ISSN 1822-8038 (online)
INTELEKTINĖ EKONOMIKA
INTELLECTUAL ECONOMICS
2025, No. 19(1), p. 62–91

DURING EXTREME MARKET CONDITIONS: A COMPARATIVE STUDY OF ADVANCED GARCH SPECIFICATIONS

Tomas Pečiulis
Vilnius Gediminas Technical University, Saulėtekio al. 11, Vilnius, Lithuania, tomas.
peciulis@vilniustech.lt
https://orcid.org/0009-0006-8837-6580

Asta Vasiliauskaitė
Mykokolo Romerio University, Ateities st. 20, Vilnius, Lithuania,
avasil@mruni.lt
https://orcid.org/0000-0003-3483-070X

DOI:10.13165/IE-25-19-1-03

Abstract

Purpose. This study investigates the performance of advanced GARCH-family models in forecasting cryptocurrency volatility during extreme market conditions. It aims to determine whether volatility in major digital assets is primarily endogenous and asset-specific, and to identify which modelling approaches offer superior predictive accuracy across different volatility regimes.

Methodology. Using a Leave-One-Crisis-Out (LOCO) cross-validation framework, we evaluate five GARCH-type models—GARCH, EGARCH, GJR-GARCH, CS-GARCH, and MS-GARCH—across three major cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), and Binance Coin (BNB). High-volatility episodes are identified using realised volatility thresholds above the 90th percentile, and hyperparameters are optimised via grid search. The analysis spans daily data from August 2017 to June 2025, focusing exclusively on threshold-based filtering after empirical evidence showed limited alignment between crypto volatility and traditional financial stress indicators such as the VIX.

Findings. The results reveal that cryptocurrency volatility is largely endogenous and decoupled from traditional market stress, supporting the "crypto exceptionalism" hypothesis. Volatility patterns are highly asset-specific, reflecting the distinct roles and market structures of each cryptocurrency. Regime-switching (MS-GARCH) and component (CS-GARCH) models consistently outperform traditional specifications in forecasting accuracy, particularly for BTC and ETH. However, convergence issues in MS-GARCH for BNB highlight the need for model-specific diagnostics and asset-tailored approaches.

Originality. This is the first study to apply a LOCO-based stress-testing framework to evaluate GARCH-family models under extreme cryptocurrency market conditions. It contributes to the literature by demonstrating the limitations of traditional financial stress indicators in crypto volatility modelling, highlighting the heterogeneous nature of digital asset dynamics, and offering a robust methodology for volatility forecasting in high-risk, rapidly evolving markets during extreme market conditions.

Keywords: Cryptocurrency volatility, GARCH models, Regime-switching, Volatility forecasting, Extreme market conditions.

JEL index: C22, C53, G17

1. Introduction

The cryptocurrency market's emergence as a distinct asset class has fundamentally challenged traditional approaches to volatility modeling and risk assessment. While conventional financial theory suggests that asset volatility should exhibit co-movement during periods of market stress, cryptocurrency markets have repeatedly demonstrated their capacity to experience extreme volatility episodes that are largely decoupled from traditional financial market indicators. This phenomenon was starkly illustrated during the March 2020 COVID-19 crash, when Bitcoin plummeted over 50% in a single day—from approximately $8,000 to $3,800—yet subsequent cryptocurrency-specific crises, such as the Terra/LUNA collapse in May 2022 and the FTX bankruptcy in November 2022, occurred independently of broader financial market stress.

Unlike traditional assets, cryptocurrency volatility is largely decoupled from conventional market stress indicators such as the VIX index, credit spreads, or equity market volatility. This disconnect suggests that the standard toolkit of volatility forecasting models, developed and validated on traditional financial assets, may be inadequate for capturing the unique dynamics of digital asset markets. The conventional approach of using broad-based financial stress indicators to identify extreme market conditions—while effective for equity and bond markets—fails to account for the idiosyncratic nature of cryptocurrency volatility clustering, which is often driven by protocol upgrades, exchange failures, regulatory announcements, and ecosystem-specific events rather than macroeconomic fundamentals.

