Vrije Universiteit Amsterdam

Academic Year 2024-2025

Financial Econometrics Case Study

Group 4

Authors:

Abe Tempelman	2747518
Jiaxuan Zhu	2838063
Yunji Eo	2735445
Robin Klaasen	2744646
Vân Lê	2676244

Emails:
a.w.r.tempelman@student.vu.nl
j.zhu8@student.vu.nl
y.eo@student.vu.nl
r.l.klaasen@student.vu.nl
k.v.le@student.vu.nl

January 22, 2025

Contents

1 Abstract

This paper explores advanced econometric strategies for modeling and forecasting financial volatility in high-frequency stock data from Cisco Systems, Inc. between 2018 and 2025. Multiple GARCH-based and GAS-based models are applied. The realized kernel is incorporated to account for market microstructure effects. 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.

2 Introduction

This study investigates advanced econometric approaches for modelling and forecasting financial volatility using high-frequency stock data from Cisco Systems, Inc. between 2018 and 2025. . Before delving into the analysis, let's first discuss what volatility is. According to Investopedia, 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 forecasting price movements, making it a vital tool for traders and analysts seeking to navigate financial markets.

First, a description of the data set is shown and the data cleaning process is explained, followed by a preliminary analysis. Followed by the implied methods, models, and the results.

The methodology starts by using the realized kernel to refine intraday volatility estimates by accounting for market microstructure effects. To capture diverse volatility dynamics, a range of GARCH-type and GAS-type models is considered, including GARCH, GJR-GARCH, EGARCH, R-GARCH-RV, and GAS under Normal, Student-t, and Skewed Student-t distributions.

Model parameters are estimated using log-likelihood maximization. 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, Root Mean Squared Error, and Mean Absolute Error, 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 outperforms another in a statistically significant way. Value-at-Risk considerations are also introduced for risk management, illustrating how volatility forecasting supports both strategic decision-making and trading activities.

3 Data Description and Preprocessing

3.1 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 was 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 data, offering insights into the price behaviour over the observation period. The in-sample dataset spans 30,514,903 records from January 5 2018 to January 3 2023. The out-of-sample dataset consists of 12,238,721 records from January 4 2023 to January 3 2025, reflecting the scope and granularity of the data.

The descriptive statistics reveal several critical aspects of the dataset. The skewness values for both datasets are positive, indicating slight asymmetry with a longer right tail in the price distribution. The kurtosis values, particularly 51.93 for the in-sample data, suggest the presence of extreme outliers and heavy tails, characteristics commonly observed in financial time series. These statistical properties highlight the importance of using advanced econometric models capable of capturing non-normality and volatility clustering, ensuring robust volatility estimation and forecasting.

The dataset also shows volatility clustering patterns, evident from the skewness and kurtosis values of the returns. Such patterns further emphasize the need to explore realized measures of volatility and advanced modeling approaches, such as GARCH and Realized GARCH, which effectively handle these temporal dynamics.

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

Metric	**In-Sample Price**	**Out-Sample Price**
Count	30,514,903	12,238,721
Mean	47.24	50.78
Standard Deviation	5.95	3.59
Minimum	32.41	44.66
25%	43.06	48.20
50% (Median)	46.06	49.92
75%	52.05	52.78
Maximum	64.28	60.21
Skewness	0.52	0.21
Kurtosis	51.93	18.61

3.2 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 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.

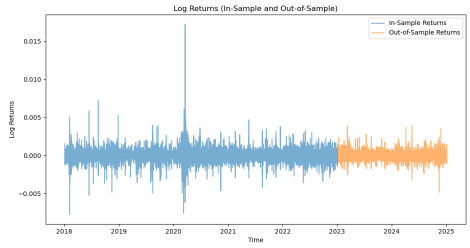
3.3 Data Preliminary Analysis

Figures 1a and 1b provide an overview of the in-sample and out-of-sample datasets for Cisco Systems, Inc. (CSCO) from January 5, 2018, to January 3, 2025. 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.

