Abstract: Accurate estimate of tail risk remains a major difficulty in quantitative finance. Standard volatility models provide practical frameworks, but their success is dependent on both model dynamics and the distributional specification applied to the innovations. This study analyzes the statistical performance of four general GARCH-family models (sGARCH, eGARCH, GJR-GARCH, and apARCH) under three distributional assumptions (Gaussian, Student's t, and skewed Student's t). The objective is to evaluate how it combined effects of structural design and distributional choice on the reliability of Value-at-Risk (VaR) and Expected Shortfall (ES) predictions. Risk projections are calculated at 95% and 99% confidence levels using adjusted closing prices of Amazon.com, Inc. (AMZN) and are checked by Kupiec's Proportion of Failures test (POF), Christoffersen's Independence test, and the Conditional Coverage test (CC). The empirical evidence shows that heavy-tailed models provide more conservative and statistically reliable estimates of tail risk, while Expected Shortfall (ES) indicates more resilience compared to Value at Risk (VaR) under volatile conditions. The results demonstrate the relationship between volatility model design and innovation distribution, contributing to the literature on statistical risk modeling and offering methodological guidance for the application of GARCH-type processes in financial econometrics and machine learning.

Keywords: GARCH-family models, Distributional assumptions, Value at Risk (VaR), Expected Shortfall (ES), Volatility Modeling

1. Introduction

Volatility modeling is a fundamental problem in quantitative finance, statistical learning, and algorithmic trading. The Autoregressive Conditional Heteroskedasticity (ARCH) model, proposed by Engle in 1982 and generalized by Bollerslev in 1986, provided the groundwork for predicting time-varying second moments in financial returns [1, 2]. Subsequently, various GARCH-family processes have been formulated to capture nonlinearities, asymmetries, and leverage effects in volatility dynamics, such as exponential GARCH of Nelson in 1991, the GJR-GARCH of Glosten, Jagannathan, and Runkle in 1993, and the asymmetric power ARCH of Ding, Granger, and Engle in 1993 [3, 4, 5]. These models are important to financial econometrics and applications in data-driven risk assessment.

Meanwhile, regulatory highlight the significance of precise tail risk evaluation. Value-at-Risk (VaR), though its extensive application in regulatory frameworks like Basel II and Basel III, has faced criticized for its absence of subadditivity and restricted sensitivity to high losses [6]. Expected Shortfall (ES) was developed by Acerbi and Tasche as a logical alternative, offering a more thorough evaluation of downside risk [7]. Both Value at Risk (VaR) and Expected Shortfall (ES) are based on the specification of conditional volatility and the supposed distribution of innovations. Empirical research indicates that Gaussian distribution typically underestimates tail risk, but heavy-tailed or skewed models provide stronger alignment with observed extremes [8].

Despite much research on volatility forecasting and risk analysis, evaluations of the combined effects of model design and distributional assumptions on tail risk predictions are limited. Most research concentrate on either model dynamics or innovation distributions independently. The interaction between structural features of volatility processes and the statistical properties of innovations is not fully understood, particularly for high-frequency and highly volatile assets. This research gap is crucial, as it directly affects the statistical reliability of risk predictions and the flexibility of algorithmic trading systems.

The present study improves the existing literature with a comparative analysis of four widely used GARCH-family models: standard GARCH (sGARCH), exponential GARCH (eGARCH), GJR-GARCH, and asymmetric power ARCH (apARCH), applying Gaussian (Normal), Student's t, and skewed Student's t distributions. The models are evaluated based on its ability to provide statistically accurate Value at Risk (VaR) and Expected Shortfall (ES) projections at 95% and 99% confidence levels, with the high frequency adjusted returns of Amazon.com, Inc. (AMZN). Standard backtesting methodologies, including Kupiec's Proportion of Failures test (POF), Christoffersen's Independence test, and the Conditional Coverage test (CC), are employed to evaluate the consistency and dependability of these predictions.

This research presents three contributions. first it discusses the interaction between volatility model design and distributional assumptions in influencing the statistical characteristics of tail risk measurements. Secondly, it offers an assessment of Expected Shortfall (ES) compared to Value at Risk (VaR) across several specifications, emphasizing their comparative stability in high-volatility environments. Third, it situates volatility modeling in the expansive field of statistical learning and computational finance, providing methodological insights useful to quantitative trading, portfolio risk management, and the integration of econometric models with machine learning frameworks.

2. Literature Review

The literature on volatility modeling and tail risk forecasting can be described into four related fields: (i) extensions of the GARCH framework, (ii) distributional assumptions, (iii) methodological progress in VaR and ES estimation, and (iv) the integration of econometric volatility models with machine learning methodologies. Collectively, these perspectives establish a foundation for analytical decisions in this work, particularly a comparative study of four typical GARCH-family processes across different distributions and their evaluation using VaR and ES backtesting.