The challenge of forecasting cryptocurrency volatility during extreme market conditions is further complicated by the heterogeneous nature of different digital assets. While Bitcoin's volatility patterns may reflect its role as a "digital gold" and store of value, Ethereum's volatility is influenced by its position as the backbone of decentralized finance (DeFi), and exchange tokens like Binance Coin (BNB) exhibit volatility patterns tied to platform-specific developments. This asset-specific heterogeneity necessitates a departure from the one-size-fits-all approach commonly employed in traditional finance, where correlation structures remain relatively stable across different equity securities or sovereign bonds.

In response to these challenges, this study employs asset-specific extreme period identification to reflect the unique nature of crypto markets. Rather than relying on external stress indicators or predetermined crisis dates, we identify high-volatility episodes using realized volatility thresholds calibrated to each cryptocurrency's individual return distribution. This approach recognizes that what constitutes an "extreme" market condition varies significantly across different digital assets and may not coincide with stress in traditional financial markets. For Bitcoin, Ethereum, and Binance Coin, we identify the three longest high-volatility periods occurring after 2019, ensuring our analysis captures recent market dynamics while providing sufficient variation in market conditions for robust model comparison.

The methodological contribution of this research lies in our application of Leave-One-Crisis-Out (LOCO) cross-validation, a technique specifically designed for stress-testing financial models under realistic forecasting constraints. Our LOCO validation reveals model performance across different cryptocurrency volatility regimes by systematically excluding each identified high-volatility episode from the training data and using it exclusively for out-of-sample evaluation. This approach simulates the realistic scenario faced by risk managers and portfolio optimizers, where future market stress episodes are unknown during model estimation. Unlike traditional time-series cross-validation techniques that may inadvertently incorporate look-ahead bias, LOCO ensures that model performance is evaluated on genuinely unseen data, providing more reliable guidance for practical applications.

The empirical analysis encompasses six advanced GARCH specifications: AR-GARCH, MS-GARCH, EGARCH, GJR-GARCH, FIGARCH, and CS-GARCH. These models represent the current state-of-the-art in volatility modelling, incorporating features such as asymmetric responses to positive and negative shocks (EGARCH, GJR-GARCH), long-memory dynamics (FIGARCH), regime-switching behaviour (MS-GARCH), and component structures that separately model short-term and long-term volatility components (CS-GARCH). The comparative evaluation of these specifications provides insights into which modelling approaches are most effective for capturing the complex volatility dynamics exhibited by cryptocurrency markets during periods of extreme stress.

Research Problem. The central research problem addresses a fundamental gap in volatility forecasting literature: existing GARCH-type models, developed and validated on traditional financial assets, may be inadequate for cryptocurrency markets due to their unique volatility dynamics and decoupling from conventional financial stress indicators. This creates a critical challenge for risk managers, portfolio optimizers, and regulators who require accurate volatility forecasts during the most turbulent market conditions—precisely when model performance matters most. The problem is compounded by the heterogeneous nature of different cryptocurrencies, each exhibiting distinct volatility patterns that may require asset-specific modelling approaches.

Specifically, the research problem can be formulated as follows: How do advanced GARCH specifications perform in forecasting cryptocurrency volatility during extreme market conditions, and which modelling approaches demonstrate superior predictive accuracy across different digital assets and volatility regimes?

To address this research problem systematically, we investigate three interrelated research questions:

RQ1: Endogeneity of Cryptocurrency Volatility *To what extent is cryptocurrency volatility endogenous and driven by crypto-specific factors rather than traditional financial market stress indicators?*

This question examines the fundamental assumption underlying most volatility forecasting frameworks—that asset volatility should exhibit co-movement during market stress. If cryptocurrency volatility is primarily driven by idiosyncratic factors (protocol changes, exchange failures, regulatory announcements), then models incorporating traditional financial stress indicators may provide little forecasting value.

