Volatility Forecasting: Optimal Models Across Diverse Financial Time Series Data

Guus Bouwens

(2701442)

July 1, 2023

Barend Spanjers (supervisor)

dr. Yicong Lin (second reader)

Bachelor Thesis: Econonometrics & Data Science

Abstract

This study investigates the performance of various GARCH models in forecasting the 1-day-ahead volatility of publicly available financial time series data, including ETFs, stocks, and cryptocurrencies. The aim is to contribute to a better understanding of the relationships between volatility, model selection, and distribution preferences, with potential implications for econometric analysis and decision-making within financial markets. The study uses in-sample performance measures like the loglikelihood measures AIC and BIC and out-of-sample performance measures like the backtesting methods VaR and ES. Additionally, the study employs the MAE, RMSE, scalar density forecasting test, and rolling window analysis to evaluate the accuracy of the models. Lastly, the Diebold-Mariano test is employed to compare equally well-performing models and determine the superior forecasting model. The data consist of ten datasets of varying volatility (low to high) for each of the financial time series categories. The research includes additional GARCH models, like E-, FI-, GJR-GARCH, AP-, T-, and H-ARCH to provide a comprehensive analysis of the model's performance under different volatility and stationarity levels.

Keywords: *volatility forecasting, GARCH models, financial time series, performance measures.*

Contents

1 Introduction

Econometric analysis, with a special emphasis on volatility forecasting, forms the backbone of financial research. Its pivotal role extends from risk management and portfolio optimization to economic modeling. By employing rigorous econometric methods to predict asset volatility, researchers can gain insightful market dynamics understanding, inform trading strategies, and drive efficient decision-making, enhancing financial market profitability.

The financial landscape, constantly in flux, has witnessed the rise of numerous models aiming to accurately forecast volatility. Prominent among these are Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. The performance of these models, however, can vary significantly depending on the specific characteristics of the underlying financial data.

In our previous work, we found that volatility forecasting with the Student's t-distribution generally yields the best results for financial time series data. Furthermore, our research highlighted that while GARCH models tend to perform well with low volatile data exhibiting some stationarity, Exponential GARCH (EGARCH) models are more effective for highly volatile, somewhat non-stationary data. Heterogenous ARCH (HARCH) models, conversely, excel with high-frequency trading data on an international scale. Although these conclusions align with existing literature, they were drawn from relatively narrow studies and warrant further investigation.

This study is designed to address the identified research gaps by employing a diverse range of GARCH models to analyze different types of financial data, specifically Exchange-Traded Funds (ETFs), stocks, and cryptocurrencies. Our objective is to identify the most effective models and distributions under various conditions of volatility and stationarity. In addition, we aim to expand on the existing literature by utilizing a more extensive dataset, a larger selection of GARCH models, and a more thorough set of performance metrics. Examples of previous studies in this field include the work by Gabriel (2012), which aimed to forecast the volatility of a stock index in the Romanian market using a selection of five GARCH models and a subset of the performance measures used in our study. The study by Troster et al. (2019) focused on forecasting Bitcoin volatility using backtesting metrics and some of the models used in our research. Wang et al. (2022) compared the accuracy of forecasts from a subset of GARCH models used in our study on two Chinese stock indexes. Our comprehensive approach, of combining many measures used in previous studies, not only enhances the robustness and generalizability of our results but also provides a more nuanced understanding of the conditions that benefit a large set of models and distributions.

In a bid to offer a more thorough comprehension of volatility forecasting, this research tackles a pivotal question: "How can we most effectively forecast volatility across various assets, discern trends in model selection among these assets, and understand the implications of these trends for econometric analysis and financial decision-making?" The findings of this research could yield substantial benefits for both financial practitioners and researchers by offering valuable insights into selecting the most suitable volatility forecasting model for specific needs.

In terms of organization, this paper begins by delivering an exhaustive review of existing literature pertinent to our research. This review subsequently informs the choice of data and methodology. Following an explanation of the chosen datasets – ten each of ETFs, stocks, and cryptocurrencies, ranked by volatility – we present a detailed methodology covering the distributions, models, and performance metrics. Special attention is given to the volatility-based ranking of assets as it is crucial to our central research question. The results section, for clarity and manageability, will concentrate on the top three models for each asset, featuring a data frame of combined performance metrics for the three best-performing models and an interpretation of their parameters. In the results section, we also illustrate the meaning of our findings. The paper concludes with an explicit answer to the research question, followed by a future research discussion and, lastly, an appendix that showcases additional information about the datasets and the results.

2 Literature review

The GARCH model, introduced by Bollerslev (1986) as an extension of the ARCH model proposed by Engle (1982), has been a cornerstone in volatility modeling. Its application has been ubiquitously observed in the field of financial time series data analysis (Wang et al., 2022), encompassing domains such as stock prices, foreign exchange rates, and other financial asset classes. Such modeling tasks carry substantial implications in terms of risk management and portfolio optimization, as deep comprehension of volatility intricacies and effective model selection can dramatically improve the precision and effectiveness of these functions (Diebold, 2015). GARCH models fundamentally aim to encapsulate time-dependent volatility by constructing the conditional variance of the series as a derivative of its historical values. GARCH models tend to exhibit robustness against model misspecification and are frequently favored due to their reduced computational demands (Francq and Zakoian, 2011). Hansen and Lunde (2011) illustrated the utility of GARCH models in modeling volatility for stationary data devoid of structural breaks. Performance evaluation of GARCH models across disparate financial time series categories has been pursued through the use of various datasets and metrics. Studies such as Troster et al. (2019) explored the efficacy of GARCH models in predicting Bitcoin returns and associated risk, while Gabriel (2012) assessed the forecasting performance of GARCH models using Romanian stock market data. Nugroho et al. (2019) carried out an empirical comparison of several GARCH models on stock returns volatility.

The inception of GARCH models and their subsequent variations have had a transformative impact on financial econometrics due to their ability to model and forecast volatility. A distinctive feature of GARCH models is their capacity to encapsulate the temporal variation in volatility, a key trait of financial return series. The EGARCH model, introduced by Nelson (1991), addresses volatility asymmetry or the leverage effect. This model allows positive and negative news to exert different impacts on volatility, a characteristic crucial when modeling financial asset returns that frequently exhibit such behavior. The Fractionally Integrated GARCH (FI-GARCH) model, introduced by Baillie et al. (1996), encapsulates the long memory property in volatility, reflecting the slow decay of autocorrelations in volatility, a common phenomenon in financial markets. Other noteworthy variants include the Glosten-Jagannathan-Runkle GARCH (GJR-GARCH) model, which introduces an additional term to capture volatility asymmetry (Glosten et al., 1993). This allows for a leverage effect, where negative shocks to returns can have a different effect on volatility than positive shocks. The TARCH model, initiated by Engle et al. (1987) and later developed by Zakoian (1994), permits regime-switching behavior in volatility. The Asymmetric Power ARCH (APARCH) model, proposed by Ding et al. (1993), enhances the basic GARCH model by enabling more flexible modeling of asymmetry and the leverage effect through the inclusion of a power term in the conditional variance equation. Lastly, the HARCH model, by Muller et al. (1997), incorporates heterogeneous characteristics of high-frequency financial time series, including lags, which helps in capturing the long memory property in volatility.. These models' unique attributes and versatility make them exceptionally useful in financial econometrics, empowering them to capture various volatility dynamics present in financial return series. The objective is to meticulously evaluate each model's performance across diverse volatility and stationarity levels (Bauwens et al., 2006).

Different error distributions, particularly the normal, Student's t, skewed Student's t, and Generalized Error Distribution (GED) have been widely utilized in financial econometrics for their unique characteristics and adaptability to various data conditions. The Normal distribution is the most straightforward and conventional choice for modeling error terms, attributed to its symmetric and bell-shaped structure (Tsay, 2005). It's characterized by two parameters, the mean and variance, which simplify interpretation and mathematical manipulation. Yet, the normal distribution tends to underestimate extreme values, limiting its practical applicability in financial markets known for 'fat-tailed' return distributions. However, a significant advantage of the Gaussian distribution is that it is consistent under misspecification. This property means that even if the true error distribution is not Gaussian when we assume a Gaussian distribution for the error term, the maximum likelihood estimate (MLE) of the parameters will still converge to their true values as the sample size increases. This robustness property of the Gaussian distribution, often referred to as 'Gaussian quasi-MLE', is a primary reason it is widely used despite its shortcomings in capturing extreme values. Indeed, the Gaussian quasi-MLE can provide reliable estimates even when the distributional assumption is violated, making it a versatile tool in financial econometrics (Todros and Hero, 2015). The Normal distribution's mathematical convenience and the robustness offered by Gaussian quasi-MLE together contribute to its continued popularity in financial econometrics, despite the known limitations. The Student's t-distribution offers a more nuanced approach for modeling errors in financial returns, which often exhibit heavy tails or kurtosis. The distribution introduces an additional degrees-of-freedom parameter, enabling the accommodation of extreme outliers more effectively than the normal distribution (Mandelbrot, 1963). The skewed Student's t-distribution extends the Student's t-distribution by incorporating skewness into the model (Fernandez and Steel, 1998). This additional flexibility makes it particularly suitable for financial returns, which often exhibit skewness alongside heavy tails. It accounts for asymmetric responses to positive and negative returns, often observed in financial markets. Lastly, the GED generalizes the normal and double exponential distributions (Nadarajah and Kotz, 2005). The GED introduces a parameter that controls the thickness of the tails, making it an excellent choice for modeling financial return distributions that differ from the normal distribution in terms of kurtosis.

In financial econometrics, the deployment of performance measures to evaluate forecasting models is a vital step in confirming their reliability and efficacy. Different measures provide diverse perspectives on a model's performance, thereby enriching our understanding of its strengths and weaknesses. The implementation of backtesting forms a cornerstone of assessing the effectiveness of forecasting models. VaR is a widely recognized risk metric employed in financial risk management, offering a robust measure of the maximum loss one could expect over a defined time period at a specific confidence level. It quantifies risk by determining the potential loss in value of a risky asset or portfolio over a defined period for a given confidence interval (Jorion, 2007). Its strength lies in its intuitive interpretation and straightforward computation. However, VaR has been criticized for its inability to capture tail risk adequately and for not accounting for the severity of losses beyond the VaR threshold (Danielsson, 2011). To complement VaR and to better encapsulate tail risk, Expected Shortfall (ES), can be used. ES provides an estimate of the expected loss given that a VaR threshold has been exceeded (Acerbi and Tasche, 2002). By focusing on extreme losses in the tail of the distribution, ES can capture the risk of rare and extreme events better than VaR, offering a more comprehensive picture of potential losses (Embrechts et al., 2005). It's important to note that both VaR and ES have been recognized under the Basel III regulatory framework for financial institutions, highlighting their relevance in the practical world of financial risk management (Basel, 2013).

In terms of forecasting accuracy, mean absolute error (MAE) and root mean square Error (RMSE) are popular metrics in time series analysis. MAE measures the average magnitude of errors in a set of forecasts, without considering their direction, thus providing a straightforward metric for average error (Willmott and Matsuura, 2005). It is less sensitive to outliers compared to RMSE and provides a linear error score where all individual differences are weighted equally. Conversely, RMSE is a quadratic scoring rule that measures the average magnitude of the error, assigning a higher weight to large errors. This means RMSE is more useful when large errors are particularly undesirable (Hyndman and Koehler, 2006). It offers a high penalization of large errors, thus representing both the variance of the forecast errors and the forecast's bias. The scalar density forecasting Test offers a robust method for evaluating the performance of a model's density forecasts, emphasizing the model's ability to predict the entire distribution of a variable, including higher moments like skewness and kurtosis. Unlike conventional error metrics, a higher score on this test indicates superior model performance (Elliott and Timmermann, 2016). Finally, the rolling window analysis offers a practical approach for evaluating the stability and performance of a model over time. It facilitates the observation of changes in a model's forecasting performance across different data sub-samples, potentially capturing the evolving dynamics in financial markets (Goval and Welch, 2004). Alternatively, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are used to compare between models for a given dataset (Akaike, 1974, Schwarz, 1978). These criteria balance the complexity of the model against its fit quality, with lower values indicating superiority. Lastly, the Diebold-Mariano Test can be used for comparison between models. It is a statistical procedure used for comparing the forecast errors of two models (Diebold and Mariano, 1995). It determines whether the forecast errors of the two models are statistically different. If they are not, we conclude that the two models do not significantly differ in their predictive accuracy. Combining these metrics offers a rigorous evaluation methodology, instrumental in identifying the model that delivers superior predictive accuracy and risk management.

The disparities in trading behaviors among diverse asset classes such as stocks, cryptocurrencies, and ETFs could potentially influence the performance of various forecasting models like GARCH models. Specifically, structural differences in trading schedules – with stocks observing a weekday trading pattern while cryptocurrencies are traded all day long – may cause subtle structural breaks leading to differential performance of these models across asset classes. Moreover, the impact of trading frequency on the liquidity of these asset classes and market microstructure nuances might also contribute to the observed variations in model performance (Hautsch, 2012). For instance, the high-frequency trading prevalent in cryptocurrency markets might necessitate models that can better capture such dynamics. Furthermore, varying degrees of market efficiency across stocks, cryptocurrencies, and ETFs could also impact model choice. Efficient markets, for instance, may call for models that can effectively capture rapid information assimilation into prices (Fama, 1970). Lastly, given the unique return distributions and volatility dynamics across stocks, cryptocurrencies, and ETFs, different GARCH-family models could be optimal for each asset class.

In summary, the combined power of such a wide range of models and performance metrics facilitates a robust and comprehensive evaluation. This holistic approach serves to identify the most effective model for volatility forecasting under different circumstances.

Before going into detail about the models and methods, it is important to introduce the assets that were investigated and to illustrate the structure of this paper.

3 Data and Descriptive statistics

Our research explores the volatility forecasting models across a broad spectrum of asset types—ETFs, stocks, and cryptocurrencies. Each asset type has unique characteristics that influence its volatility, necessitating a specific approach to forecasting.

ETFs encapsulate various sectors and volatility levels, offering a comprehensive view of diverse market segments. Stocks, chosen across different sectors, allow us to delve into the correlation between increasing volatility and optimal forecasting models, considering the influence of sector-specific factors and broader economic conditions. Cryptocurrencies, on the other hand, embody the dynamic nature of the crypto market, affected by distinct factors like market sentiment and regulatory news.

