VU VRIJE UNIVERSITEIT AMSTERDAM

Vrije Universiteit Amsterdam

Academic Year 2024-2025

Financial Econometrics Case Study

Group 4

Authors:

Abe Tempelman	a.w.r.tempelman@student.vu.nl
Jiaxuan Zhu	j.zhu8@student.vu.nl
Yunji Eo	y.eo@student.vu.nl
Robin Klaasen	r.l.klaasen@student.vu.nl
Vân Lê	k.v.le@student.vu.nl

Abstract

This paper explores advanced econometric strategies for modeling, measuring and forecasting financial volatility in high-frequency stock data from Cisco Systems, Inc. between 2018 and 2025. We apply multiple GARCH and GAS models. The realized kernel is incorporated to account for the microstructure effects of the markets. The analysis uses an in-sample period from 2018 to 2023 for model calibration and an out-of-sample period from 2023 to 2025 for forecast evaluation. For forecasting, the Realized Kernel GAS (Normal) model performs best in loss functions, while the Realized Kernel GARCH (Student-t) model better captures dynamic trends. The choice of the best model depends on whether the primary goal is minimizing differences in volatility or aligning with observed market trends.

Keywords: Realized Kernel, GARCH, GAS, Distributions, AIC and BIC, MCMC, RiskMetrics, Loss Functions, Diebold-Mariano, DTW

January 31, 2025

Contents

1 Introduction

This study investigates advanced econometric methods for modeling, measuring, and forecasting financial volatility using high-frequency stock data from Cisco Systems, Inc. between January 5, 2018, and January 3, 2025. Cisco Systems was chosen because it is a well-known technology company that focuses on networking, cybersecurity, and IT infrastructure. It has a large market capitalization and high trading volume, making it a suitable candidate for volatility analysis. Cisco also has a strong presence in Europe, including the Netherlands, where it works with local businesses and operates research facilities. Additionally, recent volatility in the tech sector makes Cisco an appropriate case study for examining high-frequency volatility modeling.

Before delving into the analysis, it is essential to establish a clear understanding of the concept of volatility. According to Investopedia (2024), volatility is "A statistical measure of the dispersion of returns for a given security or market index." Understanding financial volatility is essential because it acts as a critical measure of market uncertainty and risk. It enables investors to optimize their portfolios, institutions to manage risks effectively, and policymakers to monitor and stabilize economic conditions. Moreover, volatility plays a central role in the forecast of price movements, making it a vital tool for traders and analysts seeking to navigate financial markets.

This paper is structured as follows: first, a description of the data set is shown, and the data cleaning process is explained, followed by a preliminary analysis. Then we follow the implied methods, models, and results.

The methodology begins by using the realized kernel to refine intraday volatility estimates, accounting for market microstructure effects. A range of GARCH-type and GAS-type models is considered, including standard GARCH, GJR-GARCH, EGARCH, Realized GARCH, and Realized GAS models. These models are applied under Normal, Student-t, and Skewed Student's t-distribution. Additionally, MCMC methods are explored for distributional innovations in select cases to enhance the model's flexibility and robustness. The RiskMetrics model is used as a benchmark for forecasting performance.

The model parameters are estimated using log-likelihood maximization, with some models incorporating MCMC methods for parameter estimation. Model selection is guided by the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Forecast accuracy is evaluated using loss functions such as the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), allowing for out-of-sample comparisons between forecasts and actual observations. Since forecast errors can be close in magnitude, the Diebold-Mariano test is applied to assess whether one model statistically outperforms another. Additionally, Dynamic Time Warping (DTW) distance is introduced as a complementary metric to evaluate trend alignment and temporal dynamics, providing a broader perspective beyond pointwise differences. The weight assigned to each metric depends on whether the focus is on minimizing numerical volatility differences or capturing trend consistency.

2 Data Description and Preprocessing

2.1 Description Data Set

For this paper, the analysis focuses on Cisco Systems, Inc., identified by the stock ticker CSCO. Cisco Systems, Inc. is a leading global technology company listed on multiple exchanges. The study uses high-frequency trading data spanning January 5, 2018, to January 3, 2025. The data were obtained from the Wharton Research Data Services (WRDS) platform, which provides comprehensive access to detailed market information.

Table 1 presents the descriptive statistics of the in-sample and out-of-sample log return data, offering insights into return behavior over the observation period. The in-sample dataset consists of 1,257 daily return observations from January 5, 2018, to January 3, 2023, while the out-of-sample dataset includes 503 daily return observations from January 4, 2023, to January 3, 2025. This dataset structure provides a comprehensive view of market behavior over time.

The descriptive statistics highlight important characteristics of the log returns. Both datasets exhibit negative skewness, which suggests that large negative returns occur more frequently than large positive returns. The kurtosis values, particularly 10.31 for the in-sample data and 13.11 for the out-of-sample data, reveal extreme outliers and fat tails, a common trait in the financial return series. These statistical properties highlight the importance of using advanced econometric models capable of capturing non-normality, ensuring robust volatility estimation and forecasting.

Table 1: Descriptive Statistics (In-Sample and Out-Sample Log Returns)

Metric	**In-Sample Log Return**	**Out-Sample Log Return**
Count	1257	503
Mean	0.0164	0.0419
Standard Deviation	1.8537	1.2194
Minimum	-14.9918	-10.4567
25%	-0.8007	-0.5350
50% (Median)	0.0514	0.0665
75%	0.8954	0.6871
Maximum	12.5418	6.5635
Skewness	-0.7367	-1.1232
Kurtosis	10.3133	13.1059

2.2 Data Cleaning

The high-frequency dataset underwent a rigorous cleaning process to mitigate microstructure noise and ensure data reliability for subsequent analysis. The methodology closely adhered to the framework outlined by Barndorff-Nielsen et al. (2009), which provides a robust foundation for handling high-frequency financial data.

Firstly, as prescribed by step P1, the dataset was restricted to official trading hours (09:30 to 16:00 EST). This restriction ensured that the analysis was focused on periods of active trading, thereby minimizing distortions from after-hours or pre-market activity. Trades specific to CSCO were identified, and the NASDAQ exchange, being the most active for this stock, was prioritized in line with P3. Data from less active exchanges were excluded to maintain consistency and reduce noise.

Secondly, entries with anomalous values, such as zero prices for bids, asks, or trades, were eliminated following P2. Additionally, duplicate records were removed to enhance the overall integrity of the dataset.

For trade data, further refinements were applied. T1 was implemented to exclude corrected trades, which are often flagged as adjustments or erroneous entries. Subsequently, trades with abnormal sale conditions were filtered out using T2, ensuring that only valid and reliable transactions were retained for analysis.

To address instances of multiple trades occurring at identical timestamps, the median price was computed in accordance with T3. This aggregation step was crucial in reducing the influence of outliers and ensuring that the resulting dataset accurately represented the central tendency of trading activity.

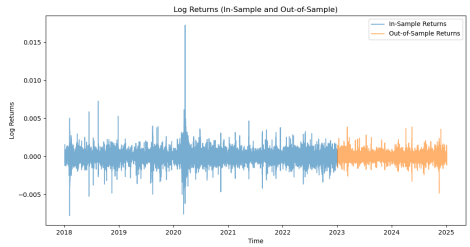
2.3 Data Preliminary Analysis

Figures 1a and 1b provide an overview of the in-sample and out-of-sample datasets for Cisco Systems, Inc. (CSCO). These visualizations highlight the key temporal dynamics of the dataset and offer insights into its structure and characteristics, including gaps resulting from non-trading days such as weekends and holidays.

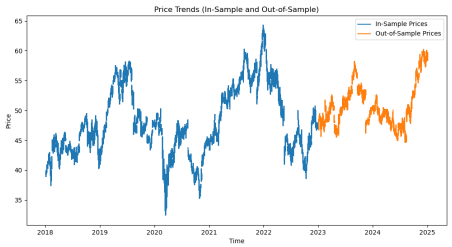
Figure 1b illustrates the temporal evolution of CSCO stock prices over the observation period. The figure reveals distinct patterns, including periods of sustained growth, declines, and volatility spikes. The visible gaps in the data correspond to non-trading days and are a natural feature of financial datasets. Separating the data into in-sample and out-of-sample segments enables rigorous model evaluation and ensures the reliability of forecasting methods.

Figure 1a depicts the logarithmic changes in prices, emphasizing the behavior of returns over time. The plot reveals significant periods of volatility clustering, where high-volatility periods are followed by similarly turbulent intervals. Notably, the large spike observed in 2020 corresponds to the extreme market volatility triggered by the COVID-19 pandemic, which led to substantial fluctuations in Cisco Systems' stock returns.

Figures 2a and 2b further analyze the temporal dependencies in the return series by presenting the autocorrelation functions (ACFs) for the in-sample and out-of-sample datasets, respectively. Figure 2a shows the in-sample autocorrelation structure, where significant short-term dependencies are evident at lower lags. Similarly, Figure 2b highlights the out-of-sample autocorrelation structure, which mirrors the in-sample behavior, confirming the consistency of temporal patterns across datasets.

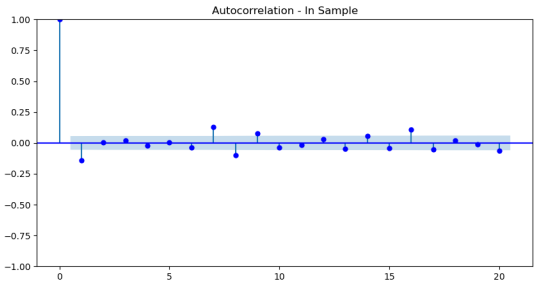


(a) Log Returns (In-Sample and Out-of-Sample). This figure illustrates the log returns calculated from high-frequency price data for both in-sample and out-of-sample datasets.

