Group Project Report

Volatility Forecast and Analysis of Ethereum

Matriculation Number	Name
A0239399H	CHEUNG Yu Jeng Ethan
A0240232Y	LUO Xinming
A0239197N	YOUNG Zhan Heng

Table of Contents

1. Introduction

The project repository can be found here.

With the increasing diversification of investment opportunities, traders worldwide continuously seek new ways to maximise their profits while managing risk. Traditional investment vehicles like index funds and exchange-traded funds offer relatively stable returns with lower risk. At the same time, equities often present higher risks but with the potential for greater rewards.

This pursuit of higher returns has led many investors to explore alternative assets beyond traditional markets, with cryptocurrencies gaining increasing attention as a new frontier in speculative trading. Major financial institutions, including JP Morgan, have established dedicated research divisions to explore the integration of cryptocurrencies into their trading strategies (J.P. Morgan, 2024). The market capitalization of all cryptocurrencies[1] has surged by approximately 338% between 1 January 2021 and 1st January 2025 (CoinGecko, 2025).

Among the various cryptocurrencies driving this rapidly evolving market, Ethereum (ETH) stands out due to its dual role as a tradable asset and a foundational technology for decentralised applications. With the second largest market capitalisation at $317.5 billion[2], ETH has slowly transcended the role of simply being another cryptocurrency. In Africa, ETH is perceived as a more stable store of value than some local currencies because the latter is prone to extreme fluctuations due to poor monetary policy and regulation (Techcabal, 2025). Moreover, with its stateful Turing-complete scripts, ETH can overcome inefficiencies in monetary transfers.

The widespread adoption and significance in Web3 development make ETH highly volatile, influenced by market speculation, regulatory shifts, and technological upgrades such as ETH 2.0. Given this volatility, accurate forecasting is essential for risk management, portfolio optimisation, and trading strategies.

This report consolidates reviews of some existing literature on cryptocurrency volatility[3] forecasting, comparing the effectiveness of traditional econometric models and modern machine learning (ML) techniques in predicting ETH's daily realised volatility (RV). By identifying robust

[1] This refers to the total market capitalization of 16,727 cryptocurrencies tracked across 1,238 exchanges. Crypto-backed tokens such as wrapped, bridged, and staked tokens are excluded from global market cap to avoid double-counting of value.
[2] As of 9 Feb 2025 (CoinGecko, 2025)
[3] Volatility is the tendency for prices to change unexpectedly (Harris, 2002)

forecasting methods, this project aims to enhance risk assessment and strategic decision-making for traders and institutions, helping them navigate the dynamic digital asset landscape more effectively.

The remainder of this report is organised as follows: Section 2 provides a literature review of the major models and key challenges in crypto volatility forecasting. Section 3 details the data source and methodology for implementing each model. Section 4 presents the main results, while Section 5 interprets these findings .

2. Current Literature and Review

Traditional Econometric Models

A characteristic of economic data under analysis is its time series nature, where each data is specifically tied to a time period. Econometric models are well-suited to this context, are easy to interpret, and have a relatively straightforward implementation compared to ML models. Two econometric models are selected as the baseline models in this project.

Built on the works of Eagle (1982) and Bollerslev (1986), the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) family of models has been widely used in financial volatility forecasting. The general GARCH(p, q) model is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

where σ_t^2 represents the conditional variance at time t, which measures the expected variability of returns based on past information. The parameter $\omega > 0$ is a constant term, ensuring that volatility remains positive even when past shocks are minimal. $\sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2$ capture the influence of past squared residuals ϵ_{t-i}^2 on current volatility and reflect the extent to which unexpected returns from previous periods contribute to fluctuations in variance. Meanwhile, $\sum_{j=1}^{p} \beta_j \sigma_{t-i}^2$

accounts for the persistence of past variances. This means that high volatility in prior periods will likely influence future volatility, contributing to the well-known volatility clustering effect observed in financial markets.

Corsi (2009) introduced the Heterogeneous Autoregressive Model (HAR-RV) as a foundational approach for predicting realized variance, incorporating daily, weekly, and monthly realised volatility components. Corsi benchmarked the HAR-RV model against traditional models such as AR(1), AR(3), and ARFIMA using datasets from USD/CHF, S&P500, and T-Bond markets. The HAR-RV model consistently achieved the lowest root mean square error (RMSE) for one-day-ahead forecasts, effectively capturing long-memory volatility behaviour and establishing itself as the baseline model to surpass. Supporting this, Bergsli et al. (2021) showed that HAR models based on realised variance (RV) outperform GARCH models that rely solely on daily data.

Machine Learning Models

ML techniques are the new trend in the economic sphere for their ability to capture complex nonlinear relationships. Basher and Sadorsky (2022) demonstrated the efficacy of random forests in accurately predicting Bitcoin price direction. Similarly, Brauneis and Sahiner (2024) showed that ML methods can enhance the accuracy of traditional econometric models, such as the HAR model, in forecasting cryptocurrency volatility. Several studies (Bucci, 2020; Kim and Won, 2018) demonstrated the effectiveness of Long Short-Term Memory (LSTM) networks in volatility forecasting, particularly in capturing the complex temporal dependencies inherent in financial time series data. Shen et al. (2021) showed that Recurrent Neural Network outperformed GARCH in average forecasting performance.

