

Article

Not peer-reviewed version

Volatility Modelling of the JSE Top40 Index: Assessing the GAS Framework Against GARCH and Hybrid GARCH-XGBoost

[Israel Maingo](#)*, [Thakhani Ravele](#), [Caston Sigauke](#)

Posted Date: 14 October 2025

doi: 10.20944/preprints202510.1019.v1

Keywords: ARMA-EGARCH-XGBoost; density forecasting; GAS model; johannesburg stock exchange; risk management; volatility modelling; value-at-risk



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Volatility Modelling of the JSE Top40 Index: Assessing the GAS Framework Against GARCH and Hybrid GARCH–XGBoost

Israel Maingo* , Thakhani Ravele  and Caston Sigauke 

Department of Mathematical and Computational Sciences, University of Venda, Private Bag X5050, Thohoyandou 0950, Limpopo, South Africa

* Correspondence: 21012713@mvula.univen.ac.za

Abstract

This paper studies the volatility dynamics of the JSE Top40 Index by estimating a univariate GAS model with time-varying location, scale, and shape parameters (identity score scaling) and comparing its density and point-forecast performance against a standalone ARMA(3,2)-EGARCH(1,1) model and a hybrid ARMA(3,2)-EGARCH(1,1)-XGBoost framework. The GAS model is estimated on 3,515 daily observations, and several conditional densities are examined. The Student-t GAS model (GAS-STD) obtains the lowest information criteria within the GAS family (AIC = 10,188.142; BIC = 10,243.626) and exhibits statistically significant persistence in location and scale dynamics. Statistical diagnostics provide evidence of correct density calibration (Normalised Log Score = 1.1932; Uniform score = 0.4417), although residual skewness remains (IID-Test skewness $p = 0.0134$). Out-of-sample analysis shows that GAS-STD performs strongly in density and risk forecasting, producing accurate 5% VaR and ES paths and passing coverage backtests (Kupiec LRuc $p = 0.8414$; DQ $p = 0.2281$). However, short-horizon point forecasts are best delivered by the hybrid ARMA-EGARCH-XGBoost model (RMSE = 0.1386), with Diebold-Mariano tests confirming a transitive ranking: Hybrid > ARMA-EGARCH > GAS-STD. Simulation experiments highlight the sensitivity of tail behaviour to degrees-of-freedom (e.g., kurtosis $\nu = 5 \approx 7.32$). Overall, GAS-STD is a strong density and risk model for the JSE Top40, while the hybrid framework excels in short-term volatility forecasting.

Keywords: ARMA–EGARCH–XGBoost; density forecasting; GAS model; Johannesburg Stock Exchange; risk management; volatility modelling; Value-at-Risk

1. Introduction

Volatility modelling has been the focus in financial econometrics due to its importance for asset pricing, portfolio decisions, and financial stability. Financial series tend to exhibit stylised features such as volatility clustering, fat tails, and asymmetric shocks (Bollerslev 1986; Engle 1982; Nelson 1991). Historical models such as Autoregressive Conditional Heteroskedasticity (ARCH) and Generalised Autoregressive Conditional Heteroskedasticity (GARCH) have been highly useful, but their restrictive assumptions prevent them from fully capturing the richness of return distributions. Developments such as Exponential GARCH (EGARCH) by Nelson (1991) and Glosten, Jagannathan & Runkle GARCH (GJR-GARCH) by Glosten et al. (1993) addressed asymmetries more effectively, and even multivariate GARCH forms have been introduced (Bauwens et al. 2006). Emerging markets are generally characterised by higher volatility because of greater exposure to macroeconomic disturbances, institutional uncertainties, and fluctuations in global capital flows. The Johannesburg Stock Exchange (JSE), being the largest stock market in Africa, offers an important setting for examining such dynamics. In particular, the JSE Top40 Index reflects the most liquid and capitalised businesses and hence serves as a benchmark for local and foreign investors (Jefferis & Smith 2005). Modelling its volatility is therefore crucial for policy, risk management, and investment. Recent developments in machine learning have

delivered alternatives to conventional econometric approaches. eXtreme Gradient Boosting (XGBoost) (Chen & Guestrin 2016) is a robust gradient boosting method that has become very popular, and combination models of GARCH with XGBoost have been shown to improve the accuracy of short-term predictions (Maingo et al. 2025). However, these hybrids often lack interpretability, limiting their utility in regulatory applications. Creal et al. (2013) proposed a promising alternative with the Generalised Autoregressive Score (GAS) model, which was later extended by Harvey (2013). GAS models learn parameters by iterating through the score of the conditional likelihood, making them highly sensitive to fresh data.

Empirical evidence indicates that the GAS models outperform conventional GARCH in density forecasting and risk measurement, particularly under heavy-tailed or asymmetric distributions (Lazar & Xue 2020). Building on these theoretical results, Ardia et al. (2019) built the R package GAS that provides practical implementation tools and demonstrated its application using examples to financial asset returns, confirming the ability of the framework to model time-varying conditional densities. Implementations in African markets augment this affirmation further. For instance, Babatunde et al. (2021) compared GARCH and GAS models in the forecasting of daily stock prices on the Nigerian Stock Exchange and found that GAS under the Student-t as well as skewed Student-t distributions outperformed GARCH in volatility forecasting but that EGARCH was good under certain distributional assumptions. In parallel, Yaya et al. (2016) analysed the Nigerian All Share Index and showed that GAS-type models, including Beta-t-EGARCH, provided a superior explanation of jumps, outliers, and asymmetry relative to traditional GARCH models. Collectively, these papers show that GAS-based models provide better density forecasts and tail risk estimation than traditional GARCH, but their application to emerging markets remains relatively rare, and comparison studies with hybrid econometric and machine learning models remain scarce.

The study gap motivating the present work is that, while volatility modelling in South Africa has been heavily studied by using GARCH-type models (Venter & Mare 2020; Maingo et al. 2025), the Generalised Autoregressive Score (GAS) framework has received little attention, and very few studies have comparatively assessed its performance with hybrid econometric-machine learning methodologies such as GARCH-XGBoost. To address this gap, the present study employs the GAS framework to estimate and forecast the volatility of the JSE Top40 Index and compares its result with GARCH and ARMA-EGARCH-XGBoost models. The article makes four significant contributions to the literature. First, it offers one of the first comprehensive uses of GAS in South African equity markets, thereby extending its use to an important emerging market. First, it undertakes systematic benchmarking of GAS with respect to both standard econometric and hybrid machine learning approaches, yielding comparative results into the relative performance of the latter two. Second, it shows that GAS is superior in density calibration and tail risk prediction, whereas hybrid models exhibit superior performance in short-horizon point prediction. Finally, through simulation tests, the paper emphasises the central role played by heavy-tailed distributions in modelling South African equities' extreme return dynamics. In total, these contributions enrich volatility modelling literature and provide recommendations for practitioners, such as investors, regulators, and policymakers who wish to pursue risk management in the emerging market setting.

2. Data and Methodology

2.1. Data

The study is going to use daily closing prices of the JSE Top40 Index for a specific time period from 31 January 2011 to 25 February 2025, totalling 3515 trading days. The data, which is freely available, can be accessed on <https://za.investing.com/indices/ftse-jse-top-40-historical-data> (accessed on 25 April 2025). The data is collected over a five-day trading week.

2.2. Methodology

2.2.1. The GAS Models

The GAS model is a flexible class of time series models that extends the traditional GARCH framework by offering a more adaptive and dynamic manner to model time-varying volatility. This model was introduced by [Creal et al. \(2013\)](#); [Harvey \(2013\)](#) and leverages the score of the conditional likelihood function, which measures how sensitive the likelihood is to implement changes in the model parameters. By utilising the score, the GAS model updates volatility at each time point, enabling it to respond more flexibly to new information and adjust volatility estimates accordingly. A major benefit of the GAS model is its robustness to outliers, as the score functions minimise the impact of extreme values by reducing the weight of isolated observations ([Harvey 2013](#); [Junior and Alagidede 2020](#); [Alanya-Beltran 2022](#)). This leads to it being well-suited for modelling financial time series data, which often exhibit fat tails and skewness ([Harvey 2013](#); [Opschoor et al. 2018](#)). Moreover, the GAS process permits for the inclusion of additional time series features, such as asymmetry and long memory effects, making it highly versatile in capturing complex dynamics in financial markets. Furthermore, the model can be estimated efficiently using maximum likelihood estimation (MLE), making it straightforward to implement ([Ardia et al. 2019](#)). This flexibility, combined with its robustness and ability to handle complicated features of financial data, makes the GAS model a powerful tool for capturing volatility persistence and modelling time-varying volatility in financial time series.

2.2.2. GAS Models Specification

Let y_t be a $N \times 1$ vector representing the dependent variable of interest at time t . The time-varying parameter vector is denoted by λ_t , while x_t represents a set of exogenous variables (also called covariates), influencing the system. Moreover, ϕ is a vector of static parameters that remain constant over time. Define $X^t = \{x_1, \dots, x_n\}$, $\Lambda^t = \{\lambda_1, \dots, \lambda_t\}$, and $Y^t = \{y_1, \dots, y_t\}$. At any given time t , the available information consists of $\{\lambda_t, F_t\}$, where F_t is defined as

$$F_t = \{Y^{t-1}, \Lambda^{t-1}, X^t\}, \text{ for } t = 1, \dots, n. \quad (1)$$

It is assumed that the dependent variable y_t is generated from an observation density function given by

$$y_t \sim p(y_t | \lambda_t, F_t; \phi). \quad (2)$$

Additionally, we assume that the process for updating the time-varying parameter λ_t follows a standard autoregressive update equation, specifically given by

$$\lambda_{t+1} = \alpha + \sum_{i=1}^p A_i s_{t-i+1} + \sum_{j=1}^q B_j \lambda_{t-j+1}, \quad (3)$$

where α is a vector of constants, the coefficient matrices A_i and B_j have dimensions that are suitable for $i = 1, \dots, p$ and $j = 1, \dots, q$, respectively. Additionally, s_t is a function of past data, defined as $s_t = s_t(y_t, \lambda_t, F_t; \phi)$. The unknown coefficients in equation 3 are functions of the parameter vector ϕ ; specifically, $\alpha = \alpha(\phi)$, $A_i = A_i(\phi)$, and $B_j = B_j(\phi)$ for $i = 1, \dots, p$ and $j = 1, \dots, q$. The method utilised by the GAS model is based on the observation density 2 for a given parameter λ_t . When an observation y_t occurs, we update the time-varying parameter λ_t to the next period, $t + 1$ using equation 3 with

$$s_t = S_t \cdot \nabla_t, \quad \nabla_t = \frac{\partial \ln p(y_t | \lambda_t, F_t; \phi)}{\partial \lambda_t}, \quad S_t = S(t, \lambda_t, F_t; \phi). \quad (4)$$

Here, $S(\cdot)$ represents a matrix function. Since the updating mechanism in equation 3 depends on the scaled score vector in equation 4, we define equations 2 through 5 as the foundation of the GAS model with orders p and q . For simplicity, we refer to this model as GAS(p, q). The use of the score to update

λ_t is straightforward. The steepest climb route improves the model's local fit in terms of likelihood or density at time t , based on the current position of parameter λ_t . This indicates the obvious direction for updating the parameter. The score is based on the whole density, not just the first- or second-order moments of the observations y_t . The GAS framework stands out among observation-driven techniques in the literature. The GAS model uses the whole density structure to transform data and update time-varying parameters λ_t . The GAS model's scaling matrix S_t provides flexibility in how the score is updated for λ_t . Changing the scaling matrix S_t yields a unique GAS model. Each of these models has unique statistical and empirical aspects that need to be examined separately. Scaling based on score variance is a common approach in several scenarios. For instance, we can provide the scaling matrix as

$$S_t = \mathfrak{I}_{t|t-1}^{-\gamma}, \quad \mathfrak{I}_{t|t-1} = E_{t-1}[\nabla_t \nabla_t'], \quad (5)$$

where the expectation E_{t-1} is computed based on the conditional distribution of $y_t|y_{1:t-1}$. The parameter γ typically takes values of 0, $\frac{1}{2}$, or 1, though other choices for S_t values are also feasible. When $\gamma = 0$, S_t becomes the identity matrix (I), implying no scaling. If $\gamma = \frac{1}{2}$, the conditional score ∇_t is multiplied by the square root of its covariance matrix $\mathfrak{I}_t$, whereas for $\gamma = 1$, it is instead multiplied by the inverse of its covariance matrix $\mathfrak{I}_t$. The scaled score s_t follows a martingale difference property concerning the distribution of $y_t|y_{1:t-1}$, ensuring that $E_{t-1}[s_t] = 0$ for every t .