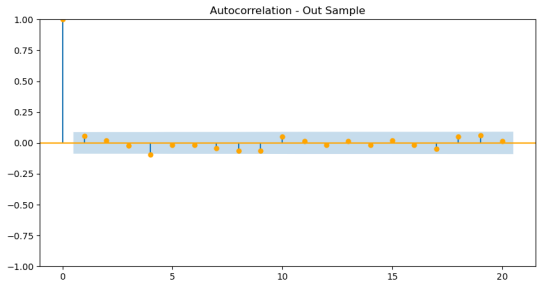


(b) Price of CSCO (In-Sample and Out-of-Sample). This figure shows the price movements of the in-sample and out-of-sample data. It shows strong price trends.

Figure 1: Side-by-side comparison of Log Returns and Price for CSCO.



(a) Autocorrelation of In-Sample Returns.



(b) Autocorrelation of Out-of-Sample Returns.

Figure 2: Autocorrelation of Returns for CSCO (In-Sample and Out-of-Sample). These plots provide insights into the temporal dependencies in the returns.

3 Methodology

This study employs econometric methods tailored to model and forecast financial volatility using high-frequency data. To address challenges such as microstructure noise and irregularities in trading data, we rely on realized measures of volatility, including the realized kernel. Our analysis further incorporates GARCH-type and GAS-type models under various distributional assumptions, such as Normal, Student-t, and Skewed Student-t, to capture features like volatility clustering, asymmetry, and heavy tails. These approaches ensure a robust framework for estimating and analyzing volatility.

3.1 Realized Kernel

The realized kernel is a robust estimator of daily volatility, designed to account for microstructure noise in high-frequency financial data. It aggregates weighted autocovariances of log returns over selected lags, using a kernel function to assign diminishing weights to further lags. We selected the Parzen kernel due to its non-negative estimates and smooth properties, which ensure that $k'(0) = k'(1) = 0$. The Parzen kernel is defined as:

$$k(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & \text{if } 0 \leq |x| \leq \frac{1}{2}, \\ 2(1 - |x|)^3 & \text{if } \frac{1}{2} < |x| \leq 1, \\ 0 & \text{if } |x| > 1. \end{cases}$$

The realized kernel estimator is computed as:

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \quad \text{where } \gamma_h = \sum_{j=|h|+1}^{n} x_j x_{j-|h|}.$$

Here, γ_h is the autocovariance at lag h, and H represents the bandwidth, determining the number of lags considered. The Parzen kernel's weighting ensures that further lags contribute less to the overall estimate, mitigating the impact of noise.

In our implementation, the bandwidth H is dynamically computed based on the data for each day. Following Barndorff-Nielsen et al. (2009) defined method, the formula for H is given by:

$$H^* = c^* \left(\frac{\omega^2}{IV}\right)^{2/5} n^{3/5},$$
$$\text{where } H^* \text{ is rounded to a integer } H.$$

where $c^* = 3.5134$ is a constant specific to the Parzen kernel, ω^2 is the noise variance, IV represents the integrated variance, and n is the number of non-zero returns. The integrated variance (IV) is approximated using sparse realized variance (RV_{sparse}), which is computed by sampling log returns at 20-minute intervals:

$$RV_{\text{sparse}} = \frac{1}{M}\sum_{m=1}^{M}\sum_{t\in\mathcal{T}_m} r_t^2,$$

where $\mathcal{T}_m$ represents the staggered sampling intervals over the trading day, and M is the number of intervals.

The noise variance ω^2 is estimated using dense realized variance (RV_{dense}), calculated over trade-based subsets:

$$\omega^2 = \frac{1}{q}\sum_{i=1}^{q}\frac{RV_{\text{dense}}^{(i)}}{2n^{(i)}},$$

where q denotes the number of trade-based subsets, and $n^{(i)}$ is the count of observations within the i-th subset. For each subset, the dense realized variance is computed as:

$$RV_{\text{dense}}^{(i)} = \sum_{t\in\mathcal{T}_i} r_t^2,$$

where $\mathcal{T}_i$ represents the subset of returns sampled at every q^{th} trade.

This dynamic computation of H ensures that the bandwidth optimally balances the reduction of microstructure noise with the retention of information from the efficient price process. Larger datasets with higher n and lower noise variance ω^2 lead to higher H, while sparser datasets or noisier data result in lower H.

3.2 Distributions

We analyze three distributions to evaluate their performance: the normal distribution, which is symmetric and lacks heavy tails, making it the simplest and least robust for extreme values; the Student's t-distribution, which is also symmetric but incorporates heavy tails to handle extreme observations; and the skewed Student's t-distribution, offering the most flexibility with asymmetry and heavy tails to capture skewed data and extreme variations. This progression allows us to determine which distribution best fits the data.

3.2.1 Normal Distributions

The probability density function is given by:

$$p(r_t) = (2\pi\sigma_t)^{-1/2} \exp\left(-\frac{(r_t - \mu_t)^2}{2\sigma_t^2}\right),$$

and the natural logarithm of the density function is:

$$\log p(r_t) = -\frac{1}{2}\log(2\pi\sigma_t) - \frac{(r_t - \mu_t)^2}{2\sigma_t^2}.$$

3.2.2 Student's t-distribution

The probability density function is given by:

$$p(r_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left((\nu-2)\pi\sigma_t^2\right)^{-1/2} \left(1 + \frac{r_t^2}{(\nu-2)\sigma_t^2}\right)^{-\frac{\nu+1}{2}}.$$

The natural logarithm of the density function is:

$$\log p(r_t) = \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\left((\nu-2)\pi\sigma_t^2\right) - \frac{\nu+1}{2}\log\left(1 + \frac{r_t^2}{(\nu-2)\sigma_t^2}\right).$$

3.2.3 Skewed Student's t-distribution

The probability density function is given by:

$$p(r_t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} (\nu-2)^{-\frac{1}{2}} \times s \times \left(\frac{2}{\xi + \frac{1}{\xi}}\right) \times \left[1 + \frac{\left(\frac{r_t}{\sigma_{t,\nu}} + m\right)^2}{\nu - 2}\right]^{-\frac{\nu+1}{2}},$$

where:

$$I_t = \begin{cases} 1 & \text{if } \frac{r_t}{\sigma_{t,\nu}} + m \geq 0 \\ -1 & \text{if } \frac{r_t}{\sigma_{t,\nu}} + m < 0 \end{cases},$$

$$m = \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \times \sqrt{\frac{\nu-2}{\pi}} \times \left(\frac{\xi - \frac{1}{\xi}}{2}\right),$$

$$s = \sqrt{\left(\xi^2 + \frac{1}{\xi^2} - 1\right) - m^2}.$$

The natural logarithm of the density function is:

$$\begin{aligned}\log p(r_t) = & \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log\left((\nu-2)\pi\sigma_t\right) \\ & + \log(s) + \log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) \\ & - \frac{\nu+1}{2}\log\left(1 + \frac{\left(s\frac{r_t-\mu_t}{\sqrt{\sigma_t}} + m\right)^2}{\nu-2}\xi^{-2I_t}\right).\end{aligned}$$

3.3 Model Types

3.3.1 GARCH

GARCH (Generalized Autoregressive Conditional Heteroskedasticity) proposed by Bollerslev, 1986 is a statistical model used to estimate and forecast the volatility of time series. It extends the ARCH (Autoregressive Conditional Heteroskedasticity) model by incorporating lagged conditional variances. The model is defined as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

In this model, σ_t^2 represents the conditional variance at time t, ω is a constant term reflecting the long-term average variance, α_i are the coefficients of the i-th lag of squared residuals (ϵ_{t-i}^2), which represents past innovations, and β_j are the coefficients of the j-th lag of conditional variance (σ_{t-j}^2). The parameters p and q denote the number of lagged conditional variances and squared residuals included in the model, respectively. To ensure that the conditional variance remains non-negative, we impose the following restrictions : $\omega > 0$, $\alpha \geq 0$, $\beta \geq 0$. We choose the GARCH(1,1) model because it is the simplest yet highly effective model to capture the clustering of volatility with minimal parameters.