2.1. Extensions of the GARCH Framework

The standard GARCH model has been widely extended to incorporate asymmetry, nonlinear effects, and long-memory behavior. For instance, the eGARCH introduces log-volatility dynamics to allow for asymmetric responses to positive and negative shocks, while the GJR-GARCH explicitly models leverage effects [3, 4]. The apARCH generalizes this further by allowing flexible power transformations of the conditional variance [5]. More recent contributions explore its variants and dynamics, showing improved performance in turbulent markets [9]. These extensions collectively form a representative set of volatility dynamics against which distributional assumptions can be systematically tested.

Recent applications have demonstrated the versatility of these extensions across different asset classes. Fiszeder and Molnár apply range-based GARCH models to cryptocurrency markets and report superior VaR and ES performance under extreme volatility [9]. García-Jorcano and Novales show that the flexibility of the power parameter in APARCH improves tail risk forecasts when combined with heavy-tailed distributions [10]. Similarly, Cerqueti and Mattera and Maciel investigate skewed and regime-switching GARCH specifications for digital assets, finding that these models significantly outperform symmetric benchmarks in capturing the pronounced asymmetry and structural breaks typical of crypto-assets [11, 12]. Together, these studies highlight that modern GARCH variants are not only theoretically richer but also empirically more reliable in high-volatility environments.

2.2. Distributional Assumptions

Even with a well-specified volatility filter, the distribution of innovations critically affects tail inference. Early studies relied on the Gaussian distribution, but empirical evidence shows that it underestimates tail probabilities. Studies by Serrano Bautista and Núñez Mora and Gerlach and Wang reveal that heavy-tailed and skewed densities, particularly skewed Student's t distribution, reduce VaR exceedances compared to normal distributions [13, 14]. Alexander and Dakos highlight that even simple EWMA models can rival complex GARCH model when paired with appropriate asymmetric and heavy-tailed innovations [15]. The literature emphasizes that model performance depends not only on conditional variance dynamics but also on distributions. Hence, the evaluation requires analyzing volatility models combined with alternative distributional choices.

2.3. VaR and ES Estimation

Risk measures derived from volatility forecasts must be statistically validated. VaR provides a quantile-based threshold, while ES captures conditional tail expectations and is coherent under axiomatic definitions of risk [7]. Recent studies extend beyond classical estimation: Kaibuchi and Stupller propose hybrid GARCH–Extreme Value Theory methods for bias correction in tail quantiles [16]. Bayer and Dimitriadis introduce regression-based backtesting for ES as an elicitable functional [17]. Qiu and Nakata and Buczyński and Chlebus demonstrate that hybrid GARCH–LSTM models can enhance VaR and ES forecasts by capturing nonlinear temporal dependencies [18, 19]. These advances highlight the importance of backtesting frameworks to ensure statistical reliability of risk forecasts.

2.4. Machine Learning and Computational Methods

A growing line of research explores the integration of econometric models with machine learning. Syuhada and Hakim's study shows that hybrid approaches, such as GA-ARMA-GARCH, perform better than purely parametric methods among multiple asset classes [20]. Recurrent and deep learning models provide additional flexibility by capturing nonlinear dependencies and long-memory features in financial time series [20]. This reflects a broader shift in computational finance toward combining interpretable parametric structures with data-driven learning, particularly for tail risk prediction where purely econometric models face limitations.

2.5. Summary and Research Gaps

The literature indicates that both the specification of GARCH-family processes and the choice of distribution exert critical influence on the reliability of tail risk forecasts. In addition, advances in backtesting and computational methods highlight the need for systematic evaluation under consistent frameworks. However, most prior work examines either model extensions or distributional assumptions in isolation, with limited attention to their joint interaction. Moreover, compared to indices and crypto, empirical evidence on individual equities, particularly those characterized by persistent volatility and heavy tails, remains scarce.

This study addresses this gap by evaluating four GARCH-family processes (sGARCH, eGARCH, GJR-GARCH, and apARCH) under Gaussian, Student's t, and skewed Student's t distribution. These models are chosen because they span the main and widely used structural extensions of GARCH (symmetric, asymmetric, logarithmic, and power-based). Also, all of these models keep mathematical consistent. The basic sGARCH serves as the benchmark with variance equations in closed form. The eGARCH captures asymmetry through log-volatility but still allows likelihood-based estimation. The GJR-GARCH incorporates threshold effects without losing closed-form variance dynamics. The apARCH adds a power parameter for greater flexibility, yet its likelihood function remains computationally feasible. The three distributions represent the standard progression from light-tailed to heavy-tailed to skewed innovations. By analyzing their joint impact on VaR and ES forecasts, the study contributes to statistical learning and risk modeling.