2.2.3. Parameter Estimation

After specifying the conditional distribution $p(y_t|\lambda_t, F_t; \phi)$, the following step is to estimate the model parameters ϕ . A common method is Maximum Likelihood Estimation (MLE). We aim to find ϕ that maximises the log-likelihood (LL) function given the observed data $Y^t = \{y_1, \dots, y_t\}$:

$$\mathcal{L}(\phi) = \sum_{k=1}^t \log p(y_k|\lambda_k, F_k; \phi) \quad (6)$$

In several GAS and GARCH-type models, λ_t is dynamically updated, and hence the likelihood is conditionally examined at each step using past information. Under practical conditions, numerical optimisation techniques are employed to solve the following equation 7:

$$\hat{\phi} = \arg \max_{\phi} \mathcal{L}(\phi), \quad (7)$$

where $\mathcal{L}(\phi) = \sum_{k=1}^t \log p(y_k|\lambda_k, F_k; \phi)$ is the log-likelihood function as shown in the equation 6.

2.2.4. Evaluation Metrics

To assess the performance of models, we use evaluation metrics such as Akaike Information Criterion (AIC), and Bayesian Information Criterion. The R software package is used to compute the value of AIC, and BIC. The formulas for AIC, and BIC are given by the following equations 8:

$$AIC = -2\mathcal{L}(\hat{\phi}) + 2k, \quad \text{and} \quad BIC = -2\mathcal{L}(\hat{\phi}) + k\log(n), \quad (8)$$

where $\mathcal{L}(\hat{\phi})$ denotes the maximised log-likelihood, k represents the number of estimated parameters in the model, and n is the sample size. The model with the lowest value of AIC and BIC is considered.

2.2.5. Forecast Accuracy Measures

To evaluate how well the prediction model performs, this study employs several accuracy metrics, including the Mean Absolute Square Error (MASE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The corresponding formulas are presented in Equations 9 to 11.

$$\text{MASE} = \frac{\frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t|}{\frac{1}{n-1} \sum_{t=2}^n |y_t - y_{t-1}|} \quad (9)$$

$$\text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (10)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (11)$$

2.2.6. Statistical Test of Predictive Accuracy

To statistically evaluate the difference in the accuracy of the two rival models' forecasts, this study applies the Diebold–Mariano (DM) test of ?. The DM test offers a statistical test for ascertaining if the difference between the accuracy of the two models is statistically significant by analysing the series of loss differentials between their forecast errors.

Let $e_{1,t}$ and $e_{2,t}$ represents the forecast errors from models 1 and 2 at time t . Given a chosen loss function $L(\cdot)$, such as the squared error loss $L(e_{j,t}) = e_{j,t}^2$ or absolute error loss $L(e_{j,t}) = |e_{j,t}|$, the loss differential is given as:

$$d_t = L(e_{1,t}) - L(e_{2,t}), \quad t = 1, 2, \dots, T \quad (12)$$

The Null and Alternative hypotheses (H_0 & H_1) of the DM test are:

$$H_0 : \mathbb{E}[d_t] = 0 \quad (\text{equal predictive accuracy}) \quad (13)$$

$$H_1 : \mathbb{E}[d_t] \neq 0 \quad (\text{different predictive accuracy}) \quad (14)$$

The sample mean of the loss differential is determined using the following mathematical formula:

$$\bar{d} = \frac{1}{T} \sum_{t=1}^T d_t \quad (15)$$

Since forecast horizons longer than one period ($h > 1$) may induce autocorrelation in d_t , the long-run variance of the loss differential must be estimated. This is fulfilled using a heteroskedasticity and autocorrelation consistent (HAC) estimator given by:

$$\hat{\sigma}_d^2 = \hat{\gamma}_d(0) + 2 \sum_{k=1}^{h-1} \hat{\gamma}_d(k), \quad (16)$$

where $\hat{\gamma}_d(k)$ denotes the sample autocovariance of d_t at lag k .

The DM statistic is then given by:

$$DM = \frac{\bar{d}}{\sqrt{\hat{\sigma}_d^2/T}} \quad (17)$$

Assuming the null hypothesis (H_0) and in conjunction with proper regularity conditions, the DM statistic is asymptotically standard normal, $DM \sim N(0, 1)$. A two-sided test rejects H_0 if $|DM| > z_{\alpha/2}$, where $z_{\alpha/2}$ is the standard normal critical value. But the asymptotic approximation can be deceptive in finite samples, particularly for small T or large forecast horizons. To address this, [Harvey et al. \(1997\)](#)

proposed a small-sample correction, which is referred to as the HLN correction. The correction scales the DM statistic down and ties it to a Student's t-distribution with $T - 1$ degrees of freedom, thereby reducing size distortions. Through this process, the DM test allows us to ascertain whether a model's one-step-ahead forecasting performance differs significantly from that of its competitor, offering a firm statistical basis for model comparison.

2.2.7. Value-at-Risk and Expected Shortfall

Let L denote the loss random variable (so larger L is worse). Let $F_L(x) = \mathbb{P}(L \leq x)$ be its cumulative distribution function (CDF). Denote the lower α -quantile by

$$q_\alpha(L) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}. \quad (18)$$

If F_L is continuous and strictly increasing then $q_\alpha(L) = F_L^{-1}(\alpha)$ (Acerbi & Tasche 2001; McNeil et al. 2005).

Value-at-Risk (VaR)

The Value-at-Risk at level $\alpha \in (0, 1)$ is defined as the α -quantile of the loss distribution:

$$\text{VaR}_\alpha(L) = q_\alpha(L) = \inf\{x \in \mathbb{R} : F_L(x) \geq \alpha\}. \quad (19)$$

(Standard definition by (Artzner et al. 1999; McNeil et al. 2005)).

For normal example: If $L \sim \mathcal{N}(\mu, \sigma^2)$ then

$$\text{VaR}_\alpha(L) = \mu + \sigma z_\alpha, \quad z_\alpha := \Phi^{-1}(\alpha), \quad (20)$$

where Φ is the standard normal CDF.

Expected Shortfall (ES) — continuous case

Expected Shortfall at level α is the conditional expectation:

$$\text{ES}_\alpha(L) = \mathbb{E}[L \mid L \geq \text{VaR}_\alpha(L)]. \quad (21)$$

For continuous distributions (with density f_L):

$$\text{ES}_\alpha(L) = \frac{1}{1 - \alpha} \int_{\text{VaR}_\alpha(L)}^{\infty} x f_L(x) dx. \quad (22)$$

If we change a variable $u = F_L(x)$ in (22). Then $du = f_L(x) dx$ and $x = F_L^{-1}(u) = \text{VaR}_u(L)$. Thus

$$\text{ES}_\alpha(L) = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_u(L) du. \quad (23)$$

For normal case closed form: If $L \sim \mathcal{N}(\mu, \sigma^2)$, then

$$\text{ES}_\alpha(L) = \mu + \sigma \frac{\varphi(z_\alpha)}{1 - \alpha}, \quad (24)$$

where φ is the standard normal PDF.

General distributions (robust definition)

For possibly discontinuous distributions, a robust definition is:

$$\text{ES}_\alpha(L) = \frac{1}{1 - \alpha} \int_\alpha^1 q_u(L) du, \quad q_u(L) = \inf\{x : F_L(x) \geq u\}. \quad (25)$$

Coherence (subadditivity of ES)

For any losses L_1, L_2 with sum $L = L_1 + L_2$,

$$\text{ES}_\alpha(L) \leq \text{ES}_\alpha(L_1) + \text{ES}_\alpha(L_2), \tag{26}$$

showing ES is subadditive (coherent) (Artzner et al. 1999; Acerbi & Tasche 2001). In contrast, VaR need not be:

$$\text{VaR}_\alpha(L_1 + L_2) > \text{VaR}_\alpha(L_1) + \text{VaR}_\alpha(L_2) \quad (\text{possible counterexamples}). \tag{27}$$

Rockafellar–Uryasev CVaR representation

Define

$$\phi_\alpha(x) := x + \frac{1}{1-\alpha} \mathbb{E}[(L-x)_+], \quad (y)_+ := \max\{y, 0\}. \tag{28}$$

Then

$$\text{ES}_\alpha(L) = \min_{x \in \mathbb{R}} \phi_\alpha(x), \tag{29}$$

and any minimizer x^* is a VaR at level α (Rockafellar & Uryasev 2000).

This is an example of a quote.

3. Empirical Results and Discussion

3.1. Fitting of the GAS Model

Table 1 presents the evaluation metrics for the GAS model under seven different conditional distributions, namely the Student-t (STD), skewed Student-t (SSTD), Gaussian, the skew-Gaussian, Asymmetric Student-t (AST) with two tail decay parameters, Asymmetric Student-t (AST1) with one tail decay parameter, and Asymmetric Laplace (ALD). The comparison is based on the AIC and the BIC, where lower values indicate a better model fit while accounting for model complexity. The GAS-STD model recorded AIC and BIC values of 10185.108 and 10228.261, respectively, which are marginally lower than the corresponding values for the GAS-SSTD, GAS-Gaussian, GAS-skew-Gaussian, GAS-AST, GAS-AST1, GAS-ALD models. These results suggest that, although both models provide a relatively similar fit to the data, the GAS-STD model offers a slightly better balance between goodness-of-fit and parsimony, making it the more preferable specification according to both criteria.

Table 1. Evaluation metrics for GAS model under seven different conditional distributions.

Evaluation Metrics		
Model	AIC	BIC
GAS-STD	10188.142	10243.626
GAS-SSTD	10190.46	10264.44
GAS-Gaussian	10250.864	10287.852
GAS-skew-Gaussian	10253.830	10309.313
GAS-AST	10203.182	10283.324
GAS-AST1	10223.508	10285.155
GAS-ALD	10327.530	10383.013

3.1.1. Parameter Estimates

Table 2 presents the parameter estimates for the estimated univariate GAS model with an STD. The model was estimated using 3,515 observations with time-varying location, scale, and shape parameters and identity score scaling. The estimates validate that the location constant term, κ_1 , is positive and statistically significant at the 1% level (with $p < 0.01$), indicating a small but persistent mean level in the series. The κ_2 and κ_3 constants are negative but statistically non-significant ($p > 0.05$),

indicating no robust evidence of a shift in the baseline levels of the shape and scale parameters. For the autoregressive terms, a_1 and b_1 are both significantly large ($p < 0.01$), indicating a high degree of persistence in location dynamics. Similarly, a_2 and b_2 are significantly large ($p < 0.01$), indicating a high degree of persistence in scale dynamics. The shape dynamics parameters, a_3 and b_3 , indicate that b_3 is crucial ($p < 0.01$) while a_3 is not statistically important ($p > 0.05$), indicating that even with persistence, short-run shocks have very little impact. That most of the persistence parameters are significant indicates that past values for the time-varying parameters play an important role in determining current levels, particularly for location and scale.

Table 2. Parameter estimates of the univariate GAS model with STD.

Parameter	Estimate	Std. Error	t-value	Pr(> t)
κ_1	0.02733955	0.007908547	3.456962	0.0002731506
κ_2	-0.003164797	0.002533474	-1.249193	0.1057973
κ_3	-0.1470870	0.1623964	-0.9057283	0.1825398
a_1	9.994037×10^{-7}	4.962956×10^{-12}	2.013727×10^5	0.0000000
a_2	0.1597487	0.02355320	6.782462	5.91×10^{-12}
a_3	0.7711878	0.9345721	0.8251774	0.2046354
b_1	0.4973487	2.701866×10^{-5}	1.840760×10^4	0.0000000
b_2	0.9782114	0.006412429	152.5493	0.0000000
b_3	0.9283452	0.07790056	11.91705	0.0000000

3.1.2. Diagnostic Evaluation of GAS-STD Model

The Probability Integral Transform (PIT) histogram plot shown in Figure 1 displays the distribution of the PIT values obtained from the out-of-sample forecasts of the GAS model. Ideally, these PIT values should be uniformly distributed on the interval $[0, 1]$ if the predictive densities are correctly specified. In the histogram, the frequencies of PIT values across the bins appear relatively even, fluctuating moderately around the expected uniform density, represented by the red reference lines. There is no discernible clustering of the PIT values at 0 and 1, indicating that the model does not underestimate or overestimate tail risks consistently. Likewise, no sharp U-shape or bell shape is observed, indicating no significant deviations caused by underdispersion or overdispersion in the predicted distribution. Some minor oscillations are observed, which are within tolerable random variation associated with finite samples. The PIT histogram validates that the GAS model produces well-calibrated density forecasts, and no strong evidence of misspecification exists for the conditional distribution of the log returns during the period under investigation.

