Beyond GARCH: Incorporating Tail Dependence into Volatility Forecasting with Copulas

Group 25X

Noam Diop Xander Pauwels Gabriela Rueda Xuanbo Zhao

May 2025

Abstract

This paper examines whether modeling tail dependence via copulas in a Realized GARCH framework improves S&P 500 volatility forecasts. We estimate a suite of GARCH and Realized GARCH models—including copula-based symmetric and asymmetric specifications, GARCH-X augmented by VIX and Economic Policy Uncertainty—benchmarked against HAR-RV and VIX. Using daily data (2000–2024), we forecast 1-, 5-, and 21-day volatility horizons. Out-of-sample results show the asymmetric Realized GARCH model delivers the most accurate point forecasts, outperforming standard GARCH and GARCH-X. While Student-t copulas capture strong in-sample tail dependence, they offer no predictive gain. We interpret these findings in light of leverage effects, high-frequency information flow, volatility premia, and policy uncertainty, and discuss implications for risk management. Future research could explore multivariate extensions, time-varying copulas, and machine-learning integration.

1 Introduction

Recent crises—most notably the COVID-19 shock—have revealed the shortcomings of traditional risk models: in March 2020 the S&P 500's volatility surged to decade-high levels, and classic GARCH forecasts failed to keep pace. Accurate volatility predictions are vital for portfolio hedging and for policymakers safeguarding financial stability.

The canonical GARCH(1,1) model captures volatility clustering but assumes symmetric shocks and Gaussian errors, often underestimating risk in downturns. Asymmetric variants (e.g. GJR-GARCH) address the leverage effect—negative returns boosting future volatility more than positive ones—and improve in-sample fit. Realized GARCH models enhance forecasts by incorporating *realized variance* (RV) from intraday data, effectively using a "volatility thermometer" to update estimates more responsively. We test both symmetric (SRGARCH) and asymmetric (ARGARCH) versions. Benchmarks include the HAR-RV model, which captures long-memory via daily, weekly, and monthly RV lags, and the VIX index, whose forward-looking premium offers a market-based forecast. We also explore GARCH-X extensions that add VIX or the Economic Policy Uncertainty (EPU) index as exogenous regressors.

Despite these advances, most models impose simplistic (often Gaussian) dependence structures between return shocks and volatility innovations. In extreme events—large negative returns coupled with volatility—like in 2020— spikes—this assumption fails. We propose that explicitly modeling *tail dependence* using copulas can better capture such joint extremes.

We embed copulas (Gaussian, Student-t, Clayton, Gumbel) into the Realized GARCH framework via a two-step IFM approach, allowing flexible, nonlinear dependence between return and volatility shocks. We then conduct an extensive out-of-sample comparison—across 1, 5, and 21-day horizons—against traditional GARCH, realized-variance models, HAR-RV, VIX, and GARCH-X. Forecast accuracy is assessed using loss functions such as Mean Squared Error (MSE) and QLIKE, for which we use the formulas of Patton (2011). Our central question is: *Does accounting for tail dependence improve S&P 500 volatility forecasts, and what are the implications for risk management and policy?*

This study builds on a growing body of literature. For example, Aslam et al. (2020) reviews volatility models, emphasizing symmetric and asymmetric GARCH variants. Huber et al. (2023) and Koop et al. (2024) evaluate volatility forecasts under different assumptions, while Ramachandran and Wang (2021) and El Brini et al. (2024) apply hybrid models to regional markets. These diverse approaches highlight the complexity of the volatility forecasting problem.

The paper is structured as follows: Section 2 describes the data, Section 3 outlines the methodology, Section 4 presents the results, and Section 5 concludes with directions for future research.

2 Data

The data employed in this study consists of six daily time series for the S&P 500 index, provided by the Erasmus School of Economics. It comprises the opening and closing prices, from which we compute close-to-close log returns to measure the logarithmic percentage change between consecutive closing values. We also include the closing level of the VIX (30-day implied volatility index), which reflects market expectations of future volatility—higher VIX values indicate greater anticipated turbulence. Additionally, the dataset contains the realized variance, calculated from high-frequency intraday returns to capture detailed daily volatility patterns. Finally, we incorporate the Economic Policy Uncertainty (EPU) index of Baker et al. (2016), which quantifies policy-related uncertainty based on the frequency of relevant terms in major newspaper articles.

3 Methodology

We consider T trading days indexed by $t = 1, \dots, T$. The observed return on day t is r_t, with unconditional mean $\mu \in \mathbb{R}$ and innovation

$$\epsilon_t = r_t - \mu.$$

The latent conditional variance is σ_t^2, and its long-run level is denoted by σ^2. We also use high-frequency data to compute realized variance x_t.