RQ2: Asset-Specific Volatility Dynamics *Do Bitcoin, Ethereum, and Binance Coin exhibit unique extreme volatility periods that reflect their distinct market roles and underlying fundamentals?*

This question investigates whether the "one-size-fits-all" approach common in traditional finance is appropriate for cryptocurrency markets. Different digital assets serve different functions (store of value, smart contract platform, exchange utility token), which may result in distinct volatility patterns requiring asset-specific modelling approaches.

RQ3: Cross-Asset Model Robustness *Which GARCH-type specifications demonstrate consistent forecasting performance across different cryptocurrencies and volatility regimes?*

This question addresses the practical challenge of model selection in cryptocurrency risk management. Given the computational and operational costs of maintaining multiple modelling frameworks, identifying specifications that perform robustly across different assets and market conditions is crucial for practical implementation.

The findings of this study have significant implications for multiple stakeholders in the evolving cryptocurrency ecosystem. For academic researchers, our results contribute to the growing literature on digital asset pricing and risk modelling by providing evidence on the relative efficacy of different volatility forecasting approaches during extreme market conditions. For practitioners, including portfolio managers, risk officers, and proprietary trading firms, our analysis offers practical guidance on model selection for volatility forecasting in cryptocurrency markets. For regulators and policymakers, our findings inform the development of appropriate risk assessment frameworks for digital assets, highlighting the inadequacy of traditional financial stability indicators for monitoring cryptocurrency market stress.

The remainder of this paper is organised as follows. Section 2 reviews the relevant literature on cryptocurrency volatility modelling. Section 3 describes our data, the identification of extreme market conditions, and the LOCO cross-validation methodology, which provides the GARCH model's specification. Section 4 reports our empirical findings, including model performance comparisons and convergence diagnostics. Section 5 discusses the implications of our results for theory and practice, Section 6 provides limitations, while Section 7 concludes with suggestions for future research directions.

2. Literature Review

The unique volatility dynamics of cryptocurrencies, characterised by extreme jumps often decoupled from traditional financial stress indicators, present significant challenges for forecasting models developed primarily for traditional assets (Ahmed et al., 2024; Naimy et al., 2021; Chernova et al., 2024). While GARCH family models remain the dominant econometric framework for modelling volatility clustering in financial time series (Bollerslev, 1986), their application and comparative performance, specifically during cryptocurrency-specific extreme events, require deeper investigation, particularly across diverse digital assets like BTC, ETH, and Binance BNB. This review critically examines the existing literature, highlighting key advancements and persistent gaps that motivate the present study›s focus on asset-specific extreme period identification and rigorous stress-testing via LOCO validation.

Research consistently confirms that cryptocurrency volatility exhibits distinct characteristics compared to traditional assets, including pronounced long memory, asymmetric leverage effects (often inverse), and heavy-tailed return distributions (Subramoney et al., 2025; Su, 2014). This distinctiveness challenges the assumption that traditional financial stress indicators like the VIX effectively capture extreme conditions in crypto markets. Studies examining volatility co-movement, such as Tzeng & Su, (2024); Ullah et al., (2020), find that while some U.S. macroeconomic indicators (e.g., consumer confidence, CPI) possess predictive power for crypto volatility, their influence is often asset-specific and period-dependent, particularly strengthening post-COVID-19. Crucially, Naimy et al., (2021) demonstrated significant differences in optimal GARCH specifications (e.g., IGARCH for world currencies vs. GJR-GARCH/CGARCH for cryptos) and highlighted the poor VaR performance of even "optimal" models for major coins like Bitcoin and Ripple during stress periods, underscoring the inadequacy of traditional frameworks. This supports the core premise of the current study that cryptocurrency volatility during crises is frequently endogenous, driven by ecosystem-specific events (exchange failures, protocol changes, regulatory shocks) rather than broad financial market stress.