However, some studies also showed that ML models do not consistently outperform traditional statistical models due to the high noise in cryptocurrency price movements. Yae and Tian (2022) compared traditional econometrics and ML methods' performances in predicting cryptocurrency returns. They found that most predictors did well for in-sample prediction but fell short for out-of-sample prediction. In addition, the results showed that the ML models underperformed against traditional econometric models, such as least absolute deviation and rank regression. The linear model even outperformed ML models such as LASSO, elastic net, random forest, and more.

Hybrid Models

Adopting hybrid models for ETH volatility forecasting is motivated by their ability to leverage the complementary strengths of traditional econometric techniques and machine learning methods, effectively capturing both linear dependencies and complex non-linear dynamics.

Amirshahi et al. (2023) integrated deep learning with GARCH-type models to forecast the volatility of 27 cryptocurrencies, revealing that the informative features extracted from GARCH forecasts significantly enhance the predictive capability of deep learning models, specifically A Deep Feed Forward Neural Network and LSTM networks. These findings were also highlighted by Kim and Won (2018), where the LSTM-GARCH hybrid model achieved a lower MSE than a single LSTM model. Ardia et al. (2018) also found strong evidence of regime shifts in Bitcoin's volatility, showing that Markov-Switching GARCH (MSGARCH) models outperform traditional single-regime GARCH models when forecasting one-day-ahead Value-at-Risk.

These findings highlight the potential of hybrid and regime-switching models in improving cryptocurrency volatility predictions, making them valuable tools for risk management and strategic decision-making.

Challenges in Cryptocurrency Volatility Forecasting

Given the focus on cryptocurrency data, we reviewed relevant literature to identify key challenges in cryptocurrency forecasting. This analysis informed adjustments to our models, ensuring their adaptability to the unique characteristics of cryptocurrencies.

Firstly, unlike stocks, which derive value from cash flows and assets, or commodities with inherent utility, cryptocurrencies lack intrinsic value. Their prices are primarily influenced by investor sentiment, market trends, and speculative trading, which makes them highly volatile and unpredictable.

Secondly, economists and prominent investors have expressed concerns that the entire cryptocurrency market constitutes a speculative bubble, posing significant risks to investors and the broader economy (*Bubbles in Cryptocurrency Markets Dwarf Any Historical Bubble*, 2024). Chen and Hafner (2019) identified several sentiment-induced bubble periods in the cryptocurrency market. Urgenc et al. (2024) demonstrated that market sentiment significantly impacts cryptocurrency volatility, primarily evidenced by the strong positive correlation between Google search volume and price volatility across nine cryptocurrencies from 2020 to 2023. It is therefore difficult to model price fluctuations using traditional econometric methods.

Thirdly, the cryptocurrency market is still relatively young compared to traditional financial markets. Lower liquidity means that large transactions, often called "whale trades," can cause dramatic price shifts, leading to abrupt changes in volatility, as seen in Chernoff and Jagtiani (2024). The relatively short history of market data and research literature available for robust model training also limits the effectiveness of long-term volatility forecasting.

Fourthly, Governments and financial regulators worldwide have taken different stances on cryptocurrencies, and policy changes often trigger volatility spikes. For instance, initial coin offerings were banned in China in 2017, leading to a potentially long-term negative impact on cryptocurrency within the region. Uncertainty regarding legal frameworks and taxation policies can lead to sudden and unpredictable market movements. This was further backed by Raza et al. (2022), who showed that financial regulation policy uncertainty is negative and significantly associated with the volatility of cryptocurrencies.

3. Methodology

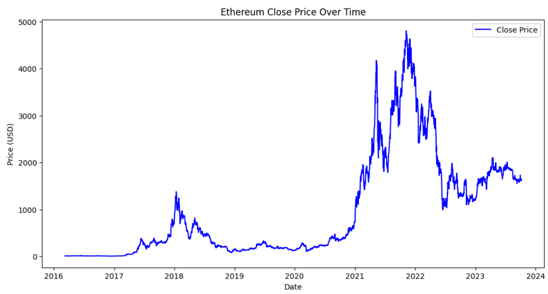
Data Source

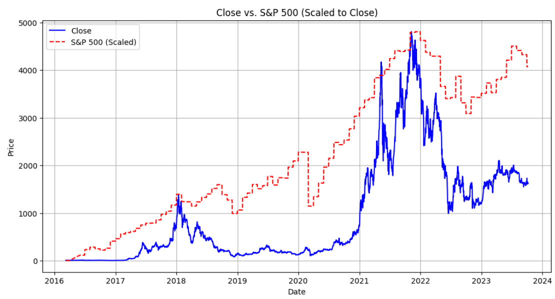
We used historical trading data for ETH at a one-minute resolution from 2016-03-09 16:04:00 to 2023-10-08 09:27:00. The dataset is sourced from Kaggle and originated from the Bitfinex exchange. It includes open, high, low, and close price data, making it highly suitable for high-frequency trading analysis and volatility forecasting.

We also obtained data from the Federal Reserve Economic Data - Monthly Database (FRED-MD). This macroeconomic dataset maintained by the Federal Reserve Bank of St. Louis provides a comprehensive collection of monthly U.S. macroeconomic indicators. It includes variables such as industrial production, inflation, interest rates, employment, and financial market conditions, which serve as key indicators of broader economic trends.