Model parameters are:

$$\omega > 0, \quad \alpha, \alpha_1, \alpha_2, \beta, \gamma \geq 0,$$

which govern variance dynamics, and in measurement equations

$$\phi,\ \tau_1,\ \tau_2 \in \mathbb{R}, \quad u_t \sim \mathcal{N}(0, s_u^2),\ s_u^2 > 0.$$

3.1 Model Specifications

We estimate four variants:

- **Symmetric GARCH(1,1) (SGARCH)**:

$$\sigma_t^2 = \sigma^2 + \alpha\,\sigma_{t-1}^2\big(\epsilon_{t-1}^2 - 1\big) + (\alpha+\beta)\big(\sigma_{t-1}^2 - \sigma^2\big).$$

- **Asymmetric GARCH(1,1) (AGARCH)**:

$$\sigma_t^2 = \sigma^2 + \alpha_1\,\sigma_{t-1}^2\Big[\epsilon_{t-1}^2\mathbf{1}_{\{\epsilon_{t-1}<0\}} - \tfrac{1}{2}\Big] + \alpha_2\,\sigma_{t-1}^2\Big[\epsilon_{t-1}^2\mathbf{1}_{\{\epsilon_{t-1}\geq 0\}} - \tfrac{1}{2}\Big] + (\alpha+\beta)\big(\sigma_{t-1}^2 - \sigma^2\big).$$

- **Symmetric Realized GARCH (SRGARCH)**:

$$\sigma_t^2 = \omega + \alpha\,(r_{t-1}-\mu)^2 + \beta\,\sigma_{t-1}^2 + \gamma\,x_{t-1}.$$

- **Asymmetric Realized GARCH (ARGARCH)**:

$$\begin{aligned}\sigma_t^2 = \omega &+ \alpha_1\,\sigma_{t-1}^2\Big[\tfrac{\epsilon_{t-1}^2\mathbf{1}_{\{\epsilon_{t-1}<0\}}}{\sigma_{t-1}^2} - 1\Big] + \alpha_2\,\sigma_{t-1}^2\Big[\tfrac{\epsilon_{t-1}^2\mathbf{1}_{\{\epsilon_{t-1}\geq 0\}}}{\sigma_{t-1}^2} - 1\Big] \\ &+ (\alpha+\beta)\big(\sigma_{t-1}^2 - \sigma^2\big) + \gamma\,x_{t-1}.\end{aligned}$$

In both realized-GARCH versions, the link between x_t and σ_t^2 is

$$x_t = \phi\,\sigma_t^2 + \tau_1\,\epsilon_t + \tau_2\,(\epsilon_t^2 - \sigma_t^2) + u_t.$$

3.2 Estimation

All models are estimated by maximum likelihood.

For the GARCH and GARCH-X variants, we assume

$$r_t \mid \mathcal{F}_{t-1} \sim t_\nu\big(\mu, \sigma_t^2\big),$$

a Student-t distribution with ν degrees of freedom (estimated jointly with the other parameters) to accommodate fat tails in returns.

For the Realized GARCH models, we maximize the joint log-likelihood of $\{\epsilon_t, x_t\}_{t=1}^T$:

$$\ell(\theta) = \sum_{t=1}^{T} \log f\big(\epsilon_t, x_t \mid \mathcal{F}_{t-1};\, \theta\big),$$

under the assumption

$$\begin{pmatrix}\epsilon_t \\ u_t\end{pmatrix} \Bigg|\, \mathcal{F}_{t-1} \sim \mathcal{N}\Big(\mathbf{0}, \begin{pmatrix}\sigma_t^2 & 0 \\ 0 & s_u^2\end{pmatrix}\Big),$$

so that r_t and the measurement error u_t are conditionally independent Gaussians.

In all cases, the BFGS algorithm is used to obtain the MLEs. For out-of-sample forecasting, parameters are re-estimated on a rolling training window (see below).

3.3 Incorporating Tail Dependence via Copulas

The novel aspect of our methodology is a copula-based extension to the Realized GARCH framework. Standard Realized GARCH assumes joint normality (or a single parametric form) for the return innovation ε_t and the measurement error u_t, which may be too restrictive in the tails. We begin by relaxing the independence assumption to allow simple linear correlation, then introduce copulas to capture richer, tail-specific dependence.

Baseline Bivariate Normal Extension. As a first step, we let

$$(\varepsilon_t, u_t) \mid \mathcal{F}_{t-1} \sim \mathcal{N}\big(\mathbf{0}, \Sigma\big), \qquad \Sigma = \begin{pmatrix} \sigma_t^2 & \rho\,\sigma_t\,s_u \\ \rho\,\sigma_t\,s_u & s_u^2 \end{pmatrix},$$

estimating a single correlation parameter ρ. This "Gaussian copula with correlation" allows linear dependence in the likelihood and provides a simple benchmark for whether any departure from independence improves fit.

Copula Extension. To flexibly model nonlinear and tail dependence, we apply Sklar's theorem: any joint distribution H with marginals F_ε and F_u can be written

$$H(\varepsilon_t, u_t) = C\big(F_\varepsilon(\varepsilon_t),\ F_u(u_t); \theta\big),$$

where $C : [0,1]^2 \to [0,1]$ is a copula with parameter(s) θ. Define

$$u_{1,t} = F_\varepsilon(\varepsilon_t), \quad u_{2,t} = F_u(u_t),$$

using the fitted marginal distributions (e.g. Student-t for ε_t, Gaussian for u_t). We then model

$$(u_{1,t}, u_{2,t}) \sim C(\cdot\,; \theta),$$

to capture any nonlinear dependence.