3. Methodology

The empirical behavior of financial return series often exhibits volatility clustering. Classical linear time series models with constant variance fail to capture, motivating the use of conditional heteroskedasticity models. This study tries to fit the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) and its extensions, including Exponential GARCH (EGARCH), Glosten–Jagannathan–Runkle GARCH (GJR-GARCH), and Asymmetric Power ARCH (APARCH), to model the volatility dynamics. This study builds model specification and distributional choice together. The goal is to evaluate behaviors of different model × distribution combinations in quantifying tail risk.

Let r_t denote the logarithmic return at time t, modeled as

$$r_t = \mu + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \sigma_t z_t \tag{2}$$

where μ is the conditional mean, σ_t the conditional standard deviation, and z_t an i.i.d. standardized innovation with zero mean and unit variance.

3.1. Conditional Variance

The standard GARCH (1,1) model captures volatility persistence by:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3}$$

where the parameters α and β measure short term shock impact and long-term persistence, respectively. While effective for many financial datasets, GARCH assumes symmetric responses of volatility to positive and negative shocks. Empirical evidence, however, shows that volatility often reacts more strongly to negative returns, a phenomenon known as the leverage effect.

To account for such asymmetry, we employ the EGARCH model, which models the log of the conditional variance as:

$$\log(\sigma_t^2) = \omega + \beta\log(\sigma_{t-1}^2) + \alpha\left(\frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \mathrm{E}|z_t|\right) + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \tag{4}$$

where γ measures the additional volatility response to negative shocks.

Similarly, the GJR-GARCH model introduces an indicator function $\mathrm{I}_{\{\varepsilon_{t-1}<0\}}$ to separately weight negative returns ensuring non-negativity and leverage effects:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \gamma\mathrm{I}_{\{\varepsilon_{t-1}<0\}}\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{5}$$

Extending further, the APARCH model introduces a power term δ :

$$\sigma_t^\delta = \omega + \alpha(|\varepsilon_{t-1}| - \gamma\varepsilon_{t-1})^\delta + \beta\sigma_{t-1}^\delta \tag{6}$$

allowing the model to capture both asymmetric effects and nonlinear transformations of volatility.

3.2. Distributional assumptions

The predictive accuracy of those models above depends critically on the assumed distribution of z_t. This study considers three distribution cases.

A Gaussian assumption, also known as the normal distribution, offers analytical convenience:

$$z_t \sim N(0,1) \tag{7}$$

but typically underestimates the probability of extreme events, leading to biased tail risk estimates. To eliminate this limitation, this study also supplements the normal distribution with the Student's t distribution and the skewed Student's t distribution.

The students' distribution introduces a degrees-of-freedom parameter v :

$$z_t \sim t_v \tag{8}$$

controlling tail thickness, improving the model's ability to replicate the leptokurtic nature of financial returns.

The skewed t distribution adds an asymmetry parameter, enhancing accuracy of prediction when return distributions are not symmetric.

3.3. Risk Quantifications

This study adopts Value-at-Risk (VaR) and Expected Shortfall (ES) as the primary measures.

Given a confidence level α , VaR is defined as:

$$VaR_{t+1}(\alpha) = \mu_{t+1} + \sigma_{t+1} q_\alpha(F_\theta) \tag{9}$$

where $q_\alpha(F_\theta)$ is the α -quantile of the assumed distribution F_θ. VaR represents the maximum expected loss over a given horizon at the specified confidence level. However, VaR provides no information about losses beyond.

To capture the expected magnitude of losses in the tail, this study also computes ES:

$$ES_{t+1}(\alpha) = \mathrm{E}[r_{t+1} | \, r_{t+1} \leq VaR_{t+1}(\alpha)] \tag{10}$$

3.4. Evaluation and Backtesting

To mimic the operational setting of risk management, this study tests rolling estimation scheme. A fixed-length estimation window of size n is moved forward by one observation at each step. Technically, the model is re-estimated using the most recent n observations, and the one-step-ahead VaR and ES are computed. This design emulates the adaptive nature of financial decision-making under evolving market conditions.

Model outputs are evaluated via statistical backtesting procedures. Firstly, the Kupiec Proportion of Failures (POF) test assesses whether empirical violation rates match the nominal level. Secondly, the Christoffersen independence test examines whether violations are serially independent. Thirdly, the Conditional Coverage (CC) test jointly evaluates both coverage and independence. For ES, we employ exceedance residual analysis to assess the accuracy of extreme loss predictions.

4. Empirical Analysis

4.1. Model Estimation Results (AMZN)

This study initiates with daily log returns derived from the adjusted close prices of a representative large-cap technology stock, here using AMZN as a case study, with data sourced from Yahoo Finance. The sample period spans from January 1, 2019, to January 1, 2025, with daily frequency. The final dataset includes 1509 observations after calculating continuously compounded returns and removing missing data.