Overall, the study emphasizes the importance of individually considering a range of volatility levels across these asset groups for effective forecasting. It suggests that the optimal forecasting model may depend on the asset type as well as the volatility level, reinforcing the need for an individualized and holistic approach to asset management. An overview of the selected assets is given in Table 1.

Table 1: Volatility of the selected financial assets

ETF		**Stock**		**Cryptocurrency**	
Asset	**Volatility**	**Asset**	**Volatility**	**Asset**	**Volatility**
EUO	1.02	JNJ	1.12	BTC	4.05
ˆDJI	1.10	MCD	1.25	LTC	6.25
VO	1.19	MDLZ	1.35	XMR	6.99
QQQ	1.34	MSFT	1.72	XRP	7.53
DBE	1.75	AMZN	2.07	NMC	8.32
UBS	2.03	BABA	2.66	DOGE	10.22
BNO	2.3	TSLA	3.63	FTC	10.95
UTSL	4.05	TUP	4.65	BLK	12.23
ERX	4.67	GME	6.73	GRC	13.97
BOIL	5.77	FCEL	7.30	XPM	22.58

The stocks and ETFs are extracted from YahooFinance directly. The cryptocurrencies are custom-assembled from raw data extracted from the Coingecko platform. The assets their full descriptive statistics and economic backgrounds can be found in appendix 8.1.

All datasets hold exactly ten years of data, counting back from the sixth of April 2023. For the ETFs and stocks this comes down to 2520 observations, due to the fact that the market is closed on the weekends. However, some of the cryptocurrencies did not exist, or were at least not publicly available, ten years ago. This is due to the nature of cryptocurrencies, they are relatively young. However, cryptocurrencies do have more observations, as they are on the market all day long. Ranging from the oldest cryptocurrency (Bitcoin) with 3627 observations to the youngest selected cryptocurrency (Gridcoin) with 2957 observations.

4 Methodology

We pre-emptively establish a significance level of 5% for all tests. This signifies that any conclusions drawn from our evaluations and measurements will be considered statistically significant if the corresponding p-value is less than 0.05. By setting this threshold, we gather evidence against the null hypothesis, as there is less than a 5% probability that the null hypothesis holds true and the observed results are merely coincidental.

4.1 Distributions

The unique characteristics of financial data, such as heavy tails and skewness, can pose a challenge for normal distribution modeling. This necessitates the use of alternative distributions, notably the Student's t-distribution ('t'), skewed Student's t-distribution ('skewt'), and Generalized Error Distribution ('GED'). These distributions use μ to represent mean return, ν to control tail heaviness, η to determine spread, and λ for skewness. For the GED, ν equals 2 for a normal distribution, less than 2 for heavier tails, and more than 2 for lighter tails. On the other hand, in the t-distribution where η is interpreted as the degrees of freedom, lower η values lead to thicker tails, and as η increases, the distribution approximates a normal one. The skewt distribution, used to address return asymmetry, additionally includes the skewness (λ) parameter. The t-distribution and the GED are part of different distribution families (location-scale and exponential, respectively), each suitable for different statistical procedures (Balakrishnan, 2019).

In GARCH models, often used to forecast financial data volatility, the distribution choice is critical as it influences the forecasting and risk measurement outcomes. Additionally, GARCH models can accommodate various data aspects, including leverage effects, long memory, and asymmetry, by modelling the log-return at time t, $e_t = h_t\epsilon_t$, where h_t a volatility term and ϵ_t a standard normal error term. However, the error term distribution significantly impacts the model. While standard GARCH models assume a standard normal distribution for ϵ_t, alternative distributions like the GED, t, or skewt are often necessary to capture the extreme values seen in financial data (Tsay, 2010).

To ascertain model distribution preferences, we employ the Probability Integral Transform (PIT) method. It calculates the assigned distribution over the residuals of the optimal model for a given asset. The Kolmogorov-Smirnov (KS) test is then used to compare the cumulative distribution function (CDF) values with a uniform distribution. The KS test, as written about by Sheskin (2003), is a non-parametric method used to compare a sample with a reference probability distribution. The KS statistic represents the maximum divergence between the distribution functions, with a smaller value indicating a closer match. The p-value quantifies the likelihood of observing a given KS statistic under the null hypothesis of identical distributions. A p-value below a common threshold (0.05) provides strong evidence to reject the null hypothesis, suggesting significant differences between the distributions. Therefore, both the KS statistic and the p-value are crucial in interpreting the KS test.

4.2 Garch-type models

Table 2: All GARCH-type models

Model	Formula	
GARCH	$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2$	(1)
GJR-GARCH	$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \lvert\epsilon_{t-i}\rvert^2 + \sum_{j=1}^{q} \beta_j h_{t-j}^2 + \sum_{k=1}^{o} \gamma_k \lvert\epsilon_{t-k}\rvert^2 I[\epsilon_{t-k} < 0]$	(2)
TARCH	$h_t^\lambda = \omega + \sum_{i=1}^{p} \alpha_i \lvert\epsilon_{t-i}\rvert^\lambda + \sum_{j=1}^{q} \beta_j h_{t-j}^\lambda + \sum_{k=1}^{o} \gamma_k \lvert\epsilon_{t-k}\rvert^\lambda I[\epsilon_{t-k} < 0]$	(3)
EGARCH	$\ln(h_t^2) = \omega + \sum_{i=1}^{p} \alpha_i (\lvert e_{t-i}\rvert - \sqrt{2/\pi}) + \sum_{j=1}^{q} \beta_j \ln(h_{t-j}^2)$	(4)
HARCH	$h_t^2 = \omega + \sum_{i=1}^{p} \alpha_{l_i} (l_i^{-1} \sum_{j=1}^{l_i} \epsilon_{t-j}^2)$	(5)
APARCH	$h_t^\delta = \omega + \sum_{i=1}^{p} \alpha_i \lvert\epsilon_{t-i}\rvert^\delta + \sum_{j=1}^{q} \beta_j h_{t-j}^\delta$	(6)
FIGARCH	$h_t = \omega(1-\beta)^{-1} + \sum_{i=1}^{q} (\theta_i \epsilon_{t-i}^2)$	(7)

Where h_t: the conditional volatility at time t, $\omega > 0$: a constant, α_i the ARCH[p] parameter related to past squared residuals, β_j the GARCH[q] parameter related to past volatility, γ_k the leverage effect parameter, $I_{\epsilon_{t-k}<0}$: an indicator function that takes the value of 1 if $\epsilon_{t-k} < 0$ and 0 otherwise, λ a power parameter which is set to 1, ϵ_{t-i}: the residual, p, q and o: the order of the GARCH models which are in our case all set to 1. Additionally, for the EGARCH model: $e_t = \epsilon_t / h_t$. For HARCH: l_i is the lag order, in our case $l_i = [1, 5, 22]$, if $l_i = 1$, the model defaults to the standard ARCH(1) model (Sheppard et al., 2023). For APARCH: δ is the power parameter. Lastly, for FIGARCH: $\theta_i = d - \beta + \phi$, where d is the fractional differencing parameter, which determines the memory of the process, and ϕ is the parameter for the long-term mean.

The constant ω and parameters α, β, γ, d, ϕ, δ are estimated using the ARCH_MODEL package in Python (Sheppard et al., 2023).

The complexity of GARCH-type models varies, each with unique parameters and assumptions. The GARCH model is the simplest, with three parameters - a constant, and two parameters for the ARCH and GARCH terms. It assumes symmetry and does not account for leverage effects. The GJR-GARCH and TARCH models are more complex than GARCH, introducing an additional term to allow for an asymmetric response to positive and negative shocks, known as the leverage effect. This increases the number of parameters to four. The TARCH model adds a slight degree of complexity by adding the λ power parameter. The EGARCH model, similar to TARCH and GJR-GARCH, introduces an asymmetric response to shocks. However, it's more complex because it uses a logarithmic form for the variance equation, allowing for more flexibility in capturing the leverage effect without the need to impose non-negativity constraints on its parameters. The HARCH model introduces heterogeneity in the aggregation of shocks. It averages past squared residuals over different horizons. The APARCH model increases complexity by allowing model parameters to vary with the data, a concept known as conditional heteroskedasticity. It includes an additional power term in the variance equation, allowing for more flexible modeling of the leverage effect and the tail behavior of the distribution. The FIGARCH model adds complexity by incorporating a fractionally differencing term and another term to capture long memory.

While complexity can enhance model performance, it's crucial to balance it against the risk of overfitting. Additionally, more complex models can significantly increase computational demands, another important consideration.

4.3 Performance metrics

4.3.1 Backtesting and Risk Measures

Backtesting was performed by estimating the models up to the last observation of 2021-03 and computing out-of-sample forecasts from 2021-04 to 2023-04.

The Value-at-Risk (VaR) is a risk measure defining a lower bound for losses, calculated as

$$\mathrm{VaR}_t(p) = -\mu - h_{t-1} F_R^{-1}(p) P_{t-1}, \tag{8}$$

where μ is the mean return, h_{t-1} is the conditional volatility, $F_R^{-1}(p)$ is the inverse CDF of returns at p = 5%, and P_{t-1} = 1 for daily returns. So it quantifies the worst expected loss over a given horizon at a given confidence level under normal market conditions.

The Violation Ratio (VR) is a measure that contrasts the observed violations of the Value at Risk (VaR) with the expected violations, calculated as

$$\mathrm{VR} = \frac{\text{Observed number of violations}}{\text{Expected number of violations}} = \frac{v_1}{p \times W}. \tag{9}$$

If the VR is greater than one, it indicates over-forecasting. Conversely, if the VR is less than one, it suggests under-forecasting.

Expected Shortfall (ES) represents the average losses when VaR is violated, computed as

$$\mathrm{ES} = \frac{1}{W \times p} \sum_{i=1}^{W \times p} q_i \mid q_i \leq -\mathrm{VaR}(p). \tag{10}$$

Where W is the number of observations, p the confidence level, and q_i refers to losses worse than VaR. It's a more conservative risk measure considering the worst-case scenario. ES is subadditive, making it coherent, whereas VaR is not.

The effectiveness of risk models isn't purely determined by their VaR or ES predictions, but by how well these predictions hold against actual incurred losses. Models must be capable of adapting swiftly to market conditions, and maintain stability even in volatile scenarios. However, accuracy is not solely about smaller deviations between predicted and actual values, but about the model's capacity to correctly estimate substantial losses (Dowd, 2005). This is why, in concurrence with the upcoming performance measures, we define the risk measures as loss functions, with VR set to 1 as the optimal outcome. This approach provides a more realistic measure of the model's performance, reflecting its ability to accurately predict losses, limit risk and adapt to changing market conditions.

However, it's important to note that the accuracy of these risk measures also matters. An overly optimistic model (predicting less risk than what truly exists) is not necessarily better even if it reports lower risk measures. So the context and accuracy of the model should be taken into account along with the VaR and ES values.

4.3.2 Accuracy Measures

The mean absolute error (MAE) and root mean squared error (RMSE) are common accuracy measures, quantifying the average discrepancy between predicted and actual values. MAE, calculated as

$$\mathrm{MAE} = \frac{1}{N} \sum_{t=1}^{N} |y_t - \hat{y}_t|, \tag{11}$$

measures the average magnitude of errors, while RMSE, calculated as

$$\mathrm{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (y_t - \hat{y}_t)^2}, \tag{12}$$

assigns a higher weight to larger errors, making it more sensitive to outliers. Where, N is the total number of observations, y_t is the actual value for the observation t, and $\hat{y}_t$ is the predicted value for the observation t. Lower values for both MAE and RMSE indicate better model fit.

The scalar density forecasting test is a tool used to assess how accurately a model is able to predict the complete distribution of returns. The procedure for this test involves calculating the logarithm of the predicted probability, denoted as p_t, given the observed return y_t at each observation t. This value is represented as s_t. Formally, s_t is computed as follows:

$$s_t = \log p_t(y_t), \tag{13}$$

After calculating s_t for all time points in the data set, the next step is to average these values. This is done by summing up all the calculated s_t values and dividing by the total number of predictions, N. Formally, this average, denoted as $\bar{s}$, is calculated as:

$$\bar{s} = \frac{1}{N} \sum_{t=1}^{N} \log p_t(y_t). \tag{14}$$

In this context, a higher value of $\bar{s}$ suggests that the model has a better fit in terms of predicting the distribution of returns. This is because a high average logarithm of the predicted probabilities indicates that the model's density forecasts align closely with the observed returns.

The rolling window analysis assesses model performance and stability over time by estimating models over sequences of rolling data windows. These windows are set to be every day, as we are forecasting the 1-day-ahead volatility. The one-step-ahead forecast $y_{t+1|t}$ is generated using a fixed window length h, given by

$$y_{t+1|t} = f(y_{t-h+1}, y_{t-h+2}, \ldots, y_t). \tag{15}$$

Model accuracy over time can be assessed using the aforementioned accuracy measures.

4.3.3 Comparison Measures

The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are criteria for model selection, comparing the quality of different models based on goodness-of-fit and complexity. The AIC is defined as

$$\text{AIC} = -2\ln(L) + 2k, \tag{16}$$

and tends to favor more complex models, while

$$\text{BIC} = -2\ln(L) + \ln(N)k \tag{17}$$

prefers simpler models. Here, L is the likelihood function, N is the number of observations, and k is the number of parameters. Lower values for both AIC and BIC indicate better models. However, they provide relative measures for model comparison, not absolute measures for individual models.

The Diebold-Mariano (DM) test is a method used to compare the forecast errors of two models. To be more specific, a forecast error at time t, denoted as $\epsilon_{A,t}$ or $\epsilon_{B,t}$, is defined as the difference between the observed value and the predicted value by model A or B, respectively. The DM test measures the performance of two predictive models by comparing the mean and standard deviation of the differences between their forecast errors. The null hypothesis for this test is that the average difference in errors is zero, whereas the alternative hypothesis posits that the average difference is not zero. The DM statistic is computed as follows:

$$d = \frac{\bar{d}}{\sqrt{Var(d)/T}}, \tag{18}$$

where d denotes the differences in the losses corresponding to the forecast errors of models A and B at time t. It's defined as $d_t = L(\epsilon_{A,t}) - L(\epsilon_{B,t})$. Here, L represents a loss function, which quantifies the penalty for inaccurate predictions. The choice of L can vary, and it's commonly the square loss function $L(\epsilon) = \epsilon^2$ (as it is in our case) or the absolute loss function $L(\epsilon) = |\epsilon|$ for regression problems. $Var(d)$ is the sample variance of these differences, T is the sample size, and $\bar{d}$ is the mean of the differences in errors. In this context, a positive d_t value means that model A has higher forecast accuracy than model B, as it incurs a smaller loss. Conversely, a negative d_t value indicates that model B is more accurate than model A (Diebold, 2015).