We consider four copula families, each with distinct tail characteristics:

- *Gaussian copula:* symmetric linear correlation, zero tail dependence ($\lambda_L = \lambda_U = 0$).
- *Student-t copula:* symmetric tail dependence, controlled by degrees of freedom ν (lower ν implies stronger $\lambda_L, \lambda_U > 0$).
- *Clayton copula:* lower-tail dependence only ($\lambda_L > 0,\ \lambda_U = 0$).
- *Gumbel copula:* upper-tail dependence only ($\lambda_U > 0,\ \lambda_L = 0$).

Having outlined the candidate dependence structures, we now describe how the copula parameters are estimated in practice.

Two-Step Estimation (IFM).

1. Fit the Realized GARCH (SRGARCH or ARGARCH) by MLE, obtain residuals $\hat{\varepsilon}_t, \hat{u}_t$ and pseudo-observations
$$u_{1,t} = \hat{F}_\varepsilon(\hat{\varepsilon}_t), \quad u_{2,t} = \hat{F}_u(\hat{u}_t).$$
2. Estimate θ by maximizing the copula log-likelihood
$$\sum_{t=1}^{T} \ln c\big(u_{1,t}, u_{2,t}; \theta\big),$$
where $c = \partial^2 C / \partial u_1 \partial u_2$. Standard errors are obtained via a two-stage bootstrap.

3.4 Forecast Evaluation Procedure

We generate volatility forecasts from each model and evaluate them against two "target variables," defined over a horizon $d \in \{1, 5, 21\}$:

$$\text{Target 1:} \quad \mathrm{TV}_{t,d}^{(1)} = \sum_{i=1}^{d} r_{t+i}^2, \tag{1}$$

$$\text{Target 2:} \quad \mathrm{TV}_{t,d}^{(2)} = 1.4 \sum_{i=1}^{d} \mathrm{rv5}_{t+i}, \tag{2}$$

where r_{t+i} is the close-to-close log-return on day $t+i$, and $\mathrm{rv5}_{t+i}$ is the five-minute realized variance on day $t+i$. The factor 1.4 scales the intraday measure to approximate total daily variance.

For each model and each d-day horizon, we produce the forecast $\hat{\sigma}^2_{t|t-d+1}(d)$ using an expanding-window rolling estimation through 2024. Forecast accuracy is measured by two loss functions:

- **MSE:**

$$\mathrm{MSE}(d) = \frac{1}{N_d} \sum_{t=d}^{T} \left(\hat{\sigma}^2_{t|t-d+1}(d) - \mathrm{TV}_{t,d}^{(j)} \right)^2,$$

  for $j = 1, 2$.

- **QLIKE:**

$$\mathrm{QLIKE}(d) = \frac{1}{N_d} \sum_{t=d}^{T} \left[\frac{\mathrm{TV}_{t,d}^{(j)}}{\hat{\sigma}^2_{t|t-d+1}(d)} - \ln \frac{\mathrm{TV}_{t,d}^{(j)}}{\hat{\sigma}^2_{t|t-d+1}(d)} - 1 \right].$$

To test whether forecast differences are significant, we apply the Diebold–Mariano test pairwise across models for each loss and each target, reporting DM statistics and p-values (with a small-sample adjustment for $d = 21$).

4 Results

4.1 In-Sample Fit of Base Models

We begin by evaluating the in-sample performance of the standard GARCH models. Table 1 presents the parameter estimates and log-likelihoods for SGARCH and AGARCH. The log-likelihood improves from $\ell_{\mathrm{SGARCH}} = -8{,}341.36$ to $\ell_{\mathrm{AGARCH}} = -8{,}234.00$, and the likelihood-ratio test

$$\mathrm{LR} = 2(\ell_{\mathrm{AGARCH}} - \ell_{\mathrm{SGARCH}}) \approx 214.72$$

far exceeds the critical χ^2 value of 3.84 (1 d.f.), leading us to reject SGARCH in favor of AGARCH. This improvement is further supported by lower AIC and BIC values. We therefore select AGARCH as the preferred base model.

Table 1: Parameter Estimates and Log-Likelihood for SGARCH and AGARCH Models

Parameter	SGARCH	AGARCH
μ	0.0605	0.0211
ω	0.0216	0.0198
α	0.1252	—
β	0.8625	—
α_1	—	0.1656
α_2	—	0.0000
Log-Likelihood	-8341.36	-8234.00
AIC	16690.72	16478.00
BIC	16717.56	16511.55

Next, we turn to the realized GARCH models. Table 2 reports the parameter estimates and log-likelihoods for SRGARCH and ARGARCH. The joint log-likelihood increases from $\ell_{\text{SRGARCH}} = -19{,}400.30$ to $\ell_{\text{ARGARCH}} = -19{,}377.19$. The associated likelihood-ratio statistic,

$$\text{LR} = 2(\ell_{\text{ARGARCH}} - \ell_{\text{SRGARCH}}) \approx 46.22,$$

again surpasses the threshold for significance, indicating that ARGARCH provides a significantly better fit due to the inclusion of the linear-shock term $\tau_1 \varepsilon_t$.