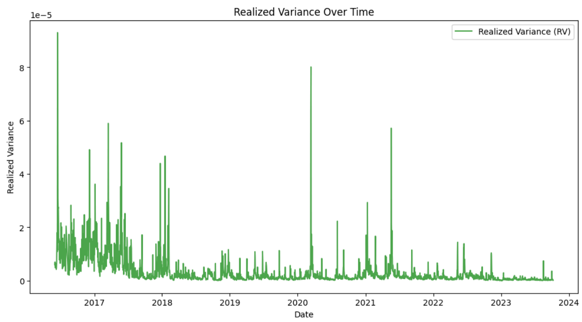
Exploratory Data Analysis

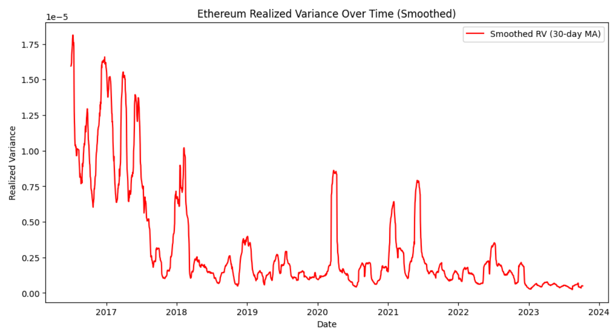
Our initial findings showed that the closing price of ETH increased with an upward trend from 2016 to 2024, with spikes in 2018, 2021, and 2022. However, since 2022, ETH has been volatile, with several gentle spikes and drops. When we overlay the scaled S&P 500 index fund, we noticed that although the closing price of ETH follows the general trend of the index fund, it is still much more volatile.





We also plotted the RV of ETH. We noticed three significant spikes in 2016, 2020, and 2021. We attributed these spikes to early crypto volatility, COVID-19, and the crypto bubble burst respectively. We used a rolling average to smooth out the price trend, and we noticed that prior to 2019, prices of ETH were much more volatile than current trends.





Data Preprocessing

The dataset did not contain any transactions between April and May 2016, likely due to the Homestead upgrade (where major improvements were made to the Ethereum network) implemented on May 14th, 2016 (A Short History of Ethereum, 2019). The dataset also lacked consistent transaction updates, meaning that the “minute-resolution” data was missing data. We handled the missing data by changing how we calculated the daily RV, which will be shared below.

We defined log returns as the natural logarithm of the ratio between the current price and the previous price:

$$r_t = \log\left(\frac{P_t}{P_{t-1}}\right) = \log P_t - \log P_{t-1}$$

where:

- $\mathbf{P_t}$ is the close price at time **t**,
- $\mathbf{P_{t-1}}$ is the close price at time **t-1**.

This transformation offers several advantages that enhance forecasting accuracy. One primary reason is that they help stabilize and normalize data distributions. Unlike simple price returns (P_t - P_t or (P_t - P_{t-1}) / P_{t-1}), which tend to be skewed and non-stationary, log returns compress extreme values and exhibit better normality properties, making them more suitable for statistical models such as GARCH[4], and deep learning-based forecasting techniques. Another key advantage is that log returns possess an additive property over time ($r_{total} = r_1 + r_2 + \ldots + r_n$). Unlike percentage returns, which require compounding calculations, log returns can be summed over multiple periods, simplifying computations in the time-series analysis.

Daily RV is defined as the sum of squared minute-level log returns within a day. As N approaches infinity (i.e., if we observe infinitely many price changes per unit of time), the estimator converges to the true quadratic variation.

To account for the inconsistent data, we normalised the RV by taking RV and then dividing it by the total observations that day, and then we multiplied the value by 1440, which is the total minutes in a day.

$$RV_t = \sum_{i=1}^{N} r_{t,i}^2 \qquad RV_{\text{normalised}} = \frac{1440}{N}\, RV_t$$

[4] This is because GARCH-type models assume stationary variance, meaning the statistical properties of the data do not change over time.

There are several assumptions in the estimation above. One key assumption is that ETH prices follow a continuous-time stochastic process, meaning price changes reflect new information efficiently. This is because cryptocurrencies trade every minute across multiple exchanges, unlike stock markets, which have fixed trading hours (e.g., NYSE: 9:30 AM – 4:00 PM EST). This makes the ETH market closer to a continuous-time model than equities, which experience overnight price gaps.

Another assumption is the absence of market microstructure noise, such as bid-ask spreads and order imbalances, which can distort high-frequency returns. The method also assumes that intraday log returns are independent and identically distributed, with no serial dependence. However, financial markets often exhibit volatility clustering, where periods of high volatility tend to follow each other. This limitation can be addressed using HAR-RV (Heterogeneous Autoregressive RV) models, which account for volatility persistence. Additionally, the computation assumes that large price jumps do not disproportionately affect RV, yet cryptocurrencies frequently experience flash crashes and sudden price spikes.

Daily returns were defined as the percentage change in the daily closing price, serving as the input for the GARCH model.

We computed rolling averages of RV to capture weekly and monthly volatility trends. Specifically, we used a 7-day window to represent a week and a 30-day window for a month. The rolling averages were calculated as follows:

$$RV_{w,t} = \frac{1}{7}\sum_{i=0}^{6} RV_{d,t-i} \quad RV_{m,t} = \frac{1}{30}\sum_{i=0}^{29} RV_{d,t-i}$$

This approach smooths short-term fluctuations in daily RV and enhances the model's ability to identify longer-term volatility patterns.