Consequently, defining "extreme market conditions" for cryptocurrencies requires methodologies beyond conventional index-based filtering. While some studies incorporate realised volatility thresholds (Ampountolas, 2022), they often lack alignment with identifiable crypto-specific crises or fail to adopt rigorous out-of-sample testing protocols designed *specifically* for stress periods. The common practice of including crisis periods within the training data or using standard rolling-window back tests introduces look-ahead bias and fails to simulate the true forecasting challenge faced by risk managers when novel, unforeseen crises erupt (Dudek et al., 2024; Queiroz & David, 2023). The LOCO methodology, while established in some traditional finance contexts for regulatory stress-testing (e.g., Basel frameworks), remains notably underutilised in cryptocurrency volatility forecasting research. Its application, as proposed in the current study, systematically isolates each major crisis for exclusive out-of-sample testing, represents a significant methodological advancement for assessing genuine model robustness under duress.

Comparative evaluations of GARCH-type models for crypto volatility are abundant but exhibit critical limitations relevant to this study. Several focus predominantly on Bitcoin (Wang, 2023; Pourrezaee & Hajizadeh, 2024; Boozary et al., 2025), neglecting the heterogeneous volatility dynamics of other major assets like Ethereum (the DeFi backbone) and BNB (an exchange token tied to platform-specific events). Studies incorporating multiple assets, such as Naimy et al. (2021) and Ampountolas (2022), often report divergent "best" models (e.g., IGARCH for Monero, GJR-GARCH/CGARCH for others) but typically evaluate performance over *general* periods, not explicitly contrasting model efficacy *during* identified extreme events. Furthermore, while advanced specifications like MS-GARCH (regime-switching), CS-GARCH (component structure), and FIGARCH (long memory) are recognised for their theoretical suitability to capture crypto stylised facts (Subramoney et al., 2025; Su, 2014), their empirical performance and crucially, their *estimation robustness* (convergence stability) during extreme crypto-specific stress, particularly for assets beyond Bitcoin, are insufficiently explored. Studies like Queiroz et al. (2023) advocate for Realised-GARCH using intraday data in out-of-sample contexts, while Subramoney et al. (2025) highlight FIAPARCH with heavy-tailed innovations, yet neither explicitly tests these models through isolated crisis episodes using a LOCO framework. The frequent convergence failures of complex models like MS-GARCH for certain assets, noted anecdotally but rarely systematically diagnosed or reported (as planned in the current study for BNB), represent a significant practical hurdle overlooked in much literature.

The rise of machine learning (ML) and deep learning (DL) models offers alternative forecasting avenues (Zubair et al., 2024; Zhang et al., 2021; Abarghouie et al., 2024; Rodrigues & Machado, 2025). While demonstrating promising accuracy, often surpassing GARCH in some studies (Wang, 2023; Dudek et al., 2024; AlMadany et al., 2024), these "black-box" models pose interpretability challenges crucial for risk management and regulatory compliance. Hybrid approaches combining econometric foundations with ML (e.g., GARCH-LSTM: AlMadany et al., 2024) are emerging but add complexity. Furthermore, ML/DL studies often rely on vast datasets and computational resources, and crucially, their performance evaluation during *isolated, unforeseen extreme events* using methodologies like LOCO is equally scarce. Bibliometric analyses by Pečiulis et al., (2024) and Ruiz Roque da Silva et al., (2022) confirm the dominance of volatility forecasting and machine learning themes but also highlight the persistent focus on major coins and the need for more sophisticated validation techniques.

Collectively, the literature reveals a critical gap: a lack of rigorous, comparative assessment of advanced GARCH specifications *specifically during isolated, cryptocurrency-defined extreme market conditions* across the major yet functionally distinct assets BTC, ETH, and BNB. Existing studies often rely on conventional crisis definitions or index thresholds, inadequately test models on genuinely unseen crisis data (lacking robust methodologies like LOCO), and insufficiently address the practical challenges of model convergence and stability during stress, especially for complex models and less-dominant coins like BNB. This study directly addresses this gap by: (1) employing an asset-specific, realized-volatility and event-aligned definition of extreme periods; (2) implementing a stringent LOCO cross-validation framework to ensure pure out-of-sample testing during crises; and (3) providing a comprehensive performance and robustness (including convergence diagnostics) evaluation of six advanced GARCH models (AR-GARCH, MS-GARCH (Haas et al., 2004), EGARCH (Nelson, 1991), GJR-GARCH (Glosten et al., 1993), FIGARCH (Baillie et al., 1996), CS-GARCH (Conrad & Loch, 2015)) across these three key cryptocurrencies during their most significant post-2019 stress events. This approach offers novel insights into model suitability for practical risk management under the most challenging conditions unique to the cryptocurrency market.