Table 2: Full Parameter Estimates and Joint Log-Likelihood for SRGARCH and ARGARCH Models

Parameter	SRGARCH	ARGARCH
μ	0.0532	0.0397
ω	0.0567	0.0574
α	0.0958	0.0938
β	0.5245	0.5230
γ	0.4660	0.4685
ξ	0.0749	0.0695
ϕ	0.6748	0.6780
τ_1	—	-0.1432
τ_2	0.1776	0.1568
s_u^2	2.3984	2.3819
$\ell(r)$ (partial)	-8085.36	-8085.36
Log-Likelihood	-19400.30	-19377.19
AIC	38816.60	38774.38
BIC	38870.28	38841.48

Several implications emerge from these results. First, the estimated mean return increases from 0.0211 in AGARCH to 0.0397 in ARGARCH, suggesting that some return drift previously captured as noise is better explained when realized variance is included.

Second, volatility persistence drops sharply—from 0.9803 in AGARCH to 0.6168 in ARGARCH—indicating faster mean reversion when high-frequency information is used. Additionally, asymmetry is handled differently: AGARCH relies on the threshold effect ($\alpha_1 = 0.1656$), while ARGARCH includes the direct shock term $\tau_1 = -0.1432$, offering a more explicit link between negative returns and realized variance. The feedback coefficients in ARGARCH are also notable: $\gamma = 0.4685$ implies that nearly 47% of next-day latent volatility is driven by current realized variance, while $\phi = 0.6780$ indicates that latent volatility explains 68% of realized variance. Finally, ARGARCH exhibits lower measurement error ($s_u^2 = 2.3819$ vs. 2.3984), indicating more precise estimates. Overall, ARGARCH clearly outperforms its counterparts, both in statistical fit and interpretability.

4.2 Out-of-Sample Analysis of Base Models

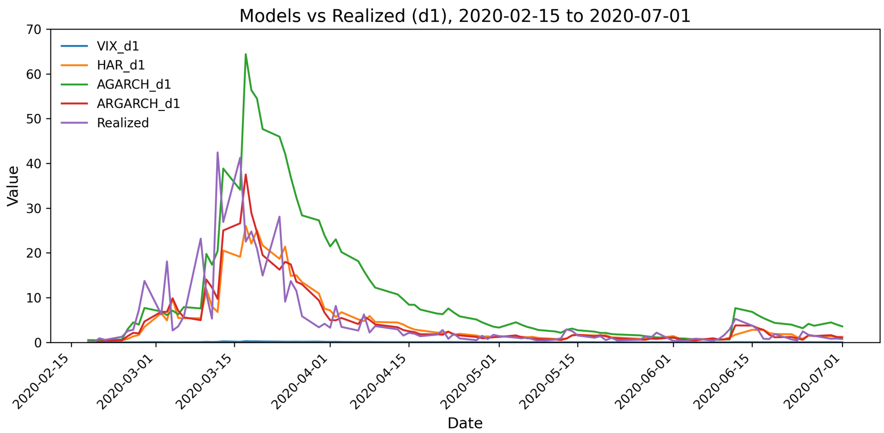

Figure 1: One-day-ahead volatility forecasts vs. actuals (Feb 15–Jul 1, 2020)

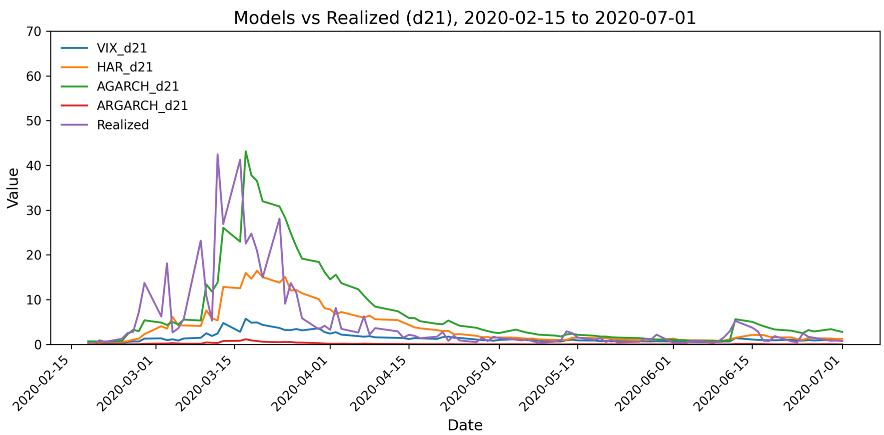

Figure 2: Twenty-one-days-ahead volatility forecasts vs. actuals (Feb 15–Jul 1, 2020)

Figures 1 and 2 illustrate one-day and twenty-one-day ahead volatility forecasts versus realized variances during the COVID-19 shock. At the daily level ($d = 1$), models reveal sharp, immediate reactions to market turbulence, whereas the monthly horizon ($d = 21$)

smooths these spikes as volatility expectations evolve. The COVID-19 shock created a rapid spike of uncertainty in March 2020, making volatility reach historic highs. While financial markets did remain sensitive to ongoing developments, aggressive policy responses helped control volatility, and levels had declined considerably by April. This short-lived peak is visible in the realized volatility, which rapidly reverted as markets partially stabilized. The VIX, which is quoted in annualized percentage points, appears to be almost unreactive in comparison with other models. This can be explained by the fact that its scale naturally compresses short-term fluctuations, making brief but sharp movements harder to discern visually.