Table 1: Descriptive statistics for daily log returns.

Min	1^{st} Qu.	Median	Mean	3^{rd} Qu.	Max
-0.1514	-0.0105	0.0009	0.0007	0.0120	0.1269

According to table 1, its empirical summary shows a small positive mean ($\approx 6.9 \times 10^{-4}$per day), a median close to zero, and wide tails (min -0.151, max 0.127), consistent with equity markets where extreme moves occur more frequently than under a Gaussian benchmark.

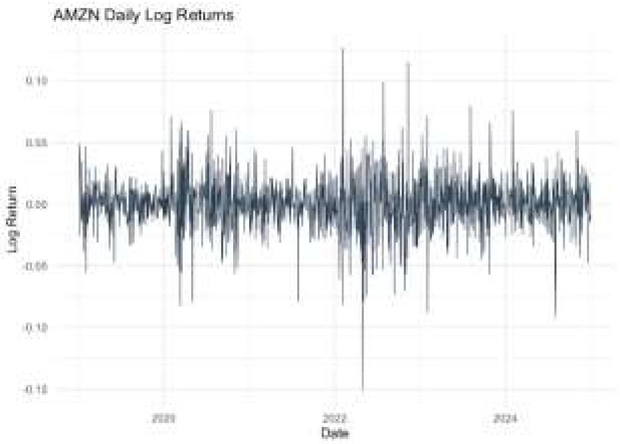

Figure 1: Daily log returns of Amazon (AMZN).

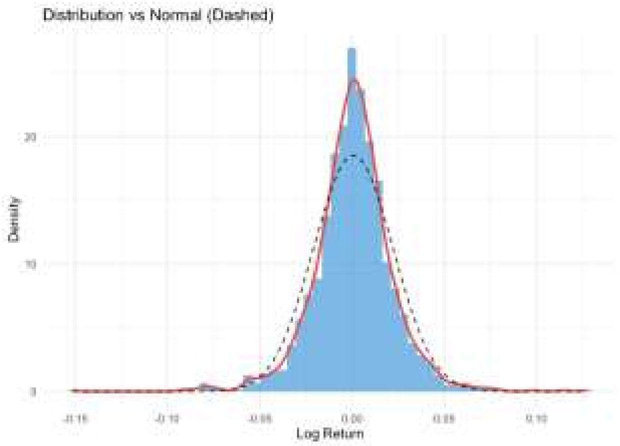

Figure 2: Empirical distribution of AMZN daily returns.

Time-series and distributional diagnostics confirm these features visually. Figure 1 shows that daily log returns fluctuate around zero, while volatility appears in clusters, suggesting the presence of time-varying conditional variance. Figure 2 further illustrates the empirical distribution of returns, which is sharply peaked and displays heavy tails.

Table 2: Serial-dependence and ARCH effect tests.

Diagnostic	Statistic	df	p-value
Ljung–Box on returns	23.216	20	0.2783
Ljung–Box on squared returns	156.760	20	<2.2e-16
ARCH-LM on returns	86.301	20	3.262e-10

Formal statistical tests reported in Table 2 support this evidence. A Ljung–Box test on returns with 20 lags yields p=0.2783, failing to reject the null hypothesis of no linear autocorrelation in r_t. When applied to squared returns, the same test strongly rejects the null (X-squared ≈ 156.8, $p \leq 2.2 \times 10^{-16}$), indicating significant dependence in the conditional variance. An ARCH LM test with 20 lags rejects the null hypothesis (Chi-squared≈86.3, $p = 3.26 \times 10^{-10}$), providing direct statistical evidence of ARCH effects in the return series.

Therefore, the evidence supports GARCH-family model as suitable models for conditional variance and emphasizes the role of distributional assumptions in measuring tail risk.

4.2. Comparative Performance

The AMZN daily log return series was fitted using four representative GARCH-family, including sGARCH, eGARCH, GJR-GARCH, and apARCH, under three distributions: Gaussian ("norm"), Student's t ("std"), and skewed Student's t ("sstd"). All models were estimated with $(p, q) = (1,1)$ variance dynamics and a constant mean. Table 3 below shows the necessary result:

Table 3: Summary of GARCH-family model estimates for AMZN returns.