It is important to keep in mind that our results for the DM test are based on the same period as the backtesting forecasting period. So the data it represents is the last 2 years.

4.3.4 Combined Metric

In order to provide a holistic evaluation of the models, we employ a weighted combination of the metrics discussed earlier. The backtesting metrics, which include VaR, ES, and VR, each carry a weight of 10%. These metrics evaluate the model's ability to forecast losses and predict extreme risks. Prediction error measures, such as MAE and RMSE, along with the scalar density forecasting test, each carry a weight of 10%. These metrics assess the accuracy of the model's forecasts. The rolling window analysis, which carries a weight of 20%, evaluates the out-of-sample forecasting performance of the model. For 1-day-ahead volatility forecasts, it justifies its higher weight due to its ability to incorporate real-time adaptation and capture time-varying patterns. Finally, complexity control measures, such as AIC and BIC, each carry a weight of 10%. These measures help prevent the overfitting of an optimal model. The combined metric is structured as a loss function, whereby a lower value denotes superior model performance. Detailed steps of its computation, implemented in Python, are elucidated in Appendix 8.3. This strategy guarantees an all-encompassing evaluation of a model by considering multiple key performance attributes, thereby optimizing its ability to manage risk.

4.3.5 Verification of Assumptions

Lastly, understanding whether the models we used comply with certain time series forecasting principles is critical. In essence, while it seems all models could potentially be employed, those not requiring normal residuals and those capable of capturing remaining autocorrelation may provide superior results. More about this can be found in Appendix 8.2.

5 Results & Analysis

For each asset the following metrics are shown: the combined metric (C) for the top three models, the parameters of those models, and if graphical analysis is possible, a graph illustrating the VaR of those models. For assets where their top three models are similar in performance, the Diebold-Mariano test is given. The parameters of the distributions are also given if the top three models all belong to the same model type, as proposed in Section 4.1. A comprehensive compilation of the performance metrics for all the top three models can be found in Appendix 8.4.

The choice of focusing on the top three models is a strategic one, aimed at offsetting potential skewness from the combined metric - an unconventional measure that may yield skewed optimal models despite its comprehensive incorporation of various crucial measures. Observing the top three models allows for a comparison of the difference in optimality between them, potentially highlighting inconsistencies in the absolute performance measure. For instance, it might reveal that a model that is relatively more computationally expensive performs only marginally better than the second or third model. A broader comparison involving more models might lead to a loss of comprehensibility. Conversely, an overly narrow focus could obscure differences in optimality and fail to flag potential inconsistencies in the combined metric. With 32 models for comparison, as detailed in Appendix 8.5, finding the right balance is essential. Therefore, the choice of the top three models strikes an optimal balance between comprehensibility and thoroughness.

As we dive into our results analysis, our first approach focuses on each asset group individually. These ten assets have been divided into three separate categories based on their varying levels of volatility: low, medium, and high. This granular analysis helps us identify patterns and trends within each group. Subsequently, we explore the broader trends that span across these groups. Finally, we point out some interesting cases and showcase whether the optimal models did indeed belong to their assigned distributions.

ETFs

While looking at Table 3, we start with the low volatility assets, including EUO, ˆDJI, and VO. For EUO, the HARCH model paired with a GED distribution outshone the rest, due to its higher ω and $\alpha[1]$ parameters, which suggest a greater long-term variance level and a responsive reaction to past shocks. The model was particularly skilled at capturing volatility dynamics, scoring lower than other contenders like the TARCH GED and APARCH GED models. When considering the ˆDJI asset, the TARCH model with a skewt distribution took the lead. Its unique parameters provided accurate long-term variance level estimation and moderate response to volatility asymmetry, thus outweighing the inferior performing GJR-GARCH skewt and TARCH GED models. For the VO asset, the APARCH model, armed with a skewt distribution, emerged as the top choice. Its unique blend of parameters - a moderately high mean, the lowest ω, and a higher $\alpha[1]$ - coupled with the

δ parameter, allowed the model to capture complex volatility dynamics effectively, outperforming the GARCH skewt and GJR-GARCH GED models.

Moving on to medium volatility assets: QQQ, DBE, and UBS. The QQQ asset saw the TARCH model with a skewt distribution perform best, given its prowess at capturing the leverage effect and volatility persistence, as suggested by its negatively skewed distribution given in Table 9. DBE also identified the TARCH with a skewt distribution as the optimal model, thanks to its ability to capture negative mean return, a moderate level of long-term variance, and high volatility persistence. However, for UBS, the HARCH model with a GED distribution was the champion, effectively capturing the asset's volatility dynamics even without parameters representing leverage effects or volatility persistence.

Table 3: ETFs: All top 3 models' combined metrics and parameters.

Asset	Model	**C**	μ	ω	$\alpha[1]$	$\gamma[1]$	$\beta[1]$	δ
EUO	HARCH GED	0.31	0.003	0.856	0.140	-	-	-
	TARCH GED	0.32	0.012	0.010	0.052	-0.014	0.955	-
	APARCH GED	0.35	0.010	0.007	0.046	-	0.954	1.538
ˆDJI	TARCH skewt	0.09	0.036	0.035	0.032	0.225	0.852	-
	GJR-GARCH skewt	0.11	0.042	0.026	0.034	0.280	0.806	-
	TARCH GED	0.12	0.055	0.037	0.047	0.222	0.836	-
VO	APARCH skewt	0.21	0.067	0.030	0.173	-	0.781	2.288
	GARCH skewt	0.22	0.068	0.032	0.182	-	0.789	-
	GJR-GARCH GED	0.24	0.071	0.038	0.075	0.230	0.771	-
QQQ	TARCH skewt	0.10	0.062	0.045	0.000	0.248	0.871	-
	TARCH GED	0.14	0.106	0.047	0.032	0.198	0.856	-
	TARCH t	0.18	0.114	0.041	0.007	0.241	0.867	-
DBE	TARCH skewt	0.32	-0.031	0.017	0.002	0.088	0.954	-
	HARCH GED	0.33	0.035	1.817	0.341	-	-	-
	TARCH GED	0.34	0.008	0.016	0.004	0.090	0.951	-
UBS	HARCH GED	0.20	0	2.247	0.322	-	-	-
	GJR-GARCH GED	0.29	0	0.385	0.083	0.096	0.750	-
	HARCH t	0.30	0.042	2.333	0.325	-	-	-
BNO	TARCH skewt	0.19	-0.025	0.025	0.008	0.093	0.945	-
	TARCH GED	0.20	0.006	0.026	0.011	0.098	0.940	-
	TARCH t	0.22	0.004	0.024	0.010	0.094	0.943	-
UTSL	TARCH skewt	0.09	0.145	0.095	0.075	0.073	0.884	-
	GJR-GARCH skewt	0.12	0.145	0.215	0.046	0.084	0.888	-
	TARCH GED	0.12	0.167	0.093	0.084	0.062	0.882	-
ERX	TARCH skewt	0.17	-0.034	0.071	0.022	0.097	0.928	-
	TARCH GED	0.18	0.002	0.081	0.029	0.101	0.919	-
	TARCH t	0.22	0.013	0.072	0.023	0.099	0.926	-
BOIL	GARCH t	0.23	-0.199	0.445	0.077	-	0.904	-
	GJR-GARCH t	0.23	-0.198	0.445	0.078	-0.002	0.905	-
	GJR-GARCH skewt	0.24	-0.159	0.435	0.077	-0.002	0.906	-

Turning our attention to high volatility assets, such as BNO, UTSL, ERX, and BOIL, the TARCH skewt model consistently delivered the best performance for BNO, UTSL, and ERX, demonstrating its proficiency in capturing both leverage effects and asymmetric return distributions. Additionally, for the BNO asset, the negative skewness parameter aids its performance, as can be seen in Table 9. The same applies to ERX, as TARCH skewt outperformed the second and third-best models, of the same model type. The BOIL asset, however, saw both the GARCH t and GJR-GARCH t models performing equally well. The GJR-GARCH skewt model was significantly outperformed (Figure 2a), despite its ability to account for skewness in the return distribution.

When taking into account all subcategories, it's clear that the best model and distribution pairing is dependent on the specific asset and the need to capture specific return distribution characteristics. Models such as TARCH, and GJR-GARCH, paired with GED, skewt, or t-distributions, often delivered stellar performances. However, the selection of the optimal model was steered by its ability to capture volatility dynamics, balance model complexity with forecasting performance, and achieve low combined scores.

Table 4: Count of each distinct model appearing in the top 3 of the ETFs asset group.

Model \ Distribution	**GED**	**skewt**	**t**	**Total**
HARCH	3	0	1	4
TARCH	7	6	3	16
APARCH	1	1	0	2
GJR-GARCH	2	3	1	6
GARCH	0	1	1	2
Total	13	11	6	30

According to Table 4, it appears that ETFs demonstrate a consistent preference for TARCH, GJR-GARCH, and HARCH models accompanied by the GED, skewt and t distribution. These models effectively account for the leverage effect and volatility clustering at different frequencies or over different time horizons, common features in ETFs. However, optimal models and their distributions shift as we transition to more volatile assets, indicating changing market dynamics and volatility patterns.

For low-volatility assets, HARCH and APARCH models emerge as preferred choices, owing to their flexibility in modeling volatility processes. As volatility increases, the optimal models shift towards TARCH and HARCH, indicating prevalent asymmetric volatility and possible increases in heavy tails in the return distribution. When we reach the most volatile assets, TARCH and GJR-GARCH models become dominant, reflecting increased complexity in the volatility process. The appearance of the skewt and t distributions suggests the presence of skewness and heavy tails in the return distribution, usually associated with higher volatility levels. This can be overseen more clearly in Table 17 in Appendix 8.5.

As ETFs are marketable securities, their prices change throughout the day as they are traded. They often demonstrate a leverage effect, where negative returns increase future volatility more than positive returns of the same magnitude. The TARCH and GJR-GARCH models, designed to capture this effect, are therefore well suited for ETFs. As the volatility of the assets increases, there seems to be a shift in the optimal model towards those that can better capture asymmetry in the volatility process. The distribution type preference also changes towards those capable of modeling heavier tails and skewness, reflecting the changing characteristics of the assets' return distributions.

Stocks

Table 5: Stocks: All top three models' combined metrics and parameters.

Asset	Model	**C**	μ	ω	$\alpha[1]$	$\gamma[1]$	$\beta[1]$	δ	ϕ	d
JNJ	TARCH GED	0.10	0.048	0.057	0.031	0.127	0.874	-	-	-
	TARCH skewt	0.12	0.048	0.054	0.023	0.149	0.875	-	-	-
	TARCH t	0.15	0.060	0.053	0.026	0.147	0.874	-	-	-
MCD	TARCH GED	0.11	0.071	0.079	0.083	0.127	0.816	-	-	-
	TARCH skewt	0.14	0.055	0.065	0.074	0.130	0.838	-	-	-
	GJR-GARCH GED	0.16	0.074	0.087	0.037	0.177	0.802	-	-	-
MDLZ	GJR-GARCH GED	0.11	0.046	0.157	0.024	0.208	0.787	-	-	-
	TARCH GED	0.11	0.049	0.119	0.059	0.152	0.808	-	-	-
	GJR-GARCH skewt	0.17	0.029	0.119	0.022	0.220	0.823	-	-	-
MSFT	GJR-GARCH GED	0.13	0.092	0.179	0.080	0.186	0.773	-	-	-
	TARCH GED	0.14	0.086	0.102	0.089	0.133	0.822	-	-	-
	GARCH GED	0.21	0.106	0.183	0.181	-	0.759	-	-	-
AMZN	GJR-GARCH GED	0.18	0.129	0.289	0.070	0.195	0.768	-	-	-
	GJR-GARCH skewt	0.22	0.134	0.279	0.052	0.227	0.780	-	-	-
	GJR-GARCH t	0.22	0.135	0.279	0.052	0.227	0.780	-	-	-
BABA	GJR-GARCH GED	0.14	0.083	0.242	0.066	0	0.882	-	-	-
	GJR-GARCH t	0.19	0.101	0.177	0.069	0	0.895	-	-	-
	TARCH GED	0.19	0.063	0.093	0.029	0.061	0.912	-	-	-
TSLA	HARCH GED	0.15	0.154	9.400	0.180	-	-	-	-	-
	FIGARCH GED	0.19	0.127	1.804	-	-	0.178	-	0.049	0.213
	TARCH GED	0.21	0.119	0.060	0.060	0.009	0.935	-	-	-
TUP	GJR-GARCH GED	0.13	0.045	0	0	0.017	0.989	-	-	-
	FIGARCH GED	0.14	0.049	0.353	-	-	0.479	-	0.374	0.251
	EGARCH GED	0.15	0.049	0.013	0.086	-	0.998	-	-	-
GME	GARCH GED	0.12	0.081	0.002	0.009	-	0.991	-	-	-
	FIGARCH GED	0.14	0.074	0.786	-	-	0.564	-	0.361	0.266
	EGARCH GED	0.15	0.085	0.005	0.046	-	1.000	-	-	-
FCEL	TARCH GED	0.12	-0.189	0.391	0.247	-0.037	0.772	-	-	-
	APARCH GED	0.12	-0.182	0.581	0.236	-	0.764	1.230	-	-
	FIGARCH t	0.14	-0.261	3.847	-	-	0.066	-	0	0.301

Firstly, while looking at Table 5, we see that the low volatility assets comprising JNJ, MCD, and MDLZ showed distinct preferences. For JNJ, the TARCH GED model stood out, accurately depicting its low average return and long-term volatility, among other metrics. The TARCH skewt and TARCH t models also made significant contributions to understanding JNJ's volatility, yet the GED prevailed. A similar pattern was observed with MCD, with the TARCH GED model providing an optimal fit. In contrast, MDLZ best matched with both the GJR-GARCH and TARCH

GED model, as Figure 2b failed to crown either one superior, exhibiting their unique low average return and high level of long-term volatility.

The medium volatility asset group showed varied responses. The GJR-GARCH GED model performed optimally for MSFT, AMZN, and BABA, highlighting unique aspects of each stock's volatility. For AMZN it becomes clear that the difference in distribution parameters aids to its significance, as shown in Table 9. TSLA, however, found its best fit with the HARCH GED model due to its high average return and long-term volatility. It also had great consistency and accuracy in the forecasts, as Figure 1b suggests.