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

3.3.2 GJR-GARCH

Another model we can use is the GJR-GARCH model proposed by Glosten et al., 1993, which is an extension of the GARCH model that incorporates asymmetric effects of positive and negative shocks on volatility. We again use the minimal parameter, and the model GJR-GARCH(1,1) is defined as:

$$\sigma_t^2 = \omega + \alpha\epsilon_{t-1}^2 + \gamma I_{t-1}\epsilon_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where:

$$I_{t-1} = \begin{cases} 1, & \text{if } \epsilon_{t-1} < 0 \text{ (negative shock)} \\ 0, & \text{if } \epsilon_{t-1} \geq 0 \text{ (positive shock)} \end{cases}$$

The model incorporates asymmetry in the volatility process described by the parameter γ. When $\gamma > 0$, the model captures the leverage effect, where negative shocks to returns lead to higher volatility than positive shocks of the same magnitude. When $\gamma = 0$, the GJR-GARCH model reduces to the GARCH model. For $\gamma < 0$, the negative shocks result in smaller future volatility. To ensure that the conditional variance remains non-negative in the GJR-GARCH model, we impose the following restrictions:

$$\omega > 0, \quad \alpha \geq 0, \quad \beta \geq 0, \quad \gamma \geq 0.$$

3.3.3 EGARCH

The Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH), proposed by Nelson (1991), is an extension of the GARCH model that captures asymmetry and leverage effects. The conditional variance for EGARCH is defined as:

$$\log(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - \mathbb{E}(|z_{t-1}|)) + \gamma z_{t-1} + \beta\log(\sigma_{t-1}^2),$$

where,

$$\mathbb{E}[|z_{t-1}|] = \begin{cases} \sqrt{\frac{2}{\pi}}, & \text{if } z_t \sim \mathcal{N}(0,1), \\ 2\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, & \text{if } z_t \sim t(0,1,\nu), \\ \frac{4\xi^2}{\xi+\xi^{-1}}\sqrt{\frac{\nu-2}{\pi}}\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, & \text{if } z_t \sim \text{Skew-}t(0,1,\nu,\xi). \end{cases}$$

Unlike GARCH, EGARCH models the logarithm of the variances, which ensures that the variance is always positive without requiring non-negativity constraints on the parameters. EGARCH is highly flexible, allowing for innovations from Normal, Student-t, or Skewed Student's t-distribution, making it effective in capturing heavy tails and volatility clustering.

3.3.4 Realized GARCH

Realized GARCH model presented by Hansen et al. (2011) is defined as:

$$r_t = \sqrt{h_t} z_t,$$

$$\log h_t = \omega + \beta \log h_{t-1} + \gamma \log x_{t-1},$$
$$\log x_t = \xi + \varphi \log h_t + \tau_1 z_t + \tau_2 \left(z_t^2 - 1\right) + u_t,$$

$$z_t \sim i.i.d.(0,1), \quad u_t \sim i.i.d.(0, \sigma_u^2),$$

where γ represents an impact of the realized measure on conditional variance, ϕ captures the sensitivity of the realized measure to the latent variance, τ_1 and τ_2 account for leverage and nonlinear effects. The variable x_t represents the realized measure such as Realized Kernel or Realized Volatility. We assume that the inclusion of this realized measure enhances the estimation of the volatility estimation for h_t.

3.3.5 GAS model

One of the more advanced econometric models is the Generalized Auto-regressive Score (GAS) model researched by Creal et al. (2013). Similar to GARCH models, the GAS model is also an observation-driven model with a time-varying parameter. The model is based on a score updating mechanism, where GAS models look at the entire conditional density of the distribution and not just at specific moments of the distribution of observations r_t. In the univariate setting, let $N \times 1$ vector r_t be the dependent variable of interest, and its conditional density is as follows:

$$r_t \sim p(r_t | f_t, \mathcal{F}_t; \theta),$$

where r_t depends on a time-varying parameter vector f_t, a vector of exogenous variables $\mathcal{F}_t$, and a vector of time-invariant parameters θ. In addition, it is summarized that $R_t = r_1, \ldots, r_t$, $\mathcal{F}_t = f_0, f_1, \ldots, f_t$, and $X_t = x_1, \ldots, x_t$ where $\mathcal{F}_t = \{R^{t-1}, \mathcal{F}^{t-1}, X^t\}$ is the set of information for time $t = 1, \ldots, n$. The mechanism for auto-regressive updating f_t is assumed to be given as follows:

$$f_{t+1} = \omega + \alpha s_t + \beta f_t, \quad \textit{with} \quad s_t = S_t \times \nabla_t \quad \textit{and} \quad \nabla_t = \frac{\partial \ln p(r_t | f_t, \mathcal{F}_t; \theta)}{\partial f_t}.$$

The updating equation f_{t+1} is computed by using the obtained R_t in order to update the time-varying parameters over time; this is done by the scaled score function s_t, where ∇_t is the first-order derivative of the log-likelihood function with respect to the parameter. In this study, $S_t = 1$ such that $s_t = \nabla_t$ and $f_t = \log(h_t)$ such that

$$\nabla_t = \frac{\partial \ln p(r_t | f_t, \mathcal{F}_t; \theta)}{\partial h_t} \times h_t.$$

The choice of the scaling matrix S_t allows for flexibility in the use of the updating equation and determines the GAS(p, q) for which p and q the orders of the GAS model as shown by Creal et al. (2013). To extend the GAS model to the realized GAS model (R-GAS), we incorporate realized measures of the daily returns. The updating equation results as follows:

$$f_{t+1} = \omega + \beta f_t + \alpha\Big(\frac{\nu_1}{2}\Big(\frac{X_t}{\exp(f_t)} - 1\Big) + \nabla_t,$$

where X_t is referred to as the Realized Kernel at time t that is mentioned in Section 3.1. The Realized Kernel is assumed to follow the Gamma($\frac{\nu_1}{2}, \frac{2\exp(f_t)}{\nu_1}$) density, and the term $\left(\frac{\nu_1}{2}\left(\frac{X_t}{\exp(f_t)} - 1\right)\right.$ is the derivative of the logarithm of the Gamma density for Realized Kernel X_t w.r.t. $f_t = \log(f_t)$. Furthermore, ∇_t is the same derivative in the GAS model, i.e. our returns R_t have some distribution $y_t \sim D(\mu_t, \exp(f_t))$.

3.3.6 RiskMetrics Model

The RiskMetrics model is a simple method for estimating volatility and is commonly used as a benchmark in financial studies. It calculates volatility using an exponentially weighted moving average (EWMA) of past squared returns. The formula is given by:

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2,$$

where σ_t^2 is the estimated volatility at time t, r_{t-1}^2 is the squared return from the previous period, λ is the smoothing parameter, typically set to 0.94 for daily data.

The RiskMetrics model is often used as a benchmark because it is simple and easy to compute. It gives more weight to recent returns and less weight to older ones, which aligns with the time-varying nature of financial markets. Although it is less sophisticated than other models, it provides a solid baseline for comparing the performance of more advanced volatility estimation methods.

3.4 Log-Likelihood, AIC and BIC

In order to estimate the parameters of all models mentioned above, we maximize the log-likelihood function, which evaluates the probability of the corresponding density distribution given the data. The log-likelihood provides a basis for estimating the model parameters efficiently. To assess model performance and avoid overfitting, we evaluate the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), whereas the BIC penalizes additional parameters introduced through the term $\log(T) \times k$, where k is the number of parameters, and AIC penalizes the additional term by $2 \times k$. The formulas for the AIC and BIC can be found in Appendix A.

3.5 Bayesian Approach

Bayesian estimation is a statistical approach that incorporates prior beliefs about the parameters of a model and updates these beliefs using observed data to produce posterior distributions. Unlike traditional estimation methods (like maximum likelihood estimation), which provide single point estimates, Bayesian estimation gives a full distribution of possible parameter values, offering a more comprehensive view of uncertainty.

The posterior distribution is used to update prior beliefs based on the observed data, providing a full distribution over the parameter space instead of a point estimate. The posterior distribution of the parameters $\Theta = (\omega, \alpha, \beta, \nu)$ is given by Bayes' Theorem:

$$p(\Theta \mid \text{data}) \propto p(\text{data} \mid \Theta) \cdot p(\Theta).$$

The acceptance probability for a candidate parameter Θ^* in the Metropolis-Hastings algorithm is defined as:

$$\alpha = \min\left(1, \frac{p(\Theta^* \mid \text{data})}{p(\Theta \mid \text{data})}\right).$$

3.6 Model Accuracy

3.6.1 Loss Functions

To evaluate the forecast accuracy of the models, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used. The loss functions measure the accuracy of the models, where we compare the out-of-sample observations and the forecast values. The lower the loss functions, the better the prediction. MSE penalizes outliers heavily by squaring deviations, making it suitable when minimizing large errors is crucial, but its sensitivity to noise or anomalies can be a drawback. RMSE offers an interpretable error in the same units as the data and emphasizes large deviations similarly. However, RMSE might not be the best choice

in scenarios where penalizing under-estimations more than over-estimations is required. MAE handles errors linearly using absolute deviations, making it more robust to outliers. MAPE measures relative percentage error for easy comparison. Hence, by looking at these four loss functions, we can evaluate the performance of the models with these characteristics in mind. The formula of RMSE, MSE, MAE, and MAPE can be found in Appendix A.

3.6.2 Diebold-Mariano test

The metric values alone cannot determine whether one model outperforms another, as these metrics are inherently random by nature. When the metrics for two models are close, it becomes important to assess whether their differences are statistically significant. The Diebold-Mariano (DM) test provides a formal method to compare the predictive accuracy of two forecasting models by evaluating the statistical significance of their differences (Diebold & Mariano, 1995). This test is particularly useful for examining whether one model significantly outperforms the other in terms of forecast accuracy. The DM test is well-suited for forecast errors that may be non-Gaussian, have a non-zero mean, or exhibit serial and contemporaneous correlation (Diebold & Mariano, 1995). In this context, the null hypothesis assumes no difference in forecast accuracy between the two models, while the alternative hypothesis posits that one model performs better. When applied at a 5% significance level, rejecting the null hypothesis indicates that the forecasting error differences are statistically significant. This allows us to conclude whether the performance of one model is demonstrably better than the other. The formula of the DM test can be found in Appendix A.

3.6.3 Dynamic Time Warping (DTW) distance

Dynamic Time Warping (DTW) is a technique used to measure similarity between two temporal sequences, allowing for non-linear alignments in the time dimension (Bringmann et al. (2024)).