Model	Dist	Conv	Pers	Skew	Shape	AIC	BIC
apARCH	norm	0	1.1696	—	—	-5.0122	-4.9910
apARCH	sstd	0	1.2297	0.9655	5.7350	-5.0723	-5.0441
apARCH	std	0	1.2222	—	5.7662	-5.0730	-5.0483
eGARCH	norm	0	0.9500	—	—	-5.0105	-4.9929
eGARCH	sstd	0	0.9829	—	5.6805	-5.0737	-5.0525
eGARCH	std	0	0.9831	0.9650	5.6547	-5.0730	-5.0483
gjrGARCH	norm	0	0.9527	—	—	-4.9960	-4.9783
gjrGARCH	sstd	0	0.9912	0.9660	5.3248	-5.0614	-5.0367
gjrGARCH	std	0	0.9904	—	5.3352	-5.0621	-5.0410
sGARCH	norm	0	0.9485	—	—	-4.9919	-4.9778
sGARCH	sstd	0	0.9943	0.9706	5.1591	-5.0600	-5.0388
sGARCH	std	0	0.9940	—	5.1637	-5.0609	-5.0432

Across all twelve model × distribution combinations, optimization converged successfully (coverage=0).

For symmetric specifications (sGARCH, eGARCH), volatility persistence was computed as $\alpha_1 + \beta_1$; for asymmetric variants (GJR-GARCH, apARCH), persistence incorporated half the leverage term. Most persistence estimates exceeded 0.93, confirming strong volatility clustering in AMZN returns. In particular, the highest persistence was observed in sGARCH–std (0.9940), suggesting near-integrated volatility under heavy-tailed assumptions. Such high persistence is common in large-cap equities and implies that shocks to volatility decay very slowly over time.

Notably, apARCH–sstd, apARCH–std, and apARCH-norm yield persistence estimates slightly exceeding unity. This phenomenon often interpreted as indicative of near-integrated or explosive volatility dynamics. While in strict theoretical terms such values would violate the covariance-stationarity condition of GARCH processes, in empirical settings this frequently reflects finite-sample effects, heavy-tailed innovations, and prolonged volatility clustering. From a risk management perspective, this implies that volatility shocks are exceptionally long-lived, enhancing long-horizon risk forecasts.

The analysis employed both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Each combination exhibits relatively minor differences, which is a typical result when modeling large sample daily return data. This indicates that all candidate models effectively capture the basic volatility dynamics of AMZN returns. The eGARCH-sstd model demonstrated the lowest AIC (-5.0730) and BIC (-5.0525), suggesting optimal fit for both heavy and light-tailed alternatives. The eGARCH-std model yielded comparable information criteria (AIC = -5.0730 and BIC = -5.0483), indicating that the inclusion of the skewness parameter offers minimal enhancement in this context. Gaussian-based specifications produced elevated AIC and BIC values, indicating the insufficiency of the normality assumption in accurately representing the heavy-tailed characteristics of AMZN returns.

Estimated degrees-of-freedom parameters (v) for std and sstd distributions fell in the range of 5.16 to 5.68, confirming the presence of heavy tails in the return distribution. In the skewed cases, skewness parameters below unity indicate a moderately heavier left tail. This is a credible feature for equity returns, where downside risk is typically greater than upside extremes.

In summary, eGARCH-sstd captures both the leverage effects inherent in equity returns and the heavier-than-normal tails observed in the empirical distribution. The skewed Student’s t remains a good secondary choice, particularly if capturing asymmetric tail risk is of interest. These results form the basis for later value-at-risk and expected shortfall analysis, where model performance will be validated through backtesting procedures.

4.3. VaR and ES Estimates

Table 4: Coverage Comparison Between VaR and ES

CI	Model	Dist	V Rate	ES Rate	Avg(ES–VaR)
95%	apARCH	norm	3.93%	1.96%	-0.0083
95%	apARCH	sstd	4.72%	1.57%	-0.0128
95%	apARCH	std	4.91%	1.77%	-0.0124
95%	eGARCH	norm	3.93%	1.77%	-0.0084
95%	eGARCH	sstd	4.91%	1.57%	-0.0130
95%	eGARCH	std	5.11%	1.57%	-0.0126
95%	gjrGARCH	norm	3.54%	1.77%	-0.0085
95%	gjrGARCH	sstd	4.91%	1.57%	-0.0134
95%	gjrGARCH	std	4.91%	1.57%	-0.0130

95%	sGARCH	norm	3.73%	1.77%	-0.0085
95%	sGARCH	sstd	4.52%	1.18%	-0.0138
95%	sGARCH	std	4.72%	1.18%	-0.0135
99%	apARCH	norm	1.38%	0.79%	-0.0067
99%	apARCH	sstd	0.98%	0.39%	-0.0152
99%	apARCH	std	0.98%	0.39%	-0.0147
99%	eGARCH	norm	1.38%	0.79%	-0.0068
99%	eGARCH	sstd	0.79%	0.39%	-0.0155
99%	eGARCH	std	0.98%	0.39%	-0.0150
99%	gjrGARCH	norm	1.18%	0.79%	-0.0069
99%	gjrGARCH	sstd	0.79%	0.39%	-0.0164
99%	gjrGARCH	std	0.79%	0.39%	-0.0158
99%	sGARCH	norm	1.18%	0.59%	-0.0069
99%	sGARCH	sstd	0.59%	0.20%	-0.0170
99%	sGARCH	std	0.59%	0.20%	-0.0166