Finally, the high volatility asset group, including TUP, GME, and FCEL, showcased unique responses. TUP and GME responded best to the GJR-GARCH GED model and the GARCH GED model respectively, while FCEL matched optimally with both the TARCH and APARCH GED model. GME it's VaR had an interesting Figure 1c, due to its high volatility at the beginning of 2021, however, the VaR of GARCH GED adapted better in the long run. We were able to crown an optimal model for FCEL by looking at Figure 2c, as the TARCH GED model proved significantly better at forecasting.

As demonstrated, TARCH and GJR-GARCH models, paired with a GED distribution especially, but also with a skewt, or t distribution, showed consistent performance across these stocks. The choice of model varied with each stock and their specific needs, highlighting the importance of flexibility in model selection. Table 6 further illustrates this analysis. It also reaffirms that stocks' optimal model selection can shift according to their respective volatility levels.

Table 6: Count of each distinct model appearing in the top 3 of the stock asset group.

Model \ Distribution	**GED**	**skewt**	**t**	**Total**
TARCH	7	2	1	10
GJR-GARCH	6	2	2	10
HARCH	1	0	0	1
FIGARCH	3	0	1	4
EGARCH	2	0	0	2
GARCH	2	0	0	2
APARCH	1	0	0	1
Total	22	4	4	30

For less volatile stocks, TARCH and GJR-GARCH models are prominent. The GARCH variants adeptly account for volatility clustering and the leverage effect, suggesting increased volatility after negative returns for these stocks. For moderately volatile stocks the GJR-GARCH model paired with a GED becomes more prevalent. The volatility of high-stakes stocks, brings the HARCH, FIGARCH, and EGARCH models to the forefront, underlining these models' proficiency in handling persistent high volatility and larger shocks. As we traverse from less volatile to more volatile stocks, there's an observable shift in the optimal forecasting model from TARCH and GJR-GARCH to HARCH, FIGARCH, and EGARCH. This shift reflects the increased importance of modeling substantial shocks and high volatility persistence in highly volatile stocks.

More volatile stocks likely exhibit greater skewness and heavier tails in their return distributions, prompting a shift in the preferred distribution. This can be overseen more clearly in Table 18 in Appendix 8.5.

The increase in volatility from stocks like JNJ to FCEL also sees an elevation in the complexity and flexibility of the optimal model. This shift indicates that as assets become more volatile, their return and volatility dynamics become more intricate, necessitating more flexible models and distributions.

In summary, more volatile assets require complex and adaptable models and distributions due to their intricate return and volatility dynamics. These findings underline the importance of model and distribution customization to capture these dynamics accurately, emphasizing the need for adaptable modeling techniques in financial forecasting.

Cryptocurrencies

While looking at Table 7, we analyze the low volatility assets, BTC, LTC, and XMR. The EGARCH GED proves optimal for BTC, indicating a small positive average return and relatively low long-term volatility. It also demonstrates BTC's moderate response to immediate shocks, which persists over significant periods. The TARCH GED model is optimal for LTC, showing a small negative average return and relatively low long-term volatility, with the unique attribute of negative shocks slightly decreasing LTC's volatility. Similarly, XMR fits best with the EGARCH GED model, indicating a small positive average return and low long-term volatility.

The next category, comprising medium volatility cryptocurrencies like XRP, NMC, and DOGE, also employs specific optimal models. XRP is best modeled by the GARCH GED model, suggesting a negative average return and moderate long-term volatility, with a moderate response to immediate shocks. NMC is aptly represented by the GJR-GARCH GED model, implying a negative average return and high long-term volatility. DOGE's volatility is optimally expressed through the EGARCH GED model, as can be seen in Figure 1d, hinting at a negative average return and low long-term volatility. All the chosen models adeptly capture the specific dynamics of each cryptocurrency's volatility.

Lastly, the high volatility category, including FTC, BLK, GRC, and XPM, aligns with unique optimal models. In the case of FTC, APARCH GED demonstrates superior performance, despite the identical results shown by FIGARCH GED. This superiority of APARCH GED, as depicted in Figure 2e, is attributed to its ability to accommodate larger shocks by integrating the leverage effect. Moreover, the inclusion of a power term for the lagged variance term in the APARCH GED model allows it to capture a broader spectrum of ARCH behaviors, which enhances the model's persistence, making it highly effective in forecasting volatility. The TARCH GED model fits BLK optimally, illustrating its response to shocks and high persistence in volatility, while indicating a unique negative leverage effect. GRC's volatility is best described by the GARCH model with a t distribution, showcasing a strong immediate response to shocks and high volatility persistence. XPM, renowned for its high volatility, is best represented by the EGARCH GED model, signifying

Table 7: Cryptocurrencies: All top three models' combined metrics and parameters.

Asset	Model	**C**	μ	ω	$\alpha[1]$	$\gamma[1]$	$\beta[1]$	δ	ϕ	d
BTC	EGARCH GED	0.11	0.163	0.126	0.287	-	0.959	-	-	-
	TARCH GED	0.12	0.160	0.162	0.156	0.006	0.841	-	-	-
	FIGARCH GED	0.12	0.170	0.272	-	-	0.552	-	0.227	0.545
LTC	TARCH GED	0.12	-0.078	0.178	0.135	-0.007	0.868	-	-	-
	APARCH GED	0.13	-0.088	0.212	0.130	-	0.870	1.167	-	-
	EGARCH GED	0.13	-0.078	0.103	0.222	-	0.973	-	-	-
XMR	EGARCH GED	0.17	0.093	0.181	0.264	-	0.955	-	-	-
	TARCH GED	0.21	0.135	0.254	0.156	-0.053	0.866	-	-	-
	APARCH GED	0.21	0.068	1.149	0.136	-	0.841	1.813	-	-
XRP	GARCH GED	0.10	-0.168	2.430	0.325	-	0.636	-	-	-
	GJR-GARCH GED	0.10	-0.166	2.448	0.330	-0.007	0.635	-	-	-
	FIGARCH GED	0.11	-0.184	1.951	-	-	0.176	-	0.261	0.479
NMC	GJR-GARCH GED	0.11	-0.205	2.386	0.186	0.116	0.727	-	-	-
	TARCH GED	0.13	-0.198	0.469	0.224	0.041	0.755	-	-	-
	FIGARCH GED	0.14	-0.176	2.956	$-$	$-$	0.393	-	0.186	0.628
DOGE	EGARCH GED	0.18	-0.204	0.174	0.383	-	0.955	-	-	-
	TARCH GED	0.22	-0.199	0.306	0.239	-0.030	0.776	-	-	-
	TARCH t	0.22	-0.202	0.378	0.251	0.010	0.744	-	-	-
FTC	APARCH GED	0.14	-0.243	0.626	0.284	-	0.698	0.814	-	-
	FIGARCH GED	0.14	-0.310	8.686	-	-	0.046	-	0.054	0.513
	TARCH GED	0.14	-0.242	1.037	0.332	-0.042	0.670	-	-	-
BLK	TARCH GED	0.12	-0.132	0.422	0.177	-0.009	0.828	-	-	-
	GJR-GARCH GED	0.13	-0.189	2.712	0.138	0.138	0.789	-	-	-
	APARCH GED	0.13	-0.141	0.636	0.176	-	0.824	1.226	-	-
GRC	GARCH t	0.11	-0.004	36.197	0.703	0.297	-	-	-	-
	GJR-GARCH t	0.12	0.001	36.328	0.728	0.296	-0.048	-	-	-
	GARCH skewt	0.13	0.212	37.257	0.709	0.291	-	-	-	-
XPM	EGARCH GED	0.13	-0.322	0.390	0.416	0.915	-	-	-	-
	GJR-GARCH GED	0.15	-0.374	4.375	0.202	0.712	0.173	-	-	-
	GARCH GED	0.15	-0.306	4.946	0.290	0.694	-	-	-	-

a strong response to shocks and high volatility persistence. Even though it overestimates the VaR at moments of high volatility, it adjusts better than the other models, as can be seen in Figure 1e. In all these cases (except GRC), the GED models manage to encapsulate the unique volatility characteristics of these cryptocurrencies, while offering optimal performance across metrics.

Table 8 highlights the importance of the TARCH, GJR-GARCH, EGARCH, GARCH, APARCH, and FIGARCH models, with the GED distribution being by far the most prevalent.

Notably, cryptocurrencies with higher volatility demonstrate a greater reliance on models such as EGARCH, FIGARCH, and TARCH. This could be seen by lower values for these models in Table 19 in Appendix 8.5. On the other hand, cryptocurrencies with relatively lower volatility, primarily prefer GARCH, EGARCH, and TARCH frameworks. This may suggest a balance between capturing basic GARCH effects and the leverage effect. However, the optimal model can change along with increasing volatility, indicating the need for flexible modeling techniques capable of

Table 8: Count of each distinct model appearing in the top 3 of the cryptocurrency asset group.

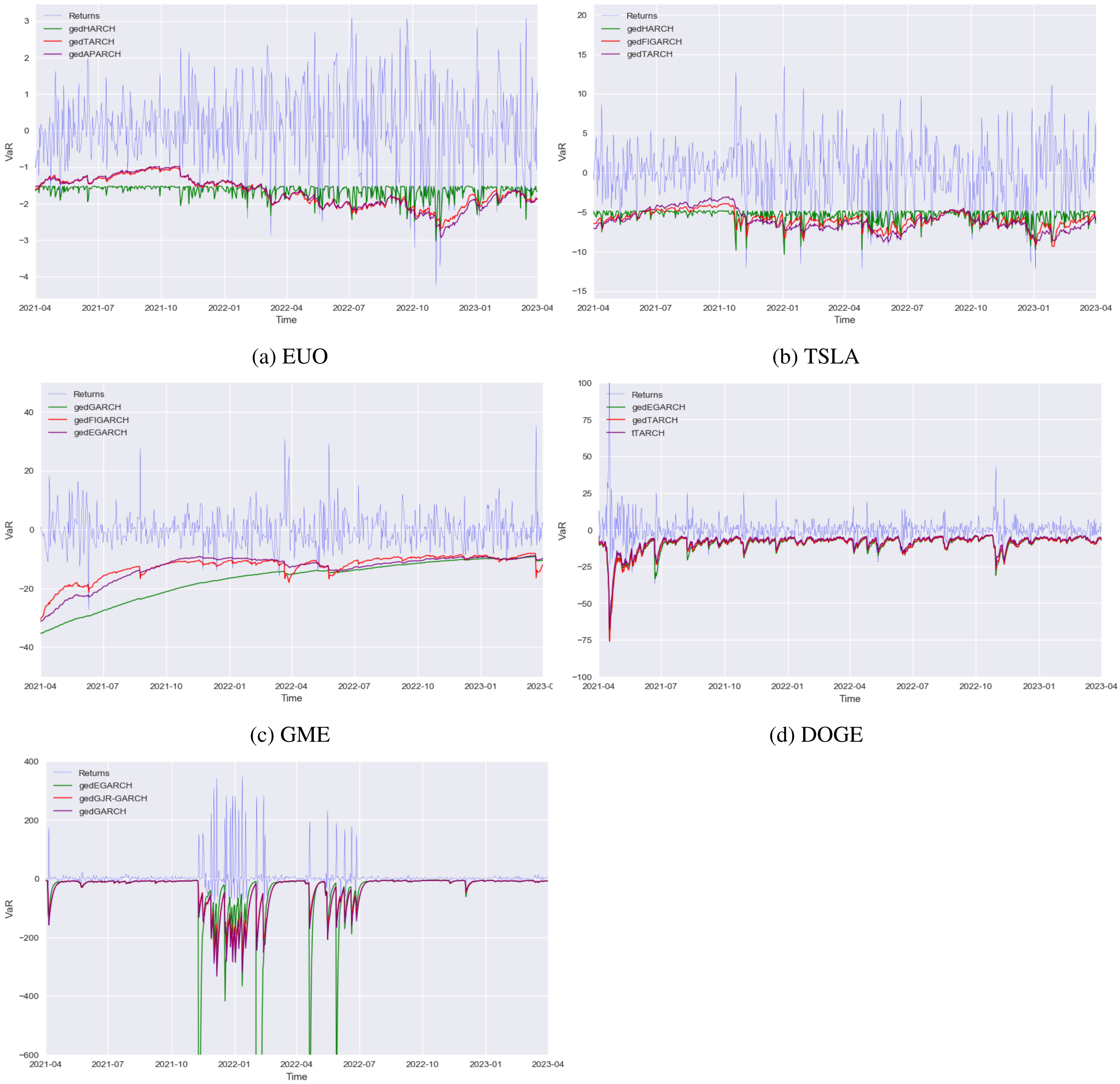
Model \ Distribution	**GED**	**skewt**	**t**	**Total**
TARCH	7	0	1	8
GJR-GARCH	5	0	0	5
EGARCH	4	0	1	5
GARCH	2	1	1	4
APARCH	4	0	0	4
FIGARCH	4	0	0	4
Total	26	1	3	30

accurately capturing the unique return and volatility dynamics of cryptocurrencies.

In summary, as volatility increases, the optimal models and their distributions for cryptocurrencies tend to exhibit increased complexity and flexibility. This progression indicates the increasing prevalence of extreme values, asymmetry, and long memory in volatility, emphasizing the unique market dynamics of these digital assets. Therefore, just as with traditional financial assets, these more complex return and volatility dynamics necessitate more flexible models and distributions to capture these dynamics accurately.

Table 9: Distribution parameters of the top three models for the ETFS: QQQ, BNO, ERX, and stocks: JNJ and AMZN.

