance analytics.

0.5 Section 1: Univariate Modeling of Log Returns

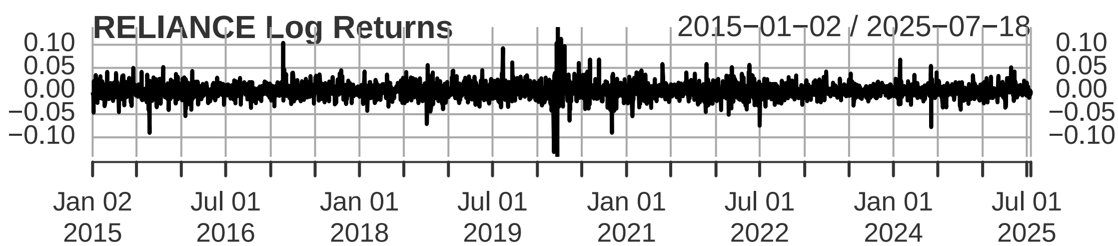
In this section, I modelled the marginal behavior of each stock — **RELIANCE**, **INFY**, and **HDFCBANK** — by analyzing their daily log-returns. The steps involve:

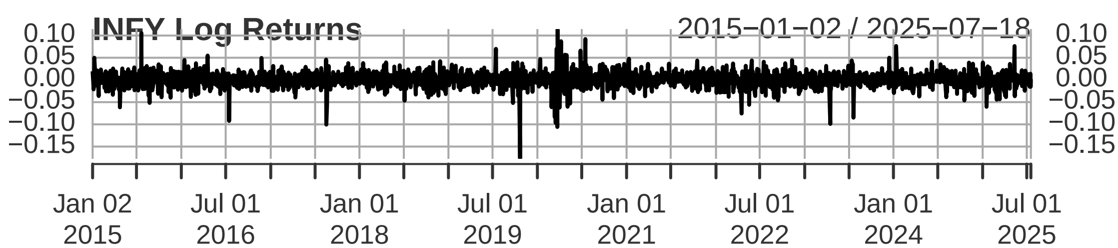
- Downloading and visualizing stock price data
- Calculating daily log-returns
- Exploring autocorrelation
- Fitting optimal **ARMA-GARCH models** to capture conditional volatility

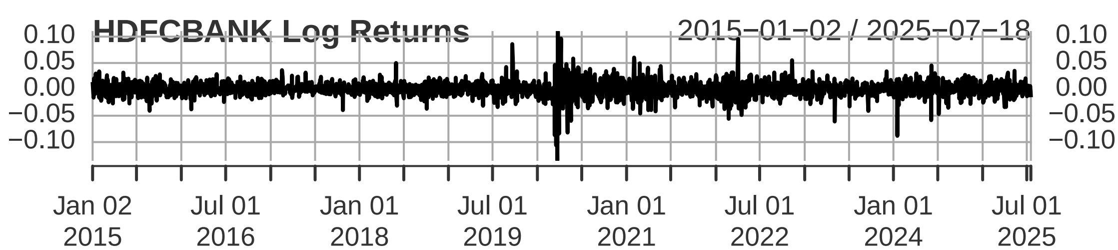
0.5.1 Data Collection and Preprocessing

I retrieve daily **adjusted closing prices** from **Yahoo Finance** for the three selected stocks using the `quantmod` package.

0.5.1.1 Adjusted Price Extraction We extract the adjusted closing prices for each stock and remove missing values. This ensures a synchronized time series across all three assets.It shows long-term price trends for all three stocks. We observe general upward movement with market volatility and short-term fluctuations.







Interpretation: The plots above show the daily log return series for each stock. All three exhibit typical financial time series behavior: High-frequency fluctuations around zero,Sudden large spikes during market stress periods (e.g., early 2020 during the COVID-19 pandemic) ,Heteroskedasticity — periods of low and high volatility alternate.