The DTW distance is then defined as the total cost along this optimal path. Formally, for two sequences

$$X = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad Y = (y_1, y_2, \ldots, y_m),$$

the DTW distance is:

$$DTW(X, Y) = \min \sum_{(i,j) \in P} d(x_i, y_j),$$

where P is the set of all possible alignments, and $d(x_i, y_j)$ is the distance between points x_i and y_j, typically calculated as the squared Euclidean distance.

4 Realized Kernels Estimation

Figure 3 compares daily RV and RK estimates for Cisco Systems, Inc. from in-sample data.

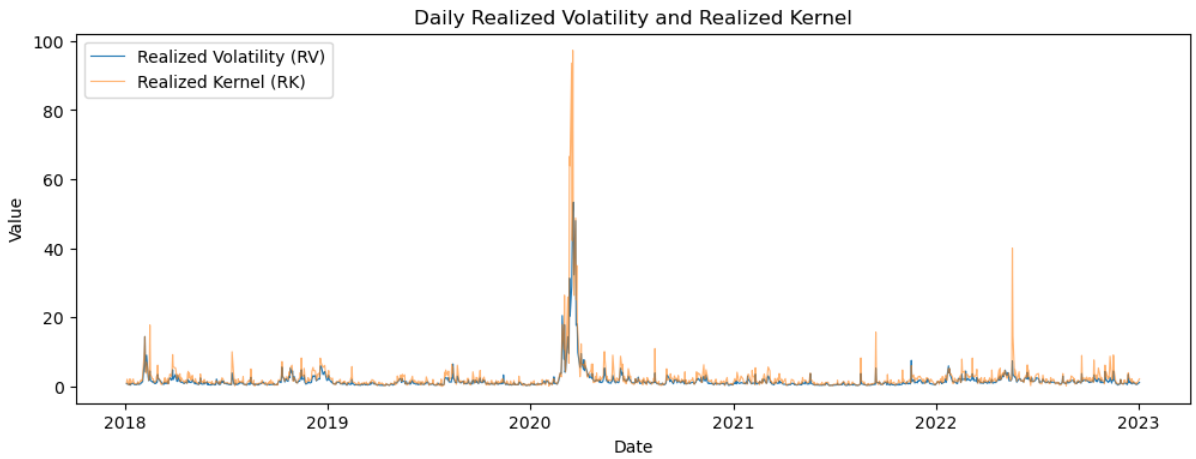


Figure 3: Daily Realized Volatility and Realized Kernel from 2018 until 2023

As seen in the figure above, notable spikes occur at the onset of 2020 and other periods of heightened market stress. As shown above, realized kernel has sharper peaks. Notable spikes in the realized kernel align with periods when daily returns are more volatile (wild) and are lower when daily returns are calmer. Table 2 highlights these differences.

Table 2: Statistics for Daily Realized Volatility (RV) and Realized Kernel (RK) from 2018 until 2023

Statistic	**RV Value**	**RK Value**
Sample Size	1258	1258
Minimum	0.2775	0.0447
Maximum	53.3245	97.2511
Average	1.7154	2.4916
Standard Deviation	3.4994	5.6092

As indicated in the table above, the realized kernel (RK) reaches a maximum of 97.2511, significantly higher than the realized volatility (RV) maximum of 53.3245, underlining the substantial difference between these two volatility measures. Both these peaks are primarily driven by the extreme market volatility observed during the 2020 peak, likely corresponding to the financial turmoil caused by the COVID-19 pandemic.

Figure 4 compares daily RV and RK estimates for Cisco Systems, Inc. from out-of-sample data.

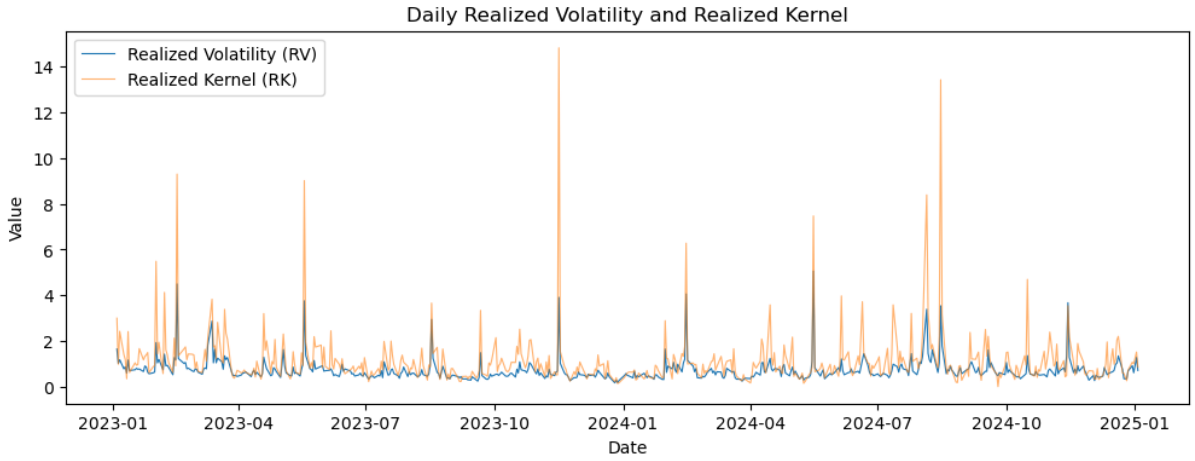


Figure 4: Daily Realized Volatility and Realized Kernel from 2023 until 2025

In the figure above, similarly as in figure 3 it is revealed that RK tends to exacerbate noise-induced jumps more effectively than RV, indicated by the smaller jumps of the RV. To clarify this, Table 3 is considered.

Table 3: Statistics for Daily Realized Volatility (RV) and Realized Kernel (RK) from 2023 until 2025

Statistic	**RV**	**RK**
Sample Size	503	503
Minimum	0.1871	0.0224
Maximum	5.0604	14.8004
Average	0.7562	1.1894
Standard Deviation	0.5226	1.2972

As seen in the table above and similarly as in Table 2 the maximum of the RK indicated by 14.8004 is much larger than the maximum of the RV indicated by 5.0604.

5 Empirical Results

5.1 Model Estimation (in-sample data)

This section discusses the estimation of our model for in-sample data estimation. Table 4 presents the estimated parameters for various models. Across all GARCH models, such as GARCH-Norm ($\beta = 0.8552$ and $\alpha = 0.1249$), the parameters α and β indicate high volatility persistence ($\alpha + \beta \approx 1$). Reflecting the clustering of volatility. Present in Skewed Student's t-distribution, ξ measures the asymmetry in the data, capturing whether the distribution is skewed to the left or right. In most models, the skewness is close to 1, suggesting a slight right-skewed distribution. Found in Student-t and Skewed Student's t-distribution, degrees of freedom ν captures the heaviness of the tails. Most models under these distributions show around 4.5 degrees of freedom, which demonstrates moderate tail heaviness.

Table 5 shows the estimated parameters for the Realized GARCH model. The Realized GARCH model shows a lower value of β (volatility persistence) than GARCH models. This confirms that the reliance on β decreases when we incorporate realized measures as an additional explanatory variable through the parameter γ. The Realized GARCH with Realized Volatility exhibited a larger value for degrees of freedom than the GARCH model (e.g., $\nu = 9.2580$ for R-GARCH-RK-SSTD compared to $\nu = 4.2981$ for GARCH-SSTD). This suggests the R-GARCH-RK models have thinner tails and a less extreme kurtosis in the return distribution than the classic GARCH models. Hence, including realized measures reduces the reliance on heavy tails. The improvement of the Realized GARCH model is evident through the criteria information, as it achieves higher log-likelihood values and lower AIC/BIC scores. This highlights that incorporating high-frequency information enhances the model's fit and robustness.

The estimates of GAS with all distributions are in line with the expectation, which is $\hat{\alpha}$ close to 0 and β closer to one. This means that the effect volatility at time $t-1$ is more important for the volatility at time t than the return at time $t-1$. GAS with Student-t and Skewed Student's t-distribution have slightly lower $\hat{\beta}$ compared to the one of Normal distribution and $\hat{\beta}$ of the Student-t and Skewed Student-t are very similar, indicating the returns are not necessarily asymmetric distributed. The estimates of $\hat{\omega}$ are all negative, while the skewed student-t and student-t have more negative $\hat{\omega}$ indicating the weight on volatility is lower when compared to the normal distribution. Furthermore, when maximizing the log-likelihood, we encountered some large outlier estimations; to ensure that the log-likelihood can be optimized correctly, we added a regularization log-term $L2 = \sum_{t=1}^{T} \exp(f_t)^2$ in the log-likelihood function to regulate the outliers. When we include the realized kernels into the realized GAS, we notice that $\hat{\alpha}$ and $\hat{\beta}$ coefficients are reduced, but similar to GAS, normal distribution still takes more weight on the volatilities compared to the other two distributions. In addition, we have parameter ν_X of the realized kernel that follows the Gamma distribution; the estimates are as expected that it is much higher, meaning the realized kernels are heavily more weighted compared to the returns.