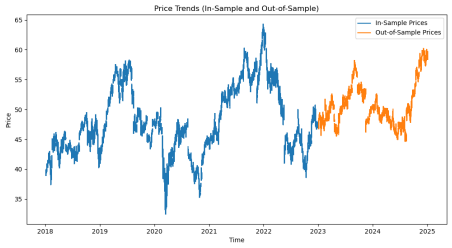
Figure 1b illustrates the temporal evolution of CSCO stock prices over the observation period. The figure reveals distinct patterns, including periods of sustained growth, declines, and volatility spikes. The visible gaps in the data correspond to non-trading days and are a natural feature of financial 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.

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 behaviour, confirming the consistency of temporal patterns across datasets.

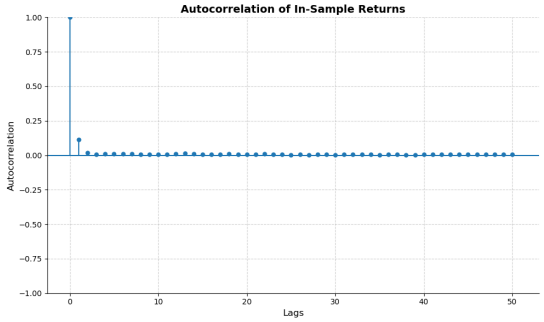


(a) Log Returns (In-Sample and Out-of-Sample). This figure illustrates the log returns calculated from high-frequency price data for both in-sample and out-of-sample datasets.

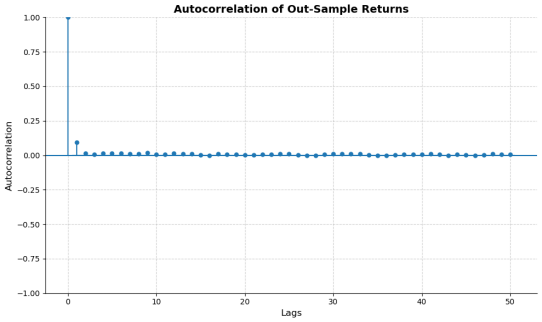


(b) Price of CSCO (In-Sample and Out-of-Sample). This figure shows the price movements of the in-sample and out-of-sample data. It shows strong price trends.

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.

4 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.

4.1 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. The Parzen kernel was chosen for this analysis due to its desirable properties, including non-negative estimates and smoothness, satisfying $k'(0) = k'(1) = 0$. The Parzen kernel is defined as:

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

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), the formula for H is given by:

$$H = c^* \left(\frac{\omega^2}{IV}\right)^{2/5} n^{3/5},$$

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.

4.2 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-t distribution, which is also symmetric but incorporates heavy tails to handle extreme observations; and the skewed Student-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.

4.2.1 Normal Distributions

The probability density function is given by:

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

and the natural logarithm of the density function is:

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

4.2.2 Student-t Distributions

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:

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

4.2.3 Skewed Student-t Distributions

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:

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

4.2.4 Markov Chain Mento Carlo (MCMC)

Markov Chain Monte Carlo (MCMC) is used to approximate complex probability distributions by generating samples from the target distribution. For GARCH-type models, we incorporate MCMC to compare parameter estimation and model performance with other methods, ensuring robustness and consistency. The acceptance probability is defined as:

$$\alpha(x, x') = \min\left(1, \frac{\pi(x')q(x|x')}{\pi(x)q(x'|x)}\right),$$

4.3 Model Types

4.3.1 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 represent 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 : $\omega > 0, \quad \alpha \geq 0, \quad \beta \geq 0$. We choose the GARCH(1,1) model because it is the simplest yet highly effective model for capturing volatility clustering with minimal parameters.

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

4.3.2 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 defines 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 indicator function term captures the asymmetric effect of negative shocks. When $\gamma > 0$, negative shocks increase volatility more than positive shocks, indicating asymmetry. When $\gamma = 0$, the model reduces to a standard GARCH(1,1) model. When $\gamma < 0$,

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 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:

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

4.3.3 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-t distributions, making it effective in capturing heavy tails and volatility clustering.

4.3.4 Realized GARCH

$$r_t = \sqrt{h_t} z_t,$$
$$h_t = \omega + \beta h_{t-1} + \gamma x_{t-1},$$
$$x_t = \xi + \phi h_t + \tau(z_t) + u_t.$$

$$\log h_t = \omega + \sum_{i=1}^{p} \beta_i \log h_{t-i} + \sum_{j=1}^{q} \gamma_j \log x_{t-j},$$
$$\log x_t = \xi + \phi \log h_t + \tau(z_t) + u_t,$$
$$z_t \sim i.i.d.(0,1), \quad u_t \sim i.i.d.(0, \sigma_u^2),$$

where $\tau(z_t)$ is called the leverage function, and we assume $\mathbb{E}[\tau(z_t)] = 0$.