The AGARCH model's one-step-ahead forecast is defined as

$$\sigma_{t+1}^2 = \omega + \alpha_1 \mathbf{1}\{\epsilon_t < 0\}\sigma_t^2\epsilon_t^2 + \alpha_2 \mathbf{1}\{\epsilon_t \geq 0\}\sigma_t^2\epsilon_t^2 + \beta\sigma_t^2,$$

which reacts immediately to the most recent shock ϵ_t and volatility σ_t^2, producing a pronounced spike during market turmoil. At a general horizon d, the AGARCH forecast evolves according to

$$\mathbb{E}_t[\sigma_{t+d}^2] = \sigma^2 + (\alpha+\beta)^{d-1}(\sigma_{t+1}^2 - \sigma^2),$$

where $\alpha = (\alpha_1 + \alpha_2)/2$ and $\alpha + \beta < 1$. For $d = 1$, the forecast fully reflects recent volatility, but for $d = 21$, the $(\alpha+\beta)^{20}$ term dampens the impact, yielding a lower, smoother forecast.

In contrast, the ARGARCH model augments this with a measurement equation,

$$x_t = \xi + \phi\sigma_t^2 + \tau_1\epsilon_t + \tau_2(\epsilon_t^2 - 1) + u_t,$$

incorporating both shocks and realized variance x_t. Its longer-horizon forecast is

$$\mathbb{E}_t[\sigma_{t+d}^2] = \sigma^2 + (\alpha + \beta + \gamma\phi)^{d-1}(\sigma_{t+1}^2 - \sigma^2),$$

where the additional $\gamma\phi$ term reflects feedback from high-frequency data. Since the realized component absorbs the immediate shock at $d = 1$, and the composite factor remains less than one, volatility reverts quickly, leading to a much flatter twenty-one-day forecast.

We assess predictive accuracy across three horizons (1, 5, 21 days), two targets (TV_1 and TV_2), and four models (AGARCH, ARGARCH, HAR-RV, VIX) using Mean Squared Error (MSE) and QLIKE loss.

Table 3: MSE Loss by Model, Horizon and Target

Model	1-day TV1	1-day TV2	5-day TV1	5-day TV2	21-day TV1	21-day TV2
AGARCH	18.7726	5.8887	418.7472	119.0885	7579.120	1962.624
ARGARCH	21.5158	4.0894	481.1651	148.9258	7885.769	2069.589
HAR-RV	20.6846	4.4434	459.7725	139.4118	7709.164	2003.608
VIX	26.6062	8.6871	498.7803	162.7800	7823.613	2047.077

Note: Circled entries are the minimum MSE in each column.

Table 4: QLIKE Loss by Model, Horizon and Target

Model	1-day TV1	1-day TV2	5-day TV1	5-day TV2	21-day TV1	21-day TV2
AGARCH	1.7740	1.1801	1.9921	1.3591	3.3602	2.6567
ARGARCH	0.8935	0.1456	1.8368	1.0960	3.3152	2.5736
HAR-RV	1.2420	0.6205	1.9374	1.2745	3.3507	2.6381
VIX	-0.0233	-0.6783	1.8010	1.0518	3.3267	2.5958

Note: Circled entries are the minimum QLIKE in each column.

Tables 3 and 4 present MSE and QLIKE losses for all model-horizon-target combinations. AGARCH consistently delivers the lowest MSE when targeting TV_1 (squared returns), while ARGARCH outperforms for one-day forecasts of TV_2 (scaled realized variance). HAR-RV and VIX models perform worse in squared-error terms, with VIX producing the largest MSEs across all settings. Under QLIKE loss, however, the VIX model dominates at shorter horizons, showing negative values, indicating superior near-term calibration. For $d = 21$, ARGARCH takes the lead under QLIKE, reflecting the benefit of incorporating realized measures and asymmetry. These results suggest that MSE favors AGARCH for return-based volatility proxies, while QLIKE highlights the predictive strength of VIX and ARGARCH depending on the horizon.

Table 5: Full-Sample Forecast Losses

Model	MSE	QLIKE
GARCH(1,1)	4.236162	0.6137283
GARCH–X (VIX)	4.236081	0.6137282
GARCH–X (EPU)	4.160420	0.6131362
RealGARCH	2.956101	0.5500002

Table 6: Diebold–Mariano Test Results (MSE and QLIKE Losses)

			MSE loss		QLIKE loss	
DM	Model 1	Model 2	DM stat	p-value	DM stat	p-value
1	GARCH(1,1)	GARCH–X (VIX)	0.91	0.362	6.92	0.000
2	GARCH(1,1)	GARCH–X (EPU)	3.71	0.000	–2.95	0.003
3	GARCH(1,1)	RealGARCH	1.52	0.129	16.17	0.000
4	GARCH–X (VIX)	GARCH–X (EPU)	3.72	0.000	–2.95	0.003
5	GARCH–X (VIX)	RealGARCH	1.52	0.129	16.17	0.000
6	GARCH–X (EPU)	RealGARCH	1.46	0.144	16.31	0.000

Table 5 reports full-sample forecast losses, with RealGARCH achieving the lowest

MSE and QLIKE. Table 6 further confirms RealGARCH's superiority under QLIKE, significantly outperforming all other models. GARCH–X (EPU) also improves over plain GARCH under MSE, although not significantly better than RealGARCH. These results reinforce the advantage of using realized volatility and suggest that adding EPU helps reduce squared-error loss, while QLIKE favors models like RealGARCH that account for high-frequency information.