Table 3 presents the average density forecast backtesting scores for the GAS-STD model. The Normalised Log Score (1.1932) indicates an acceptable overall forecast accuracy. The uniform score (0.4417) suggests reasonable calibration of PIT values, while the centre (0.1279) and tails (0.0744) scores reflect balanced performance across the distribution. Left-tail (0.2054) and right-tail (0.2363) scores show slightly better accuracy in the left tail. In conclusion, the model demonstrates adequate density forecasting ability with consistent performance across central and extreme regions.

Table 3. Average Backtest Scores for Density Forecast Evaluation of the GAS-STD Model.

Metric:	NLS	Uniform	Center	Tails	Tail_L	Tail_R
Value:	1.1932	0.4417	0.1279	0.0744	0.2054	0.2363

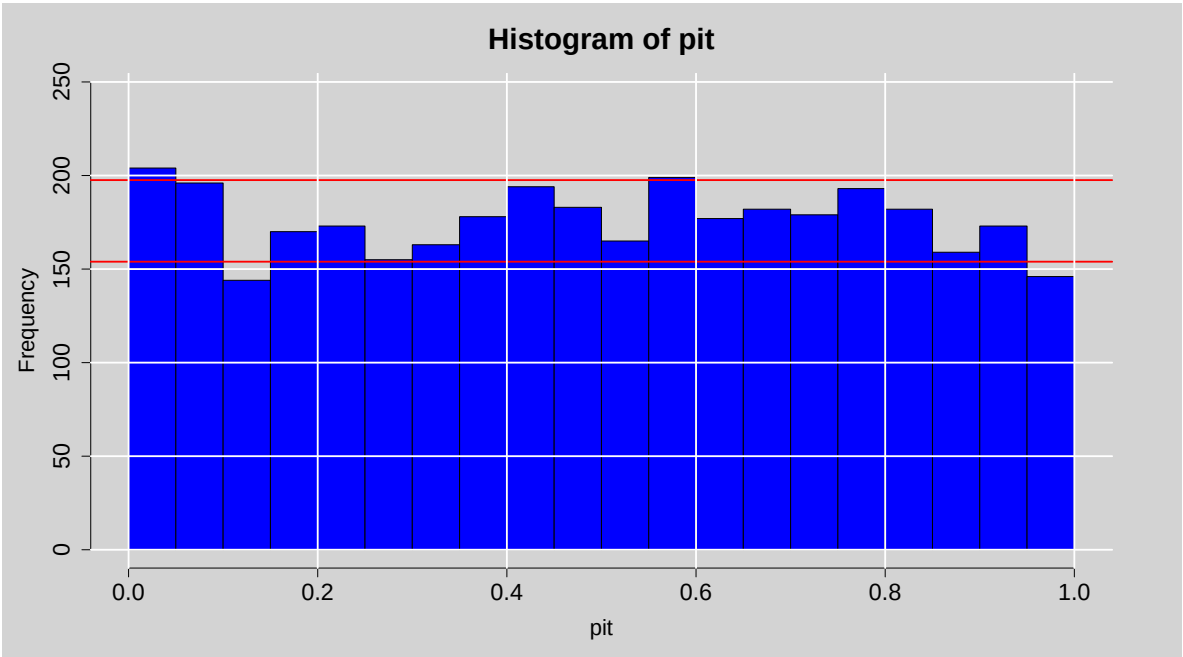


Figure 1. Histogram plot of the PIT.

Table 4 presents the PIT test results under IIDTest. The test takes into account whether the PITs are independent and identically distributed as $U(0,1)$. Test 1 (mean) obtains a borderline value ($p = 0.0510$), which suggests that the uniformity condition is narrowly satisfied. Test 2 (variance) obtains a non-significant value ($p = 0.5351$), which suggests no sign of conditional heteroskedasticity of the PITs. Test 4 (kurtosis) fails to reject the null ($p = 0.2264$), indicating that the model does a satisfactory job in modelling the tails. However, Test 3 (skewness) is significant ($p = 0.0134$), which indicates some residual asymmetry or dependence in the PITs. These results indicate that while the model does a good job in modelling variance and kurtosis, it might require refinement to handle skewness in the predictive distribution. But the GAS-STD model is better due to its smaller values of AIC and BIC, indicating general model performance for all that was obtained in Test 3 as a slight skewness.

Table 4. Lagrange Multiplier (LM) Tests for the First Four Conditional Moments of the PITs.

	Test 1	Test 2	Test 3	Test 4
Statistic	31.32747	18.79682	36.51653	24.37492
Critical Value	31.41043	31.41043	31.41043	31.41043
p-value	0.0510	0.5351	0.0134	0.2264

Figure 2 shows the time-varying dynamics of the filtered parameters of the estimated GAS model, namely the location, scale, and shape. The first panel plots the location parameter, which is quite constant throughout the sample, varying around a fixed level with imperceptible changes. This stability means that the conditional mean of the series does not change dramatically over time and implies that the majority of the dynamics are represented by higher-order parameters. The second panel shows the scale parameter, which is a measure of conditional volatility. In this case, there is a considerable clustering of volatility, with apparent periods of increased variability interspersed with more tranquil periods. Notice a strong volatility spike midway through the sample, which corresponds to a time of market instability or shocks in the underlying data. The final panel graphs the shape parameter, which determines the heaviness of the tails of the Student-t distribution. The shape evolves significantly over time with pronounced plunges for episodes of very fat-tailed behaviour when the probability of outliers or aberrant returns is higher. The plot shows that while the mean does not change, volatility and tail behaviour exhibit significant time variation.

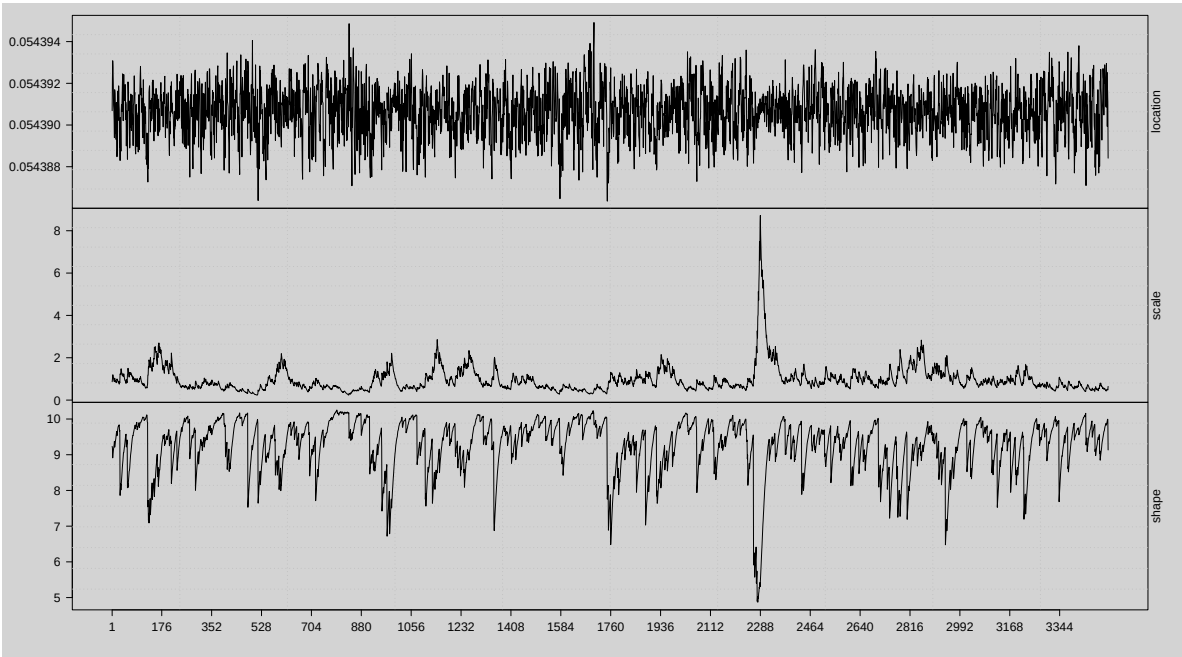


Figure 2. Time-varying parameter estimates of the fitted univariate GAS model with STD, showing the evolution of the conditional location, scale, and shape parameters.

3.1.3. Out-of-Sample Forecasts (Rolling Forecasts) of the fitted GAS-STD Model

Table 5 presents the first ten rolling forecasts produced by the GAS-STD model. Results include the conditional mean (location) being fairly stable over the forecasting period, changing in the range of 0.0525. Conditional volatility (scale), however, changes more and spans from 0.6308 to 0.8311, reflecting dynamic model responses to changing market uncertainty. The estimated shape parameter is always considerably greater than 9, indicating that the fitted Student-t distribution has moderately heavy tails. This suggests the ability of the model to handle fat-tailed behaviour in the return distribution, as essential in financial risk modelling. The forecasts indicate, however, that, although the level of returns is stable around its mean value, volatility and tail heaviness exhibit considerable dynamics over time.

Table 5. First ten rolling forecasts of the fitted GAS-STD model.

Horizon	Location	Scale	Shape
$T + 1$	0.0525	0.6464	9.7908
$T + 2$	0.0525	0.6403	9.8535
$T + 3$	0.0525	0.7544	9.6885
$T + 4$	0.0525	0.7429	9.7573
$T + 5$	0.0525	0.6863	9.7715
$T + 6$	0.0525	0.6421	9.7972
$T + 7$	0.0525	0.6308	9.8567
$T + 8$	0.0525	0.8311	9.2013
$T + 9$	0.0525	0.7653	9.2525
$T + 10$	0.0525	0.7087	9.3046

The conditional location plot in Figure 3 illustrates the time-varying expected return of the JSE Top40 Index over the rolling forecast horizon. While the first ten forecast points remain constant at approximately 0.0525, the long series of 250 points makes minimal fluctuations, with small upward and downward deviations from the mean level. These small movements are modest short-run trends in the

expected return, which capture subtle directional changes in market sentiment without pronounced directional movements. Typically, the conditional mean is quite stable, consistent with low mean daily returns typical for equity indices.

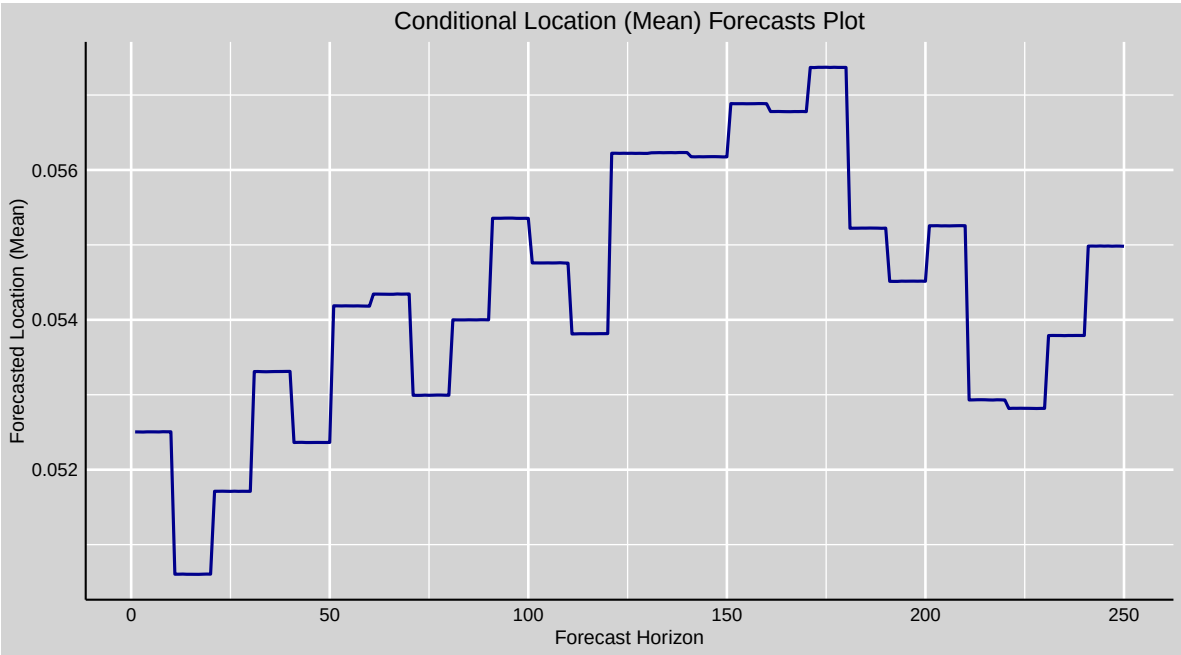


Figure 3. Forecasts plot of the forecasted conditional location (mean) of the fitted GAS-STD model.