A natural specification for the measurement equation implies that:

$$\log r_t^2 = \log h_t + \log z_t^2.$$

4.3.5 GAS model

One of the more advanced econometric models is the generalized auto-regressive score (GAS) model (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 observation r_t. In the univariate setting, let $N \times 1$ vector r_t is 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:

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

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 GAS model (Creal et al., 2013). To extend the GAS model, 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 the Realized Kernel at time t that mentioned in Section 4.1. 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 GAS model, i.e. our returns R_t have some distribution $y_t \sim D(\mu_t, \exp(f_t))$.

4.4 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 BIC penalized additional parameters through the term $\log(T) \times k$ and AIC penalizes the additional term by $2k$. The formula of the AIC and BIC can be found in Appendix A.

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

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

4.6 Model Accuracy

4.6.1 Loss Functions

To evaluate the forecast accuracy of the models, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Aboslute Error (MAE) 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 are 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. Hence, by looking at these three loss functions, we can

evaluate the performance of the models with these characteristics in mind. The formula of RMSE, MSE and MAE can be found in Appendix A.

4.6.2 Diebold-Mariano test

The RMSE values alone cannot determine whether one model outperforms another, as these metrics are inherently random by nature. When the RMSE values 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.

5 Realized Kernels Estimation

The following graph compares daily RV and RK estimates for Cisco Systems, Inc. from in-sample data.

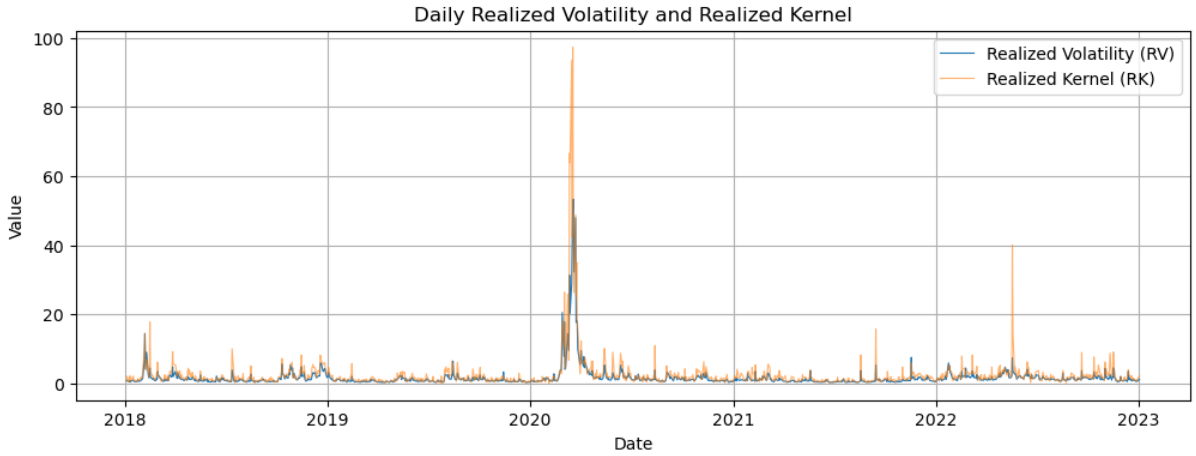


Figure 3: Daily Realized Volatility and Realized Kernel from 2018 until 2023

As seen in the figure above, notable spikes occur at the onset of 2020 and other periods of heightened market stress. A closer look at these peaks reveals that RK tends to moderate noise-induced jumps more effectively than RV, indicated by the smaller jumps of the RV. To show these differences better, the following table is considered.

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.

The following graph compares daily RV and RK estimates for Cisco Systems, Inc. from out-of-sample data.