A rolling-window evaluation of Value-at-Risk (VaR) and Expected Shortfall (ES) demonstrates clear differences across combinations. At the 95% confidence level, specifications with fat-tailed distributions (std, sstd) produce coverage rates close to the nominal 5% tail probability, while Gaussian-based models consistently underestimate risk (see Table 4). At the 99% level, eGARCH–std delivers the most accurate forecasts (0.98%), while normal-based apARCH proves overly aggressive (1.38%) and sstd-based models excessively conservative (<0.8%).

Complementary ES analysis shows consistently lower violation rates than VaR, confirming its greater conservatism and robustness. For instance, the sGARCH–std model reduced violations from 4.72% (VaR) to 1.18% (ES) at 95% level (Table 4). Overall, eGARCH with heavy-tailed distributions yields the most reliable VaR forecasts, while ES systematically provides stronger coverage of extreme losses.

4.4. Visualization

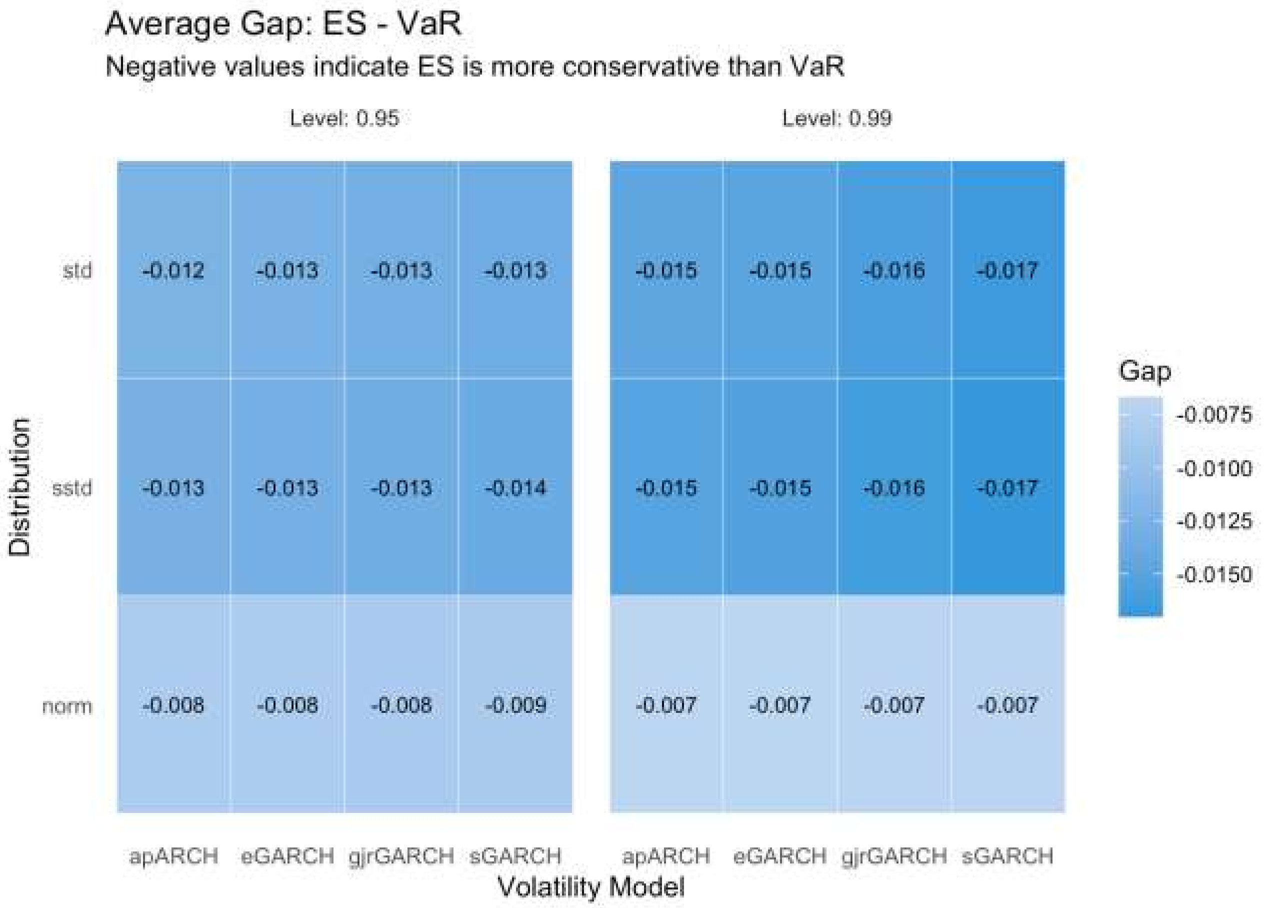

Figure 3: Average gap between ES and VaR

The heatmap displays the average value of ES − VaR (in return units) for each GARCH-family model and distributional assumption. Darker blue shading indicates that ES lies further below VaR, reflecting stronger tail conservatism.