The conditional scale (volatility) plot in Figure 4 displays the time-series volatility of the returns and indicates considerable changes over the forecast period. Volatility initially falls a bit, then registers peaks, which correspond to instances of increased uncertainty and possibly market tension. The remaining points show moderate levels of volatility, which correspond to fairly stable market conditions. These trends highlight the model’s strength in detecting short-run movements in risk, which is important for accurate risk determination and portfolio administration.

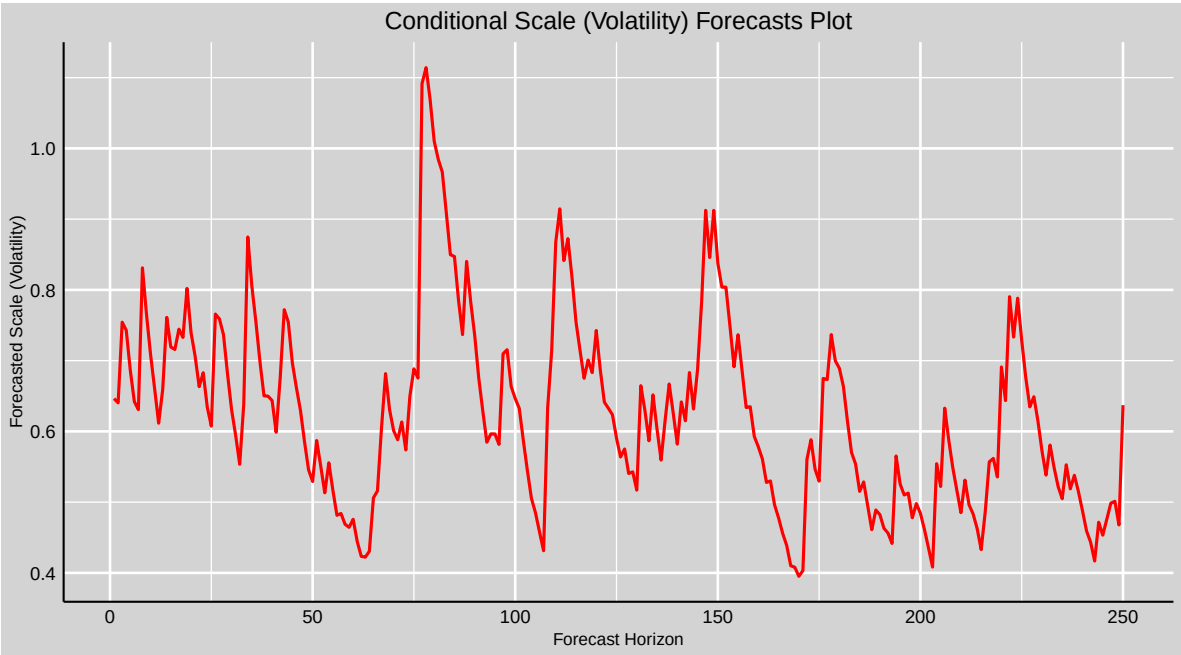


Figure 4. Forecasts plot of the forecasted conditional scale (volatility) of the fitted GAS-STD model.

The shape parameter curve in Figure 5 of the conditional shape, which reflects the Student’s t-distribution’s degrees of freedom, trends slightly above 9.7 with some drops between 8 and 10. The distributions converge to normal as the shape parameter grows larger, and smaller values represent heavier tails. Volatility observed means that the return distributions predicted are generally well-approximated by a nearly normal distribution but periodically encounter bouts of heavier tails, which is valuable for the extreme return event modelling in risk analysis.

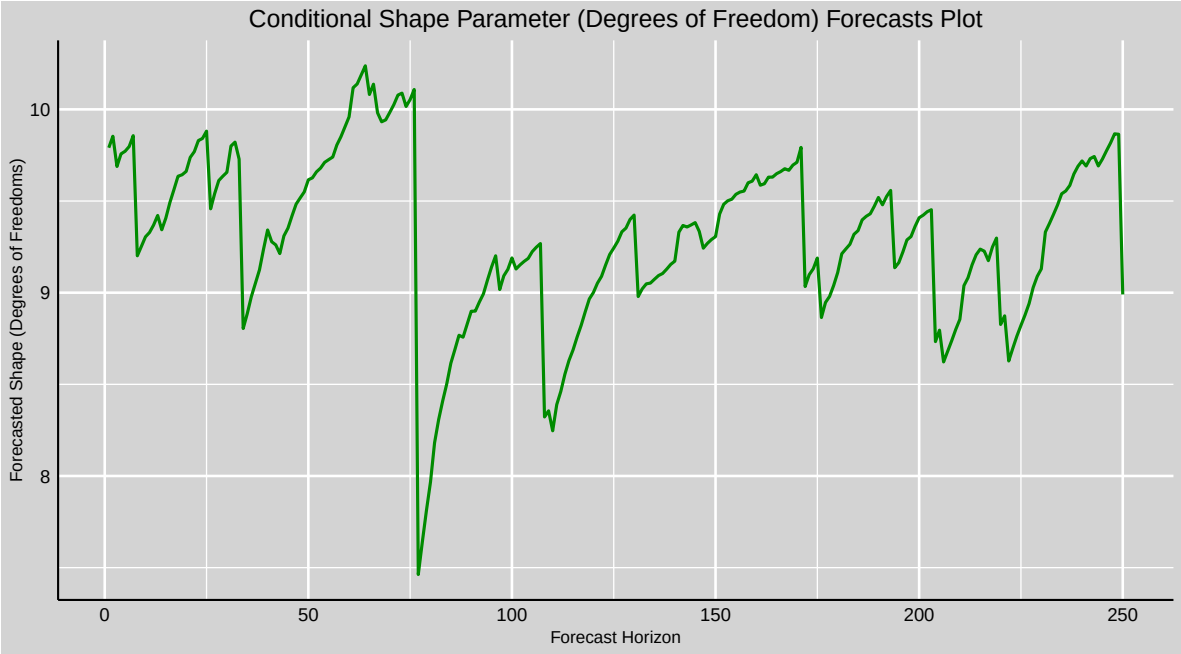


Figure 5. Forecasts plot of the forecasted conditional scale (volatility) of the fitted GAS-STD model.

Figure 6 shows the realised volatility against the forecasted conditional scale (volatility). The smoothed and more stable character of the forecasts of the conditional volatility is evident but succeeds in describing the overall patterns and volatility clusters fairly satisfactorily over the course of time. The surges in the higher volatility for the realised data near horizons 40, 100, and 200 are reflected in corresponding surges in the forecasted scale, though with lower amplitude. Such behaviour is consistent with the mission of volatility models, which anticipate explaining persistent dynamics and volatility clustering but not replicating every individual spike. These results indicate that the GAS-STD model provides an adequate explanation of conditional mean and volatility, as well as a stronger explanation of volatility dynamics than of return levels.

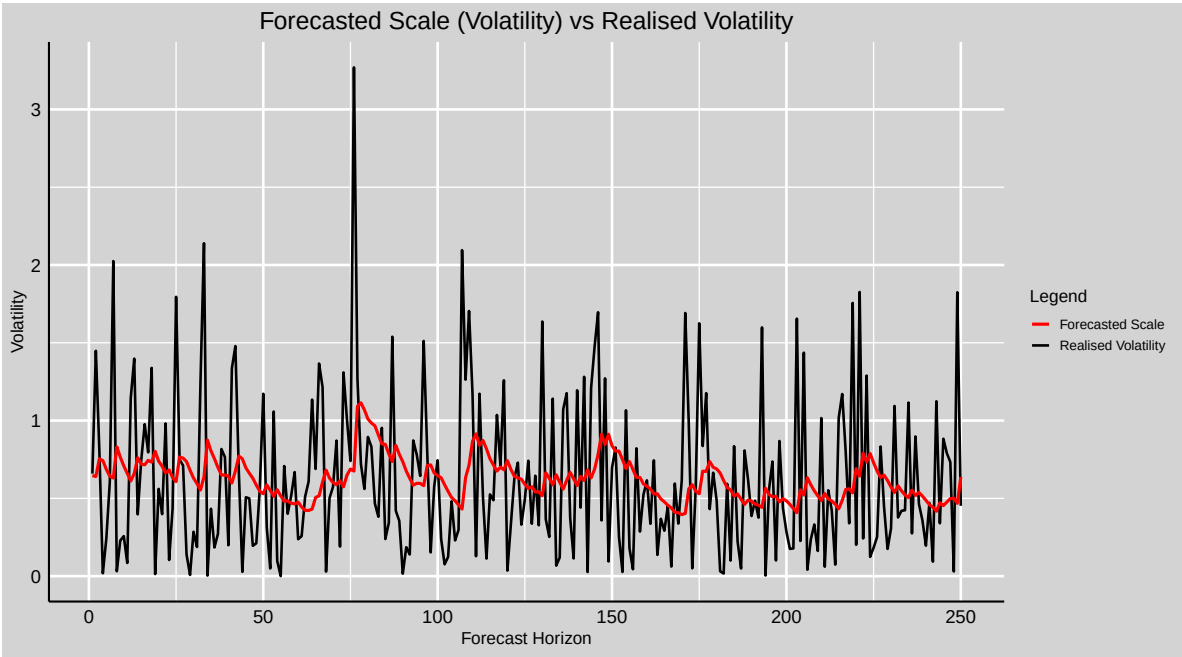


Figure 6. Plot of the forecasted conditional scale (volatility) versus realised volatility of the fitted GAS-STD model.

Table 6 displays some of the rolling risk forecasts of the 5% Value-at-Risk (VaR) and Expected Shortfall (ES) from the GAS-STD model. Both VaR and ES are invariably negative in the results, which reflects the model’s risk estimation in the return series from the downside. VaR estimates from approximately -1.0925 to -1.4672 indicate the estimated tail risk at the 5% confidence level, while ES estimates from 1.4704 to -1.9787 indicate the mean loss in the tail 5% of situations. As expected, ES is more negative than VaR, confirming its characteristic as a coherent risk measure giving a closer estimate of tail risk than VaR. Figure 7 further illustrates these numerical results by displaying the full rolling paths of the 5% VaR and ES across the forecast horizon. These risk measure exhibit strong time variation, with fluctuations that correspond to periods of heightened volatility. As theory would expect, ES is below VaR for all forecast horizons, as it accounts for the severity of losses beyond the VaR level. Growing differences between the two measures at more volatile times (e.g., past forecast horizons of 80, 150, and 220) again illustrate how ES is reacting more sensitively to extreme negative returns. The Table 6 and Figure 7 provide a neat image of the tail risk behaviour of the GAS-STD model in favour of the idea that it is essential to utilise both VaR and ES for complete risk assessment.

Table 6. Rolling Risk Forecasts from the 5% VaR and ES.