As noted by Anderson et al. (2001), the logarithms of the RV of equities are approximately normal. We further applied logarithmic transformation to the RV measures as follows:

$$\ln RV_{d,t} = \log(RV_{d,t}) \quad \ln RV_{w,t} = \log(RV_{w,t}) \quad \ln RV_{m,t} = \log(RV_{m,t})$$

Feature Selection

The dataset, which was obtained from Kaggle, initially comprised over 120 covariates, excluding lagged variables. Training the models and then generating 500 predictions using the full set of variables was not computationally feasible to most users, ourselves included. To mitigate this challenge, a least absolute shrinkage and selection operator (LASSO) regression was used to identify the most critical features[5]. The variables selected are:

1. **UEMPMEAN**: Average monthly level of unemployment in the United States, used to gauge employment trends and sometimes consumer confidence.
2. **IPNMAT**: Reflects the activity in the non-durable materials or manufacturing sectors and gives a rough idea of the supply chain industry.
3. **AAAFM**: Monthly yield of AAA-rated corporate bonds.
4. **Volume**: Total volume of ETH traded daily, used as a proxy for market liquidity.
5. **BUSLOANS**: Total volume of business loans or credit extended to firms, giving a rough idea of investment levels and economic expansion
6. **USCONS**: Measures consumer spending in the US economy, which is a key factor in economic growth.

Selecting volume as one of the key variables contrasted with the findings of Balcilar et al. (2017), who concluded that volume was not a significant predictor in predicting the RV of Bitcoin, another cryptocurrency. This discrepancy could be attributed to the one-minute resolution of our data and the difference in behaviour between ETH and Bitcoin.

After identifying the six variables, we lagged the covariates by one period. In addition, we lagged the daily RV by one day, one week, and one month to capture the temporal dynamics.

Models

We adopted steps outlined in Dudek et al. (2024), to train and test the model.

1. Calculate the daily, weekly, and monthly RV of ETH.
2. Apply logarithmic transformation.

[5] A LASSO regression is most suitable here as it will be able to perform variable selection by using an ordinary least squares regression with L1 penalty, effectively shrinking down redundant variables to zero.

3. Constructed an 80/20 train/test split or 60/20/20 train/validation/test split, depending on the model.
4. Train the model with the respective prediction model.
5. Optimise the model using the validation set if need be.
6. Generate a one-step-ahead prediction.
7. Update the training sample using a rolling window procedure, meaning that the data used to input the model will be the test data, not the predicted data.
8. Update T = T+1 and repeat steps 4-7.
9. Generate a list of one-day ahead forecasts and calculate the root mean square error (RMSE).

We trained the respective models: GARCH, HAR-RV, RF, and LSTM. Additionally, we trained a hybrid GARCH-LSTM model. The former two models served as a benchmark against our ML models. Our hypothesis for this was that the ML models would outperform our baseline models by a statistically significant p-value.

Generalised Autoregressive Conditional Heteroskedasticity (GARCH)

We used the widely adopted GARCH(1,1) model as a baseline model.

$$r_t = \mu + \epsilon_t$$

$$\epsilon_t = \sigma_t e_t$$

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

In this model, we assume that ETH daily return at time t is conditionally normally distributed, where F_{t-1} denotes the information set available at time t−1.

$$r_t|\mathcal{F}_{t-1} \sim N(\mu, \sigma_t^2)$$

We used an 80/20 train/test split to train and test our GARCH model. We first fitted the model on the training set to get the predicted values of $\alpha \geq 0$ and $\beta \geq 0$. We checked that $\alpha+\beta<1$ to ensure the stationarity of the variance process. Then, we used r_{t-1} -μ instead of ϵ_t when predicting the conditional volatility for days in the test set.

Heterogeneous Autoregressive (HAR) Model

The HAR model uses a daily, weekly, and monthly lag to generate a one-step-ahead forecast. We followed the specifications used in Dudek et al. (2024).

$$\ln RV_{d,t} = \gamma_0 + \gamma_1 \ln RV_{d,t-1} + \gamma_2 \ln RV_{w,t-1} + \gamma_3 \ln RV_{m,t-1} + \varepsilon_t .$$

We used a one-day, one-week, and one-month lag for the RV to predict the RV for the next day. The model developed by Carsi (2009) works well with high-frequency data, which aligns with cryptocurrency's high-frequency nature. The model was easy to implement, and we used it to generate a one-step-ahead forecast for all the models.

Random Forest (RF)

Random forest (RF) is an extremely popular method for forecasting financial assets, such as returns and volatility, since it is easy to implement and produces readable output.

We used an 80/20 train/test split to train and then test our RF model. In our approach, we used the variables selected in the LASSO regression and made the RF consider p/3 variables at each split, which is common practice for regression trees.

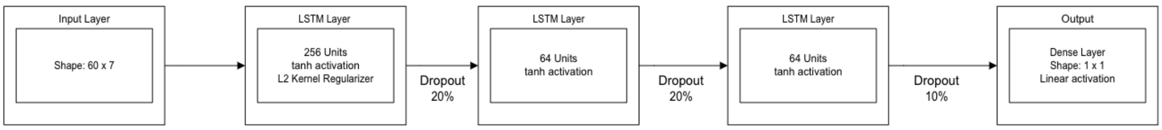
We specified the number of estimators to 100 and set the random state to 42 to ensure reproducibility. The model utilised all available CPU cores by setting *n_jobs* = -1 and limited the maximum number of features considered for splitting at each node to one-third of the total features max_features = ⌊*n_features*/3⌋. This configuration aimed to balance computational efficiency and model performance while mitigating the risk of overfitting.