4.3 Bivariate Normal and Copulas Extensions

4.3.1 Bivariate Normal Extensions

We re-estimated the joint SRGARCH and ARGARCH models by allowing the return shock r_t and the measurement-equation residual u_t to follow a bivariate normal with correlation ρ.

Table 7: Comparison of SRGARCH and ARGARCH Models

Model	LogLik	AIC	BIC
SRGARCH-indep	-19400.39	38818.77	38879.16
SRGARCH-corr	-19377.36	38774.72	38841.82
ARGARCH-indep	-19377.36	38774.72	38841.82
ARGARCH-corr	-19377.36	38776.72	38850.53

Table Table 7 shows that adding a Gaussian correlation ("corr") to SRGARCH raises the log-likelihood by 23 (–19400.39→–19377.36) and cuts AIC/BIC by about 44/38 points, confirming significant linear return–variance dependence. In ARGARCH, whose measurement equation already includes a linear term, adding corr leaves the log-likelihood unchanged; its slight AIC/BIC increase simply reflects the extra parameter (p=0).

4.3.2 Copula-Based Extension

We constructed uniform "pseudo-observations" from the standardized residuals of each marginal model and fitted four copulas.

Copula	SRGARCH LogLik	SRGARCH AIC	ARGARCH LogLik	ARGARCH AIC
Gaussian	90.37	-178.75	57.55	-113.10
Clayton	-2.54	7.09	171.17	-340.34
Gumbel	-0.00	2.00	41.23	-80.47
Student-t	**229.98**	**-457.97**	**296.23**	**-410.47**

Note: Because copula densities $c(u_1, u_2; \theta)$ can exceed one (unlike marginal densities on the original scale), their logarithms can be positive or negative.

Table 8: Copula fit statistics for SRGARCH and ARGARCH

- SRGARCH: the Student-t copula (df = 4) is overwhelmingly preferred (AIC = –457.97 vs –178.75 for Gaussian), capturing both linear and tail dependence.

- ARGARCH: likewise the Student-t copula dominates (AIC = –410.47).
- Clayton and Gumbel copulas perform poorly in SRGARCG. But Clayton shows some interesting details for the ARGARCH, which can be expected due to it's assymetric properties. Further investigation may search for a copula that carries the properties of both Clayton and Student-t

Table 9: In-sample Fit of Realized-GARCH Models with Student-t Copula

Model	LogLik	AIC	BIC
SRGARCH + Student-t Copula	-19173.17	38368.34	38442.15
ARGARCH + Student-t Copula	-19170.98	38365.96	38446.48

Perhaps the most intriguing result comes when we evaluate the copula-augmented Realized GARCH forecasts. Table 9 reports the in-sample log-likelihood, AIC and BIC for the SRGARCH and ARGARCH models augmented with a Student-t copula. Both copula-augmented specifications deliver substantial improvements over their Gaussian-correlation counterparts, with log-likelihood gains of roughly 200 points and AIC/BIC reductions of over 400 points. Yet, despite its superior in-sample fit, this model did not translate into better out-of-sample forecast accuracy. In fact, its forecast losses were slightly worse than the standard ARGARCH in our evaluation. For instance, for the 1-day horizon the ARGARCH with Student-t copula yielded an MSE of 23.92 (versus about 20.68 for HAR and 26.61 for VIX on the same scale), and for the 21-day horizon its MSE was also higher than ARGARCH without copula (these numbers are from internal calculations; qualitatively, copula-RGARCH was not the best in any category).

We found that the copula-based model tended to overfit the noise: by flexibly capturing every extreme co-movement in-sample, it likely chased patterns that did not repeat out-of-sample. In other words, while tail dependence undeniably exists, using it for point forecasting requires care. The copula model, with its additional parameters (e.g. degrees of freedom ν), may have been too finely tuned to past extremes, and since extreme events are rare, the model had difficulty generalizing. This underperformance persisted across both MSE and QLIKE metrics in the out-of-sample tests.

One way to think about it: the copula helps model the full distribution of volatility (useful for stress testing or computing VaR of volatility itself), but when forecasting the expected value of volatility, simpler models that capture only the first two moments (mean and linear correlation/leverage) performed as well or better.

From a practical perspective, this finding carries an important lesson: model complexity can improve in-sample fit but might not yield out-of-sample benefits. Risk modelers and forecasters should be cautious about using highly flexible models for routine forecasting without strong evidence of predictive gain. It may be that tail-dependence models are more useful for scenario analysis—e.g. asking "if a crash happens, how much volatility spike should we expect?"—rather than for everyday volatility prediction.

5 Conclusion

This study aimed to enhance S&P 500 volatility forecasts by integrating tail-dependence modeling into a broad set of volatility models. Using daily returns (2000–2024), realized variance, VIX, and an economic policy uncertainty index, we compared symmetric and asymmetric GARCH, Realized GARCH, GARCH-X (with VIX or EPU), HAR-RV, and VIX. We then augmented the leading Realized GARCH model with both Gaussian correlation and flexible copulas.