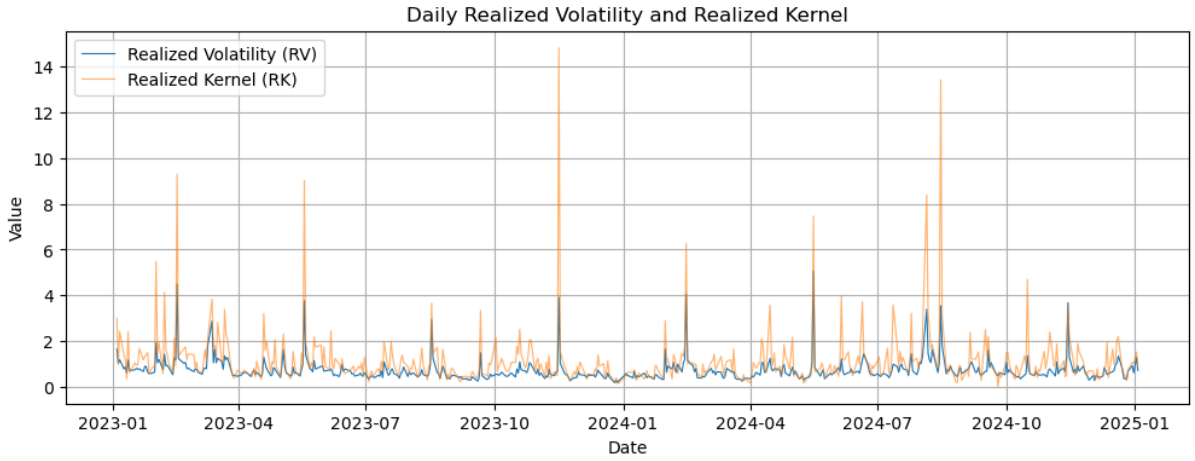


Figure 4: Daily Realized Volatility and Realized Kernel from 2023 until 2025

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

6 Empirical Results

6.1 Model Estimation (in-sample data)

Table 4: Model Parameters Summary with Standard Errors

Model	$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\nu}$	$\hat{\xi}$	$\hat{\gamma}$
GARCH-Norm	7.0e-5	0.298763	0.503746	-	-	-
	(1.6e-5)	(0.026368)	(0.077771)	-	-	-
GARCH-STD	1.6e-5	0.107159	0.832159	5.091224	-	-
	(1.21e-4)	(0.170811)	(0.639317)	(0.111329)	-	-
GARCH-SSTD	2.4e-5	0.169666	0.749812	5.434060	0.990095	-
	(2.4e-5)	(0.047529)	(0.131227)	(0.637671)	(0.014603)	-
GJR-GARCH-Norm	4.0e-5	0.086252	0.700000	-	-	0.196988
	(2.0e-5)	(0.023980)	(0.099712)	-	-	(0.033043)
GJR-GARCH-STD	2.0e-5	0.122195	0.870883	2.634221	-	0.113248
	(1.38e-4)	(0.288583)	(0.516043)	(0.145108)	-	(0.383421)
GJR-GARCH-SSTD	1.15e-4	0.163677	0.327354	4.800448	0.989911	0.163677
	(2.8e-5)	(0.066651)	(0.114405)	(0.591757)	(0.013764)	(0.095028)
EGARCH-Norm	-1.0e-2	0.106456	0.997679	-	-	-0.083096
	(0.310337)	(0.036228)	(0.036893)	-	-	(0.019204)
EGARCH-STD	0.0e0	0.192730	0.992214	3.773932	-	-0.083674
	(5.81e-4)	(5.7e-5)	(6.5e-5)	(3.9e-5)	-	(0.178856)
EGARCH-SSTD	-3.884e-2	0.152480	0.962354	4.537990	0.990873	-0.114808
	(0.690254)	(0.074649)	(0.085880)	(0.523738)	(0.013665)	(0.049007)
R-GARCH-RV-Norm	3.525e-1	0.158700	0.299000	-	-	0.406100
	-	-	-	-	-	-
R-GARCH-RV-STD	3.044e-1	0.156300	0.242600	4.801900	-	0.518000
	-	-	-	-	-	-
R-GARCH-RV-SSTD	1.0e-8	1.000000	0.852500	2.065900	1.0e-8	1.000000
	-	-	-	-	-	-
GAS-Norm	-1.0601	0.461402	0.759775	-	-	-
	1.0000	1.0000	1.0000	-	-	-
GAS-STD	-0.413730	0.336559	0.949933	4.518381	-	-
	(19.031862)	(8.102799)	(2.374743)	(76.821024)	-	-
GAS-SSTD	-0.412969	0.333556	0.950000	4.535421	0.972791	-
	(7.482522)	(11.275188)	(0.894981)	(29.337421)	(6.536779)	-
R-GAS-Norm	-	-	-	-	-	-
	-	-	-	-	-	-
R-GAS-STD	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)
R-GAS-SSTD	-	-	-	-	-	-
	(-)	(-)	(-)	(-)	(-)	(-)