Long Short-Term Memory (LSTM) Neural Network

Our input features include log-transformed daily, weekly, and monthly RV, along with the variables selected through LASSO. To capture temporal dependencies, we created sequences using a 60-day lookback window for each target RV.

We then performed hyperparameter tuning via random search to find the best model specification. This consisted of the number of LSTM layers, the number of nodes for each layer, the dropout after each layer, and the activation function used.

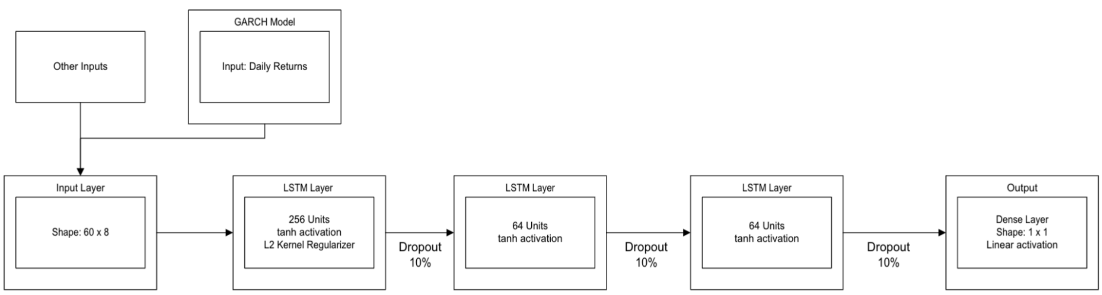
The optimal model configuration uses a tanh activation function and consists of three layers with 256, 64, and 64 units, respectively. Dropout regularization was applied with dropout rates of 0.2 in the first two layers and 0.1 in the final layer. Additionally, an L2 kernel regularizer was added to the first LSTM layer to further control overfitting.



GARCH-LSTM Hybrid Neural Network

Inspired by the works of Amirshahi et al. (2023), the GARCH-LSTM hybrid model included the GARCH output as an additional feature while retaining all input features from the LSTM model. We used the in-sample GARCH predictions with the training set and the GARCH forecast in the test set. This configuration aims to leverage the GARCH model's strength in modelling time-varying volatility while incorporating deep learning techniques to capture complex temporal patterns.

With hyperparameter tuning, we found that the model configuration identified includes a tanh activation function with a three-layer architecture. The model comprises 256 units in the first layer, followed by two layers with 64 units each. To enhance generalization and prevent overfitting, dropout regularization was applied uniformly at a rate of 0.1 across all layers. Including an L2 kernel regularizer in the first LSTM layer also improved the convergence of the model.



4. Analysis of Results

Selection of Metrics

To assess the performance of the different volatility forecasting models, we consider four key metrics: **Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), R^2 (coefficient of determination), and Quasi-Likelihood**. Each metric provides unique insights into the accuracy and effectiveness of the models.

Root Mean Squared Error (RMSE) measures the absolute difference between predicted and actual RV, with higher penalties for larger errors. This makes it a crucial metric for volatility forecasting, where extreme fluctuations must be captured accurately.

Mean Absolute Percentage Error (MAPE) represents relative error, indicating how much the predictions deviate from actual values in percentage terms. However, MAPE may be less reliable in evaluating the extreme nature of volatility.

R^2 (coefficient of determination) measures how well a model explains the variance in realised volatility. A higher R^2 suggests that the model provides better explanatory power.

Quasi-Likelihood (Q-LIKE) assesses how well a model fits the probabilistic structure of volatility. It is an asymmetric measure that penalises the underestimation of volatility, and therefore, it is especially relevant in financial risk assessment applications that rely on conservative estimates. Hence, under-estimation incurs higher costs.

The quasi-likelihood (Q-Like) loss function is computed using the formula:

$$Q\text{-Like} = \mathbb{E}\left[\frac{y_t}{\hat{y}_t} - \log\left(\frac{y_t}{\hat{y}_t}\right) - 1\right]$$

where y_t represents the actual values and $\widehat{y}_t$ denotes the predicted values.

The test data's start date differs across our models because certain methods (e.g., the LSTM) require a 30-day lookback window. Consequently, we aligned all forecast dates to begin on the latest available start date among the models, namely June 26, 2022.

Each metric contributes to understanding model performance from different perspectives, ensuring a comprehensive evaluation. The results can be found in Table 1 below.

Model	RMSE	MAPE (%)	R2	Q-LIKE
GARCH	0.001569	266.07	0.0018	0.4838
GARCH_LSTM	0.001414	58.53	0.1887	0.3962
LSTM	0.001291	129.12	0.3237	0.3050
HAR	0.001170	73.61	0.4444	0.1957
RF	0.001236	56.71	0.3806	0.2051

Table 1: Metric comparison of all models

We also plotted out the actual RV and compared it against the forecast RV of each model. The plots can be found in Figure 1 below.