Asset	Model	ν	η	λ
QQQ	TARCH skewt	-	5.443	-0.219
	TARCH GED	1.230	-	-
	TARCH t	-	4.887	-
BNO	TARCH skewt	-	6.479	-0.079
	TARCH GED	1.381	-	-
	TARCH t	-	6.421	-
ERX	TARCH skewt	-	7.806	-0.073
	TARCH GED	1.431	-	-
	TARCH t	-	7.591	-
JNJ	TARCH GED	1.168	-	-
	TARCH skewt	-	4.745	-0.043
	TARCH t	-	4.694	-
AMZN	GJR-GARCH GED	1.078	-	-
	GJR-GARCH skewt	-	3.946	-0.002
	GJR-GARCH t	-	3.946	-



(a) EUO

(b) TSLA

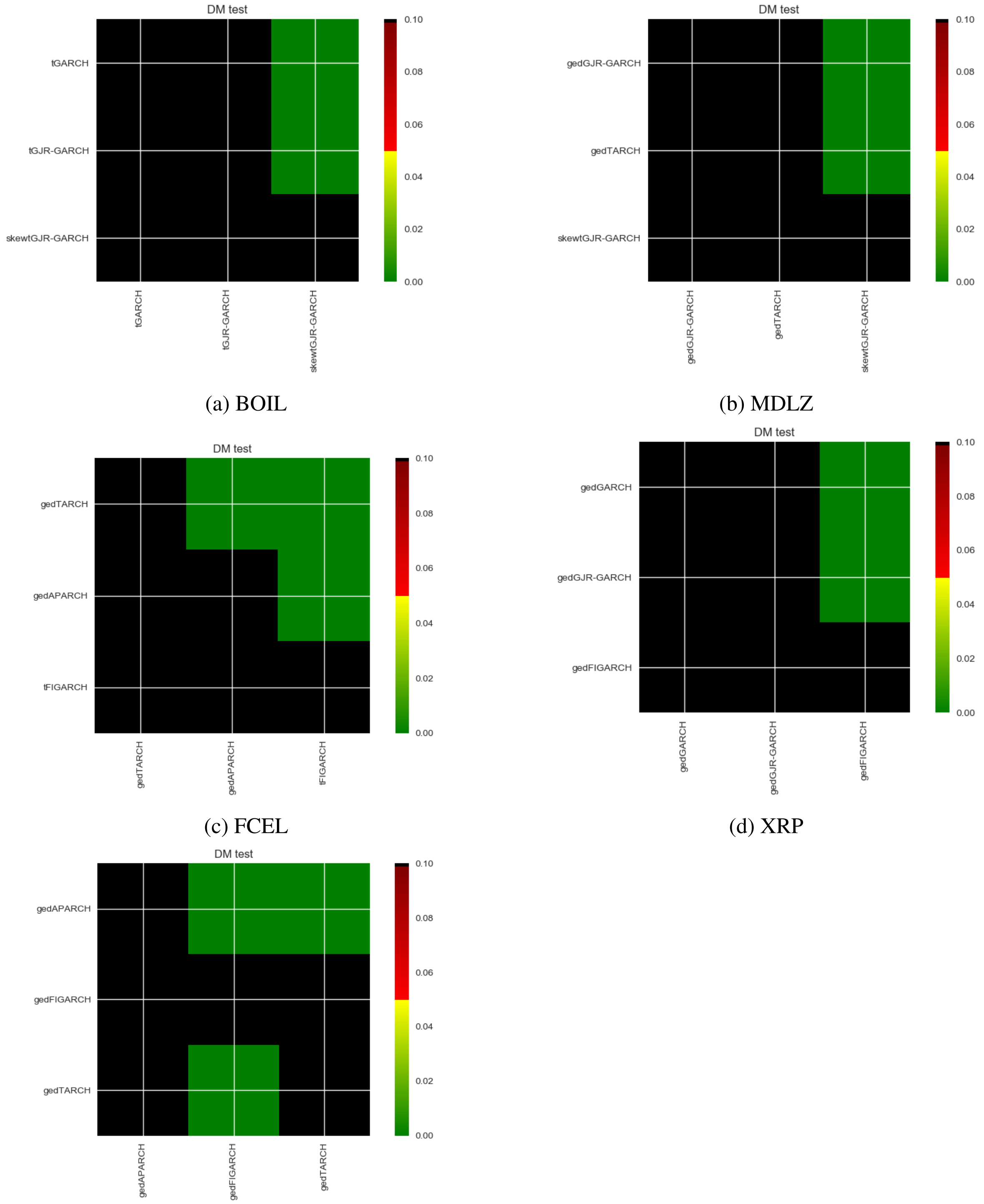
(c) GME

(d) DOGE

(e) XPM

Figure 1: The returns and its top three models' VaR for the ETF: EUO, the stocks: TSLA and GME, and the cryptocurrencies: DOGE and XPM.

When it comes to multivariate analysis, the interpretation of the DM test results can be visualized in a heatmap, as shown in Figure 2. The closer the p-values are to zero (dark green), the more significant the difference is between the forecasts of a model on the x-axis (worse) and the forecasts of a model on the y-axis (better). Black color indicates p-values above the color map limit, i.e., p-values larger than or equal to 0.10. Therefore, green indicates that the model on the x-axis probably performs worse, and anything darker than yellow is inconclusive (Jesus Lago, 2021).



(a) BOIL

(b) MDLZ

(c) FCEL

(d) XRP

(e) FTC

Figure 2: The Diebold-Mariano test for the top three models for the ETF: BOIL, the stocks: MDLZ and FCEL, and the cryptocurrencies: XRP and FTC.

Table 10: KS stats and p-values for the PIT method for all assets their optimal models.

ETF			Stock			Cryptocurrency		
Model	**KS**	**p-value**	**Model**	**KS**	**p-value**	**Model**	**KS**	**p-value**
H GED	0.015	0.60	T GED	0.017	0.43	E GED	0.213	0.00
T skewt	0.059	0.00	T GED	0.020	0.28	T GED	0.253	0.00
AP skewt	0.050	0.00	GJR GED	0.043	0.00	E GED	0.302	0.00
T skewt	0.047	0.00	GJR GED	0.102	0.00	G GED	0.255	0.00
T skewt	0.112	0.00	GJR GED	0.130	0.00	GJR GED	0.269	0.00
H GED	0.113	0.00	GJR GED	0.171	0.00	E GED	0.266	0.00
T skewt	0.150	0.00	H GED	0.231	0.00	AP GED	0.341	0.00
T skewt	0.258	0.00	GJR GED	0.176	0.00	T GED	0.325	0.00
T skewt	0.279	0.00	G GED	0.235	0.00	G t	0.349	0.00
G t	0.319	0.00	T GED	0.299	0.00	E GED	0.317	0.00

The abbreviations, H, T, AP, G, GJR and E stand for the models: HARCH, TARCH, APARCH, GARCH, and EGARCH, respectively.

Overall

Choosing the best forecasting model is highly contingent on the individual traits of each asset, volatility being one of those key factors. It is evident that as asset volatility increases, there is a noticeable shift in favor of models that can better accommodate asymmetry, leverage effects, and heavy tails in the return distribution. It is, therefore, imperative to align the choice of forecasting model with each asset's unique features and dynamics.

Despite studying a range of assets, no single model continually outperforms others. However, models such as TARCH (skewt or GED) and GJR-GARCH (GED) show broad applicability across all asset types. Conversely, HARCH (GED) effectively applies to select ETFs and stocks, while EGARCH (GED) performs well with cryptocurrencies and stocks with high volatility. This variability underscores the fact that different assets possess distinct volatility characteristics, each requiring a unique model for an accurate depiction.

Increasing the volatility value of an asset generally indicates an escalated risk associated with it. While this affects the model selection, with certain models performing better with high-volatility data, it does not necessarily dictate a hierarchy of model efficiency. But still, why does the shift in preferred models occur as volatility increases?

The increased complexity in return and volatility dynamics seen in high-volatility assets explains the shift towards more flexible models. Medium to high volatility assets may display stronger leverage effects, higher volatility persistence, and more extreme shocks, requiring models that can handle these complexities, like TARCH, GJR-GARCH, or EGARCH. For example, the EGARCH model allows for both asymmetry and leverage in the volatility process, making it apt for high-volatility assets as these characteristics become more prominent. Moreover, the EGARCH model allows the conditional variance to respond asymmetrically to past shocks. As volatility increases even more, models like EGARCH and FIGARCH that capture long-memory effects in volatility pose to be more relevant. Yet, overall, TARCH and GJR-GARCH keep on being great performers.

Interestingly, the findings illustrate that GARCH sometimes emerges as the optimal model for assets with high volatility. This is exemplified in the case of GME, which had a brief period of intense volatility influenced by social media-driven trading activity. However, this event didn't significantly impact GME's data for the majority of the time, just majorly its historic volatility. When forecasting volatility, it's crucial to analyze the entire period's returns graph, not just the backtesting period. As seen in Figure 1c, GARCH, despite initial inaccuracies, outperformed other models in the long run, possibly due to its lack of reliance on the additional parameters that more complex models use. This is why in this rare case, higher volatility does not correspond with a more complex model.

With regard to the combined metric utilized in our study, it appears to hold up robustly, despite its unconventional nature. In situations where one model marginally outperforms others, the use of this absolute combined metric as a determinant seems to introduce no error. This assertion is reinforced by the fact that the DM test never identifies any other model than the optimal model as significantly superior in terms of forecasting, even considering potential biases introduced by the combined metric towards certain used measures. While this does not suggest that there is no scope for further optimization, it confirms the metric's consistent accuracy and utility when comparing various datasets and models.

Lastly, our statistical evaluation using the KS test, as detailed in Table 10, demonstrated that many of our top models produced low p-values. This often suggests a potential misalignment between the data generated by these models and the theoretical distribution we assumed. This divergence could indicate that the predictions made by the model may not accurately reflect the distribution of real-world data. As discussed in section 4.3.5, all models rejected the assumption of normality, which aligns with the fact that no optimal models utilized the normal distribution. Interestingly, we observed a trend where the p-value tends to decrease and the KS statistic increases as the asset's volatility rises. This pattern is consistent across all asset groups, models, and distributions, highlighting the inherent complexity of encapsulating volatility forecasting into a distribution, especially as volatility increases. This trend supports the hypothesis that higher volatility significantly impacts distribution preferences. It's important to note that while low p-values are often interpreted as evidence against the null hypothesis, they do not definitively indicate model inadequacy. They could also suggest the presence of unexplained variance or complex patterns in the data that the model does not capture. Despite these considerations, these findings call for further exploration to better understand the causes of this deviation and how it might influence the performance of our models.

6 Conclusion

This research addressed critical questions in volatility forecasting across a diverse range of assets, examining their unique characteristics and volatility dynamics. The study's aim was to identify the most effective forecasting method for different asset types, discern trends in the selection of volatility forecasting models among distinct asset groups, and understand how these trends relate to the respective assets' characteristics and volatility.

In terms of the selection of an optimal forecasting model, the study underscored that it is heavily reliant on the unique characteristics of each asset, particularly its volatility level. As asset volatility increases, a distinct shift was noted toward models that accommodate asymmetry, leverage effects, and heavy tails in the return distribution. This demonstrated the criticality of aligning the choice of forecasting model with each asset's distinct features and dynamics.

The research did not pinpoint a single model that persistently outperforms others across the studied assets, which included ETFs, stocks, and cryptocurrencies. However, TARCH (skewt or GED) and GJR-GARCH (GED) models displayed wide applicability across different asset types. Conversely, HARCH (GED) proved effective for certain ETFs and stocks, while EGARCH (GED) was optimal for cryptocurrencies. This variability emphasizes the distinctive volatility characteristics inherent to different assets, necessitating a unique model for accurate representation. When examining distribution preferences, it becomes evident that the GED distribution consistently outperforms across all asset categories, with particularly strong results in the realm of cryptocurrencies. The skewt distribution also shows promising performance for ETFs, almost matching the efficacy of the GED distribution. However, its performance declines considerably when applied to stocks and cryptocurrencies, in that sequence. Similarly, the t distribution displays an even poorer performance compared to the skewt distribution, this in particular does not align with our previous small-scale research. It's important to note that the normal distribution doesn't find favor with any high-performing models, suggesting that it probably is not an optimal choice for modeling the return distributions of these asset categories.

The escalating intricacies in return and volatility dynamics in high-volatility assets validate the progression toward models with greater flexibility and complexity. Notably, the EGARCH model, which accommodates asymmetry and leverage effects in the volatility process, becomes a prominent choice. In the same vein, FIGARCH, arguably the most complex model in the set, emerges more frequently as the optimal model in correspondence with rising asset volatility. These models aptly address the nuanced patterns exhibited by high-volatility assets. There is a clear shrinkage in the range of optimal models as volatility increases, indicating a more pronounced need for these flexible models in high-volatility scenarios.

The findings of this research provide a significant contribution to the existing body of literature by shedding light on the intricate relationship between asset-specific characteristics, primarily volatility, and the selection of forecasting models. It has improved the understanding of volatility forecasting, providing valuable insights for academia and professionals in the financial market.

In conclusion, while the research identifies a narrow range of potentially optimal volatility forecasting models for a given asset using only its historical volatility, the practical implementation of this goal is complex and reliant on various factors. The study emphasizes the importance of understanding an asset's specific characteristics, the context of the modeling task, and the state of the art in financial modeling for the successful application of these models. This study has significantly enriched the existing literature, offering a deeper and more nuanced understanding of volatility forecasting models and their applicability across different asset types.

7 Discussion

Our extensive analysis reveals the complex interplay between model optimality and asset volatility. Interestingly, as volatility increases, the range of optimal models contracts. However, this relationship is far from straightforward, suggesting the need for additional nuanced analysis. Also, the impact of heightened volatility varies across asset classes and individual assets. Even after examining 30 different assets, a clear-cut link between asset volatility and model effectiveness remains elusive, indicating the possibility of multifaceted or non-linear correlations that merit further exploration.

Considering the multifaceted nature of asset market dynamics, adopting a wider variety of models and distributions may enhance our analysis. Expanding the dataset to cover a broader range of assets could potentially uncover subtle interrelationships not immediately apparent in smaller datasets. Despite the lack of a definitive link between model optimality and volatility in our current analysis, the intricacies inherent in financial market forecasting open up a rich field for more detailed investigations.

7.1 Prospects for Future Research

Our study has made considerable headway in discerning the relationship between volatility levels and optimal forecasting models across a broad array of financial assets, including stocks, ETFs, and cryptocurrencies. Despite the complexity introduced by the unique characteristics and market dynamics of each asset type, our expansive approach yielded comprehensive insights and opened up exciting prospects for future research.

Future investigations could strive to pinpoint the optimal degree of skewness and tail length in the error distributions for specific assets. This notion stems from our observation where, in certain scenarios, the Student's t-distribution outperformed the skewed-t distribution, despite the latter being a generalized version of the former. In these cases, introducing skewness did not improve, and sometimes even degraded, the model's performance. Meanwhile, the GED distribution consistently yielded strong results. Understanding this optimality of skewness and tail length could enhance the precision of volatility forecasting models.