Our findings show that models exploiting asymmetry and high-frequency data perform best. ARGARCH was the top performer, substantially beating classic GARCH in accuracy and statistical tests. Adding leverage (AGARCH) improved GARCH, and incorporating realized variance (RGARCH) yielded the largest gains. HAR-RV and VIX benchmarks lagged, though VIX edged out others on QLIKE at ultra-short horizons. Including EPU added modest MSE improvements but remained inferior to realized-variance models. In line with **?**, we confirm that allowing for asymmetry in a GARCH(1,1) specification yields a markedly better fit than its symmetric counterpart. However, by incorporating realized variance, our Realized GARCH framework demonstrates that it is indeed possible to outperform even the best asymmetric GARCH(1,1) models.

Introducing a Student-t copula delivered the best in-sample fit—evidence of symmetric tail dependence—but did not improve out-of-sample forecasts, which slightly underperformed standard ARGARCH. This suggests that while copulas capture joint extremes useful for risk analysis, they can overfit rare events and are less reliable for point forecasts.

For practitioners, the key message is to leverage high-frequency data and asymmetry for routine volatility forecasting, and to use VIX primarily as a stress indicator. Regulators should incorporate realized-variance models in stress tests and monitor policy uncertainty alongside market volatility. Copula models, although not superior for point forecasts, offer valuable insights into extreme co-movements for scenario analysis.

Future work could extend to multivariate volatility, time-varying copulas, alternative realized measures, and machine-learning hybrids that respect realized volatility and asymmetry. Explicit modeling of overnight returns is another promising avenue. In balancing complexity and robustness, our results underline that richer data and simple asymmetry deliver reliable forecasts, while tail-dependence models are best reserved for risk-management applications.

References

Aslam, F., Mohti, W., and Ferreira, P. (2020). Stock market volatility and return analysis: A systematic literature review. *Entropy*, 22(5):522.

El Brini, S., Bahri, A., and Chouaibi, J. (2024). Modeling and forecasting historical volatility using econometric and deep learning approaches: Evidence from the moroccan and bahraini stock markets. *Journal of Risk and Financial Management*, 17(7):300.

Huber, F., Koopman, S. J., and Lucas, A. (2023). A false discovery rate approach to optimal volatility forecasting model selection. *International Journal of Forecasting*.

Koop, S., Deistler, M., and Gruber, D. (2024). Dynamics in realized volatility forecasting: Evaluating garch models and deep learning algorithms across parameter variations. *Computational Economics*.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Ramachandran, N. and Wang, L. (2021). Sector volatility prediction performance using garch models and artificial neural networks. *arXiv preprint arXiv:2110.09489*.

A Additional Forecast Figures

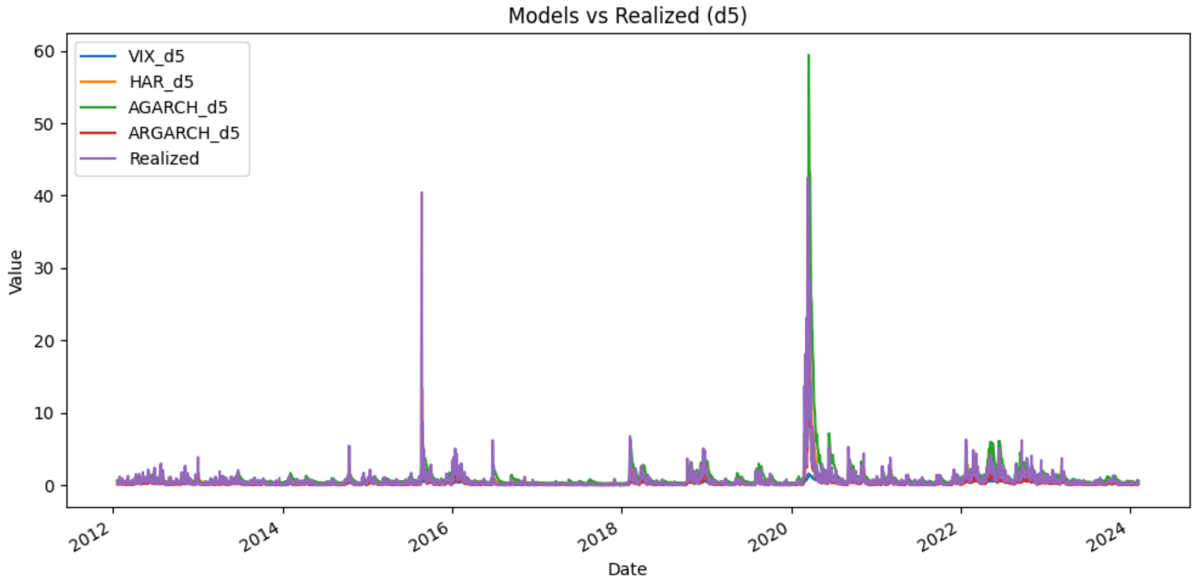

Figure 3: Five-days-ahead volatility forecasts vs. actuals (2012-2024)