VaR (5%)	ES (5%)
-1.1215	-1.5098
-1.1098	-1.4935
-1.3192	-1.7742
-1.2974	-1.7443
-1.1943	-1.6069
-1.1137	-1.4993
-1.0925	-1.4704
-1.4672	-1.9787
-1.3460	-1.8160
-1.2418	-1.6760

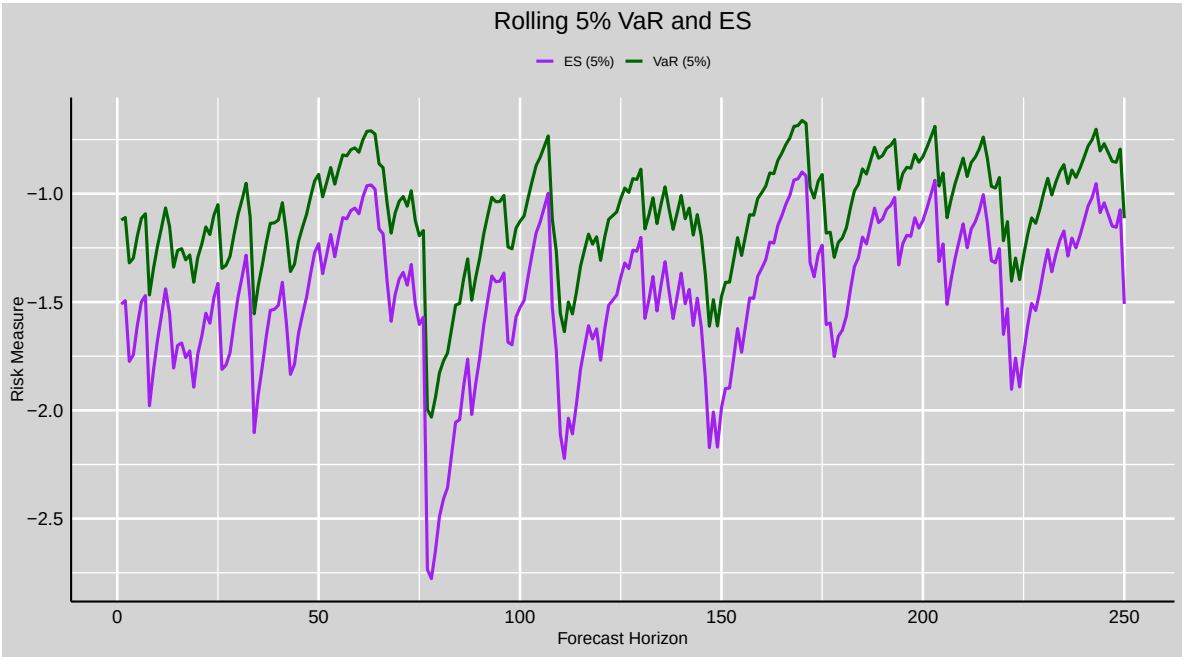


Figure 7. Rolling Risk Forecasts plot from the 5% VaR and ES.

Table 7 shows the forecast measures of performance for both the conditional location (mean) and conditional scale (volatility) of the estimated GAS-STD model. It is revealed from the findings that there are lower errors in scale forecasts than in the location forecasts based on lower RMSE (0.5373 vs. 0.8055) and MAE (0.4197 vs. 0.6233) values. This suggests that the model is more appropriate to predict volatility dynamics than return levels, which is consistent with the typical dynamics of financial time series where volatility is more persistent and hence more predictable than returns. However, the MAPE measures for the two series are very high (184.5542% for location and 1083.5356% for scale), as one would anticipate, because returns and volatility occasionally assume values close to zero, thereby inflating percentage-error measures. Finally, the MASE values indicate that both forecasts are quite good relative to a naïve baseline, with values close to but below unity (0.7026 for the location and 0.7464 for the scale). These results indicate that the GAS-STD model produces optimal volatility predictions with excellent persistence in its volatility, whereas conditional mean predictions are poorer due to the inherent volatility uncertainty of financial return levels.

Table 7. Forecast accuracy metrics for conditional location and scale of the fitted GAS-STD model.

Accuracy Measure	Location	Scale
MASE	0.7026	0.7464
RMSE	0.8055	0.5373
MAE	0.6233	0.4197

The backtest results of the 5% Value-at-Risk (VaR) model in Table 8 are discovered to have adequate predictive power. The Kupiec Unconditional Coverage Test (LRuc) yielded a test statistic of 0.0400 and a p-value of 0.8414, failing to reject the null hypothesis of correct unconditional coverage. This is a pointer that the observed rate of violations is statistically consistent with the expected rate at the specified confidence level. The Christoffersen Conditional Coverage Test (LRcc) produced a statistic of 3.5026 and a p-value of 0.1735, not rejecting once more the null hypothesis of adequate conditional coverage. It therefore follows that VaR violations are both time-series independent and at the correct frequency. The Actual/Expected Exceedance Ratio (AE) was 0.9673, very close to unity, which testifies that the number of exceedances is very close to the model’s theoretical expectation. The Acerbi-Szekely (AD) tests provided ADmean = 0.4723 and ADmax = 1.6313, which are both low

values signifying no systematic underestimation or overestimation of the tail losses. The Dynamic Quantile (DQ) test statistic value of 9.3557 and p-value of 0.2281 also signify the absence of significant autocorrelation in exceedances and that the model is good in the detection of the time dependence of the tail events. Finally, the observed loss function value of 0.1097 is relatively low, indicating high predictive precision in the left tail of the distribution. Overall, these diagnostic results individually suggest that the VaR model reflects high statistical fitness and credibility in quantifying downside risk at the 5% level of the out-of-sample period.

Table 8. Backtest Results for the 5% VaR Model.

Test	Statistic	p-value
Kupiec Unconditional Coverage (LRuc)	0.0400	0.8414
Christoffersen Conditional Coverage (LRcc)	3.5026	0.1735
Actual/Expected (AE) Ratio	0.9673	–
ADmean	0.4723	–
ADmax	1.6313	–
Dynamic Quantile (DQ)	9.3557	0.2281
Loss Function	0.1097	–

3.1.4. Simulation-Based Tail Risk Analysis

In this research, there were a total of 10,000 observations simulated from the GAS-STD model for every fixed value of shape parameter or degrees of freedom, $\nu = (5, 8, 10, 15, 20, 25, 30)$. Table 9 displays kurtosis of simulated data, and it is clear that lower values of ν create series with heavier tails, as indicated by higher kurtosis (e.g., 7.32 for $\nu = 5$), while higher values of ν create distributions with kurtosis closer to the Gaussian standard of 3 (e.g., 3.14 for $\nu = 30$). This finding is also indicated by the QQ plots in Figure 8, comparing the quantiles of the simulated data with those of a standard normal distribution. For lower values of ν , i.e., 5, 8, and 10, the QQ plots indicate strong deviations from the 45-degree line at the tails, confirming strong leptokurtic behaviour. By contrast, larger values of ν , especially 20, 25, and 30, fall exactly on the theoretical line, showing distributions that are close to normal. These evidence suggests that smaller degrees of freedom produce greater tail heaviness, with $\nu = 5$ being the best choice when extreme tail behaviour is assigned greatest significance. This option represents the largest leptokurtosis, particularly for the simulation of financial asset returns and estimation of the risk of rare extreme market events.

Table 9. Kurtosis values of simulated data generated from the GAS-STD model under fixed shape parameter or degrees of freedom ($\nu = 5, 8, 10, 15, 20, 25, 30$).

Degrees of Freedom (ν):	$\nu = 5$	$\nu = 8$	$\nu = 10$	$\nu = 15$	$\nu = 20$	$\nu = 25$	$\nu = 30$
kurt value:	7.3197	4.4795	3.9495	3.4740	3.3005	3.1968	3.1401

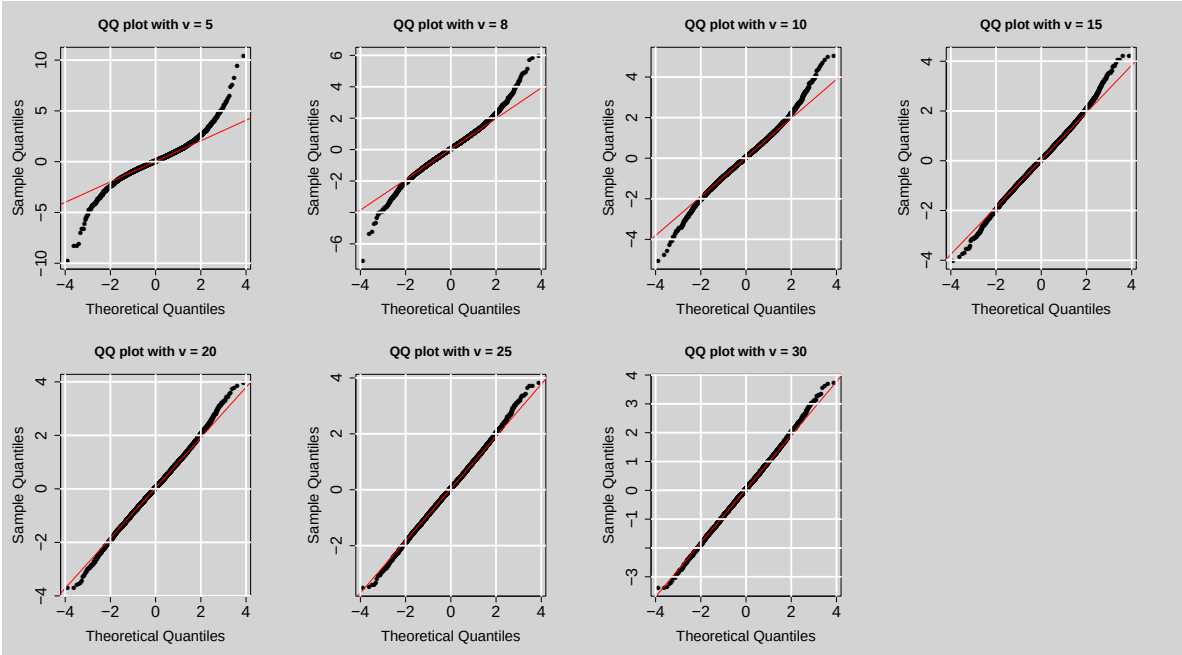


Figure 8. QQ plots of simulated data from the GAS-STD model under fixed shape parameter or degrees of freedom values ($\nu = 6, 8, 10, 15, 20, 25, 30$).

Figure 9 represents the histogram of simulated data together with the estimated 5% Value-at-Risk (VaR) and Expected Shortfall (ES) thresholds. Most of the distribution is concentrated around zero, indicating typical behaviour of financial return series that fluctuate around the mean. However, fat tails also exist, as indicated by the presence of extreme negative observations in the left tail. The vertical dashed red lines represent the 5% VaR and ES levels, which were calculated to be ~ 1.8032 and ~ 2.6702 , respectively. The 5% VaR provides the cutoff point after which the worst 5% of losses lie, whereas the ES is an estimate of the average loss given that it crosses this threshold. The locations of these thresholds also readily demonstrate that although infrequent instances of catastrophic downside loss are themselves fairly frequent, their size is substantial, with mean losses (ES) even larger than the VaR cutoff. This finding confirms the presence of leptokurtosis in the simulated data and the necessity of modelling fat-tailed distributions to optimally describe the risk profile of financial returns.

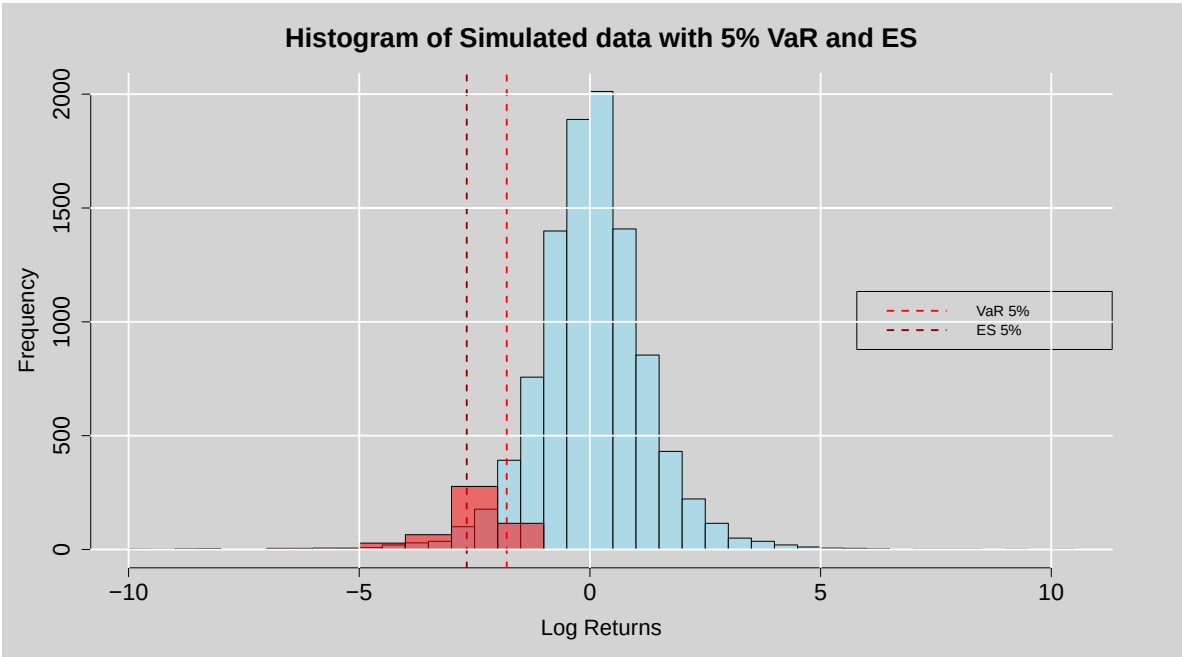


Figure 9. Histogram plot of the simulated data with 5% Var and ES.