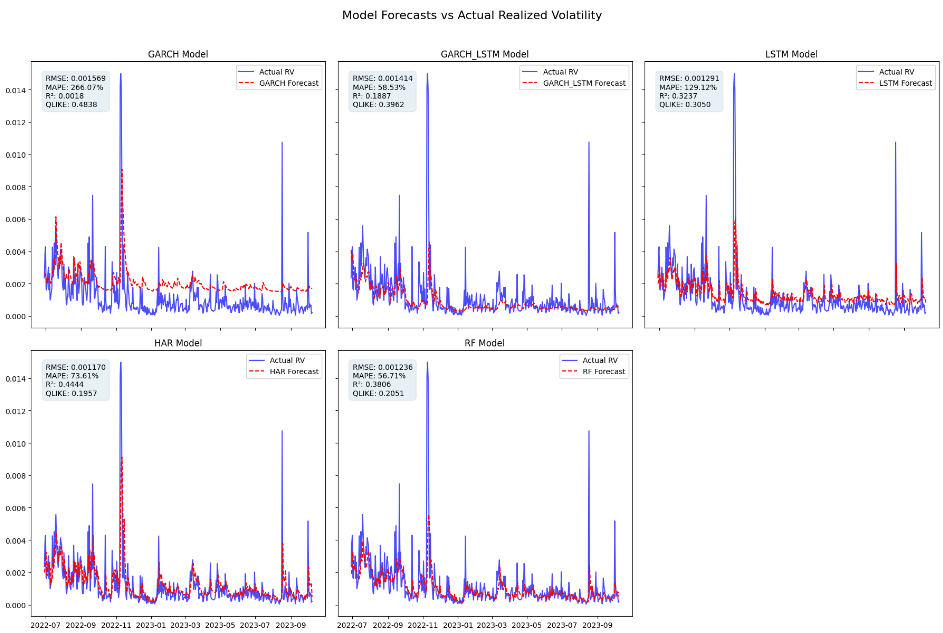


Figure 1: Actual RV vs Forecast RV

Diebold-Mariano Test

To further evaluate the forecasting ability of the models, we conducted the **Diebold-Mariano (DM) test**, a statistical test designed to compare the predictive accuracy of two competing models (Diebold and Mariano, 2002). Unlike RMSE and R^2, which provide aggregate measures of error, the DM test assesses whether the forecast errors of one model are significantly different from those of another over time. This allows us to determine whether one model consistently outperforms another in a statistically meaningful way. We will be using the dm_test package created by the GitHub user *johntwk*.

The DM test is particularly useful in volatility forecasting because it evaluates short-term predictive accuracy rather than overall fit. It helps identify whether a model's forecasts are systematically better or worse than a competing model's at each time step. This is crucial when forecasting financial volatility, where timely and responsive predictions are more important than long-term trends. The results of the DM test can be found in Table 2 below.

Model 1	Model 2	DM Statistic	P-Value	Conclusion (95% sig. level)
GARCH	GARCH_LSTM	11.89	1.29E-28	GARCH_LSTM outperforms GARCH.
GARCH	LSTM	17.67	9.47E-54	LSTM outperforms GARCH.
GARCH	HAR	15.32	3.76E-43	HAR outperforms GARCH.
GARCH	RF	14.75	1.21E-40	RF outperforms GARCH.
GARCH_LSTM	LSTM	-5.407	1.03E-07	GARCH_LSTM outperforms LSTM.
GARCH_LSTM	HAR	1.954	0.0513	Similar predictive accuracy.
GARCH_LSTM	RF	4.325	1.87E-05	RF outperforms GARCH_LSTM.
LSTM	HAR	9.867	5.76E-21	HAR outperforms LSTM.
LSTM	RF	9.646	3.51E-20	RF outperforms LSTM.
HAR	RF	3.378	0.000793	RF outperforms HAR.

Table 2: Diebold-Mariano Test Results

GARCH

The GARCH model exhibits the weakest performance across all four evaluation metrics. It produces the highest RMSE (0.001569), indicating significant absolute forecast errors. More notably, it struggles with percentage-based accuracy, as seen in its exceptionally high MAPE (266.07%). This suggests that the model frequently misestimates RV, particularly in volatile regimes with large changes.

Its R^2 value (0.0018) is effectively zero, implying that the model provides almost no explanatory power for RV. Furthermore, GARCH's QLIKE score (0.4838) is the highest of all models, meaning it is highly prone to underestimating volatility.

These weaknesses are visually confirmed in the forecast plots, where GARCH consistently lags behind actual RV, particularly during sharp market swings. We also observe qualitatively that the model's responsiveness to spikes falls as the forecast period progresses. This occurs because of GARCH's autoregressive, mean-reverting nature, which causes past volatility shocks to decay over time, leading to an underestimation of sustained volatility shifts. Every other model significantly outperforms GARCH via the DM test. These findings suggest that GARCH's rigid parametric structure is inadequate for capturing ETH's highly dynamic and nonlinear volatility patterns.

GARCH-LSTM

The hybrid GARCH-LSTM model introduces a deep learning component to traditional GARCH dynamics, resulting in measurable improvements over all four metrics. With significantly reduced RMSE (0.001414), MAPE (58.53%), and higher R^2 (0.1887). Its QLIKE (0.3962) is also reduced, meaning it is less prone to severe underestimations.

While these improvements are evident, GARCH-LSTM still lags behind the other models, particularly the LSTM. The forecast plots reveal that it still struggles to capture extreme fluctuations accurately. The declining performance over time also reveals the model places excessive importance on the GARCH output feature. Interestingly, the DM test shows that GARCH-LSTM significantly outperforms GARCH and LSTM, but when compared to HAR, their difference is not statistically significant at the 5% level. This result is somewhat surprising given HAR's better aggregate error metrics and may suggest that GARCH-LSTM occasionally achieves comparable forecasting accuracy over short intervals.