5.1.1 Bayesian Approach

Parameter	Posterior Mean	95% Credible Interval
Ω	0.000031	[0.000014, 0.000060]
α	0.1956	[0.1125, 0.3062]
β	0.7311	[0.5606, 0.8396]
ν	4.0944	[3.8560, 4.3201]

We employed a Bayesian approach to the classic GARCH(1,1) model. We observed that the posterior distributions for all parameters (Omega, Alpha, Beta, and Nu) are well-formed, with credible intervals providing a measure of uncertainty. The acceptance rate of 35.96% is within the optimal range (20-40%), indicating efficient sampling. The trace plots mostly show good mixing, indicating a well-behaved Markov chain. However, degree of freedom (ν) seems to drift a bit, suggesting that the chain might not be exploring its full range as effectively as the others. This could mean the sampler needs a larger step size or more iterations to fully capture the uncertainty in the degree of freedom. Still, the Bayesian approach gives solid parameter estimates with well-calibrated uncertainty, making it a reliable tool for inference in this GARCH model.

5.2 Model Forecasting (out-of-sample performance)

This section discusses the forecasting of the daily volatility using the estimation results above and evaluates the forecast performances of the out-of-sample data.

5.2.1 Forecasting Results

Figure 5 shows all the forecasting results for all 20 models. The meaning of model abbreviations can be seen in Appendix. B.

Significant volatility spikes occur around late 2023, likely corresponding to major market events. Most models follow similar patterns, but the size of the spikes varies, showing differences in how each model estimates volatility. The realized kernel is also included for comparison.

5.2.2 Model Performance Metrics

In general, the lower the MSE, RMSE, MAE, and MAPE, the better the accuracy. Table 7 presents the MSE, RMSE, MAE, MAPE, and score values (sum of four metrics), excluding realized volatility models because they are simply used to achieve comprehensiveness without regard to comparative expressiveness.

According to all four loss functions, the best-performing model is the realized kernel GAS model with the normal distribution. Among the GARCH models, models with realized kernel are significantly better than traditional GARCH models (GARCH, EGARCH, GJR-GARCH), which aligns with the purpose of using realized kernel to replace return volatility.

Among the GAS models, Student-t and skewed Student-t perform worse than the normal distribution, suggesting that the financial returns in this case may not exhibit as pronounced fat tails as expected. Including realized kernels always improves forecasts, suggesting that realized kernels provide additional information beyond stock returns. Still, for distributions that struggle with large outliers or higher volatility, realized kernels can provide added stability. This advantage may be especially useful in multivariate GAS models, as indicated by Gorgi et al. (2019).

Specifically, the R-GAS-Norm model performs best in terms of MSE, RMSE, and overall score, while the GAS-Norm model shows the best performance in MAE. For MAPE, the GAS-STD and GAS-SSTD models perform better, likely due to their ability to handle data with heavy tails and asymmetry.

We used the performance of the RiskMetrics model as a benchmark. Models that performed worse than RiskMetrics were excluded from further consideration. Although we initially aimed to optimize the distribution method through MCMC, the results indicate that models using MCMC do not largely outperform those using traditional distribution methods. Consequently, we did not apply the MCMC method within the realized kernel and GAS models, focusing instead on simpler, more interpretable approaches.

5.2.3 Diebold-Mariano Test Results

To check if the forecasting between the two competing models is statistically significantly different, the Diebold-Mariano test is conducted for MSE, RMSE, MAE and MAPE loss functions. The results of the DM test can be found in Table 8 and Table 9. We obtained quite a few significant results. We only tested the models which are ranked higher than our benchmark (RiskMetrics).

To further illustrate the outcome of the test, the result in black (positive test result) means that model 2 is better than model 1, and the red result (negative test result) means that model 1 is better than model 2.

For MSE, we got 37 rejections of the null hypothesis out of 45, and 42 rejections for both RMSE and MAE, which signifies that the forecasting performance between the 2 models is significant. We can conclude that R-GAS-Norm performs significantly better than R-GAS-STD, R-GAS-SSTD, GAS-STD, and GAS-SSTD (p-value < 0.001, with negative DM Statistic). The difference with RK-GARCH-Norm is not significant, indicating similar performance with these models. This aligns with that the MSE for the R-GAS-Norm is not much smaller than that of the RK-GARCH-Norm.

But for MAPE, we got 29 rejections of the null hypothesis out of 45, suggesting that the differences in forecasting performance between the two models are less pronounced for this metric. So even though the MAPE value for GAS-STD and GAS-SSTD is smaller than R-GAS-Norm, we can't conclude that they're better.

5.2.4 DTW Distance

Unlike traditional metrics (MSE, RMSE, MAE, and MAPE), which evaluate pointwise differences and tend to favor smoother sequences and heavily penalize large deviations, DTW accounts for shifts in time by aligning sequences optimally to minimize the distance. It provides a more comprehensive measure of how well a model captures the dynamic structure of the realized kernel.

By plotting the best model identified using previous metrics, the forecasting performance does not appear satisfactory (Fig. 6). This is because traditional metrics such as MSE and MAE tend to favor smoother sequences by heavily penalizing large fluctuations, which can lead to suboptimal rankings for models better aligned with the dynamic structure of the data. After incorporating the DTW distance, we improved the forecasting performance measurement by considering trend alignment and dynamic patterns. In Table 10, with DTW included, RK-GARCH-STD emerges as the top-ranked model, which aligns well with the visual observations from the plot, where it closely follows the realized kernel's fluctuations.

6 Conclusion

In this case study, we modelled, measured, and forecasted volatility using Cisco Systems, Inc. stocks from January 5, 2018, to January 3, 2025. Alongside basic models like Standard GARCH, Realized GARCH, and Standard GAS, we introduced Exponential GARCH, GJR-GARCH, and realized GAS as innovations. The RiskMetrics model was used as a benchmark and, for distribution innovations, we applied the normal distribution, Student's t-distribution, skewed Student's t-distribution, and Markov Chain Monte Carlo (MCMC).

Each model has its own strengths and weaknesses in volatility forecasting and modeling. Their performance in model fitting was evaluated using AIC, BIC, and log-likelihood values. For in-sample model performance, RK-GARCH-STD is the best for these three values.

We evaluated the forecast accuracy of the models using several loss functions, including MSE, RMSE, MAE, and MAPE, along with the Diebold-Mariano test. The R-GAS-Norm model showed the best performance based on traditional error metrics, as these metrics heavily penalize large fluctuations and favor smoother predictions. However, after incorporating DTW distance, which accounts for trend alignment and dynamic patterns, RK-GARCH-STD was the top-ranked model. The choice between R-GAS-Norm and RK-GARCH-STD depends on the objective: minimizing numerical differences in volatility or better capturing the overall trend and structure of the realized kernel. Regardless of the difference, models with realized kernel perform better than those without the realized kernel under the same distribution. This suggests that the realized kernel effectively incorporates high-frequency information, improving the performance of models.

Future research should apply MCMC estimation to GAS and GARCH models with realized kernels and explore more efficient computation methods for realized kernels.

References

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., & Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *The Econometrics Journal*, *12*(3), C1–C32. https://doi.org/10.1111/j.1368-423X.2008.00275.x

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Bringmann, K., Fischer, N., van der Hoog, I., Kipouridis, E., Kociumaka, T., & Rotenberg, E. (2024). Dynamic dynamic time warping. *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 208–242.

Creal, D., Koopman, S. J., & Lucas, A. (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*, *28*(5), 777–795.

Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, *13*(3), 253–263.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*, 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

Gorgi, P., Hansen, P. R., Janus, P., & Koopman, S. J. (2019). Realized wishart-garch: A score-driven multi-asset volatility model. *Journal of Financial Econometrics*, *17*(1), 1–32.

Hansen, P. R., Huang, Z., & Shek, H. H. (2011). Realized garch: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *26*, 877–906.

Investopedia. (2024). Volatility [Accessed: 31 January 2025]. https://www.investopedia.com/terms/v/volatility.asp

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*, 347–370. https://doi.org/10.2307/2938260

A Formulas

$$
\begin{aligned}
\text{AIC} &= 2k - 2\log L\big(y_1, \ldots, y_T; \hat{\theta}_T\big), \\
\text{BIC} &= \log(T)\, k - 2\log L\big(y_1, \ldots, y_T; \hat{\theta}_T\big), \\
\text{RMSE} &= \sqrt{\frac{1}{H}\sum_{t=T+1}^{T+H}\Big(\hat{\sigma}_t^2 - \sigma_t^2\Big)^2}, \\
\text{MSE} &= \frac{1}{H}\sum_{t=T+1}^{T+H}\Big(\hat{\sigma}_t^2 - \sigma_t^2\Big)^2, \\
\text{MAPE} &= \frac{1}{H}\sum_{t=T+1}^{T+H}\left|\frac{\hat{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2}\right| \times 100, \\
\text{MAE} &= \frac{1}{H}\sum_{t=T+1}^{T+H}\Big|\hat{\sigma}_t^2 - \sigma_t^2\Big|, \\
\text{DM} &= \frac{\bar{d}}{\hat{\sigma}_d/\sqrt{H}} \xrightarrow{d} N(0,1)
\end{aligned}
$$