0.5.2 MultiStock ACF & PACF Analysis

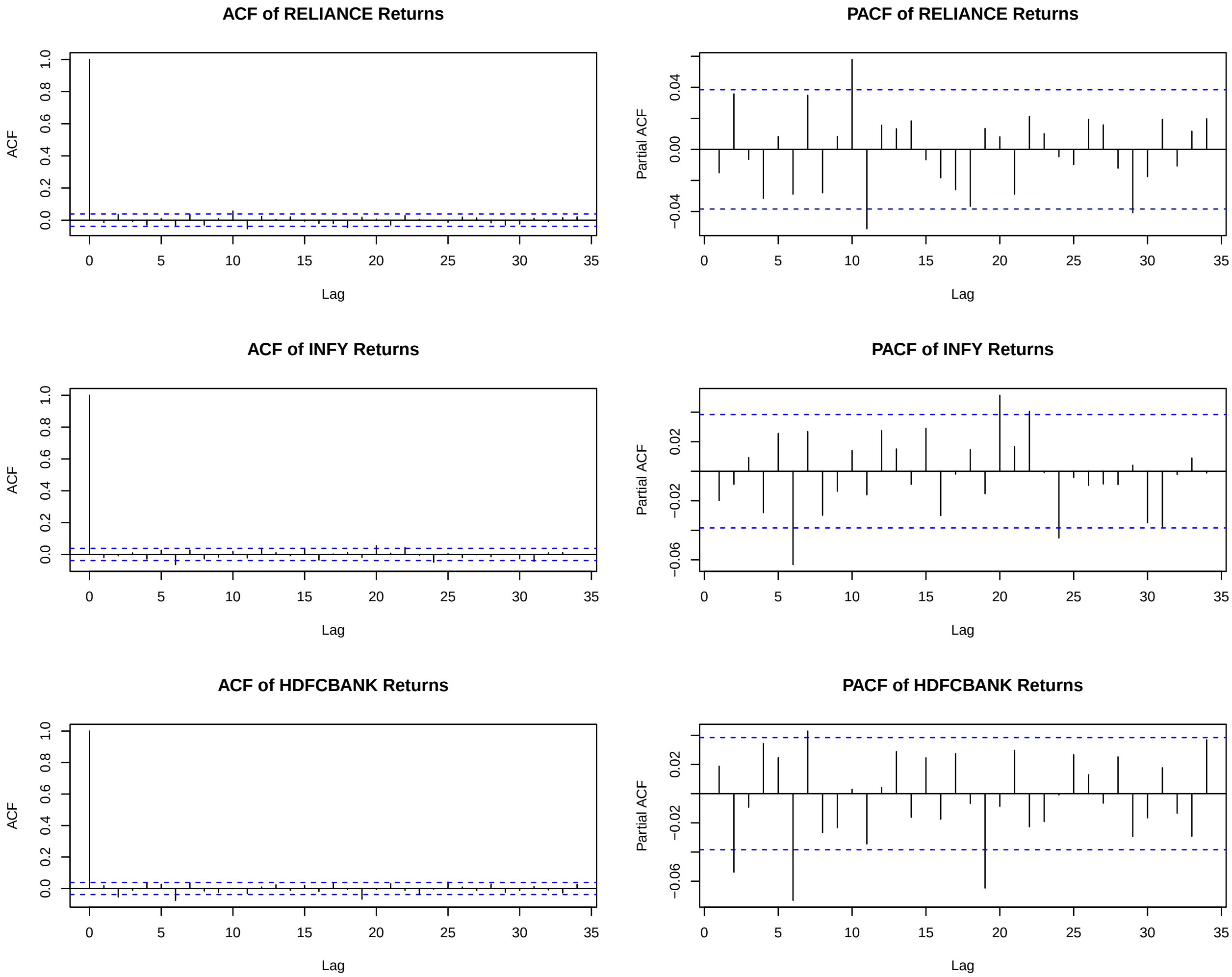


Figure 1: ACF and PACF for Stocks

Interpretation: As shown in Figure, the ACF and PACF plots for RELIANCE log-returns suggest:

- The ACF (left) drops sharply after lag 0, indicating weak autocorrelation, typical for return series.
- The PACF (right) shows small spikes around lag 1 and beyond, hinting at short-memory dynamics.
- These patterns support fitting a low-order ARMA model such as ARMA(1,1) or ARMA(2,1) to capture the serial dependence.

0.5.2.1 ARMA Modeling – RELIANCE Returns

Interpretation: The automatic model selected by auto.arima() was a simple ARIMA(0,0,0) with a non-zero mean, which suggests white noise.To improve structure, a manual ARMA(1,1) model was tested. It slightly improved the log-likelihood (6851.62 vs 6850.77) and retained the same error variance.**Residual Check**: The Ljung-Box test on residuals gave a p-value = 0.0055, indicating autocorrelation remains. This implies that even ARMA(1,1) might be insufficient, and a better model may be needed in the next step.**Conclusion**: We proceed to test higher-order ARMA models and move toward volatility modeling with GARCH to capture heteroskedasticity.