The estimates of GAS with the student-t distribution and skewed student-t distribution are in line with the expectation, which is $\hat{\alpha}$ is close to 0 and β close to 1. This means that the effect volatility of $t-1$ is more important for the volatility at t than the return at $t-1$. GAS with normal distribution has much high α and has similar weight as $\hat{\beta}$, meaning normal distribution take both the volatility and return of the previous t equally into account. Gaussian distribution has a negative $\hat{\omega}$, while the skewed student-t and student-t have positive $\hat{\omega}$ indicating the weight on volatility is higher when compared to the normal distribution. Furthermore when maximizing the log-likelihood, we encountered some large outlier estimation, 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.

Table 5: Criteria information for various GARCH and GAS models

Model	Log-likelihood	AIC	BIC
GARCH-Norm	-2134.3007	4274.6014	4290.0156
GARCH-STD	-2057.9238	4123.8477	4144.3999
GARCH-SSTD	-2057.9231	4125.8461	4151.5365
EGARCH-Norm	-2131.6903	4271.3806	4291.9329
EGARCH-STD	-2058.3484	4126.6968	4152.3872
EGARCH-SSTD	-2634.6643	5281.3285	5312.1569
GJR-GARCH-Norm	-2133.1577	4274.3155	4294.8678
GJR-GARCH-STD	-2056.9828	4123.9656	4149.6559
GJR-GARCH-SSTD	-2056.9826	4125.9652	4156.7937
R-GARCH-RV-Norm	-2055.0842	4118.1684	4138.7207
R-GARCH-RV-STD	-2037.0890	4084.1779	4109.8683
R-GARCH-RV-SSTD	-2601.1058	5214.2117	5245.0401
R-GARCH-RK-Norm	-2088.2805	4184.5609	4205.1132
R-GARCH-RK-STD	-2034.8932	4079.7864	4105.4767
R-GARCH-RK-SSTD	-2599.4669	5210.9338	5241.7622
GAS-Norm	2305.4111	-4604.8222	-4589.4128
GAS-STD	3482.7337	-6957.4674	-6936.9215
GAS-SSTD	3483.0437	-6956.0874	-6930.4040
R-GAS-Norm	-	-	-
R-GAS-STD	-	-	-
R-GAS-SSTD	-	-	-

Please note the loglik, AIC and BIC of all GARCH models are not available yet.

6.2 Model Forecasting (out-of-sample performance

Table 6: RMSE, MSR, and MAE values of the GARCH and GAS models with different distributions of out-sample data

Model	RMSE	MSE	MAE
GARCH-Norm			
GARCH-STD			
GARCH-SSTD			
EGARCH-Norm			
EGARCH-STD			
EGARCH-SSTD			
GJR-GARCH-Norm			
GJR-GARCH-STD			
GJR-GARCH-SSTD			
R-GARCH-RV-Norm			
R-GARCH-RV-STD			
R-GARCH-RV-SSTD			
R-GARCH-RK-Norm			
R-GARCH-RK-STD			
R-GARCH-RK-SSTD			
GAS-Norm	0.012516	0.000157	0.008598
GAS-STD	0.012184	0.000148	0.008374
GAS-SSTD	0.012184	0.000148	0.008374
R-GAS-Norm	-	-	-
R-GAS-STD	-	-	-
R-GAS-SSTD	-	-	-