Figure 3 reports the average difference between Expected Shortfall (ES) and Value-at-Risk (VaR) across model with distribution pairs. The heatmap illustrates that ES consistently lies below VaR, with larger gaps observed under heavy-tailed distributions. This result indicates that VaR provides a reasonably accurate point estimate of tail quantiles when models are well calibrated. ES, however, is more robust in capturing extreme losses. The findings support the use of ES as a risk metric to VaR in portfolio risk management, particularly in market conditions marked by excess kurtosis and volatility clustering.

4.5. Backtesting Outcomes

This study performs a backtesting analysis to assess the adequacy of forecasts across various configurations and confidence levels. Three standard tests are employed: the Kupiec Proportion of Failures (POF) test, which assesses the statistical consistency of the observed violation rate with the nominal coverage; the Christoffersen independence test, which analyses the serial independence of VaR violations; and the Conditional Coverage (CC) test, which integrates the two to evaluate overall adequacy.

Tests are conducted at the 95% and 99% confidence levels, which correspond to the 5% and 1% lower tails, respectively. Inference relies on the null hypothesis, which claims that the model maintains accurate coverage and p-values exceeding 0.05 are regarded as a sign of statistical acceptance.

Table 5: Result under 95% CI

Model	Dist	EmpRate	p(POF)	p(Ind)	p(CC)
apARCH	norm	0.0393	0.2503	0.2004	0.2274
apARCH	sstd	0.0472	0.7660	0.8931	0.9481
apARCH	std	0.0491	0.9269	0.8219	0.9709
eGARCH	norm	0.0393	0.2503	0.2004	0.2274
eGARCH	sstd	0.0491	0.9269	0.8219	0.9709
eGARCH	std	0.0511	0.9112	0.7530	0.9458
gjrGARCH	norm	0.0354	0.1106	0.2501	0.1445
gjrGARCH	sstd	0.0491	0.9269	0.8219	0.9709
gjrGARCH	std	0.0491	0.9269	0.8219	0.9709
sGARCH	norm	0.0373	0.1704	0.2243	0.1867
sGARCH	sstd	0.0452	0.6127	0.9659	0.8790
sGARCH	std	0.0472	0.7660	0.8931	0.9481

Table 5 demonstrates that, at the 95% confidence level, heavy-tailed distributions consistently exhibit better statistical performance compared to the normal distribution. Models including apARCH–std, eGARCH-sstd, gjrGARCH–std, and gjrGARCH–sstd demonstrate conditional coverage with p-values nearing 0.97, indicating precise violation rates and independence of exceedances. The normal distribution consistently yields conservative Value at Risk (VaR) estimates, with empirical violation rates are 4.91%. Although these deviations are statistically acceptable ($p \approx 0.82$–0.93), they suggest a tendency to overestimate risk. The apARCH and eGARCH models, when combined with heavy-tailed innovations, demonstrate strong performance, as indicated by CC p-values greater than 0.94, which implies a significant reliability in addressing the moderate tail risk environment.

Table 6: Result under 99% CI

Model	Dist	EmpRate	p(POF)	p(Ind)	p(CC)
apARCH	norm	0.0138	0.4208	0.6583	0.6558
apARCH	sstd	0.0098	0.9679	0.7525	0.9507
apARCH	std	0.0098	0.9679	0.7525	0.9507
eGARCH	norm	0.0138	0.4208	0.6583	0.6558

eGARCH	sstd	0.0079	0.6139	0.8011	0.8530
eGARCH	std	0.0098	0.9679	0.7525	0.9507
gjrGARCH	norm	0.0118	0.6934	0.7049	0.8612
gjrGARCH	sstd	0.0079	0.6139	0.8011	0.8530
gjrGARCH	std	0.0079	0.6139	0.8011	0.8530
sGARCH	norm	0.0118	0.6934	0.7049	0.8612
sGARCH	sstd	0.0059	0.3133	0.8503	0.5909
sGARCH	std	0.0059	0.3133	0.8503	0.5909

Table 6 indicates that at the 99% confidence level, eGARCH–std, apARCH–sstd, and apARCH–std closely align with the theoretical 1% target, resulting in notably high POF p-values (~0.968) and CC p-values (~0.951). This accuracy illustrates their capacity to maintain a balance between coverage and the statistical independence of violations. In comparison, normal-distribution variants exhibit a slightly aggressive tendency, with violation rates approximately at 1.38%. sGARCH–std and sGARCH–sstd exhibit conservative behavior, demonstrating only three violations, which accounts for approximately 0.59%. Despite these deviations passing statistical backtests ($p > 0.05$), they indicate that heavy-tailed assumptions yield more reliable predictions for extreme risk events.