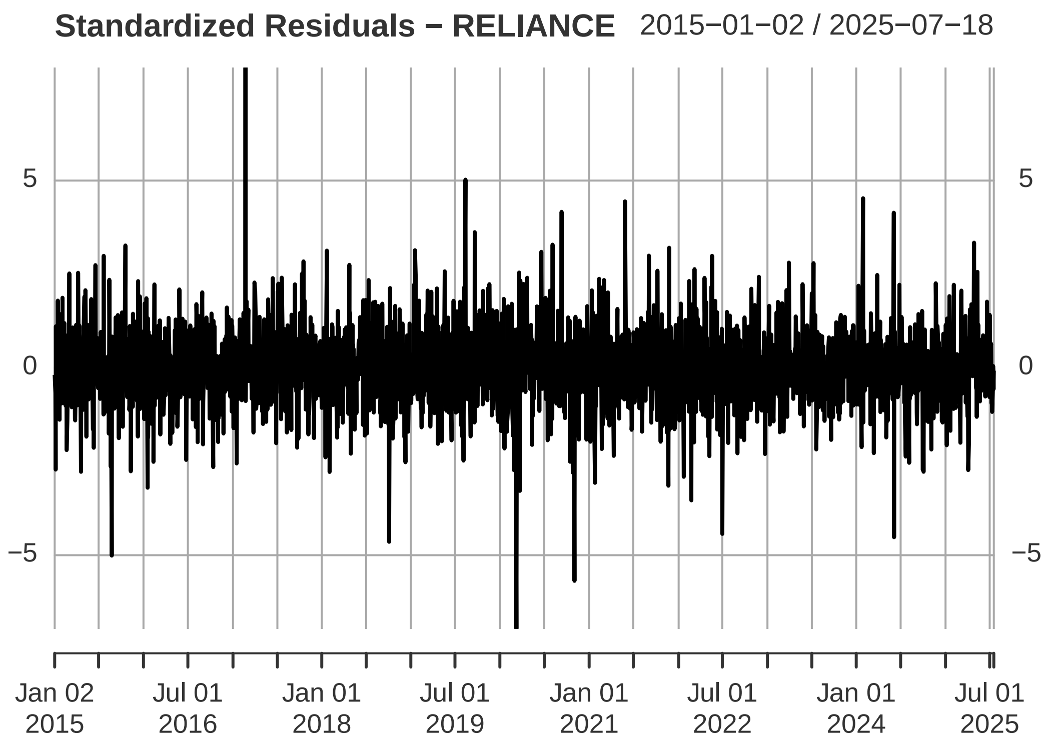
Residual Diagnostics – ARMA(0,0) Model (RELIANCE)

The ACF plot of residuals from the ARMA(0,0) model reveals mild autocorrelation across several lags.The Ljung-Box test yields a **p-value of 0.0024**, which strongly suggests that the residuals are not white noise.This confirms that the ARMA(0,0) model fails to fully capture the serial dependence in RELIANCE's return series.**Conclusion**: A more complex ARMA model is necessary to model the autocorrelation structure before proceeding to volatility modeling (GARCH).

Model Refinement: ARMA(3,2) for RELIANCE Returns

The improved model selected is ARMA(3,2), which includes higher-order AR and MA terms to better capture autocorrelation.The model has a lower AIC (-13706.83) than previous candidates, indicating improved fit.The Ljung-Box test on residuals now yields a p-value of 0.433, suggesting the residuals no longer exhibit significant autocorrelation.This validates that the ARMA(3,2) model adequately models the mean dynamics of RELIANCE returns.**Conclusion**: ARMA(3,2) is a suitable model for filtering serial correlation, and we can now proceed to model volatility using a GARCH specification.