B Model abbreviations

GARCH-STD: GARCH model using Student-t distribution.
GJR-GARCH-STD: Glosten-Jagannathan-Runkle GARCH model using Student-t distribution.
EGARCH-STD: Exponential GARCH model using Student-t distribution.
RK-GARCH-STD: GARCH model using realized kernel (RK) and Student-t distribution.
RV-GARCH-STD: GARCH model using realized volatility (RV) and Student-t distribution.
GARCH-SSTD: GARCH model using Skewed Student-t (SSTD) distribution.
GJR-GARCH-SSTD: Glosten-Jagannathan-Runkle GARCH model using Skewed Student-t distribution.
EGARCH-SSTD: Exponential GARCH model using Skewed Student-t distribution.
RK-GARCH-SSTD: GARCH model using realized kernel (RK) and Skewed Student-t distribution.
RV-GARCH-SSTD: GARCH model using realized volatility (RV) and Skewed Student-t distribution.
GARCH-Norm: GARCH model using normal distribution.
GJR-GARCH-Norm: Glosten-Jagannathan-Runkle GARCH model using normal distribution.
EGARCH-Norm: Exponential GARCH model using normal distribution.
RK-GARCH-Norm: GARCH model using realized kernel (RK) and normal distribution.
RV-GARCH-Norm: GARCH model using realized volatility (RV) and normal distribution.
GARCH-MCMC: GARCH model estimated using Markov Chain Monte Carlo (MCMC) methods.
GAS-Norm: Generalized Autoregressive Score (GAS) model using normal distribution.
GAS-STD: Generalized Autoregressive Score (GAS) model using Student-t distribution.
GAS-SSTD: Generalized Autoregressive Score (GAS) model using Skewed Student-t distribution.
R-GAS-Norm: Realized kernel-based GAS model using normal distribution.
R-GAS-STD: Realized kernel-based GAS model using Student-t distribution.
R-GAS-SSTD: Realized kernel-based GAS model using Skewed Student-t distribution.
RiskMetrics: A simple volatility model used as a benchmark, typically assumes an exponentially weighted moving average of past variances.

C Empirical Results

Table 4: Model Parameters Summary with Standard Errors

Model	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$	$\hat{\xi}$	$\hat{\gamma}$	$\hat{\nu}_X$
GARCH-Norm	0.1000	0.1249	0.8552	-	-	-	-
	(0.1142)	(0.0159)	(0.0742)	-	-	-	-
GARCH-STD	0.2213	0.1592	0.7817	4.3381	-	-	-
	(0.6533)	(0.0051)	(0.0131)	(1.1901)	-	-	-
GARCH-SSTD	0.2216	0.1592	0.7821	4.2981	0.9899	-	-
	(0.2796)	(0.0515)	(0.1386)	(0.2979)	(0.0132)	-	-
GJR-GARCH-Norm	0.3957	0.0862	0.7000	-	-	0.1971	-
	(0.1956)	(0.0240)	(0.0997)	-	-	(0.0330)	-
GJR-GARCH-STD	0.1819	0.0571	0.8120	4.4610	-	0.1645	-
	(0.7752)	(0.0091)	(0.0119)	(1.3442)	-	(0.0102)	-
GJR-GARCH-SSTD	0.1818	0.0572	0.8126	4.4196	0.9902	0.1640	-
	(0.2522)	(0.0560)	(0.1307)	(0.2487)	(0.0134)	(0.0859)	-
EGARCH-Norm	0.0856	0.2077	0.9286	-	-	-0.1132	-
	(0.0561)	(0.0421)	(0.0642)	-	-	(0.0206)	-
EGARCH-STD	0.0713	0.1525	0.9623	4.5825	-	-0.1149	-
	(0.9313)	(0.0897)	(0.1151)	(0.0622)	-	(1.4464)	-
EGARCH-SSTD	0.3079	0.1525	0.9624	4.5377	0.9909	-0.1148	-
	(0.1780)	(0.0743)	(0.0853)	(0.2115)	(0.0136)	(0.0489)	-
GAS-Norm	-0.004371	0.013448	0.847255	-	-	-	-
GAS-STD	-0.254834	0.097716	0.686688	4.594240	-	-	-
GAS-SSTD	-0.255328	0.098121	0.686433	4.594199	1.011564	-	-
R-GAS-Norm	-0.017745	0.004539	0.680486	-	-	-	10.289941
R-GAS-STD	-0.343348	0.001916	0.502762	4.551239	-	-	10.304009
R-GAS-SSTD	-0.343139	0.002080	0.503294	4.550155	1.00496	-	9.480303

Table 5: Realized GARCH Models Parameters with Standard Errors

Model	$\hat{\omega}$	$\hat{\beta}$	$\hat{\gamma}$	$\hat{\nu}$	$\hat{\xi}$	$\hat{\phi}$	$\hat{\sigma}_u$	$\hat{\tau}_1$	$\hat{\tau}_2$
	0.0999	0.8572	0.1733	-	-0.1237	1.0168	0.0000	-0.3149	0.8225
	(0.7425)	(0.7068)	(0.6207)	-	(1.0908)	(1.3395)	(1.0761)	(1.1533)	(1.7708)
	0.5627	0.3330	0.6832	4.0870	-0.3464	0.7712	0.0053	-0.0978	0.3466
	(0.7218)	(1.7490)	(2.9721)	(6.8926)	(5.8928)	(1.7916)	(20.9799)	(1.7769)	(4.9376)
	0.2276	0.7000	0.3197	4.9002	-1.9901	0.3001	0.4363	-0.5363	-1.3206
	(0.9312)	(1.0000)	(0.8524)	(10.2065)	(22.3347)	(12.1437)	(21.7737)	(9.0744)	(22.2360)
R-GARCH-RK-Norm	0.0997	0.7637	0.2336	-	-0.2299	0.9551	0.0000	-0.0699	0.1633
	(1.0000)	(1.0000)	(1.0000)	-	(1.0000)	(1.0000)	(1.0000)	(22.3785)	(1.0000)
R-GARCH-RK-STD	0.2775	0.4111	0.5007	9.5081	-0.1642	1.0407	0.0000	-0.0612	-0.0001
	(1.0682)	(0.9890)	(0.9969)	(1.0000)	(0.9843)	(0.9997)	(1.0000)	(8.1231)	(0.9774)
R-GARCH-RK-SSTD	0.1192	0.7011	0.2859	9.2580	-0.3064	0.9795	0.0016	-0.3830	-0.0745
	(1.6017)	(1.5048)	(0.8434)	(27.7021)	(7.4348)	(1.0770)	(10.3583)	(11.2515)	(4.4415)

Table 6: Criteria information for various GARCH and GAS models

Model	**Log-likelihood**	**AIC**	**BIC**
GARCH-Norm	-2423.9562	4853.9125	4869.3219
GARCH-STD	-2317.7402	4643.4803	4664.0263
GARCH-SSTD	-2317.1373	4644.2746	4669.9570
EGARCH-Norm	-2394.3986	4796.7972	4817.3431
EGARCH-STD	-2302.1877	4614.3754	4640.0578
EGARCH-SSTD	-2301.6993	4615.3986	4646.2175
GJR-GARCH-Norm	-2401.7283	4811.4567	4832.0026
GJR-GARCH-STD	-2312.5113	4635.0226	4660.7050
GJR-GARCH-SSTD	-2311.9425	4635.8850	4666.7039
R-GARCH-RK-Norm	-2387.0278	4790.0556	4831.1475
R-GARCH-RK-STD	-2294.4192	4606.8385	4653.0668
R-GARCH-RK-SSTD	-2309.3148	4638.6297	4689.9945
GAS-Norm	-22220.6382	44447.2764	44462.6859
GAS-STD	-8584.7785	17177.5571	17198.1030
GAS-SSTD	-8582.9529	17175.9058	17201.5883
R-GAS-Norm	-6268.3751	12544.7502	12565.2961
R-GAS-STD	-3323.0813	6656.1627	6681.8451
R-GAS-SSTD	-3323.0242	6658.0484	6688.8673