Figure 10 plots the time series plot of the simulated data across 10,000 observations with the 5% VaR level plotted in red. The simulated data exhibit large fluctuations about the zero mean, with some sharp spikes indicating instances of very high variation. The red points that fall below the 5% VaR level are those instances in which losses exceeded the risk threshold, an instance of very high negative returns. These excesses are randomly distributed throughout the sample, suggesting the existence of tail risk in the simulated distribution. The frequency and diversity of these breaches suggest that the model captures both the middle dynamics of return behaviour and fairly accurately reflects the fat-tailed nature of financial returns, which is fundamental for risk management applications.

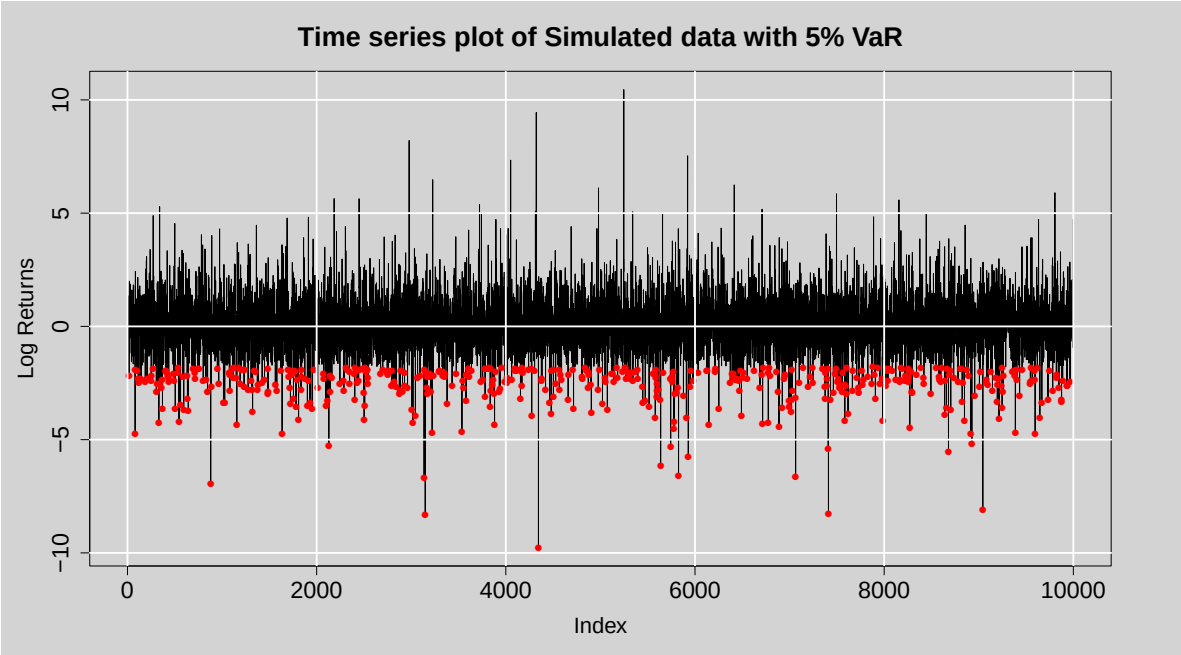


Figure 10. Time series plot of the simulated data with 5% VaR.

3.1.5. Forecasts Comparisons of the Standalone ARMA(3,2)-EGARCH(1,1) Model, GAS-STD Model and ARMA(3,2)-EGARCH(1,1)-XGBoost Hybrid Model

Table 10 presents the forecast accuracy statistics of the ARMA(3,2)-EGARCH(1,1), GAS-STD, and ARMA(3,2)-EGARCH(1,1)-XGBoost models. For all the measures, the ARMA(3,2)-EGARCH(1,1)-XGBoost model performs better than the remaining models, with the lowest MASE, RMSE and MAE. The ARMA(3,2)-EGARCH(1,1) model comes in second overall, and the GAS-STD model has comparatively higher errors in most measures. These results imply that inclusion of the XGBoost module drastically improves predictive power, catching nuances the fully parametric models may not.

Table 10. Forecast Accuracy Measures for ARMA(3,2)-EGARCH(1,1), GAS-STD, and ARMA(3,2)-EGARCH(1,1)-XGBoost Models.

Accuracy Measure	ARMA(3,2)-EGARCH(1,1)	GAS-STD	ARMA(3,2)-EGARCH(1,1)-XGBoost
MASE	0.6827	0.7464	0.0534
RMSE	1.0845	0.5373	0.1386
MAE	0.8176	0.4197	0.0595

The Diebold-Mariano (DM) test was employed to formally examine whether in the rival models there are significant differences in the predictive accuracy. The results in Table 11 reject the null hypothesis of equal predictive accuracy for all the pairwise comparisons and provide evidence of statistically significant differences in forecast performance. Under the working convention $d_t = L(model1) - L(model2)$ (so a negative DM statistic indicates the first named model has lower loss), a comparison of Model A (ARMA(3,2)-EGARCH(1,1)-XGBoost) with Model B (GAS-STD) yields a DM

statistic of -6.5331 ($p = 0.0000$), and Model A is demonstrated to achieve significantly lower forecast loss than Model B. Moreover, Model A’s DM statistic against Model C (ARMA(3,2)-EGARCH(1,1)) is -2.2154 ($p = 0.02825$), demonstrating that Model A also significantly dominates Model C. Crucially, the pairwise test between Model C and Model B now yields a positive DM statistic of 2.4541 ($p = 0.01481$); under the sign convention above, this positive value demonstrates that Model B has greater forecast loss than Model C, and so Model C is significantly better than Model B. Taken together, these findings establish the transitive ranking of predictive performance: Model A (ARMA(3,2)-EGARCH(1,1)-XGBoost) $>$ Model C (ARMA(3,2)-EGARCH(1,1)) $>$ Model B (GAS-STD), where all pairwise differences are statistically significant at conventional levels. This ranking shows that the hybrid model has the best forecasts, followed by the pure ARMA–EGARCH specification, while the GAS-STD model is the weakest of the three.

Table 11. Diebold–Mariano (DM) test results for pairwise model comparisons.

Comparison	DM Statistic	p-value	Conclusion
Model A vs Model B	-6.5331	0.0000	Model A is significantly better than Model B
Model A vs Model C	-2.2154	0.02825	Model A is significantly better than Model C
Model B vs Model C	-2.4541	0.01481	Model C is significantly better than Model B

The empirical scores plot displayed in Figure 11 presents a comparison between the ARMA(3,2)–EGARCH(1,1) model and the hybrid ARMA(3,2)–EGARCH(1,1)–XGBoost model based on Murphy diagrams. The results confirm that the hybrid model has lower empirical scores over a wider parameter space compared to the standalone ARMA(3,2)–EGARCH(1,1). The implication is that the hybrid model provides more accurate volatility forecasts, particularly for the right tail of the distribution, where it provides a more sustained and smoother reduction in scores. Conversely, the ARMA(3,2)–EGARCH(1,1) model by itself does not have such a steep peak but a more limited fall, which reflects worse performance in explaining volatility dynamics at longer horizons. Figure 12 also substantiates this conclusion by presenting the difference between the two models’ scores. The negative values of the score differences across most of the parameter domain indicate that the hybrid model outperforms the standalone ARMA(3,2)–EGARCH(1,1). The confidence band reinforces the robustness of this finding, as it remains below zero for a substantial portion of the parameter range. Only in the very narrow region around the origin do the two models exhibit comparable performance, after which the hybrid model clearly dominates. These results highlight the predictive gain obtained by combining the traditional ARMA(3,2)–EGARCH(1,1) specification with the nonlinear learning ability of XGBoost, confirming that the hybrid framework provides superior predictive accuracy in volatility forecasting.

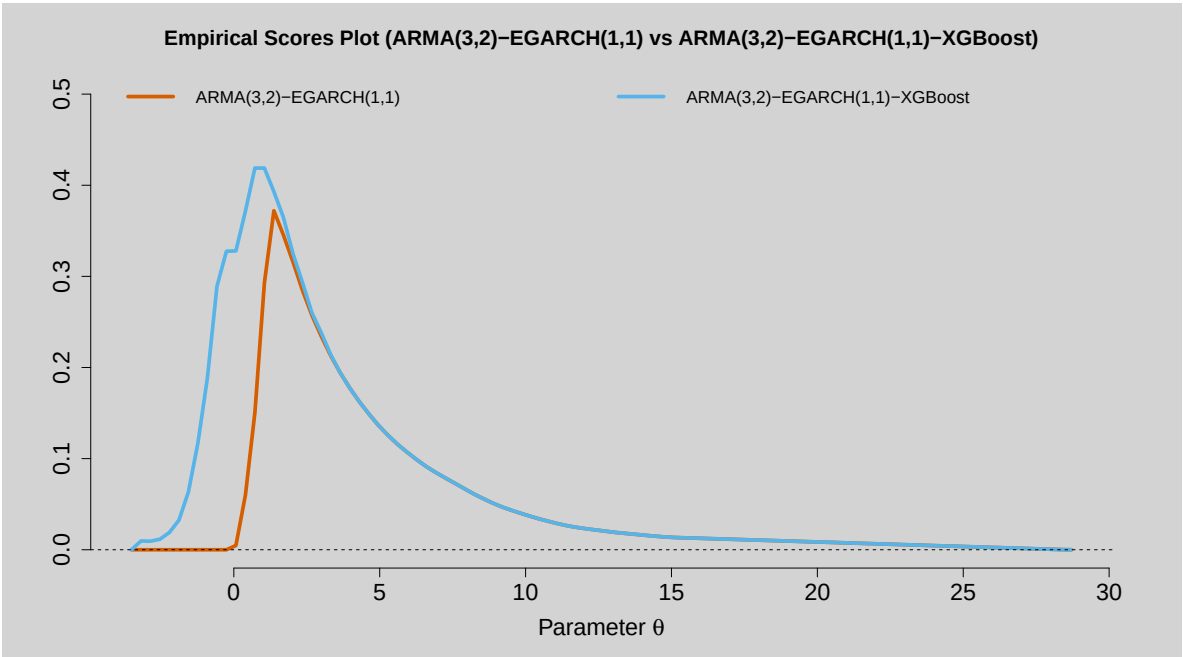


Figure 11. Empirical Scores Plot of ARMA(3,2)-EGARCH(1,1) versus ARMA(3,2)-EGARCH(1,1)-XGBoost.