LSTM

The standalone LSTM model performs better than both GARCH and GARCH-LSTM in terms of RMSE (0.001291) and R^2 (0.3237), indicating improved absolute error and explanatory power. However, its MAPE (129.12%) is still quite high, suggesting that it struggles with percentage-based accuracy, particularly when volatility is low. This is a common issue with deep learning models that are optimised for absolute error but can struggle in relative percentage terms when the magnitude of the target variable varies significantly. In terms of QLIKE (0.3050), LSTM performs moderately well, indicating it is better at avoiding severe underestimation than GARCH but still not as strong as HAR or RF.

The forecast plots suggest that while LSTM is adept at recognizing broad volatility trends, it still mispredicts sudden spikes, contributing to its higher MAPE. The DM test results confirm that HAR and RF significantly outperform LSTM, reinforcing that while deep learning alone offers some predictive strength, it does not match the robustness of structured volatility models like HAR or tree-based approaches like RF.

HAR

The HAR model emerges as the top performer in terms of absolute error and explanatory power. It achieves the lowest RMSE (0.001170) and the highest R^2 (0.4444), indicating it explains nearly half of the variance in RV.

Additionally, HAR's QLIKE (0.1957) is the lowest, meaning it is least prone to dangerous underestimations of volatility, making it particularly useful for risk management applications. Its MAPE (73.61%) is lower than all models but RF.

The forecast plots confirm HAR's strength, showing that it consistently tracks volatility patterns and captures both high and low fluctuations more effectively than the other models. Interestingly, while HAR outperforms GARCH-LSTM in nearly every metric, their DM test yielded a statistically insignificant difference. This could indicate that while HAR provides better overall accuracy, GARCH-LSTM occasionally produces forecasts that are just as accurate in certain periods.

RF

The RF model is another strong performer, balancing accuracy across all metrics. It achieves a relatively low RMSE (0.001236), making it the second-best model after HAR in terms of absolute error. It also records the lowest relative forecasting errors: MAPE (56.71%).

Its R^2 (0.3806) is slightly lower than HAR's but still much better than the deep learning and GARCH-based models. Additionally, its QLIKE (0.2051) is second only to HAR's, indicating that it does well in avoiding volatility underestimations.

The DM test results highlight RF's consistency, outperforming all models, meaning its forecasting errors are systematically smaller over time. The forecast plots show that RF adapts well to different volatility regimes, likely due to its non-parametric nature, which allows it to capture complex volatility structures flexibly.

5. Conclusion

Overall, the HAR and RF models emerge as the best options for forecasting Ethereum volatility, though each has distinct strengths. **HAR** provides the best overall goodness-of-fit, with the lowest RMSE, highest R^2, and the lowest QLIKE. It is particularly effective at capturing volatility persistence at multiple time scales, which contributes to its strong performance. On the other hand, **RF** excels in relative accuracy (lowest MAPE) and demonstrates the most consistent superiority in forecasting errors over time, as shown by its strong DM test results.

GARCH struggled with predictive accuracy, supporting the findings of Dudek et al. (2024). Its assumption of a well-defined conditional variance process limits adaptability to complex volatility structures.

The LSTM-based models underperformed, likely due to their tendency to smooth predictions and fail to capture volatility spikes effectively. This corroborates the discussions of Dudek et al. (2024), who suggests that the chaotic nature of cryptocurrency volatility and a lack of discernible patterns give the deep learning methodologies no clear advantage. Nevertheless, deep learning models may require additional features (e.g., news sentiment) to maximize their strengths.

In conclusion, traditional time-series models like HAR continue to be highly effective for volatility forecasting, while machine learning models like RF offer promising alternatives that balance flexibility and predictive strength.

Challenges encountered

Implementing a rolling window forecast was extremely challenging, given the need to pay attention to small details such as the order of adding new observations to training data, etc… In our initial implementation, the RF and HAR models reported suspiciously good results, with RMSE values being half of the other models such as GARCH and LSTM. We then noticed data leakage in the respective models and promptly appended the models.

The lack of publicly available datasets for ETH and other cryptocurrency pricing made it challenging to develop the models. One major issue is the inconsistency of the trading data, which means that the initial RV calculated was not accurate. This inconsistency also disrupts the training process, as some of the models might disproportionately weight dates with incomplete data, thus skewing the performance and reliability. We had to take extra precautions when it came to cleaning and preprocessing the ETH data, and we will take additional steps to prevent such things from happening in our individual components.

Areas for Improvement and Future Plans

Chan and Hefner (2019) concluded that online sentiment is crucial in predicting cryptocurrency volatility. However, we did not have sufficient time and resources to implement online sentiments to make a volatility forecast.

The nature of cryptocurrency means that it is extremely volatile, as shown by our results of ETH. We wish to explore this more by considering other models such as regime-switching, which will help capture abrupt changes in the cryptocurrency market.

Adopting a rolling window approach can potentially improve the forecast accuracy of the GARCH and GARCH-LSTM models, as the rolling window method allows for continuous updating of GARCH parameters and thus helps the models adapt to potential shifts in volatility regimes.