3.Methodology

This section outlines the empirical strategy employed to evaluate the performance of various GARCH-type models in forecasting cryptocurrency volatility. The methodological framework is designed to ensure consistency, robustness, and comparability across assets and model specifications. It begins with a detailed description of the data sources, preprocessing steps, and return transformations. Subsequently, the section introduces the suite of volatility models under consideration, including both standard and advanced GARCH variants. Each model is estimated using maximum likelihood techniques, and their predictive accuracy is assessed through a rigorous out-of-sample evaluation.

3.1. Data description

Daily closing prices for BTC, ETH, and BNB were retrieved from Yahoo Finance using the yfinance Python package. The sample period extends from 11 August 2017 to 31 December 2024. This timeframe was deliberately selected to ensure the inclusion of all three cryptocurrencies from their earliest common availability, thereby maximizing data coverage while preserving consistency across assets. Although Bitcoin has a longer trading history, aligning the start date with the earliest available data for BNB mitigates survivorship bias and enhances the comparability of results across assets, as emphasized in prior literature (Grobys & Sapkota, 2020; Stambaugh, 2011).

To prepare the data for volatility modelling, daily prices were transformed into continuously compounded returns using the natural logarithm of price relatives. This transformation is standard in financial econometrics, as it stabilizes the variance and facilitates the assumption of conditional normality in GARCH-type models (Tsay, 2010). The resulting return series were then used to compute realized volatility, which serves as a non-parametric benchmark for evaluating model performance. Realized volatility was estimated using a 30-day rolling window of standard deviations, scaled by the square root of 30 to annualize the measure. Formally, for each day $\hat{\sigma}_t$, the realized volatility is computed as:

$$\hat{\sigma}_t = \sqrt{30} \cdot \mathrm{SD}(r_{t-29},\ldots,r_t)$$

where r_t denotes the daily log return. This approach captures the local variability in returns and provides a robust proxy for latent volatility, which is particularly valuable in assessing the out-of-sample forecasting accuracy of competing GARCH specifications.

For model estimation and evaluation, the dataset was partitioned into a training set comprising 80% of the observations and a test set comprising the remaining 20%. This division facilitates both in-sample calibration and out-of-sample validation, allowing for a rigorous comparison of model performance under realistic forecasting conditions. The two best-performing models, as determined by their average RMSE, MAE, and AIC across assets and periods, were subjected to deeper analysis in the test set to assess their robustness and predictive reliability.

Table 1. Major Cryptocurrency Market Events (2020–2025)

Period	**Event Description**	**Market Impact**
March 12–13, 2020	COVID-19 Crash ("Black Thursday")	Bitcoin fell over 50% in a single day, from approximately \$8,000 to \$3,800, amid a global liquidity crisis and widespread asset sell-offs.
May 2022	Terra/LUNA Collapse	The algorithmic stablecoin UST lost its peg, triggering a "death spiral" that wiped out over \$50 billion in market value and destabilized the broader DeFi ecosystem
November 2022	FTX Collapse	One of the largest centralized crypto exchanges filed for bankruptcy, leading to sharp declines in BTC and ETH prices due to contagion fears and loss of investor confidence.
February 2024	Mt. Gox Repayment & Hack Fallout	The long-awaited release of BTC from Mt. Gox repayments, combined with a major hack, led to panic selling and a sharp market correction.
Q1 2025	Bybit Hack, ETF Outflows, and Fed Policy Uncertainty	BTC dropped from \$109,600 to \$74,500 (−32%) amid a confluence of negative catalysts, including a major exchange hack, institutional ETF outflows, and macroeconomic uncertainty.