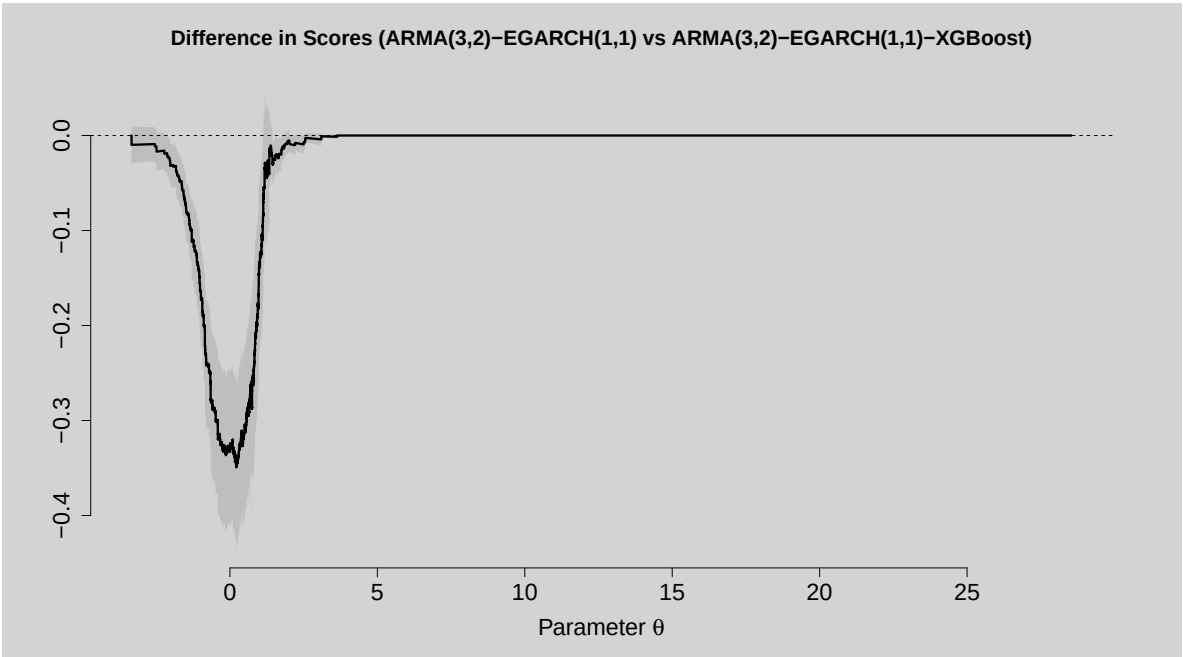


Figure 12. Difference in Scores Plot of ARMA(3,2)-EGARCH(1,1) versus ARMA(3,2)-EGARCH(1,1)-XGBoost.

Figures 13 and 14 show the predictive capability of the GAS-STD model compared to the ARMA(3,2)-EGARCH(1,1)-XGBoost model. The empirical scores plot shows that both models reach their peak in roughly the same region but that the ARMA(3,2)-EGARCH(1,1)-XGBoost scores higher across the entire parameter set, which means better predictive capability. Difference in scores plot confirms this, as the curve is below zero for nearly all periods, indicating that hybrid ARMA(3,2)-EGARCH(1,1)-XGBoost model is better in all horizons than GAS-STD, particularly the short horizon. At higher horizons, differences tend towards zero, indicating similar performance between both models.

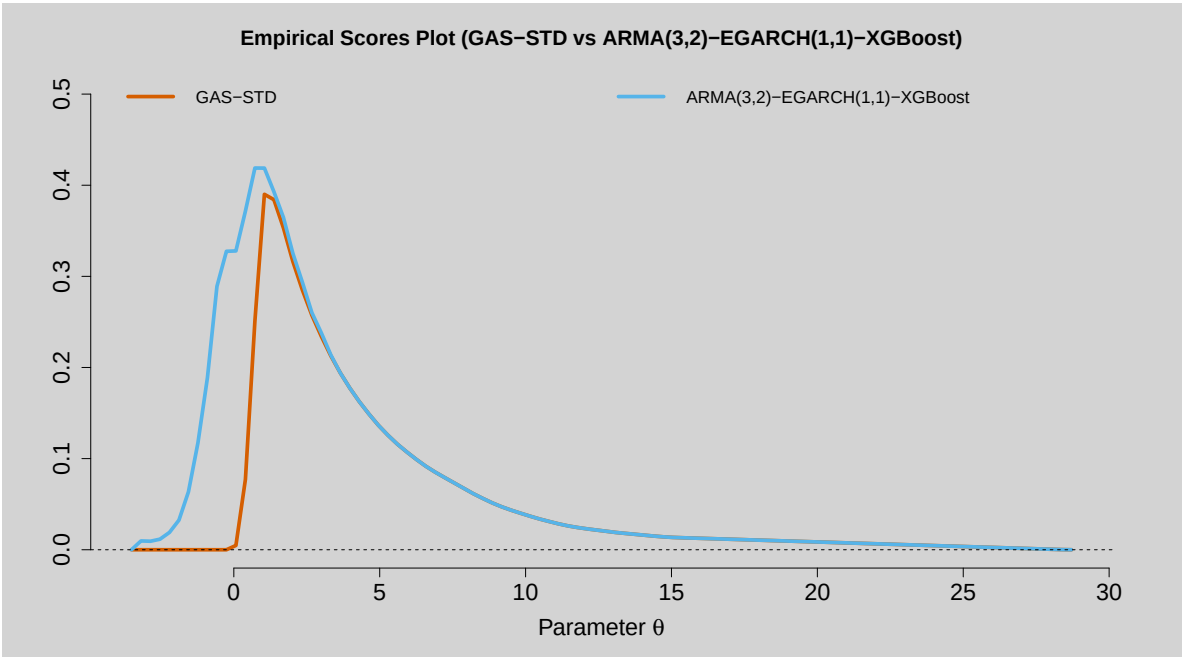


Figure 13. Empirical Scores Plot of GAS-STD versus ARMA(3,2)-EGARCH(1,1)-XGBoost.

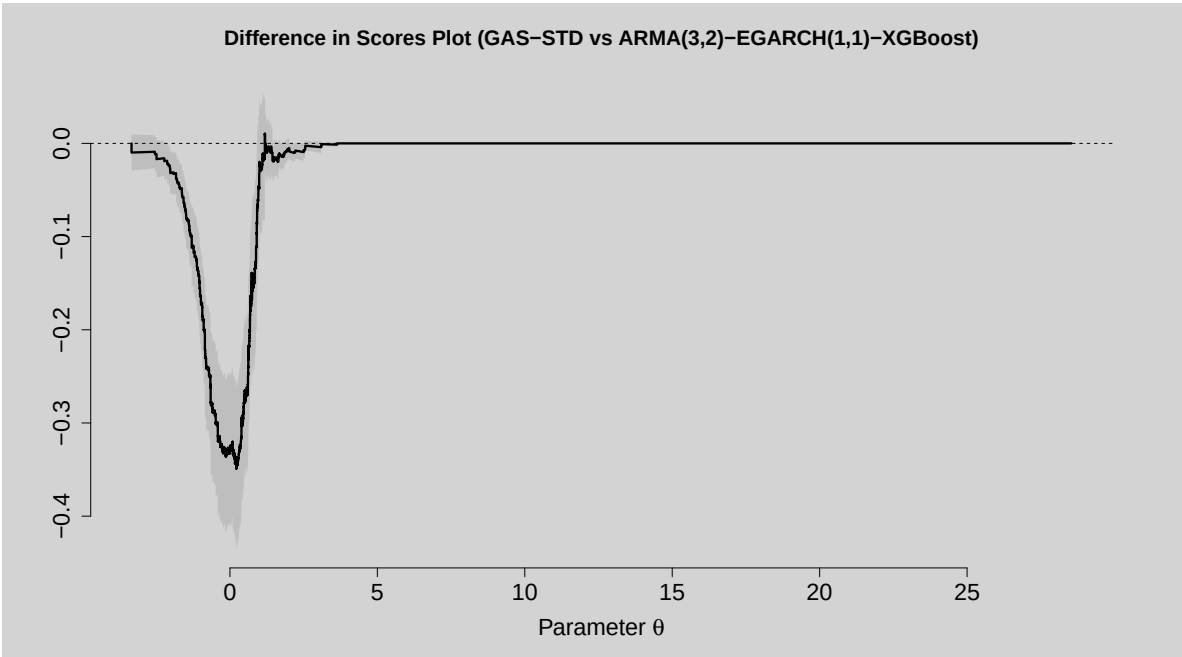


Figure 14. Difference in Scores Plot of GAS-STD versus ARMA(3,2)-EGARCH(1,1)-XGBoost.

Figure 15 presents the empirical scores plot comparing the performance of the ARMA(3,2)-EGARCH(1,1) model to that of the GAS-STD model. The plots validate that both models begin with the same trajectory, but with the ARMA(3,2)-EGARCH(1,1) having greatly reduced scores within the parameter space, reflecting better predictability, especially for large parameter values of θ . On the contrary, the GAS-STD model increases more steeply and subsequently converges rapidly towards zero, indicative of a diminished ability to forecast beyond the short-run horizon. The plot of the score difference in Figure 16 lends additional support to this conclusion. The positive differences indicated in the difference curve, particularly at the lower range of θ , attest that ARMA(3,2)-EGARCH(1,1) is superior to the GAS-STD model. The confidence band supports the strong evidence for the result, since it remains positive for the majority of the relevant domain. Beyond the small range, the differences converge and finally centre on zero, demonstrating comparable long-horizon performance by both

models. The overall results confirm that the ARMA(3,2)–EGARCH(1,1) model is more predictive in the short to medium term than the GAS-STD model, with both models demonstrating similar performance over longer horizons.

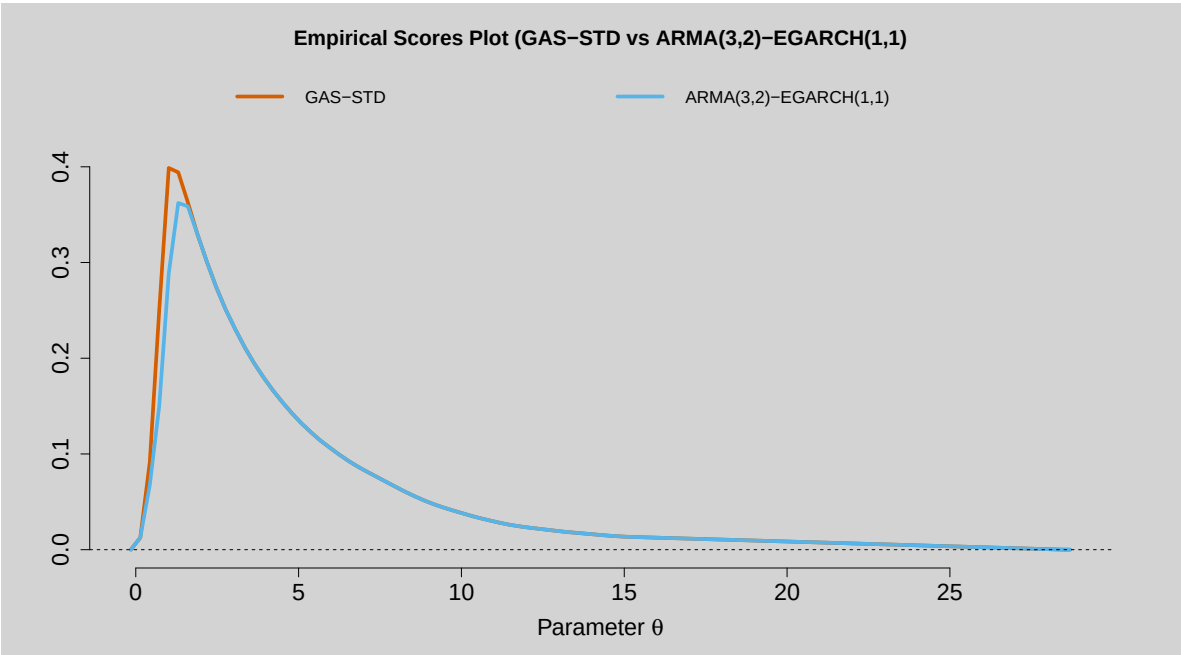


Figure 15. Empirical Scores Plot of GAS-STD versus ARMA(3,2)-EGARCH(1,1).