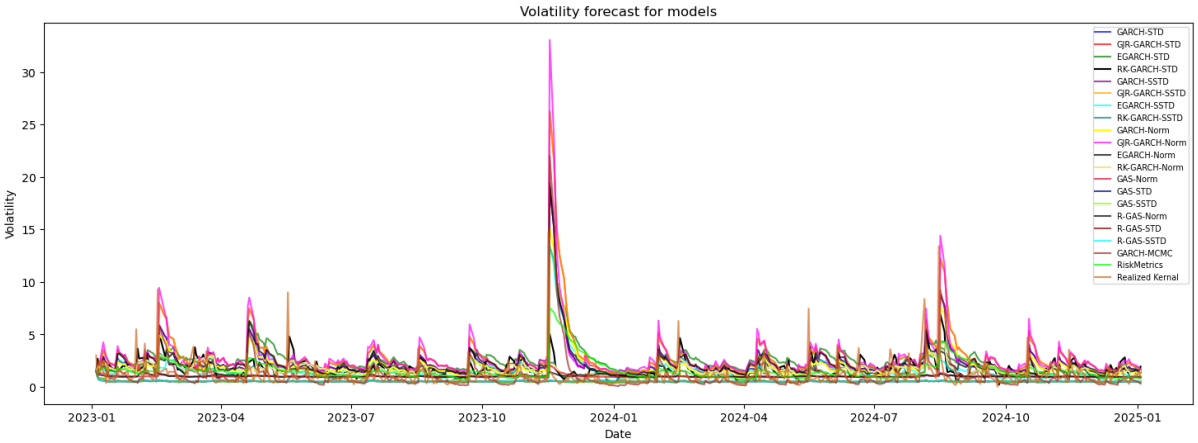


Figure 5: Daily volatility forecast for all the models

Table 7: Model Performance Metrics

Model	MSE	RMSE	MAE	MAPE	Score
R-GAS-Norm	**1.6773**	**1.2951**	0.6096	0.7042	**4.2862**
GAS-Norm	1.7153	1.3097	**0.6069**	0.6876	4.3195
R-GAS-STD	2.0966	1.4480	0.7107	0.5121	4.7674
R-GAS-SSTD	2.0967	1.4480	0.7108	0.5121	4.7676
GAS-STD	2.1300	1.4594	0.7247	**0.5120**	4.8261
GAS-SSTD	2.1304	1.4596	0.7249	**0.5120**	4.8268
RK-GARCH-SSTD	1.8197	1.3490	0.7846	1.0186	4.9719
RK-GARCH-Norm	1.8140	1.3468	0.7924	1.0566	5.0098
RK-GARCH-STD	2.1356	1.4614	0.9167	1.1481	5.6617
RiskMetrics	2.6247	1.6201	0.9195	1.3332	6.4975
EGARCH-SSTD	3.0966	1.7597	0.9854	1.3200	7.1618
GARCH-MCMC	4.1746	2.0432	0.9325	0.9296	8.0798
GARCH-Norm	4.0374	2.0093	1.2280	1.7982	9.0730
EGARCH-Norm	4.2359	2.0581	1.2746	1.8139	9.3826
GARCH-STD	4.4745	2.1153	1.3117	1.9109	9.8124
GARCH-SSTD	4.4898	2.1189	1.3163	1.9188	9.8438
EGARCH-STD	4.8090	2.1929	1.6376	2.4146	11.0541
GJR-GARCH-STD	8.4320	2.9038	1.8573	2.7158	15.9089
GJR-GARCH-SSTD	8.4458	2.9062	1.8617	2.7232	15.9370
GJR-GARCH-Norm	8.8822	2.9803	1.8486	2.6489	16.3601

Table 8: MSE Loss Differential (DM Stat.) and Significance

	R-GAS-Norm	GAS-Norm	R-GAS-STD	R-GAS-SSTD	GAS-STD	GAS-SSTD	RK-GARCH-SSTD	RK-GARCH-Norm	RK-GARCH-STD	GARCH-MCMC
R-GAS-Norm	—	-1.9328*	-6.5457***	-6.5462***	-6.4391***	-6.4392***	-1.7211*	-1.5750	-3.9589***	-3.3403***
GAS-Norm	—	—	-7.7217***	-7.7222***	-7.5468***	-7.5461***	-1.1487	-1.0288	-3.4257***	-3.2632***
R-GAS-STD	—	—	—	-6.7821***	-3.9436***	-3.9610***	2.1471**	2.0994**	-0.2510	-2.1416**
R-GAS-SSTD	—	—	—	—	-3.9374***	-3.9549***	2.1479**	2.1002**	-0.2502	-2.1413**
GAS-STD	—	—	—	—	—	-3.3001***	2.3216**	2.2658**	-0.0354	-2.0426**
GAS-SSTD	—	—	—	—	—	—	2.3241**	2.2682**	-0.0328	-2.0414**
RK-GARCH-SSTD	—	—	—	—	—	—	—	0.5521	-5.0328***	-2.8600***
RK-GARCH-Norm	—	—	—	—	—	—	—	—	-4.7621**	-2.8539***
RK-GARCH-STD	—	—	—	—	—	—	—	—	—	-2.0168**
GARCH-MCMC	—	—	—	—	—	—	—	—	—	—

Table 9: MAPE Loss Differential (DM Stat.) and Significance

	R-GAS-Norm	GAS-Norm	R-GAS-STD	R-GAS-SSTD	GAS-STD	GAS-SSTD	RK-GARCH-SSTD	RK-GARCH-Norm	RK-GARCH-STD	GARCH-MCMC
R-GAS-Norm	—	0.6852	0.0938	0.0953	-0.2230	-0.2184	-5.1546***	-5.0308***	-9.0518***	-2.5465**
GAS-Norm	—	—	-0.3981	-0.3966	-0.3335	-0.3290	-5.5274***	-5.8116***	-9.5920***	-2.8209***
R-GAS-STD	—	—	—	-1.0198	-1.2525	-1.2315	-4.2369***	-4.4393***	-7.1392***	-2.0134**
R-GAS-SSTD	—	—	—	—	-1.2435	-1.2226	-4.2381***	-4.4406***	-7.1403***	-2.0106**
GAS-STD	—	—	—	—	—	-0.2864	-4.0406***	-4.6431***	-6.7509***	-1.8956*
GAS-SSTD	—	—	—	—	—	—	-4.0437***	-4.6464***	-6.7532***	-1.7284*
RK-GARCH-SSTD	—	—	—	—	—	—	—	-0.7297	-5.0282***	2.9073***
RK-GARCH-Norm	—	—	—	—	—	—	—	—	-3.8749***	3.1254***
RK-GARCH-STD	—	—	—	—	—	—	—	—	—	5.0197***
GARCH-MCMC	—	—	—	—	—	—	—	—	—	—

* indicates significance at the 10% level.

** indicates significance at the 5% level.

*** indicates significance at the 1% level.

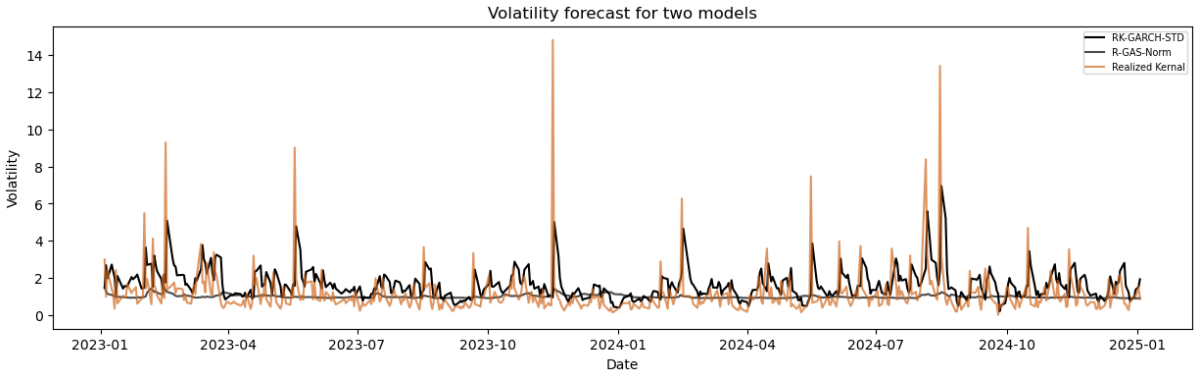
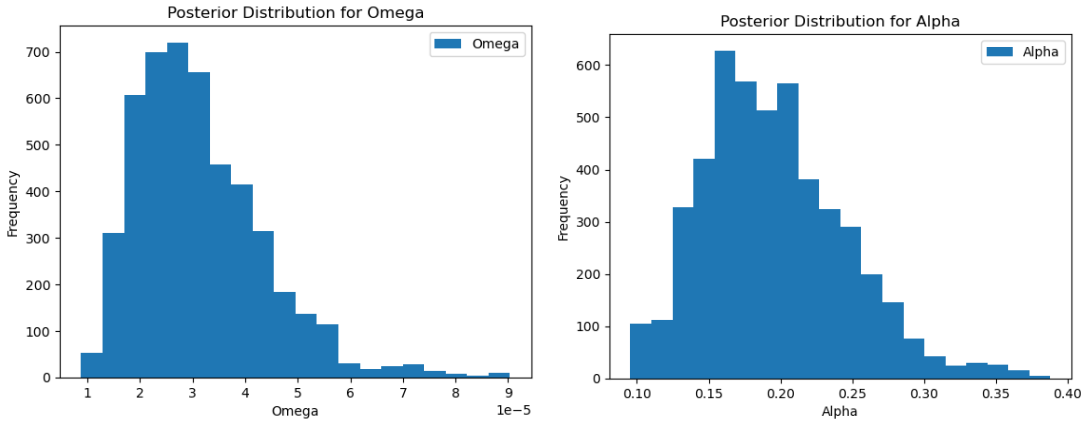