5. Conclusion

This work focused on VaR and ES estimates evaluating four GARCH-family models (sGARCH, eGARCH, GJR-GARCH, and apARCH) under normal, Student's t, and skewed Student's t distributions.

The findings indicate that tail risk projections are significantly influenced by distributional choice, with heavy-tailed innovations generally providing better coverage than the normal distribution. When return distributions were asymmetric, skew Student's t yielded incremental benefits. While symmetric models only functioned well under heavy-tailed assumptions, asymmetric specifications (eGARCH, apARCH) better captured leverage effects and increased the accuracy of risk forecasts.

Backtesting confirmed that unconditional coverage alone can be misleading, as some models with correct violation rates failed independence or conditional coverage tests. At higher confidence levels, ES estimations were better at capturing severe losses because they were more cautious in heavy-tailed circumstances.

Overall, the best reliable risk forecasting results were obtained by combining heavy-tailed innovations with asymmetric volatility models. These results emphasize the necessity of optimizing distributional assumptions and model dynamics simultaneously in financial risk management.

Reference

[1] *Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica, 50(4), 987–1007.*

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

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

[4] *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. The Journal of Finance, 48(5), 1779–1801.*

[5] *Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). A Long Memory Property of Stock Market Returns and a New Model. Journal of Empirical Finance, 1(1), 83–106.*

[6] *Jorion, P. (2007). Value at Risk: The New Benchmark for Managing Financial Risk. McGraw-Hill.*

[7] *Acerbi, C., & Tasche, D. (2002). Expected Shortfall: A Natural Coherent Alternative to Value at Risk. Economic Notes, 31(2), 379–388.*

[8] *McNeil, A. J., Frey, R., & Embrechts, P. (2015). Quantitative Risk Management: Concepts, Techniques, and Tools (Revised Edition). Princeton University Press.*

[9] *Fiszeder, P., Małecka, M., & Molnár, P. (2024). Robust estimation of the range-based GARCH model: Forecasting volatility, value at risk and expected shortfall of cryptocurrencies. Economic Modelling, 141, 106887.*

[10] *García-Jorcano, L., & Novales, A. (2021). Volatility specifications versus probability distributions in VaR forecasting. Journal of Forecasting, 40(2), 189–212.*

[11] *Cerqueti, R., Giacalone, M., & Mattera, R. (2020). Skewed non-Gaussian GARCH models for cryptocurrencies volatility modelling. Information Sciences, 527, 1–26.*

[12] *Maciel, L. dos S. (2021). Cryptocurrencies value-at-risk and expected shortfall: Do regime-switching volatility models improve forecasting? International Journal of Finance & Economics, 26(3), 4840–4855.*

[13] *Serrano Bautista, R., & Núñez Mora, J. A. (2021). Value-at-risk predictive performance: a comparison between the CaViaR and GARCH models for the MILA and ASEAN-5 stock markets. Journal of Economics, Finance and Administrative Science, 26(52), 197–221.*

[14] *Gerlach, R., & Wang, C. (2020). Semi-parametric dynamic asymmetric Laplace models for tail risk forecasting, incorporating realized measures. International Journal of Forecasting, 36(2), 489–506.*

[15] *Alexander, C., & Dakos, M. (2023). Assessing the accuracy of exponentially weighted moving average models for Value-at-Risk and Expected Shortfall of crypto portfolios. Quantitative Finance, 23(3), 393–427.*

[16] *Kaibuchi, H., Kawasaki, Y., & Stupfler, G. (2022). GARCH-UGH: a bias-reduced approach for dynamic extreme Value-at-Risk estimation in financial time series. Quantitative Finance, 22(7), 1277–1294.*

[17] *Bayer, S., & Dimitriadis, T. (2022). Regression-based Expected Shortfall backtesting. Journal of Financial Econometrics, 20(3), 437–471.*

[18] *Qiu, Z., Lazar, E., & Nakata, K. (2024). VaR and ES forecasting via recurrent neural network-based stateful models. International Review of Financial Analysis, 92, 103102.*

[19] *Buczyński, M., & Chlebus, M. (2023). GARCHNet: Value-at-Risk forecasting with GARCH models based on neural networks. Computational Economics, 1–31.*

[20] *Syuhada, K. I. A., Tjahjono, V., & Hakim, A. R. (2023). Improving Value-at-Risk forecast using GA-ARMA-GARCH and AI-KDE models. Applied Soft Computing, 148, 110885.*