<u>References</u>

Amirshahi, B., & Lahmiri, S. (2023). Hybrid deep learning and garch-family models for forecasting volatility of Cryptocurrencies. *Machine Learning with Applications*, *12*, 100465. https://doi.org/10.1016/j.mlwa.2023.100465

Andersen, T. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, *61*(1), 43–76. https://doi.org/10.1016/s0304-405x(01)00055-1

Ardia, D., Bluteau, K., & Rüede, M. (2019). Regime changes in Bitcoin Garch volatility dynamics. *Finance Research Letters*, *29*, 266–271. https://doi.org/10.1016/j.frl.2018.08.009

Balcilar, M., Bouri, E., Gupta, R., & Roubaud, D. (2017). Can volume predict Bitcoin returns and volatility? A quantiles-based approach. *Economic Modelling*, *64*, 74–81. https://doi.org/10.1016/j.econmod.2017.03.019

Basher, S. A., & Sadorsky, P. (2022). Forecasting bitcoin price direction with random forests: How important are interest rates, inflation, and market volatility? *Machine Learning with Applications*, *9*, 100355. https://doi.org/10.1016/j.mlwa.2022.100355

Bergsli, L. Ø., Lind, A. F., Molnár, P., & Polasik, M. (2022). Forecasting volatility of bitcoin. *Research in International Business and Finance*, *59*, 101540. https://doi.org/10.1016/j.ribaf.2021.101540

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1

Brauneis, A., & Sahiner, M. (2024). Crypto Volatility Forecasting: Mounting a HAR, sentiment, and machine learning horserace. *Asia-Pacific Financial Markets*. https://doi.org/10.1007/s10690-024-09510-6

Chen, C. Y.-H., & Hafner, C. M. (2019). Sentiment-induced bubbles in the cryptocurrency market. *Journal of Risk and Financial Management*, *12*(2), 53. https://doi.org/10.3390/jrfm12020053

Chernoff, A., & Jagtiani, J. (2024). Beneath the crypto currents: The hidden effect of crypto "whales." *Working Paper (Federal Reserve Bank of Philadelphia)*. https://doi.org/10.21799/frbp.wp.2024.14

Corsi, F. (2008). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, *7*(2), 174–196. https://doi.org/10.1093/jjfinec/nbp001

Crypto market cap charts. CoinGecko. (n.d.-a). https://www.coingecko.com/en/global-charts

Diebold, F. X., & Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, *20*(1), 134–144. https://doi.org/10.1198/073500102753410444

Dockx, A., Drews, E., Kennedy, K., Rathjen, T., Sharma, N., Stepanian, J., & Toprak, B. (2024). The evolution of digital assets: Challenges, opportunities, and J.P. Morgan solutions. https://www.jpmorgan.com/content/dam/jpm/cib/complex/content/securities-services/regulatory-solutions/evolution-of-digital-assets.pdf

Dudek, G., Fiszeder, P., Kobus, P., & Orzeszko, W. (2024). Forecasting cryptocurrencies volatility using statistical and Machine Learning Methods: A Comparative Study. *Applied Soft Computing*, *151*, 111132. https://doi.org/10.1016/j.asoc.2023.111132

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987. https://doi.org/10.2307/1912773

Ethereum Price: ETH Live Price Chart, Market Cap & News Today. CoinGecko. (n.d.-b). https://www.coingecko.com/en/coins/ethereum

Fakhfekh, M., & Jeribi, A. (2020). Volatility dynamics of crypto-currencies' returns: Evidence from asymmetric and long memory GARCH models. *Research in International Business and Finance*, *51*, 101075. https://doi.org/10.1016/j.ribaf.2019.101075

Harris, L. (2002). *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press USA - OSO.

Johntwk. (n.d.). *Diebold-mariano-test/dm_test.py at master · johntwk/diebold-mariano-test*. GitHub. https://github.com/johntwk/Diebold-Mariano-Test/blob/master/dm_test.py

Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. *Expert Systems with Applications*, *103*, 25–37. https://doi.org/10.1016/j.eswa.2018.03.002

Partner. (2025, January 23). *The role of Ethereum in Africa's emerging Tech Ecosystem*. TechCabal. https://techcabal.com/2025/01/23/the-role-of-ethereum-in-africas-emerging-tech-ecosystem/

Raza, S. A., Khan, K. A., Guesmi, K., & Benkraiem, R. (2023). Uncertainty in the financial regulation policy and the boom of Cryptocurrencies. *Finance Research Letters*, *52*, 103515. https://doi.org/10.1016/j.frl.2022.103515

Shen, Z., Wan, Q., & Leatham, D. J. (2021). Bitcoin return volatility forecasting: A comparative study between Garch and RNN. *Journal of Risk and Financial Management*, *14*(7), 337. https://doi.org/10.3390/jrfm14070337

A short history of ethereum. Consensys. (2019, May 13). https://consensys.io/blog/a-short-history-of-ethereum

Urgenc, O. A., & Tas, O. (2024). Social media sentiment and its effects on cryptocurrency price volatility. *Pressacademia*. https://doi.org/10.17261/pressacademia.2024.1915

Yae, J., & Tian, G. Z. (2022). Out-of-sample forecasting of cryptocurrency returns: A comprehensive comparison of predictors and algorithms. *Physica A: Statistical Mechanics and Its Applications*, *598*, 127379. https://doi.org/10.1016/j.physa.2022.127379