Source: The authors

The events listed in the Table 1 were selected based on their systemic relevance, magnitude of market disruption, and alignment with the literature on financial contagion, structural breaks, and crisis-induced volatility. Each event represents a significant exogenous shock to the cryptocurrency ecosystem, either through macro-financial contagion (e.g., COVID-19), protocol failure (e.g., Terra/LUNA), institutional collapse (e.g., FTX), or large-scale asset redistribution (e.g., Mt. Gox repayments). These events are consistent with the typology of financial crises described by Reinhart & Rogoff, (2009), and their inclusion is motivated by the need to assess model robustness under extreme market conditions.

The Terra/LUNA collapse in particular has been extensively analysed in academic literature as a case of financial fragility in decentralised systems. Liu et al., (2023)describe it as the first major run in crypto markets, highlighting the role of pseudonymous transparency and reflexive investor behaviour in accelerating the collapse. Similarly, the FTX bankruptcy in late 2022 marked a critical turning point in centralised exchange risk, with widespread implications for market structure and investor trust (Fu et al., 2023).

To evaluate the robustness of volatility forecasting models under extreme market conditions, this study adopts a LOCO framework. LOCO is a variant of cross-validation specifically designed for time series and financial stress testing. In this approach, each identified high-volatility period—interpreted as a "crisis"—is systematically excluded from the training data and used exclusively for out-of-sample testing. This ensures that the model is not exposed to any information from the crisis period during training, thereby simulating a realistic forecasting scenario where future shocks are unknown.

Table 2. LOCO Training and Test Periods for Longest High-Volatility Episodes (2019–2025)

Asset	Training Start	Training End	Training Duration (days)	Start Date	End Date	Test Duration
BTC	2017-08-01	2021-02-04	1,283 days	2021-02-05	2021-07-12	158 days
	2017-08-01	2020-03-11	953 days	2020-03-12	2020-04-17	37 days
	2017-08-01	2020-09-11	1,137 days	2020-09-12	2020-10-05	24 days
ETH	2017-08-01	2021-05-14	1,382 days	2021-05-15	2021-07-08	55 days
	2017-08-01	2021-01-20	1,268 days	2021-01-21	2021-03-09	48 days
	2017-08-01	2019-06-26	694 days	2019-06-27	2019-08-12	47 days
BNB	2017-08-01	2021-01-02	1,250 days	2021-01-03	2021-02-19	48 days
	2017-08-01	2021-05-13	1,381 days	2021-05-14	2021-06-29	47 days
	2017-08-01	2020-03-11	953 days	2020-03-12	2020-04-17	37 days

Source: The authors

We identify the three longest high-volatility periods for each of the three major cryptocurrencies: BTC, ETH, and BNB, using realised volatility thresholds above the 90th percentile. The analysis is restricted to the post-2019 period to reflect recent market dynamics. For each high-volatility episode, the training set begins at the start of the dataset (August 1, 2017) and ends one day prior to the onset of the crisis. For example, the longest BTC high-volatility period spans from February 5, 2021, to July 12, 2021, lasting 158 days, with a corresponding training period of 1,283 days ending on February 4, 2021. Similarly, ETH experienced a 55-day high-volatility stretch from May 15 to July 8, 2021, preceded by a 1,382-day training window. BNB's longest episode occurred from January 3 to February 19, 2021, with a 48-day test period and a 1,250-day training span (See Table 2).

We did not include event-based or index-based filtering since crypto markets have idiosyncratic volatility patterns unrelated to traditional financial stress (see results section).

This LOCO setup allows for rigorous out-of-sample validation, ensuring that model performance is assessed on genuinely unseen data. It also facilitates comparative analysis across assets and timeframes, offering insights into how models generalise across different types of market stress.

Despite the theoretical appeal of incorporating event-based filtering into the modelling framework, our empirical analysis revealed a lack of consistent overlap between these major market events and the periods of statistically defined high volatility. Specifically, realized volatility—computed using a 30-day rolling standard deviation—did not consistently spike in alignment with the event dates across all assets. This suggests that cryptocurrency volatility is often driven by idiosyncratic, endogenous factors rather than by identifiable macro or institutional shocks.