0.5.2.2 Volatility Modeling: GARCH(1,1) with Student-t Errors for RELIANCE



Interpretation:

A GARCH(1,1) model was fitted with an ARMA(3,2) mean process and Student-t distribution to account for heavy tails in returns.The GARCH terms are highly significant: (ARCH) = 0.056, (GARCH) = 0.906, indicating strong volatility clustering, a typical feature of financial returns.The shape parameter (~5.43) confirms the presence of heavy tails.The standardized residuals plot shows reduced volatility clustering, supporting that the conditional heteroscedasticity has been modeled adequately.Ljung-Box and ARCH LM tests on residuals and squared residuals return high p-values, indicating no remaining serial correlation or ARCH effects.The model passes the stability tests, confirming good specification and parameter stability.**Conclusion**: The GARCH(1,1) model with Student-t innovations fits the RELIANCE returns well and produces uncorrelated, homoscedastic residuals ready for copula-based dependence modeling.

0.5.2.3 Transformation to Pseudo-Observations – RELIANCE To model dependence using copulas, we first need to convert the standardized residuals into uniform pseudo-observations on the [0,1] interval.This is done via rank transformation, which ensures the marginal distribution of each series is uniform

— a prerequisite for copula modeling.

0.5.3 Autocorrelation Analysis of INFY Returns

Interpretation: As shown in Figure, the ACF and PACF plots for INFY log-returns reveal:

- The ACF shows a sharp drop after lag 1, suggesting weak autocorrelation beyond immediate lags.
- The PACF contains several significant spikes, indicating potential autoregressive components.
- These results suggest that INFY returns may benefit from an ARMA model to adequately remove linear dependence before proceeding with GARCH modeling.

0.5.3.1 ARMA Modeling and Residual Diagnostics – INFY **Interpretation:** The automatically selected model for INFY is an ARIMA(0,0,0) with a constant mean (essentially a white noise model).Although the residual ACF plot appears flat, the Ljung-Box test p-value = 0.00344 indicates significant autocorrelation remains.This implies that the ARIMA(0,0,0) model does not sufficiently capture the dynamics of the INFY return series.**Conclusion:** We should fit a more complex ARMA(p,q) model to better eliminate autocorrelation and prepare the series for GARCH modeling.

0.5.3.2 Improved ARMA Model Selection for INFY **Interpretation:** The refined model selected is an ARIMA(3,0,2) with a non-zero mean.It significantly improves log-likelihood (6935.38) and AIC (-13856.76) compared to the initial ARIMA(0,0,0).Importantly, the Ljung-Box test p-value = 0.3671 indicates no significant autocorrelation in residuals — a strong sign that the model adequately captures serial dependence.**Conclusion:**This ARMA(3,2) model provides a much better fit for INFY log returns, ensuring the residuals are approximately white noise and making it suitable for GARCH modeling next.

0.5.3.3 ARMA-GARCH Modeling and Pseudo-Observations for INFY **Interpretation:** We fitted an ARMA(3,2) + GARCH(1,1) model with Student-t errors to capture both serial dependence and volatility clustering in INFY log returns.Model diagnostics confirm a good fit:Ljung-Box p-values > 0.9 on residuals and squared residuals indicate no autocorrelation or ARCH effects remain.All GARCH parameters are statistically significant or nearly so, and the Student-t shape parameter (~4.51) confirms heavier tails than Gaussian.The histogram of pseudo-observations (uniform[0,1]) shows a nearly flat distribution — a sign of properly modeled margins, suitable for copula modeling.

0.5.4 Autocorrelation Analysis: HDFCBANK Returns

As shown in Figure, the ACF and PACF plots for HDFCBANK log-returns reveal:

- A sharp drop in ACF after lag 1, indicating weak serial correlation beyond the first lag.
- Several scattered spikes in the PACF, suggesting short-term autoregressive structure.
- These features support fitting an ARMA-GARCH model to capture autocorrelation and volatility clustering.

0.5.4.1 ARMA(5,2) Model Estimation and Residual Diagnostics for HDFCBANK The ARMA(5,2) model selected by auto.arima() provides a good fit to the HDFCBANK return series, as reflected by a high log-likelihood and low AIC value.The ACF plot of residuals shows no significant autocorrelation, suggesting the residuals are white noise.The Ljung-Box test returns a high p-value (0.736), failing to reject the null hypothesis of no autocorrelation. This confirms the model has adequately captured the serial structure in the data.Overall, the ARMA(5,2) specification seems statistically valid and provides a suitable mean equation for the subsequent GARCH modeling.