Alongside the former prospect, future research should explore potential sources of the low p-values for the KS test observed in our study, perhaps due to data complexity or model assumptions. It's recommended to investigate other statistical tests, metrics, or different model architectures to better fit our data. Exploring techniques to handle model uncertainty and variability, such as bootstrapping, Bayesian methods, or ensemble modeling, could also be valuable. Ultimately, these efforts aim to enhance our model's predictive accuracy and reliability, providing robust insights for decision-making.

Further, the models' performance could be analyzed over varying forecast horizons. While our study was primarily confined to daily forecasts, venturing into 1-week-ahead or 1-month-ahead forecasts could yield insights pertinent to diverse applications and decision-making contexts.

A more narrow focus on individual asset classes could also be beneficial. This approach could unearth more nuanced findings specific to each asset type, potentially revealing unique influences on the relationship between volatility and optimal forecasting models. Furthermore, dissecting broad asset classes into subcategories might yield more granular insights into how volatility interacts with specific asset groups.

Another potential area for exploration involves enhancing the model selection and evaluation process. Future research could encompass newer or less commonly used models and distributions, potentially offering better fits for specific asset types. Our experience with the GAS model, although challenging due to inconsistencies (as can be seen in Appendix 8.5), indicates the potential for future research to resolve these irregularities, expanding the range of situations where this model could be applicable.

The combined metric, our composite measure used to compare various models, also presents an area for refinement. While useful and seemingly robust, this measure may not fully encapsulate the subtleties of model performance. Some areas with doubt are the VaR and ES metrics, as they are not optimally categorized as loss functions. Future research could delve into optimizing this approach by exploring different combinations of metrics, reassessing the inclusion or exclusion of specific metrics, and potentially applying different weightings or standardizations based on relevance.

Finally, the relationship between volatility and optimal model selection could be further elucidated with the analysis of larger datasets. The rapid expansion of financial data provides an opportunity for future research to explore this interplay on a larger scale, reducing the impact of outliers such as GME and potentially uncovering patterns that may not be discernible in smaller datasets.

In conclusion, our study, while significant, marks just the beginning of the journey to understanding volatility forecasting. The promise for future research in this area is immense and the potential findings will likely expand our understanding of this complex domain, leading to more effective risk management strategies and investment decisions.

8 Appendix

Click on the following section numbers to go back to the data descriptives 3 or to the results 5.

8.1 Appendix A - Full descriptive statistics

Table 11: ETFs: Descriptive Statistics

	Count	Mean	Std	Min	25%	50%	75%	Max
EUO	2520	0.020	1.016	-6.292	-0.578	0.000	0.621	5.184
ˆDJI	2520	0.039	1.102	-12.927	-0.378	0.063	0.539	11.365
VO	2520	0.046	1.188	-12.389	-0.448	0.098	0.635	10.884
QQQ	2520	0.073	1.342	-11.979	-0.467	0.121	0.742	8.471
DBE	2520	0.006	1.745	-14.048	-0.870	0.089	0.950	8.395
UBS	2520	0.044	2.030	-22.457	-0.845	0.021	0.962	32.693
BNO	2520	0.012	2.301	-21.913	-1.087	0.064	1.168	18.898
UTSL	2520	0.106	4.046	-34.846	-1.665	0.181	1.928	40.665
ERX	2520	0.031	4.674	-60.105	-2.184	0.092	2.385	40.333
BOIL	2520	-0.146	5.771	-38.438	-3.196	-0.251	2.865	38.892

First, we consider the ProShares UltraShort Euro (EUO), characterized by a relatively low volatility of 1.02. The volatility of EUO is largely shaped by factors such as macroeconomic indicators, monetary policies, and market sentiment toward the European and U.S. economies. Following EUO, we have the Dow Jones Industrial Average (ˆDJI) with a volatility of 1.10, and the Vanguard Mid-Cap ETF (VO) with a volatility of 1.19. Both are influenced by economic conditions, including GDP growth, unemployment rates, inflation data, and Federal Reserve policy shifts. Moving into the realm of sector-specific ETFs, we look at the Invesco QQQ Trust (QQQ). With a volatility of 1.34, QQQ is particularly sensitive to sector-specific developments and broader market trends, especially in the technology sector. Simultaneously, the Invesco DB Energy Fund (DBE), with a volatility level of 1.75, reflects the dynamics of global energy markets and broader economic conditions. Next, we delve into multi-asset representation with the UBS ETRACS ETNs, carrying a moderate volatility of 2.03 and encapsulating various asset classes with an added layer of credit risk. The United States Brent Oil Fund (BNO), tracking Brent crude oil, has a volatility of 2.30, making it responsive to global events and conditions in the oil market. Taking a step towards high-volatility instruments, we consider the Direxion Daily Utilities Bull 3X Shares (UTSL) and the Direxion Daily Energy Bull 2X Shares (ERX). These leveraged ETFs carry volatilities of 4.05 and 4.67, respectively, making them inherently more volatile than their non-leveraged counterparts. They are influenced by sector-specific events and market conditions. Lastly, we include the ProShares Ultra Bloomberg Natural Gas ETF (BOIL) in our portfolio. With a volatility of 5.77, BOIL is primarily impacted by fluctuations in natural gas prices and the leveraged nature of the fund. These chosen ETFs, from the relatively low volatility of EUO to high volatility ones like

BOIL and ERX, collectively offer a comprehensive array of volatility levels for this study, helping to discern the trend of optimal volatility forecasting models across varying volatility levels.

Table 12: Stocks: Descriptive Statistics

	Count	Mean	Std	Min	25%	50%	75%	Max
JNJ	2520	0.045	1.124	-10.038	-0.476	0.040	0.606	7.998
MCD	2520	0.060	1.254	-15.875	-0.502	0.073	0.626	18.125
MDLZ	2520	0.051	1.350	-11.430	-0.606	0.054	0.708	11.281
MSFT	2520	0.114	1.724	-14.739	-0.664	0.083	0.983	14.217
AMZN	2520	0.103	2.070	-14.049	-0.860	0.107	1.129	14.131
BABA	2520	0.037	2.664	-13.342	-1.259	-0.032	1.300	36.764
TSLA	2520	0.233	3.633	-21.063	-1.527	0.135	1.923	24.395
TUP	2520	-0.017	4.649	-45.629	-1.259	0.016	1.138	67.705
GME	2520	0.251	6.729	-60.000	-1.757	0.000	1.664	134.836
FCEL	2520	0.084	7.304	-47.222	-3.101	-0.201	2.640	129.167

Starting with the realm of blue-chip stocks, we have Johnson & Johnson (JNJ) and McDonald's Corporation (MCD). These stable stalwarts, exhibiting relatively low volatility levels of 1.12 and 1.25 respectively, offer a solid baseline for our analysis and symbolize firms with a long-standing market presence. Johnson & Johnson, a global healthcare corporation, has shown consistent performance, while McDonald's, thanks to its expansive global presence and brand recognition, has maintained relative stability. In the category of moderate volatility, we explore the performances of tech giants Microsoft Corporation (MSFT) and Amazon Inc. (AMZN). Over the past decade, these companies have shown a moderate degree of volatility, with MSFT exhibiting a volatility of 1.72 and AMZN at 2.07. The inclusion of these firms acknowledges their robust market positions, consistent growth, and occasional periods of high volatility due to disruptive innovations. Alibaba Group Holding Limited (BABA), another significant player in the tech industry hailing from the emerging market of China, provides a broader context with a slightly higher volatility level of 2.66, reflecting the influences of both global and Chinese economies and their regulatory landscapes. For the high-volatility category, we turn our gaze to stocks known for their substantial price fluctuations. This category includes Tesla, Inc. (TSLA), Tupperware Brands Corporation (TUP), GameStop Corp. (GME), and FuelCell Energy, Inc. (FCEL). With volatilities ranging from 3.63 to 7.30, these firms represent diverse sectors. Tesla, renowned for its innovation in sustainable energy and automotive industries, has shown exceptional growth despite its high volatility. Tupperware, a premium food storage manufacturer, has experienced significant volatility due to industry shifts and its own financial performance. GameStop, which recently underwent substantial volatility due to the "Reddit rally" and other industry changes, and FuelCell Energy, a firm in the clean energy sector subject to significant volatility due to policy changes and renewable energy trends, round out this category.

Table 13: Cryptocurrencies: Descriptive Statistics.

	Count	Mean	Std Dev	Min	25%	Median	75%	Max
BTC	3627	0.229	4.045	-35.190	-1.363	0.178	1.847	33.256
LTC	3627	0.268	6.245	-42.144	-2.266	-0.080	2.176	93.254
XMR	3239	0.355	6.986	-43.178	-2.684	0.125	3.085	123.930
XRP	3526	0.378	7.526	-59.884	-2.219	-0.087	2.130	141.396
NMC	3626	0.308	8.323	-54.030	-2.570	-0.113	2.361	144.622
DOGE	3395	0.512	10.217	-60.955	-2.550	-0.227	2.158	338.901
FTC	3619	0.414	10.947	-69.575	-4.258	-0.320	3.696	166.421
BLK	3316	0.653	12.233	-57.098	-3.812	-0.197	3.600	163.884
GRC	2957	0.729	13.974	-78.067	-3.298	-0.024	3.277	363.808
XPM	3552	1.437	22.584	-78.519	-3.771	-0.289	3.116	345.584

At the lower end of the volatility scale, we've chosen to include Bitcoin (BTC, volatility 4.05), Litecoin (LTC, volatility 6.25), and Monero (XMR, volatility 6.99). Bitcoin, as the flagship cryptocurrency, offers a benchmark for our investigation due to its extensive trading history and widespread acceptance. In contrast, Litecoin and Monero represent the realm of alternative cryptocurrencies, each with unique technological frameworks and use cases. For example, Litecoin, often regarded as the silver to Bitcoin's gold, was designed to have faster transaction confirmation times than Bitcoin, which can lead to differences in volatility. Monero, on the other hand, is distinguished by its stringent privacy features, a factor that can influence its volatility in ways distinct from other cryptocurrencies. Moving up the volatility scale, our study incorporates cryptocurrencies like Ripple (XRP, volatility 7.53), Namecoin (NMC, volatility 8.32), and Dogecoin (DOGE, volatility 10.22). These digital currencies, noted for their higher volatility measurements, shed light on the nuances of forecasting models under varying degrees of volatility. Dogecoin, a cryptocurrency originally started as a joke, stands out for its extreme price swings often propelled by social media-driven hype, offering a compelling case study on the impacts of non-economic factors on volatility. At the extreme end of the volatility spectrum, we've included cryptocurrencies like Feathercoin (FTC, volatility 10.95), BlackCoin (BLK, volatility 12.23), Gridcoin (GRC, volatility 13.97), and Primecoin (XPM, volatility 22.58). These smaller market-cap cryptocurrencies, while less commonly traded, are characterized by high levels of volatility. Feathercoin, a derivative of Litecoin, exhibits volatility influenced by both the broader crypto market and the specific dynamics of smaller cryptocurrencies. BlackCoin, one of the first cryptocurrencies to use a pure Proof of Stake protocol, is sensitive to market sentiment, regulatory changes, and technological advancements. Gridcoin's volatility, while influenced by the broader crypto market, is also affected by the dynamics of the scientific research it rewards. Primecoin, with its unique proof-of-work system that involves searching for prime number chains, introduces a fascinating link between mathematical research and crypto volatility.

8.2 Appendix B - Assumptions

When working with time series data and specifically GARCH models, there are several critical considerations. Firstly, the data must be stationary, i.e., its statistical properties must remain constant over time. This stationarity is typically verified using unit root tests, including the Augmented Dickey-Fuller (ADF) or the Phillips-Perron (PP) tests. Moreover, the model's errors must be independent and identically distributed (i.i.d), a condition checked using statistical independence tests like the Ljung-Box (LB) test, Autocorrelation function (ACF), and the Durbin-Watson (DW) test. In addition, the error term structure of the GARCH model must conform to specific assumptions. For instance, in the GARCH(1,1) model, it's assumed that the error term arises from a mix of normal and i.i.d errors and past residuals, a condition that can be verified using the Jarque-Bera test. While the assumption of normality is common, it's noteworthy that even non-normal time series data can still be modeled and forecasted using GARCH models with appropriate modifications. Finally, the model must exhibit finiteness, or more specifically, it should have finite variance. This is often ensured by selecting a particular distribution for the model's errors, such as Gaussian or Student's t-distributions, each of which inherently possess finite variance. It's worth noting that, while these assumptions must be met in order to use the GARCH model, some of them can be relaxed to some extent by using more general versions of the model (Bollerslev, 1986).

In our case, the p-values for the Jarque-Bera test and the Ljung-Box test indicate that the residuals do not follow a normal distribution (rejecting null hypothesis of normality) and there is evidence of autocorrelation in the residuals, respectively. Furthermore, all the models seem to have been able to achieve stationarity based on the ADF and PP tests. Lastly, the DW statistics are all very close to 2, which suggests there is no first-order linear autocorrelation in the sample data.

8.3 Appendix C - Computation combined metric

Algorithm 1 Computation of the combined metric for each model and distribution

for each model m in vol_list **do**
  for each distribution d in dist_list **do**
    $Scale_{m,d} \leftarrow Scale_{m,d}^{-1}$ {We invert 'Scaling' to standardize it as a loss function}
    $VR_{m,d} \leftarrow |1 - VR_{m,d}|$ {We adjust VR to optimize its value when equal to 1}
    $M_{m,d} \leftarrow (M_{m,d} - \min(M_{m,d}))/(\max(M_{m,d}) - \min(M_{m,d}))$ {We normalize metrics to range [0, 1], for all metrics M}
    $C_{m,d} \leftarrow 0.1(AIC_{m,d} + BIC_{m,d} + ES_{m,d} + VR_{m,d} + VaR_{m,d} + Scale_{m,d} + MAE_{m,d} + RMSE_{m,d}) + 0.2 \cdot Roll_{m,d}$ {Compute combined metric C for all models pairs m, d}
  end for
end for
Sort C by $C_{m,d}$ in ascending order {This is an interpretation of $C.sort\_values(by = ['C'_{m,d}], ascending = True)$}
return C

8.4 Appendix D - Full performance metrics of top three models

Table 14: ETFs: All top three models' complete metrics.