This finding aligns with recent research emphasising the unique behavioural and structural dynamics of crypto markets, where volatility is frequently decoupled from traditional financial stress indicators (Corbet et al., 2019; Y. Liu et al., 2022). As a result, we opted not to include event-based filtering in the final model architecture. Instead, we relied on a data-driven approach using realised volatility thresholds to define high-volatility regimes. This ensures that the model evaluation remains grounded in observable market behaviour rather than ex-post event classification, thereby enhancing the objectivity and replicability of our results.

2.2 Model architecture

In preparation for model estimation, a comprehensive suite of diagnostic tests was conducted to assess the statistical properties of the return series (See Table 3). Stationarity was evaluated using the ADF test, which confirmed the absence of unit roots in all series, thereby validating the use of GARCH-type models. To examine the presence of volatility clustering—a hallmark of financial time series—the Ljung-Box test was applied to squared returns with 10 lag, complemented by the ARCH-LM test, which specifically tests for autoregressive conditional heteroskedasticity. Both tests provided strong evidence of time-varying volatility, justifying the use of conditional variance models.

Table 3. Data diagnostic tests

Diagnostic Objective	Test Applied	Reference	Notes
Stationarity	Augmented Dickey-Fuller (ADF) Test	(Dickey & Fuller, 1979)	Assesses the presence of unit roots in return series
Autocorrelation	Ljung-Box Test (10 lags) on squared returns	(Ljung & Box, 1978)	Detects autocorrelation in squared returns
Volatility Clustering	ARCH-LM Test (10 lags)	(Engle, 1982)	Tests for autoregressive conditional heteroskedasticity
Normality	Jarque-Bera Test	(Jarque & Bera, 1987)	Evaluates skewness and kurtosis relative to normal distribution
Structural Breaks	Mean 30-day rolling variance		Identifies multiple structural changes in the variance

Source: The authors

Normality was assessed using the Jarque-Bera test (Jarque & Bera, 1987), which revealed significant departures from Gaussianity, consistent with the heavy tails and skewness commonly observed in cryptocurrency returns. Structural breaks were investigated using the Mean 30-day rolling variance identified several statistically significant shifts in the return-generating process, further motivating the inclusion of regime-switching and component-based GARCH models in the analysis.

To estimate the suite of GARCH-type models, this study employed the rug arch package in R and the arch package in Python, which provides a flexible and robust framework for specifying, estimating, and forecasting a wide range of univariate GARCH models. The mean equation was modelled using an ARMA(p,q) process by checking different combinations of hyperparameters (up till 10^{th} lag). This approach ensures that the conditional mean dynamics are adequately captured prior to modelling the conditional variance.

For the volatility component, a grid search was conducted over ARCH and GARCH lag orders, with p,q∈[1,5]$p,q\in[1,5]$, to identify the optimal lag structure for each model specification. This exhaustive search strategy allows for a systematic exploration of the model space while maintaining computational feasibility. To account for the heavy tails and potential asymmetry commonly observed in cryptocurrency return distributions, four alternative error distributions were considered: normal, Student's t, skewed Student's t, and the generalised error distribution (GED). This flexibility in distributional assumptions enhances the robustness of the volatility forecasts.

Model in-sample selection was guided by the Akaike Information Criterion (AIC). The criteria weas computed for each candidate model, and the specification with the lowest AIC was retained for further analysis. To evaluate out-of-sample forecasting performance, a rolling window cross-validation procedure was implemented. Specifically, models were re-estimated over a moving training window, and forecasts were generated for the subsequent period. Forecast accuracy was assessed using two standard loss functions: the root mean squared error (RMSE) and the mean absolute error (MAE). This rolling evaluation framework provides a realistic approximation of real-time forecasting conditions and allows for a robust comparison of model performance across different market regimes (see Table 4).

Table 4. Model Estimation and Validation Procedures

Component	**Methodology**	**Reference(s)**
Estimation Framework	rug arch package in R arch package in python	(Ghalanos, 2022)
Mean Equation	ARMA(p,q) orders selected via	
Volatility Equation	Grid search over ARCH/GARCH lags p,q∈[1,5]p,q∈[1,5]	—
Error Distributions	Normal, Student's t, Skewed t, GED	—
Model Selection	AIC for balancing fit and parsimony	(Akaike, 1974)

Source: The authors

Together, these procedures ensure that the model estimation process is both statistically rigorous and empirically grounded, aligning with best practices in financial econometrics and the specific challenges posed by high-frequency cryptocurrency data.