Figure 6: Daily volatility forecast for two the models

Table 10: Overall Model Performance

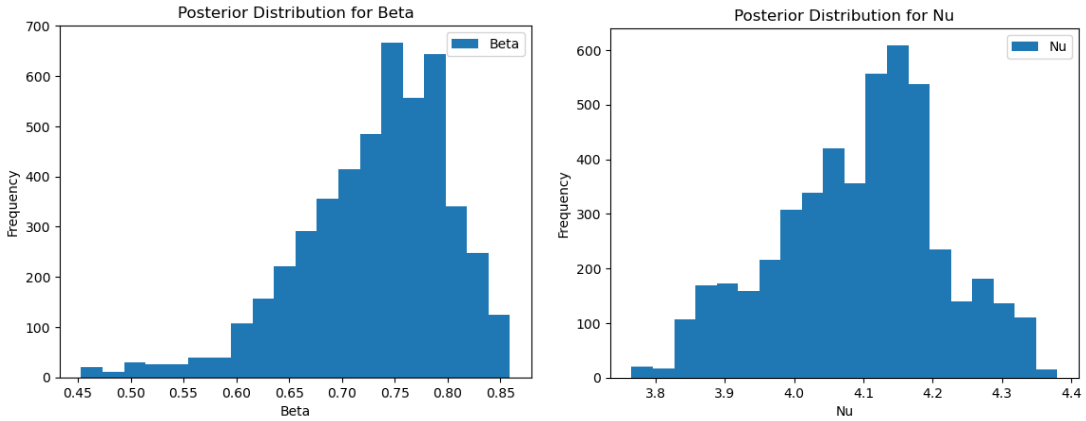
Model	**MSE**	**RMSE**	**MAE**	**MAPE**	**DTW Distance**	**Score**
RK-GARCH-STD	2.135623	1.461377	0.916653	1.148057	18.141753	23.803463
RK-GARCH-SSTD	1.819745	1.348979	0.784579	1.018624	23.235776	28.207703
GARCH-MCMC	4.174607	2.135623	1.461377	0.916653	28.963986	28.626179
RK-GARCH-Norm	1.813989	1.346844	0.792369	1.056644	24.033994	29.043840
EGARCH-SSTD	3.096630	1.759725	0.985422	1.319975	24.734025	31.895776
GAS-Norm	1.715279	1.309686	0.606921	0.687630	27.738003	32.057519
R-GAS-Norm	1.677341	1.295122	0.609589	0.704173	27.829693	32.115918
RiskMetrics	2.624715	1.620097	0.919522	1.333152	25.651595	32.149081
GAS-STD	2.129966	1.459440	0.724681	0.512032	29.050768	33.876888
GAS-SSTD	2.130380	1.459582	0.724856	0.511985	29.052538	33.879341
R-GAS-SSTD	2.096699	1.447998	0.710779	0.512128	29.290675	34.058279
R-GAS-STD	2.096583	1.447958	0.710724	0.512121	29.291474	34.058860
GARCH-Norm	4.037445	2.009340	1.228017	1.798203	25.475333	34.548338
GARCH-STD	4.474492	2.115299	1.311721	1.910900	24.810577	34.622989
GARCH-SSTD	4.489785	2.118911	1.316336	1.918766	24.873156	34.716953
EGARCH-Norm	4.235905	2.058131	1.274615	1.813908	25.725046	35.107606
EGARCH-STD	4.809002	2.192944	1.637619	2.414564	29.256591	40.310719
GJR-GARCH-STD	8.431951	2.903782	1.857324	2.715793	31.241536	47.150386
GJR-GARCH-SSTD	8.445835	2.906172	1.861735	2.723211	31.284231	47.221183
GJR-GARCH-Norm	8.882221	2.980305	1.848648	2.648898	34.279715	50.639787

C.1 Bayesian Approach



(a) Posterior Distribution for Ω

(b) Posterior Distribution for α



(c) Posterior Distribution for β

(d) Posterior Distribution for ν

Figure 7: Posterior Distributions of Estimated Parameters

Figure 8: Trace Plot for All Parameters

D GAS Derivation

D.1 Normal Distribution

The derivation of the Normal distribution can be found in Canvas paper (8). The scaled score function is:

$$\nabla_t = \left(-\frac{1}{2h_t} + \frac{(r_t-\mu_t)^2}{2h_t^2}\right) \times h_t \;=\; -\frac{1}{2} + \frac{(r_t-\mu_t)^2}{2\exp(f_t)}.$$

Hence, the updating equation becomes:

$$f_{t+1} \;=\; \omega + \beta f_t + \alpha\left(-\frac{1}{2} + \frac{(r_t-\mu_t)^2}{2\exp(f_t)}\right).$$

The maximum log-likelihood estimate is derived from:

$$L(\theta) \;=\; \sum_{t=1}^{T}\Big(-\tfrac{1}{2}\log\big(2\pi\,\exp(f_t)\big) \;-\; \tfrac{(r_t-\mu_t)^2}{2\exp(f_t)}\Big) \quad \text{with respect to} \quad \theta=(\omega,\alpha,\beta).$$

D.2 Student-t Distribution

The derivation of the Student-t distribution can be found in Canvas paper (8). The scaled score function is:

$$\nabla_t \;=\; -\tfrac{1}{2} + \tfrac{\nu+1}{2}\cdot\tfrac{(r_t-\mu_t)^2}{(\nu-2)\exp(f_t)+(r_t-\mu_t)^2}.$$

We maximize the log-likelihood:

$$L(\theta) \;=\; \sum_{t=1}^{T}\Big[\log\Gamma\Big(\tfrac{\nu+1}{2}\Big) \;-\; \log\Gamma\Big(\tfrac{\nu}{2}\Big) \;-\; \tfrac{1}{2}\log\big((\nu-2)\pi\exp(f_t)\big) \;-\; \tfrac{\nu+1}{2}\log\Big(1+\tfrac{(r_t-\mu_t)^2}{(\nu-2)\exp(f_t)}\Big)\Big],$$

with respect to $\theta=(\omega,\alpha,\beta,\nu)$.

D.3 Skewed Student-t Distribution

Given $\log p(r_t)$ of the Skewed Student-t distribution, we derive the log-pdf with respect to h_t. First, define

$$A=(\nu-2)\,\pi\,h_t, \quad B=1+\frac{\big(s\frac{r_t-\mu_t}{\sqrt{h_t}}+m\big)^2}{\nu-2}\,\xi^{-2I_t}.$$

Then

- $\frac{\partial\log(A)}{\partial h_t}=\frac{1}{h_t}$,
- $\frac{\partial\log(B)}{\partial h_t}=\frac{1}{B}\cdot\frac{\partial B}{\partial h_t}$.

Let $z=s\,\frac{r_t-\mu_t}{\sqrt{h_t}}+m$. We compute

$$\begin{aligned}
\frac{\partial z}{\partial h_t} &= -\tfrac{1}{2}\,s\cdot\frac{r_t-\mu_t}{h_t^{3/2}},\\
\frac{\partial(z^2)}{\partial h_t} &= 2z\,\frac{\partial z}{\partial h_t},\\
\frac{\partial B}{\partial h_t} &= \frac{2z}{\nu-2}\left(-\tfrac{1}{2}s\frac{r_t-\mu_t}{h_t^{3/2}}\right)\xi^{-2I_t} \;=\; -\frac{z\,s\,(r_t-\mu_t)}{(\nu-2)\,h_t^{3/2}}\,\xi^{-2I_t},\\
\frac{\partial\log(B)}{\partial h_t} &= \frac{1}{B}\left(-\frac{z\,s\,(r_t-\mu_t)}{(\nu-2)\,h_t^{3/2}}\,\xi^{-2I_t}\right).
\end{aligned}$$

Combining these,

$$\nabla_t = \left(-\tfrac{1}{2}\,\tfrac{\partial \log(A)}{\partial h_t} \;-\; \tfrac{\nu+1}{2}\,\tfrac{\partial \log(B)}{\partial h_t}\right) h_t \;=\; -\tfrac{1}{2} + \tfrac{(\nu+1)\,z\,s\,(r_t-\mu_t)\,\xi^{-2I_t}}{2\,B\,(\nu-2)\,\sqrt{\exp(f_t)}} = -\tfrac{1}{2} + \tfrac{(\nu+1)\,z\,s\,((r_t-\mu_t)/\sqrt{\exp f_t})}{2\,(\nu-2)\,\xi^{2I_t}+z^2}$$

The log-likelihood to maximize is:

$$\begin{aligned} L(\theta) = \sum_{t=1}^{T} & \log\Gamma\Big(\frac{\nu+1}{2}\Big) \;-\; \log\Gamma\Big(\frac{\nu}{2}\Big) \;-\; \tfrac{1}{2}\,\log\big((\nu-2)\,\pi\,\exp(f_t)\big) \\ & + \;\log(s) + \;\log\Big(\frac{2}{\xi+\frac{1}{\xi}}\Big) \;-\; \tfrac{\nu+1}{2}\,\log\Big(1 + \frac{(s\,\frac{r_t-\mu_t}{\sqrt{\exp(f_t)}} + m)^2}{\nu-2}\,\xi^{-2I_t}\Big). \end{aligned}$$

D.4 Realized GAS

For the Realized GAS model, the scaled score functions and maximum log-likelihood of the Normal and Student-t distributions follow those in Canvas paper (8). For the Skewed Student-t distribution, the scaled score function is:

$$\nabla_t = \frac{\nu}{2}\Big(\frac{X_t}{\exp(f_t)} - 1\Big) - \frac{1}{2} + \frac{(\nu+1)\,z\,s\,(r_t-\mu_t)\,\xi^{-2I_t}}{2\,B\,(\nu-2)\,\sqrt{\exp(f_t)}},$$

with the log-likelihood:

$$\begin{aligned} L(\theta) \;=\; \sum_{t=1}^{T} & -\log\Gamma\Big(\tfrac{\nu}{2}\Big) \;-\; \Big(\tfrac{\nu}{2}\Big)\,\log\Big(\tfrac{2\,\exp(f_t)}{\nu}\Big) \;+\; \Big(\tfrac{\nu}{2}-1\Big)\,\log(X_t) \;-\; \Big(\tfrac{\nu\,X_t}{2\,\exp(f_t)}\Big) \\ & + \;\log\Gamma\Big(\tfrac{\nu+1}{2}\Big) \;-\; \log\Gamma\Big(\tfrac{\nu}{2}\Big) \;-\; \tfrac{1}{2}\,\log\big((\nu-2)\,\pi\,\exp(f_t)\big) \;+\; \log(s) \;+\; \log\Big(\tfrac{2}{\xi+\frac{1}{\xi}}\Big) \\ & - \;\tfrac{\nu+1}{2}\,\log\Big(1 + \tfrac{(s\,\frac{r_t-\mu_t}{\sqrt{\exp(f_t)}}+m)^2}{\nu-2}\,\xi^{-2I_t}\Big). \end{aligned}$$