Asset	Model	AIC	BIC	ES	VR	μVaR	Roll	Scaling	MAE	RMSE
EUO	HARCH GED	5311.62	5333.93	0.36	0.67	-1.65	0.63	2701.81	1.10	1.38
	TARCH GED	5277.48	5310.94	0.45	0.77	-1.67	0.72	2632.74	1.11	1.40
	APARCH GED	5275.37	5308.83	0.42	0.74	-1.67	0.72	2631.68	1.12	1.41
ˆDJI	TARCH skewt	4397.98	4437.02	0.34	0.28	-1.63	0.61	2191.99	0.97	1.55
	GJR-GARCH skewt	4414.33	4453.37	0.01	0.04	-1.61	0.61	2200.17	0.98	1.58
	TARCH GED	4423.47	4456.93	0.36	0.28	-1.57	0.61	2205.73	0.97	1.55
VO	APARCH skewt	4756.62	4795.66	0.17	0.42	-2.02	0.76	2371.31	1.01	1.58
	GARCH skewt	4755.20	4788.66	0.18	0.53	-2.02	0.77	2371.60	1.01	1.58
	GJR-GARCH GED	4776.19	4809.65	0.02	0.07	-1.91	0.76	2382.10	1.01	1.60
QQQ	TARCH skewt	5272.15	5311.19	0.51	0.11	-2.63	0.93	2629.08	1.16	1.71
	TARCH GED	5340.14	5373.60	0.39	0.35	-2.39	0.93	2664.07	1.14	1.68
	TARCH t	5317.97	5351.43	0.37	0.39	-2.34	0.93	2652.98	1.16	1.71
DBE	TARCH skewt	6833.68	6872.71	1.47	0.92	-2.90	1.15	3409.84	1.77	2.36
	HARCH GED	7121.23	7143.53	0.13	0.11	-2.78	1.14	3556.61	1.86	2.35
	TARCH GED	6865.59	6899.05	1.37	1.02	-2.82	1.15	3426.79	1.77	2.36
UBS	HARCH GED	7351.63	7373.93	0.20	0.11	-2.97	1.32	3671.81	2.02	2.81
	GJR-GARCH GED	7305.74	7339.20	0.77	0.39	-3.01	1.60	3646.87	2.01	2.79
	HARCH t	7283.00	7305.31	0.41	0.49	-2.69	1.32	3637.50	2.04	2.83
BNO	TARCH skewt	7775.32	7814.36	1.32	0.99	-3.46	1.40	3880.66	2.30	3.16
	TARCH GED	7799.04	7832.50	1.30	0.95	-3.45	1.40	3893.52	2.31	3.17
	TARCH t	7779.38	7812.84	1.45	0.99	-3.32	1.40	3883.69	2.31	3.17
UTSL	TARCH skewt	4506.33	4540.12	1.16	0.70	-5.57	2.22	2246.16	3.65	5.82
	GJR-GARCH skewt	4501.23	4535.02	1.05	0.70	-5.43	2.21	2243.61	3.64	5.87
	TARCH GED	4512.05	4541.01	1.23	0.74	-5.50	2.25	2250.02	3.65	5.84
ERX	TARCH skewt	10702.72	10741.75	1.61	0.63	-6.49	2.41	5344.36	4.79	6.77
	TARCH GED	10717.22	10750.68	1.61	0.60	-6.48	2.40	5352.61	4.80	6.79
	TARCH t	10705.85	10739.31	1.52	0.77	-6.27	2.40	5346.92	4.80	6.78
BOIL	GARCH t	11313.81	11341.70	3.30	0.81	-11.88	4.75	5651.91	5.60	7.07
	GJR-GARCH t	11315.81	11349.27	3.32	0.81	-11.88	4.75	5651.90	5.60	7.07
	GJR-GARCH skewt	11314.48	11353.52	2.97	1.02	-11.57	4.76	5650.24	5.60	7.07

Table 15: Stocks: All top three models' complete metrics.

Asset	Model	AIC	BIC	ES	VR	μVaR	Roll	Scaling	MAE	RMSE
JNJ	TARCH GED	5303.67	5337.13	0.48	0.53	-1.67	0.64	2645.84	1.14	1.55
	TARCH skewt	5261.38	5300.41	0.48	0.56	-1.62	0.62	2623.69	1.15	1.56
	TARCH t	5261.24	5294.70	0.50	0.60	-1.58	0.63	2624.62	1.15	1.56
MCD	TARCH GED	5391.20	5424.66	0.52	0.35	-1.73	0.62	2689.60	1.20	1.72
	TARCH skewt	5339.00	5378.04	0.50	0.46	-1.68	0.62	2662.50	1.21	1.73
	GJR-GARCH GED	5402.53	5435.99	0.31	0.32	-1.68	0.62	2695.27	1.20	1.75
MDLZ	GJR-GARCH GED	6159.57	6193.03	0.30	0.25	-1.95	0.68	3073.79	1.45	1.91
	TARCH GED	6157.40	6190.86	0.59	0.35	-1.94	0.68	3072.70	1.45	1.91
	GJR-GARCH skewt	6117.64	6156.67	0.37	0.35	-1.88	0.68	3051.82	1.49	1.96
MSFT	GJR-GARCH GED	6788.38	6821.84	0.27	0.18	-2.78	1.09	3388.19	1.69	2.31
	TARCH GED	6770.90	6804.36	0.60	0.39	-2.85	1.10	3379.45	1.68	2.27
	GARCH GED	6799.16	6827.04	0.46	0.67	-2.75	1.09	3394.58	1.70	2.32
AMZN	GJR-GARCH GED	7514.89	7548.35	0.75	0.18	-3.58	1.56	3751.44	2.02	2.67
	GJR-GARCH skewt	7452.45	7491.49	0.63	0.39	-3.35	1.56	3719.23	2.04	2.69
	GJR-GARCH t	7450.45	7483.91	0.63	0.39	-3.35	1.56	3719.23	2.04	2.69
BABA	GJR-GARCH GED	6678.30	6710.50	1.38	0.99	-4.82	2.41	3333.15	2.37	2.94
	GJR-GARCH t	6669.87	6702.07	1.24	1.16	-4.76	2.42	3328.94	2.38	2.95
	TARCH GED	6677.83	6710.03	1.62	1.06	-4.61	2.42	3332.91	2.37	2.94
TSLA	HARCH GED	9952.88	9975.19	0.88	0.88	-5.45	2.42	4972.44	3.76	4.68
	FIGARCH GED	9842.53	9875.99	1.20	1.09	-6.00	2.52	4915.26	3.68	4.70
	TARCH GED	9845.15	9878.61	1.64	0.88	-6.13	2.54	4916.58	3.68	4.69
TUP	GJR-GARCH GED	8539.10	8572.56	5.88	0.77	-8.28	4.67	4263.55	3.26	5.33
	FIGARCH GED	8552.01	8585.47	5.30	0.39	-8.10	3.71	4270.00	3.31	5.62
	EGARCH GED	8520.89	8548.77	5.70	0.85	-7.98	3.96	4255.44	3.30	5.50
GME	GARCH GED	9555.39	9583.27	6.51	0.18	-18.42	4.49	4772.69	3.67	5.02
	FIGARCH GED	9563.45	9596.91	2.18	0.32	-15.01	4.16	4775.72	3.70	5.06
	EGARCH GED	9536.81	9564.69	5.31	0.53	-14.56	4.04	4763.41	3.72	5.10
FCEL	TARCH GED	11863.57	11897.03	1.66	0.18	-10.36	3.31	5925.78	7.18	10.22
	APARCH GED	11864.23	11897.69	1.18	0.14	-10.23	3.32	5926.12	7.20	10.30
	FIGARCH t	11739.12	11772.58	0.42	0.04	-9.49	3.41	5863.56	7.33	10.70

Table 16: Cryptocurrencies: All top three models' complete metrics.

Asset	Model	AIC	BIC	ES	VR	μVaR	Roll	Scaling	MAE	RMSE
	EGARCH GED	14429.81	14459.50	2.04	0.65	-5.86	2.24	7209.90	4.16	5.55
BTC	TARCH GED	14433.68	14469.31	2.04	0.70	-5.76	2.22	7210.84	4.12	5.51
	FIGARCH GED	14427.68	14463.31	0.92	0.61	-5.64	2.16	7207.84	4.21	5.65
	TARCH GED	16206.60	16242.23	3.19	0.68	-8.35	3.05	8097.30	6.08	8.33
LTC	APARCH GED	16204.64	16240.26	3.11	0.68	-8.26	3.05	8096.32	6.09	8.36
	EGARCH GED	16191.84	16221.53	3.03	0.68	-8.52	3.09	8090.92	6.18	8.43
	EGARCH GED	15231.39	15260.34	3.52	0.44	-8.61	3.11	7610.70	7.45	10.00
XMR	TARCH GED	15236.90	15271.64	3.55	0.61	-8.54	3.14	7612.45	7.46	10.03
	APARCH GED	15226.97	15261.70	2.75	0.48	-8.62	3.12	7607.48	7.51	10.27
	GARCH GED	15863.74	15893.25	0.82	0.05	-9.11	3.72	7926.87	6.95	10.49
XRP	GJR-GARCH GED	15865.73	15901.13	0.90	0.05	-9.11	3.72	7926.86	6.95	10.50
	FIGARCH GED	15771.23	15806.64	0.85	0.06	-8.86	3.63	7879.61	6.96	10.63
	GJR-GARCH GED	17379.82	17415.44	1.08	0.07	-9.83	3.85	8683.91	7.93	11.51
NMC	TARCH GED	17378.06	17413.68	2.73	0.22	-9.65	3.83	8683.03	7.78	11.08
	FIGARCH GED	17368.35	17403.98	1.72	0.15	-9.86	3.79	8678.17	8.12	11.99
	EGARCH GED	15019.94	15049.19	2.06	0.34	-408.37	7.87	7504.97	6.98	10.56
DOGE	TARCH GED	15029.86	15064.97	3.31	0.44	-11.09	6.43	7508.93	6.82	10.37
	TARCH t	14948.02	14983.13	2.80	0.58	-9.76	6.38	7468.01	6.85	10.36
	APARCH GED	19371.42	19407.03	3.30	0.15	-15.17	5.46	9679.71	11.27	15.25
FTC	FIGARCH GED	19372.67	19408.29	3.01	0.14	-13.75	4.88	9680.34	11.59	16.25
	TARCH GED	19373.30	19408.91	3.26	0.15	-15.00	5.39	9680.65	11.32	15.46
	TARCH GED	17301.96	17336.89	2.93	0.32	-13.77	5.33	8644.98	11.47	16.65
BLK	GJR-GARCH GED	17312.57	17347.50	0.46	0.10	-13.73	5.27	8650.29	11.97	17.61
	APARCH GED	17296.41	17331.34	2.73	0.32	-13.62	5.28	8642.20	11.59	16.95
	GARCH t	14341.18	14369.50	0.42	0.02	-12.46	8.26	7165.59	12.50	23.18
GRC	GJR-GARCH t	14343.11	14377.10	0.88	0.05	-12.48	8.25	7165.56	12.53	23.31
	GARCH skewt	14337.30	14371.29	1.03	0.10	-11.83	8.32	7162.65	12.54	23.19
	EGARCH GED	18808.63	18838.18	2.46	0.24	-286.96	27.55	9399.31	10.99	0.13
XPM	GJR-GARCH GED	18809.31	18844.77	3.24	0.16	-33.94	27.52	9398.65	10.86	14.56
	GARCH GED	18818.46	18848.02	6.08	0.07	-36.62	27.52	9404.23	11.00	14.91

8.5 Appendix E - Full combined metrics of all models

This part shows the difference in performance between all the tested models for all assets. It is ordered such that horizontally it is visible for which asset a certain model performs best, and vertically it can be seen how all the models are ranked in performance for a given asset. It also indicates why GAS was not included in the results.

Table 17: ETFs: All models' combined metrics.

	EUO	ˆDJI	VO	QQQ	DBE	UBS	BNO	UTSL	ERX	BOIL
GARCH normal	0.45	0.31	0.31	0.29	0.49	0.76	0.49	0.25	0.46	0.32
GARCH t	0.44	0.34	0.32	0.29	0.50	0.38	0.45	0.28	0.39	0.23
GARCH skewt	0.48	0.28	0.22	0.20	0.44	0.38	0.41	0.20	0.34	0.24
GARCH GED	0.37	0.27	0.29	0.22	0.42	0.33	0.41	0.22	0.37	0.25
EGARCH normal	0.48	0.37	0.34	0.30	0.47	0.71	0.33	0.17	0.41	0.36
EGARCH t	0.47	0.43	0.36	0.35	0.47	0.42	0.29	0.18	0.34	0.27
EGARCH skewt	0.50	0.37	0.26	0.24	0.43	0.42	0.25	0.13	0.30	0.29
EGARCH GED	0.40	0.34	0.32	0.24	0.40	0.40	0.25	0.14	0.33	0.29
FIGARCH normal	0.46	0.35	0.38	0.30	0.48	0.59	0.52	0.29	0.38	0.44
FIGARCH t	0.44	0.40	0.42	0.31	0.42	0.35	0.46	0.30	0.41	0.48
FIGARCH skewt	0.46	0.32	0.30	0.21	0.35	0.30	0.42	0.24	0.36	0.47
FIGARCH GED	0.37	0.31	0.36	0.24	0.40	0.53	0.44	0.24	0.35	0.43
HARCH normal	0.52	0.58	0.64	0.62	0.48	0.30	0.58	0.87	0.55	0.58
HARCH t	0.45	0.63	0.57	0.81	0.50	0.31	0.55	0.74	0.41	0.47
HARCH skewt	0.46	0.60	0.52	0.67	0.46	0.31	0.52	0.71	0.40	0.48
HARCH GED	0.31	0.37	0.49	0.42	0.33	0.20	0.40	0.61	0.36	0.45
APARCH normal	0.43	0.30	0.30	0.28	0.45	0.82	0.30	0.28	0.41	0.33
APARCH t	0.42	0.33	0.31	0.28	0.42	0.42	0.30	0.28	0.37	0.27
APARCH skewt	0.45	0.27	0.21	0.18	0.37	0.42	0.27	0.21	0.32	0.29
APARCH GED	0.35	0.26	0.29	0.21	0.38	0.38	0.26	0.24	0.35	0.28
GJR-GARCH normal	0.47	0.20	0.32	0.28	0.39	0.60	0.30	0.17	0.33	0.34
GJR-GARCH t	0.46	0.17	0.26	0.29	0.41	0.34	0.31	0.21	0.31	0.23
GJR-GARCH skewt	0.49	0.11	0.24	0.23	0.37	0.34	0.27	0.12	0.27	0.24
GJR-GARCH GED	0.39	0.15	0.24	0.24	0.34	0.29	0.25	0.15	0.28	0.26
TARCH normal	0.40	0.15	0.30	0.20	0.38	0.58	0.24	0.14	0.22	0.32
TARCH t	0.38	0.17	0.33	0.18	0.37	0.37	0.22	0.16	0.22	0.31
TARCH skewt	0.40	0.09	0.29	0.10	0.32	0.37	0.19	0.09	0.17	0.33
TARCH GED	0.32	0.12	0.30	0.14	0.34	0.35	0.20	0.12	0.18	0.28
GAS normal	0.60	0.50	0.61	0.56	0.69	0.80	0.80	0.62	0.70	0.62
GAS t	0.20	0.49	0.54	0.29	0.23	0.18	0.19	0.22	0.20	0.15
GAS skewt	0.24	0.41	0.35	0.21	0.43	0.20	0.18	0.47	0.25	0.10
GAS GED	0.23	0.50	0.50	0.46	0.33	0.20	0.11	0.34	0.11	0.23

Table 18: Stocks: All models' combined metrics.