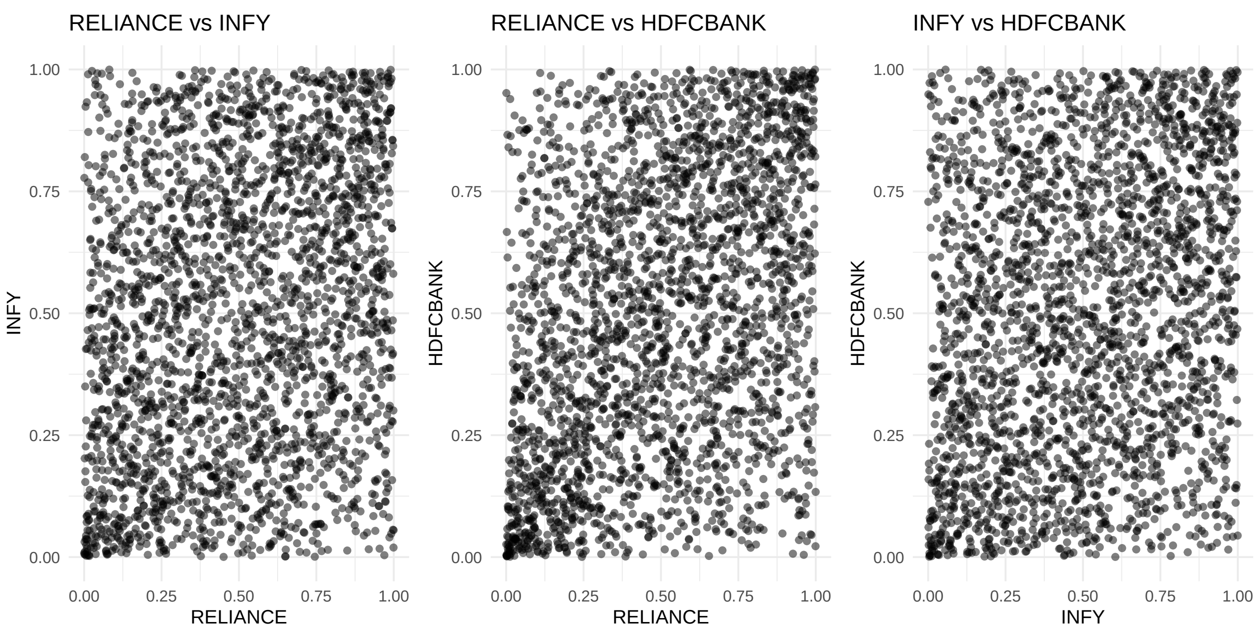
0.5.4.2 GARCH(1,1) Modeling with Student-t Innovations for HDFCBANK

The ARMA(5,2) structure was selected for the mean equation, as manually specified following earlier auto.arima suggestions. This structure aims to capture any remaining short-term autocorrelation in the returns.The GARCH(1,1) model shows strong evidence of volatility clustering, with significant GARCH parameters:$\alpha_1 = 0.0556$, $\beta_1 = 0.9191$, both highly significant. The persistence ($\alpha + \beta \approx 0.975$) confirms long memory in volatility.The innovations are modeled using the Student-t distribution, with estimated shape parameter 5.22, capturing heavy tails in the residuals.The standardized residuals plot shows mostly stable fluctuations around zero, supporting the adequacy of the fitted model.Ljung-Box tests on standardized and squared residuals show high p-values across lags (e.g., p = 0.999 for lag 34), suggesting no remaining serial correlation or ARCH effects.The pseudo-observations histogram appears close to uniform over [0,1], which is ideal for the upcoming copula-based dependence modeling. **Conclusion:The ARMA(5,2)–GARCH(1,1)–Student-t model** is well-fitted to HDFCBANK log-returns, producing standardized residuals that are iid and heavy-tailed—ideal inputs for copula construction.

0.6 Section b-Dependence Modelling

0.6.1 Visualizing Pairwise Dependence via Pseudo-Observation Scatterplots*

To gain an initial understanding of the joint dependence between the standardized innovations of the three financial return series (RELIANCE, INFY, and HDFCBANK), we plot scatterplots of their pseudo-observations. These plots allow for a nonparametric visual inspection of the dependence structure before any copula modeling.



Interpretation of Scatterplots

RELIANCE vs INFY: The scatterplot appears fairly uniform, with slight clustering near the corners.This suggests mild tail dependence, particularly in joint extreme movements (lower-left and upper-right corners).