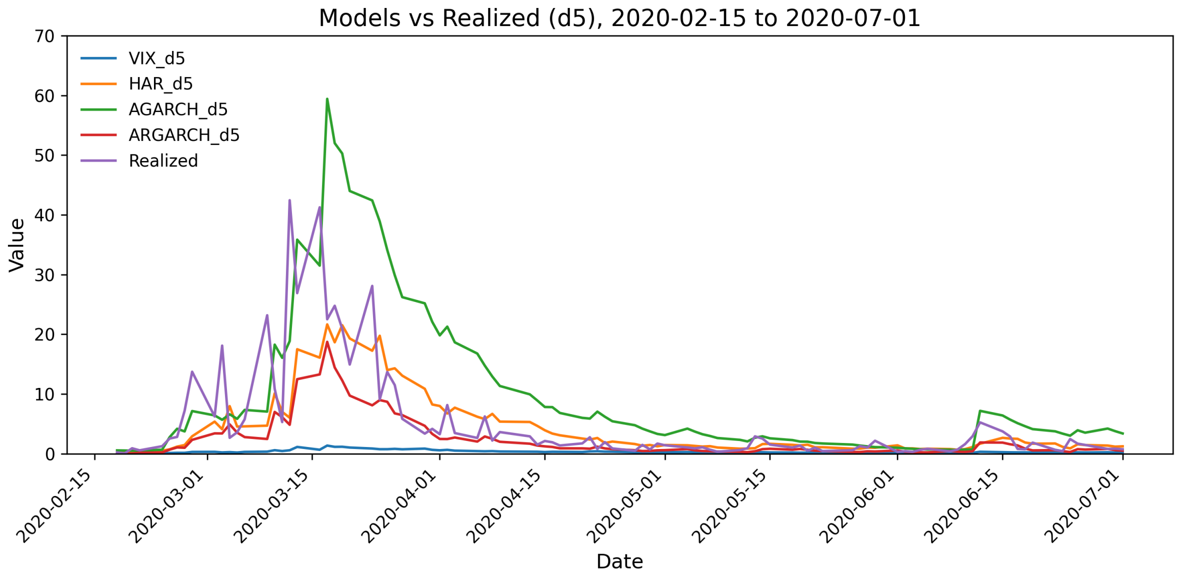

Figure 4: Five-days-ahead volatility forecasts vs. actuals (Feb 15–Jul 1, 2020)

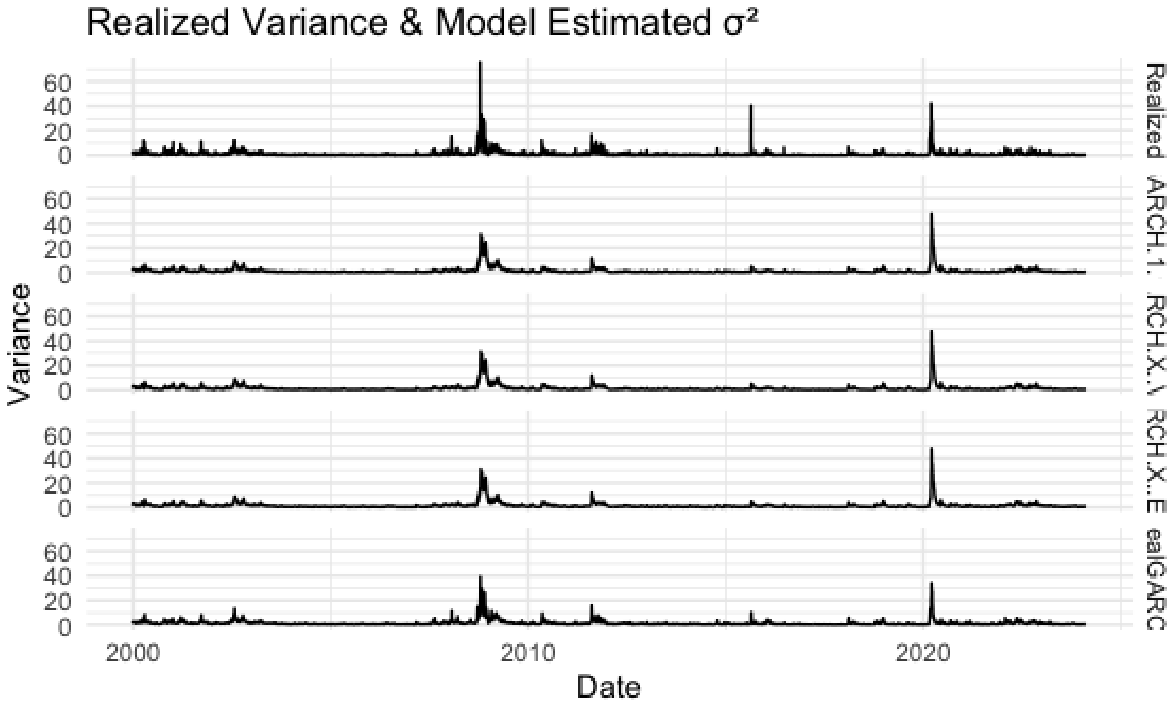

Figure 5: One-day-ahead volatility forecasts of four models (2000-2008)