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

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

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

A Formulas

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

B Empirical Results

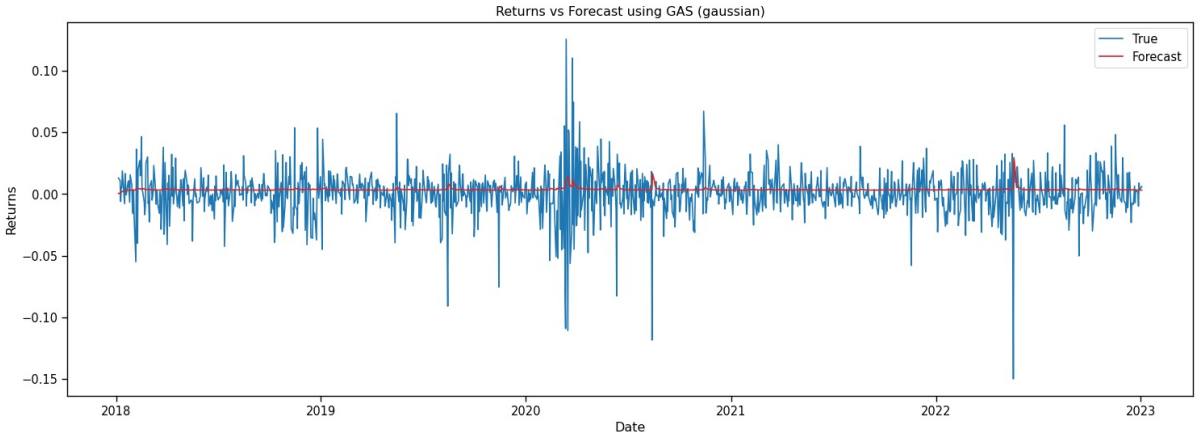


Figure 5: GAS fit plot of the estimates with the normal distribution

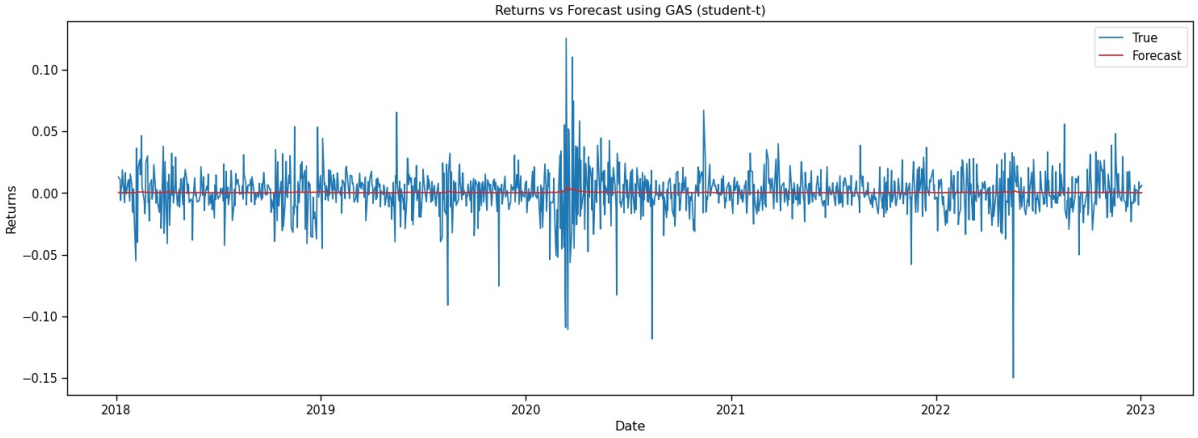


Figure 6: GAS fit plot of the estimates with the student-t distribution