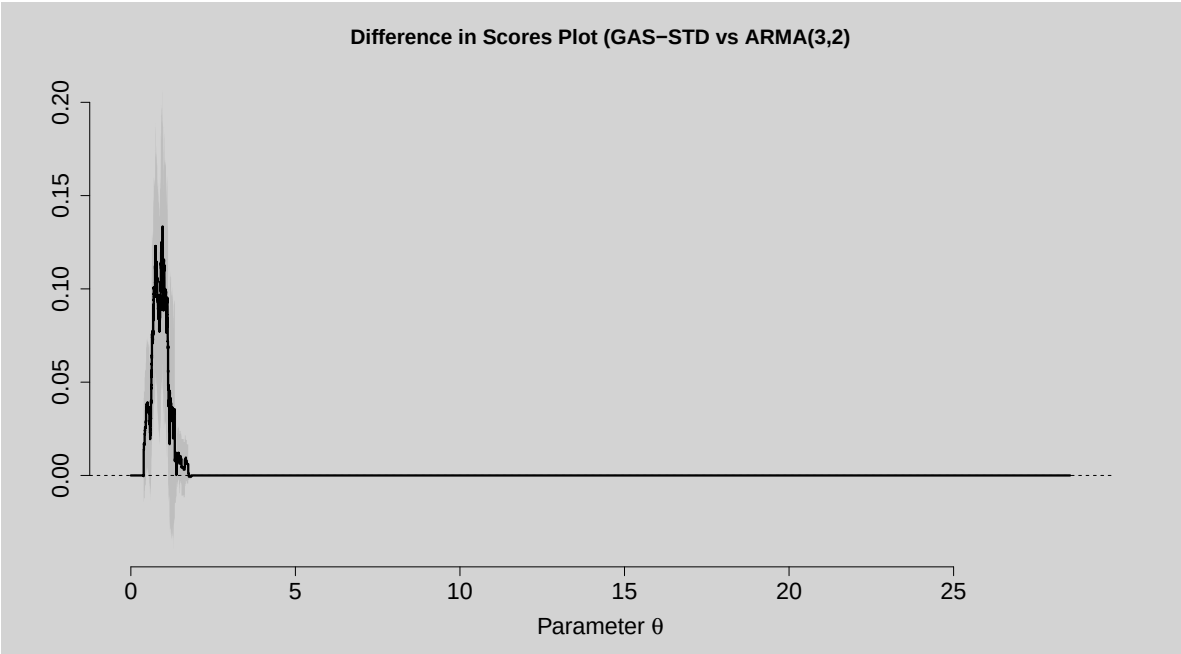


Figure 16. Difference in Scores Plot of GAS-STD versus ARMA(3,2)-EGARCH(1,1).

4. Discussion

The empirical results in Section 3 demonstrate that the Student-t distribution of the GAS model (GAS-STD) provides the best fit for the JSE Top40 Index returns among the conditional distributions considered. Table 1 comparison measures show that the GAS-STD produces the lowest AIC (10188.142) and BIC (10243.626) values, which means that it maximises goodness-of-fit and parsimony compared to the Gaussian, skew-Gaussian, or asymmetric forms. The parameter estimates also detect considerable persistence of location and scale dynamics with coefficients a_2 and b_2 statistically significant at 1%.

This confirms that past volatility has a strong effect on current volatility, a common feature of financial return series. The findings of the shape parameter are that the return distribution is time-varying in terms of tail heaviness and that there are indications of episodes of extreme kurtosis during times of market distress. A diagnostic examination of the estimated GAS-STD model also confirms that it is a density forecast. Both the PIT histogram in Figure 1 and the uniform scores in Table 3 indicate that the model produces well-calibrated predictive densities, although skewness tests do identify an unexplained residual asymmetry. Consequently, density backtests such as Normalised Log Score (1.1932) and uniform PIT score (0.4417) in Table 3 attain good forecast accuracy through balanced performance in the centre and tails of the distribution. Forecasting evaluation further indicates that GAS-STD outperforms in the capture of volatility dynamics over mean returns. Rolling forecasts show the conditional mean is generally level, such as in the martingale property of asset returns, while forecast volatility shows clear clustering that tracks very closely with realised volatility. Accuracy measures for the forecasts verify that RMSE (0.5373) for volatility forecast is substantially lower than for mean return levels (which is 0.8055). These findings underscore that the GAS model works best when the modelling objective is forecasting volatility rather than mean returns prediction. Risk management analysis incorporating Value-at-Risk (VaR) and Expected Shortfall (ES) concludes that the GAS-STD yields plausible downside risk estimates. Both 5% VaR and ES forecasts dynamically adapt to clusters of volatility, and backtests confirm good performance: the Kupiec test for unconditional coverage ($p = 0.8414$) and the Christoffersen test for conditional coverage ($p = 0.1735$) fail to reject good coverage, and the Dynamic Quantile test ($p = 0.2281$) further confirms good dynamic exceedance behaviour. Such findings demonstrate the GAS-STD to be an effective method of quantifying tail risk in equity indices. These are supported by simulation studies in demonstrating the impact of the shape parameter ν on tail behaviour. Small values of ν ($\nu = 5$) yield levels of kurtosis greater than 7 that characterise heavy tails, whereas larger ones ($\nu = 30$) converge towards the Gaussian benchmark point. This is the kind of behaviour that captures the leptokurtic nature of financial returns, an essential requirement when modelling extreme risk. Comparison with other models, however, shows that although GAS-STD is superior in density and risk forecasting, hybrid model that include machine learning surpass it in point forecast accuracy. The ARMA(3,2)–EGARCH(1,1)–XGBoost model has much lower RMSE (0.1386) than GAS-STD, and DM tests in Table 11 confirm that the hybrid forecasts are better. Murphy diagrams in Figures 13, 14, 11, 12, 15, and 16 also show that the hybrid model surpasses GAS-STD and standalone ARMA(3,2)–EGARCH(1,1) at all forecast horizons. These results highlight that model choice depends on the purpose of forecasting: GAS-STD for risk measures and density, and hybrid model for short-horizon volatility forecasts.

5. Conclusions

This study focused on volatility modelling of the JSE Top40 Index using the GAS framework as the primary method, while GARCH and the hybrid ARMA(3,2)–EGARCH(1,1)–XGBoost models were used as benchmarks for comparison. It is shown that GAS-STD provides the most suitable specification within the GAS family and excels in density calibration, tail risk forecasting, and VaR/ES backtesting. Its attractiveness lies in its capability to model volatility persistence and heavy-tailed distributions, making it a suitable match for financial risk management. However, the hybrid ARMA(3,2)–EGARCH(1,1)–XGBoost is more accurate than GAS and ARMA(3,2)–EGARCH(1,1) in terms of point forecasting accuracy, as corroborated by forecast error measures and DM tests in Table 11. The results suggest that the GAS model ought to be preferred for density-based applications such as risk estimation and stress testing, whereas the hybrid approach is more appropriate for short-term forecasting exercises. Future research should examine skewed GAS distributions, hybrid GAS models, and multivariate extensions to facilitate cross-market volatility spillovers.

Author Contributions: Conceptualization, I.M., T.R. and C.S.; methodology, I.M.; software, I.M.; validation, I.M., T.R. and C.S.; formal analysis, I.M.; investigation, I.M., T.R. and C.S.; data curation, I.M.; writing—original draft preparation, I.M.; writing—review and editing, I.M., T.R. and C.S.; visualization, I.M.; supervision, T.R. and C.S.;

project administration, T.R. and C.S.; funding acquisition, I.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2025-2026 NRF MSc Postgraduate Scholarship: REF NO: PMDS240701235994.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data were obtained from the Wall Street Journal Markets website <https://za.investing.com/indices/ftse-jse-top-40-historical-data> (accessed on 25 April 2025).

Acknowledgments: The support of the 2025–2026 NRF MSc Postgraduate Scholarship towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF. In addition, the authors thank the anonymous reviewers for their helpful comments on this paper.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the study’s design, in the collection, analyses, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
DM	Diebold-Mariano
EGARCH	Exponential Generalised Autoregressive Conditional Heteroskedasticity
ES	Expected Shortfall
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
GAS	Generalised Autoregressive Score
JSE	Johannesburg Stock Exchange
MAE	Mean Absolute Error
MASE	Mean Absolute Scaled Error
PIT	Probability Integral Transform
RMSE	Root Mean Square Error
STD	Student- <i>t</i> distribution
VaR	Value at Risk
XGBOOST	Extreme Gradient Boosting

References

Tsay, R.S. (2010). *Analysis of Financial Time Series, Third Edition*. John Wiley & Sons. Available online: [CrossRef]

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. Available online: [CrossRef]

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, 50(4), 987–1007. Available online: [CrossRef]

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59(2), 347–370. Available online: [CrossRef]

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(5), 1779–1801. Available online: [CrossRef]

Bauwens, L., Laurent, S., & Rombouts, J.V.K. (2006). Multivariate GARCH Models: A Survey. *Journal of Applied Econometrics*, 21(1), 79–109. Available online: [CrossRef]

Jefferis, K., & Smith, G. (2005). The Changing Efficiency of African Stock Markets. *South African Journal of Economics*, 73(1), 54–67. Available online: [CrossRef]

- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Available online: [\[CrossRef\]](#)
- Maingo, I., Ravele, T., & Sigauke, C. (2025). A Fusion of Statistical and Machine Learning Methods: GARCH-XGBoost for Improved Volatility Modelling of the JSE Top40 Index. *International Journal of Financial Studies*, 13(3), 155. Available online: [\[CrossRef\]](#)
- Creal, D., Koopman, S.J., & Lucas, A. (2013). Generalized Autoregressive Score Models with Applications. *Journal of Applied Econometrics*, 28(5), 777–795. Available online: [\[CrossRef\]](#)
- Harvey, A.C. (2013). *Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series*. Cambridge University Press: New York, USA. Available online: [\[CrossRef\]](#)
- Lazar, E., & Xue, X. (2020). Forecasting Risk Measures Using Intraday Data in a Generalized Autoregressive Score (GAS) Framework. *International Journal of Forecasting*, 36(3), 1057–1072. Available online: [\[CrossRef\]](#)
- Venter, P.J., & Mare, E. (2020). GARCH Option Pricing Models in a South African Equity Context. *ORiON: Journal of the Operational Research Society of Southern Africa*, 36(1), 1–17. Available online: [\[CrossRef\]](#)
- Maingo, I., Ravele, T., & Sigauke, C. (2025). Volatility Modelling of the Johannesburg Stock Exchange All Share Index Using the Family GARCH Model. *Forecasting*, 7(2), 16. Available online: [\[CrossRef\]](#)
- Blasques, F., Koopman, S.J., & Lucas, A. (2014). Maximum Likelihood Estimation for Generalized Autoregressive Score Models. *Tinbergen Institute Discussion Paper*. Available online: [\[URL\]](#)
- Koopman, S.J., Lucas, A., & Scharth, M. (2016). Predicting Time-Varying Parameters with Generalized Autoregressive Score Models. *Journal of Econometrics*, 193(1), 19–38. Available online: [\[CrossRef\]](#)
- Ardia, D., Boudt, K., & Catania, L. (2019). Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software*, 88, 1–28. Available online: [\[CrossRef\]](#)
- Babatunde, O., Folorunso, S., & Saliu, F. (2021). Comparative Forecasting Performance of GARCH and GAS Models in the Stock Price Traded on Nigerian Stock Exchange. *International Journal of Mathematical Modelling & Computations*, 11(2 SPRING). Available online: [\[URL\]](#)
- Yaya, O.S., Bada, A.S., & Atoi, V.N. (2016). Volatility in the Nigerian Stock Market: Empirical application of Beta-t-GARCH variants. *CBN Journal of Applied Statistics*, 7(2), 27–48. Available online: [\[URL\]](#)
- Junior, P.O., & Alagidede, I. (2020). Risks in emerging markets equities: Time-varying versus spatial risk analysis. *Physica A: Statistical Mechanics and its Applications*, 542, 123–474. Available online: [\[CrossRef\]](#)
- Alanya-Beltran, W. (2022). Modelling stock returns volatility with dynamic conditional score models and random shifts. *Finance Research Letters*, 45, 102–121. Available online: [\[CrossRef\]](#)
- Opschoor, A., Janus, P., Lucas, A., & Van Dijk, D. (2018). New HEAVY models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 36(4), 643–657. Available online: [\[CrossRef\]](#)
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13(2), 281–291. Available online: [\[CrossRef\]](#)
- Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3), 203–228. Available online: [\[URL\]](#)
- Acerbi, C., & Tasche, D. (2001). On the coherence of expected shortfall. *arXiv / preprint*. Published 2002 in *Journal of Banking & Finance*. Available online: [\[URL\]](#)
- McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press. Available online: [\[URL\]](#)
- Rockafellar, R. T., & Uryasev, S. (2000). Optimization of Conditional Value-at-Risk. *Journal of Risk*, 2(3), 21–41. Available online: [\[URL\]](#)
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), 73–84. Available online: [\[CrossRef\]](#)
- Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862. Available online: [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.