RELIANCE vs HDFCBANK: The scatter shows a slightly stronger pattern of co-movement, especially along the diagonal.Indicates moderate positive dependence, potentially with stronger lower-tail co-movement.

INFY vs HDFCBANK: This pair appears visually similar to RELIANCE–INFY, with scattered points and slight density in tail regions.Suggests weak to moderate symmetric dependence. These plots serve as a precursor to formal tail dependence and copula fitting, guiding expectations about the types of copulas (e.g., t, Clayton, Gumbel) that may be appropriate in the next steps.

0.6.2 Dependence Measure Estimation: Kendall's Tau, Spearman's Rho & Tail Dependence

In this section, we quantify the strength and nature of dependence between the standardized returns of RELIANCE, INFY, and HDFCBANK using both rank-based correlation metrics and empirical tail dependence.

Kendall's Tau & Spearman's Rho These nonparametric measures are computed from the pseudo-observations derived after GARCH filtering. They assess monotonic relationships between asset pairs:

Pair	Kendall's Tau	Spearman's Rho
RELIANCE–INFY	0.15	0.22
RELIANCE–HDFCBANK	0.24	0.35
INFY–HDFCBANK	0.14	0.21

All values are positive, confirming weak to moderate positive dependence.The strongest dependence is observed between RELIANCE and HDFCBANK, consistent with visual inspections.The metrics provide a global summary, complementing localized insights from scatterplots.

0.6.3 Empirical Tail Dependence Estimation (λ_l, λ_u) Between Asset Pairs

Pair	λ_u (Upper Tail)	λ_l (Lower Tail)
RELIANCE–INFY	0.092	0.193
RELIANCE–HDFCBANK	0.146	0.285
INFY–HDFCBANK	0.116	0.200

Interpretation: Lower tail dependence is consistently stronger than upper tail, indicating that extreme losses tend to co-occur more than extreme gains among the stocks.RELIANCE–HDFCBANK shows the strongest tail dependence in both tails, again confirming earlier findings of stronger dependence between these two stocks.These insights support the choice of asymmetric copula families like Clayton (lower tail focus) or Gumbel (upper tail focus) for modeling joint distributions.

Pair	Clayton	Gumbel	Gaussian	**t-Copula**
RELIANCE – INFY	72.44	51.41	68.41	**77.62**
RELIANCE – HDFCBANK	188.05	146.65	181.68	**200.44**
INFY – HDFCBANK	66.18	51.05	65.04	**74.89**

Interpretation: In all three cases, the t-copula has the highest log-likelihood-Clayton is second-best, reflecting strong lower tail dependence ,Gumbel is worst in all cases $\rightarrow$ confirms weak upper tail behavior and Gaussian is not terrible, but doesn't account for tail risk

Interpretation: Very high p-value (0.97) $\rightarrow$ excellent fit

Suggests t-copula describes this pair's joint behavior almost perfectly

Pair	CvM Stat	p-value	Fit Verdict
RELIANCE – INFY	0.693	0.73	Valid
RELIANCE – HDFCBANK	1.268	0.52	Valid
INFY – HDFCBANK	0.070	0.97	Excellent

0.6.4 (Dependence Modeling) for all 3 stock pairs using copula-based techniques

Here's a quick recap of what i've achieved:

Pair	Best Copula	Log-Likelihood	GOF p-value	Key Tail Feature
RELIANCE–INFY	t-copula	77.62	0.73	Lower tail dependence ((λ_l) = 0.193)
RELIANCE–HDFCBANK	t-copula	200.44	0.52	Strongest tail dependence overall
INFY–HDFCBANK	t-copula	74.89	0.97	Similar to REL–INFY ((λ_l) = 0.193)

Visual Diagnostics: Empirical vs Simulated Pseudo-Obs Plots (all pairs)

Density Contours from Fitted Copula Densities-These plots show that your simulated t-copulas closely reproduce the observed dependence structure in the pseudo-observations — a very strong validation.

Interpretation: The stocks do not move independently. There is weak to moderate dependence, especially in extreme losses (lower tail). RELIANCE–HDFCBANK exhibit the strongest co-movement in extreme losses, suggesting systemic risk. t-copula consistently performs best — makes sense due to its flexibility in capturing both symmetric and tail dependence.