	JNJ	MCD	MDLZ	MSFT	AMZN	BABA	TSLA	TUP	GME	FCEL
GARCH normal	0.42	0.37	0.44	0.41	0.47	0.54	0.45	0.30	0.48	0.33
GARCH t	0.36	0.31	0.36	0.32	0.31	0.41	0.46	0.26	0.21	0.21
GARCH skewt	0.35	0.29	0.37	0.34	0.32	0.45	0.49	0.24	0.20	0.31
GARCH GED	0.29	0.23	0.24	0.21	0.24	0.33	0.26	0.17	0.12	0.16
EGARCH normal	0.36	0.30	0.45	0.41	0.53	0.53	0.43	0.28	0.31	0.41
EGARCH t	0.30	0.23	0.38	0.42	0.36	0.34	0.47	0.20	0.23	0.21
EGARCH skewt	0.29	0.22	0.38	0.44	0.36	0.37	0.50	0.19	0.21	0.31
EGARCH GED	0.22	0.17	0.24	0.26	0.29	0.29	0.26	0.15	0.15	0.15
FIGARCH normal	0.42	0.32	0.46	0.51	0.56	0.48	0.37	0.30	0.33	0.26
FIGARCH t	0.44	0.29	0.39	0.36	0.40	0.40	0.38	0.18	0.16	0.14
FIGARCH skewt	0.42	0.28	0.39	0.38	0.40	0.43	0.41	0.17	0.16	0.19
FIGARCH GED	0.29	0.20	0.26	0.27	0.33	0.30	0.19	0.14	0.14	0.15
HARCH normal	0.82	0.80	0.71	0.79	0.64	0.60	0.46	0.67	0.66	0.73
HARCH t	0.57	0.53	0.53	0.55	0.59	0.45	0.52	0.43	0.40	0.59
HARCH skewt	0.56	0.53	0.54	0.58	0.60	0.47	0.59	0.42	0.38	0.61
HARCH GED	0.54	0.44	0.40	0.42	0.37	0.33	0.15	0.34	0.29	0.35
APARCH normal	0.38	0.30	0.43	0.39	0.54	0.52	0.45	0.50	0.36	0.32
APARCH t	0.31	0.22	0.36	0.39	0.37	0.35	0.43	0.21	0.21	0.16
APARCH skewt	0.30	0.22	0.37	0.41	0.38	0.39	0.45	0.19	0.21	0.25
APARCH GED	0.23	0.17	0.23	0.23	0.32	0.29	0.25	0.15	0.18	0.12
GJR-GARCH normal	0.32	0.27	0.27	0.35	0.44	0.37	0.46	0.25	0.59	0.36
GJR-GARCH t	0.24	0.23	0.19	0.23	0.22	0.19	0.44	0.19	0.24	0.22
GJR-GARCH skewt	0.21	0.22	0.17	0.23	0.22	0.21	0.47	0.18	0.24	0.30
GJR-GARCH GED	0.17	0.16	0.11	0.13	0.18	0.14	0.25	0.13	0.16	0.17
TARCH normal	0.26	0.21	0.27	0.34	0.46	0.41	0.41	0.30	0.36	0.31
TARCH t	0.15	0.17	0.19	0.25	0.26	0.22	0.37	0.19	0.20	0.15
TARCH skewt	0.12	0.14	0.18	0.25	0.26	0.25	0.40	0.17	0.19	0.24
TARCH GED	0.10	0.11	0.11	0.14	0.23	0.19	0.21	0.15	0.18	0.12
GAS normal	0.80	0.80	0.80	0.50	0.45	0.61	0.81	0.80	0.80	0.60
GAS t	0.18	0.19	0.22	0.13	0.47	0.31	0.25	0.21	0.20	0.11
GAS skewt	0.21	0.21	0.21	0.29	0.59	0.49	0.08	0.20	0.25	0.12
GAS GED	0.18	0.21	0.21	0.28	0.41	0.49	0.20	0.10	0.20	0.21

Table 19: Cryptocurrencies: All models' combined metrics.

	BTC	LTC	XMR	XRP	NMC	DOGE	FTC	BLK	GRD	XPM
GARCH normal	0.30	0.39	0.40	0.35	0.35	0.47	0.49	0.36	0.41	0.42
GARCH t	0.23	0.23	0.29	0.19	0.23	0.39	0.28	0.23	0.11	0.22
GARCH skewt	0.23	0.25	0.31	0.21	0.26	0.41	0.32	0.27	0.13	0.23
GARCH GED	0.13	0.17	0.22	0.10	0.16	0.37	0.16	0.15	0.17	0.15
EGARCH normal	0.28	0.35	0.35	0.38	0.31	0.33	0.58	0.33	0.41	0.40
EGARCH t	0.31	0.38	0.24	0.49	0.33	0.47	0.50	0.42	0.54	0.32
EGARCH skewt	0.31	0.40	0.26	0.45	0.34	0.46	0.48	0.42	0.57	0.32
EGARCH GED	0.11	0.13	0.17	0.18	0.14	0.18	0.20	0.14	0.23	0.13
FIGARCH normal	0.28	0.40	0.45	0.33	0.39	0.43	0.43	0.35	0.41	0.38
FIGARCH t	0.22	0.22	0.32	0.13	0.22	0.34	0.21	0.19	0.19	0.18
FIGARCH skewt	0.22	0.24	0.35	0.16	0.24	0.37	0.26	0.23	0.21	0.19
FIGARCH GED	0.12	0.15	0.26	0.11	0.14	0.39	0.14	0.18	0.14	0.16
HARCH normal	0.48	0.46	0.68	0.40	0.47	0.71	0.52	0.69	0.41	0.81
HARCH t	0.68	0.62	0.62	0.17	0.32	0.44	0.44	0.53	0.16	0.45
HARCH skewt	0.67	0.64	0.50	0.17	0.33	0.44	0.46	0.55	0.15	0.45
HARCH GED	0.20	0.21	0.32	0.14	0.15	0.35	0.16	0.26	0.17	0.27
APARCH normal	0.30	0.35	0.40	0.35	0.35	0.50	0.47	0.36	0.41	0.41
APARCH t	0.28	0.22	0.27	0.18	0.44	0.24	0.41	0.23	0.23	0.25
APARCH skewt	0.27	0.24	0.29	0.18	0.26	0.31	0.41	0.23	0.23	0.26
APARCH GED	0.13	0.13	0.21	0.14	0.14	0.25	0.14	0.13	0.17	0.19
GJR-GARCH normal	0.28	0.39	0.44	0.36	0.32	0.50	0.48	0.34	0.41	0.41
GJR-GARCH t	0.24	0.22	0.30	0.19	0.22	0.31	0.28	0.20	0.12	0.15
GJR-GARCH skewt	0.23	0.24	0.32	0.21	0.23	0.35	0.35	0.25	0.16	0.17
GJR-GARCH GED	0.13	0.16	0.24	0.10	0.11	0.31	0.16	0.13	0.15	0.15
TARCH normal	0.28	0.36	0.40	0.37	0.30	0.43	0.47	0.33	0.41	0.43
TARCH t	0.23	0.17	0.25	0.18	0.21	0.22	0.29	0.20	0.18	0.22
TARCH skewt	0.22	0.20	0.28	0.20	0.24	0.25	0.35	0.25	0.19	0.24
TARCH GED	0.12	0.12	0.21	0.15	0.13	0.22	0.14	0.12	0.18	0.18
GAS normal	0.84	0.86	0.95	0.80	0.95	0.86	1.00	0.70	1.00	0.82
GAS t	0.03	0.23	0.06	0.08	0.02	0.06	0.03	0.14	0.01	0.01
GAS skewt	0.03	0.05	0.12	0.07	0.08	0.04	0.06	0.10	0.03	0.01
GAS GED	0.20	0.20	0.20	0.20	0.20	0.20	0.19	0.04	0.11	0.20

References

C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking & Finance*, 26(7):1487–1503, 2002.

H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, 1974.

Richard TT Baillie, Tim Bollerslev, and Hans Ole Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1):3–30, 1996.

N. Balakrishnan, editor. *Handbook of Distributions for Applications in Statistics, Biostatistics, and Quality Control*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. Chapman and Hall/CRC, 2019. ISBN 9780367383537.

Committee Basel. Basel iii: The liquidity coverage ratio and liquidity risk monitoring tools. https://www.bis.org/publ/bcbs238.htm, 2013.

L. Bauwens, S. Laurent, and J. V. Rombouts. Multivariate garch models: a survey. *Journal of Applied Econometrics*, 21(1):79–109, 2006.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327, 1986.

J. Danielsson. *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in R and Matlab*. John Wiley & Sons, 2011.

F. X. Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold-mariano tests. *Journal of Business & Economic Statistics*, 33(1):1–19, 2015.

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

Zhuanxin Ding, Clive WJ Granger, and Robert F Engle. A long memory property of stock market returns and a new model. *Journal of empirical finance*, 1(1):83–106, 1993.

Kevin Dowd. *Measuring Market Risk*. Wiley Finance Series. Wiley, 2005. ISBN 9780470013038.

Graham Elliott and Allan Timmermann. *Economic Forecasting*. Princeton University Press, Princeton, NJ, 2016.

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*. Springer, 2005.

R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of uk inflation. *Econometrica*, pages 987–1008, 1982.

Robert F Engle, David M Lilien, and Russell P Robins. Estimating time varying risk premia in the term structure: The arch-m model. *Econometrica: Journal of the Econometric Society*, pages 391–407, 1987.

Eugene F Fama. *Efficient capital markets: A review of theory and empirical work*, volume 25. Wiley Online Library, 1970.

C. Fernandez and M. F. J. Steel. On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371, 1998.

Christian Francq and Jean-Michel Zakoian. Garch models: Structure, statistical inference and financial applications. 2011.

Anton Sorin Gabriel. Evaluating the forecasting performance of garch models. evidence from romania. *Procedia - Social and Behavioral Sciences*, 62:1006–1010, 2012. ISSN 1877-0428. doi: https://doi.org/10.1016/j.sbspro.2012.09.171. URL https://www.sciencedirect.com/science/article/pii/S1877042812036129. World Conference on Business, Economics and Management (BEM-2012), May 4–6 2012, Antalya, Turkey.

L. R. Glosten, R. Jaganathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of finance*, 48(5):1779–1801, 1993.

Amit Goval and Ivo Welch. A comprehensive look at the empirical performance of equity premium prediction. Working Paper 10483, National Bureau of Economic Research, May 2004. URL http://www.nber.org/papers/w10483.

Peter Reinhard Hansen and Asger Lunde. Forecasting volatility using high-frequency data. 2011.

Nikolaus Hautsch. Econometrics of financial high-frequency data. *Springer Science & Business Media*, 27, 2012.

R. J. Hyndman and A. B. Koehler. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4):679–688, 2006.

Rafal Weron. Jesus Lago, Bart De Schutter. Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access bench-mark. *Applied Energy*, 293:116983, 2021.

Philippe Jorion. *Value at risk: the new benchmark for managing financial risk*. The McGraw-Hill Companies, Inc., 2007.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4): 394–419, 1963.

U. A. Muller, M. M. Dacorogna, R. D. Dave, R. B. Olsen, O. V. Pictet, and J. E. von Weizsacker. Volatilities of different time resolutions—analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2-3):213–239, 1997.

S. Nadarajah and S. Kotz. On the exponential power distribution. *Statistics*, 39(4):345–356, 2005.

D. B. Nelson. Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59 (2):347–370, 1991.

A. B. Nugroho, B. Susanto, and M. Rizkiyah. Empirical comparison of garch type models on forecasting volatility of stock returns. *Jurnal Ilmiah Wahana Akuntansi*, 14(1):1–16, 2019.

G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, 1978.

Kevin Sheppard, Stanislav Khrapov, Gábor Lipták, mikedeltalima, Rob Capellini, alejandro cermeno, Hugle, esvhd, Snyk bot, Alex Fortin, and et al. bashtage/arch: Release 6.1.0. May 2023. doi: 10.5281/zenodo.7975104.

David J Sheskin. *Handbook of Parametric and Nonparametric Statistical Procedures*. CRC Press, 2003.

Koby Todros and Alfred O. Hero. Measure-transformed quasi maximum likelihood estimation. *arXiv preprint arXiv:1511.00237*, 2015. URL https://arxiv.org/abs/1511.00237.

Victor Troster, Aviral Kumar Tiwari, Muhammad Shahbaz, and Demian Nicolás Macedo. Bitcoin returns and risk: A general GARCH and GAS analysis. *Finance Research Letters*, 30(C):187–193, 2019. doi: 10.1016/j.frl.2018.09.014. URL https://ideas.repec.org/a/eee/finlet/v30y2019icp187-193.html.

Ruey S Tsay. *Analysis of Financial Time Series*. John Wiley Sons, 2005.

Ruey S. Tsay. *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics. Wiley, 2010. ISBN 9780470414354.

Yuling Wang, Yunshuang Xiang, Xinyu Lei, and Yucheng Zhou. Volatility analysis based on garch-type models: Evidence from the chinese stock market. *Economic Research-Ekonomska Istraživanja*, 35(1):2530–2554, 2022. doi: 10.1080/1331677X.2021.1967771. URL https://doi.org/10.1080/1331677X.2021.1967771.

C. J. Willmott and K. Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.

Jean-Michel Zakoian. Threshold heteroskedastic models. *Journal of Economic Dynamics and control*, 18(5):931–955, 1994.