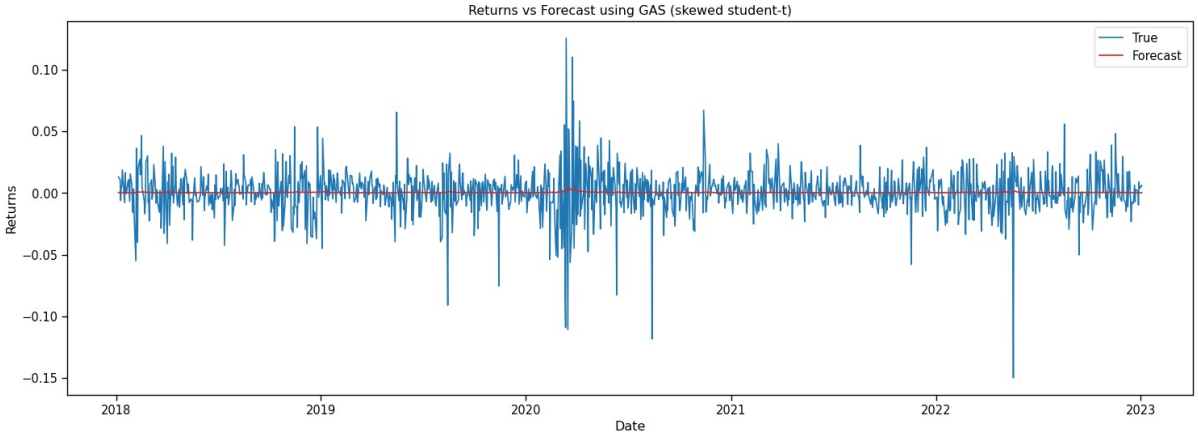


Figure 7: GAS fit plot of the estimates with the skewed student-t distribution

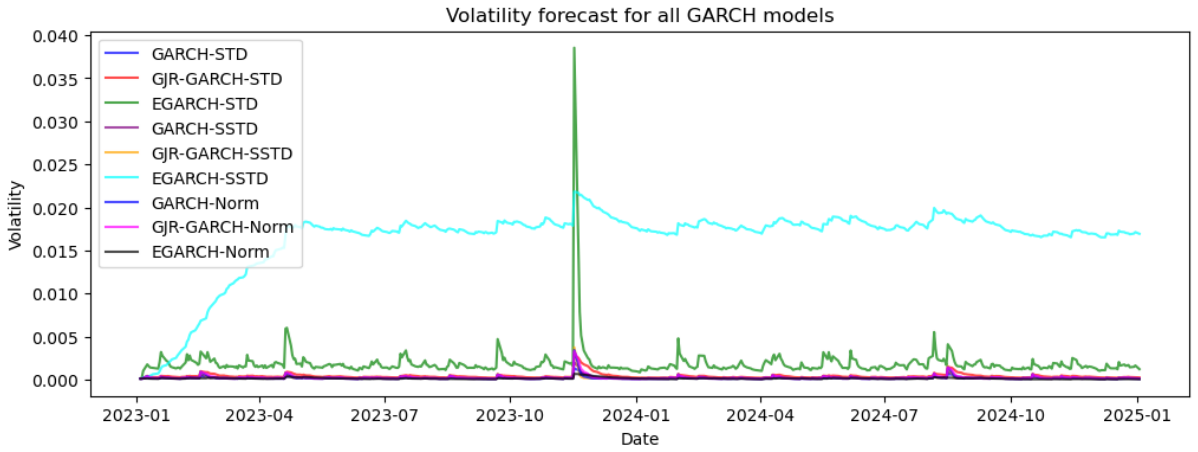


Figure 8: Volatility forecast for all GARCH models

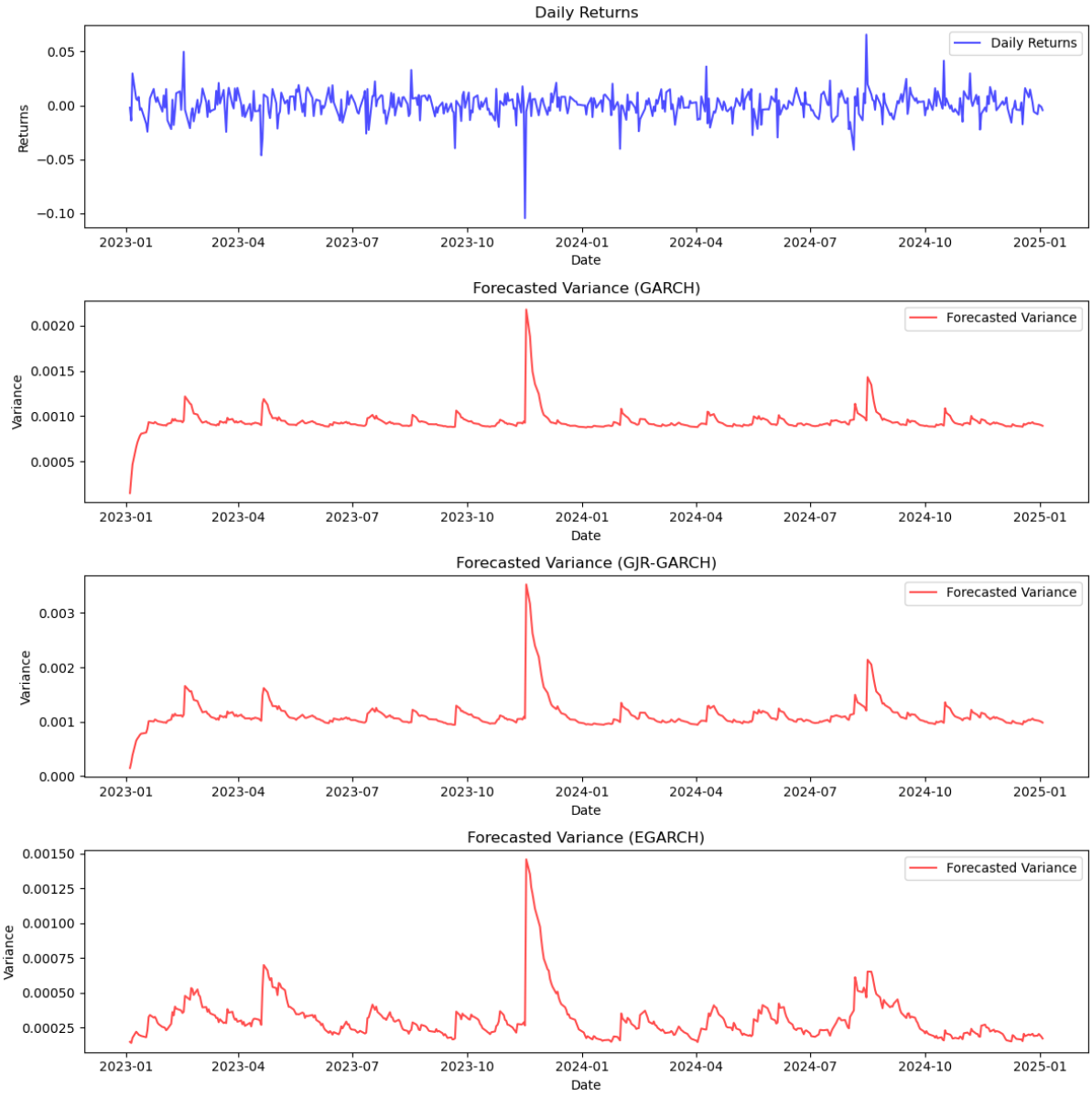


Figure 9: Volatility forecast for MCMC

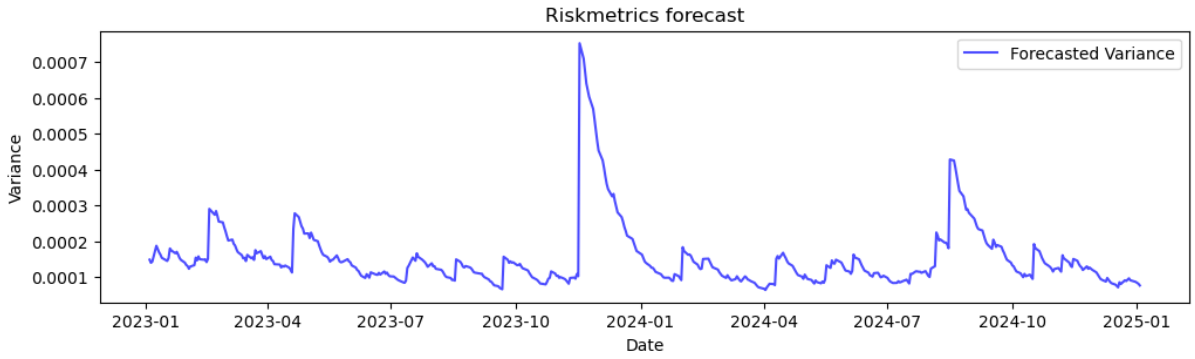


Figure 10: Volatility forecast for Riskmetrics

C GAS Derivation

C.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:

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

C.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 $\theta=(\omega,\alpha,\beta,\nu)$.

C.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,

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

The log-likelihood to maximize is:

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

C.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:

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