We use this GARCH model:

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

where:
σ_t^2 is the conditional variance at time *t*;
Ω is a constant volatility baseline;
α: ARCH term;
β: GARCH term (persistence of volatility);

The GARCH model extends the ARCH framework by incorporating lagged conditional variance, enabling parsimonious modelling of volatility clustering. Defined by Equation (1), it captures persistence in volatility through the GARCH term (β) and responsiveness to shocks via the ARCH term (α). The intercept (ω) represents baseline volatility. While computationally efficient and widely applicable, GARCH assumes symmetric responses to positive and negative shocks, limiting its utility in markets with leverage effects. It remains a baseline choice for volatility forecasting in the absence of asymmetry or structural breaks.

We used this EGARCH model:

$$ln(\sigma_t^2) = \omega + \alpha \left|\frac{\varepsilon_{-1}}{\sigma_{t-1}}\right| + \gamma \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}}\right) + \beta\, ln(\sigma_{t-1}^2) \tag{2}$$

where:
γ is a leverage effect term (asymmetric response to shocks).

The EGARCH model, formalized in Equation (2), introduces asymmetry via the leverage parameter (γ), which differentiates the impact of positive and negative shocks on volatility. By modelling the logarithm of conditional variance, EGARCH ensures non-negativity without parameter constraints. This model is particularly suited for financial markets where "bad news" amplifies volatility more than "good news." However, the log-transformation complicates direct interpretation of coefficients, and estimation requires robust numerical methods.

We used this GJR-GARCH model:

$$\sigma_t^2 = \omega + (\alpha + \gamma\, I_{\varepsilon_{t-1}<0})\varepsilon_{t-1}^2 + \beta\, \sigma_{t-1}^2 \tag{3}$$

where:
ε_{t-1}^2 is a dummy = 1 if $\varepsilon_{t-1} < 0$, else 0.

The GJR-GARCH in Equation (3) incorporates asymmetry through a dummy variable, which activates an additional volatility response (γ) to negative shocks. This threshold-based approach explicitly quantifies the differential impact of market downturns, making it ideal for equity or crisis-prone markets. While intuitive, GJR-GARCH may overfit in small samples due to its discrete treatment of shocks and is less flexible than EGARCH in capturing smooth asymmetry.

We used this FIGARCH model:

$$(1 - \beta L)\sigma_t^2 = \omega + (1 - \beta L - \alpha L)(1 - L)^d \varepsilon_t^2 \tag{4}$$

where:
L is a lag operator;
D is a fractional integration parameter.

The FIGARCH in Equation (4) addresses long memory in volatility by employing a fractional differencing parameter (d) within the lag operator (L) framework. This allows volatility shocks to decay hyperbolically rather than exponentially, accommodating prolonged persistence observed in macroeconomic or commodity markets. However, its computational complexity and sensitivity to misspecification of the fractional parameter limit its practicality for high-frequency data.

We used this MSGARCH model:

$$\sigma_{\mathrm{t}}^2 = \omega_{s_t} + \alpha_{s_t}\varepsilon_{t-1}^2 + \beta_{s_t}\sigma_{t-1}^2 \quad (5)$$

where:
s_t It is an unobserved state (regime) at *t*.

MSGARCH in Equation (5) allows parameters (ω,α,β) to shift across unobserved regimes, capturing abrupt volatility changes caused by structural breaks or policy shifts. This model is indispensable for analyzing crises or regime-dependent markets but demands large datasets for stable regime identification and imposes heavy computational burdens due to latent state estimation.

We used this CS-GARCH model:

$$\sigma_{\mathrm{t}}^2 = q_{\mathrm{t}} + \alpha\,(\varepsilon_{\mathrm{t}-1}^2 - q_{\mathrm{t}-1}) + \beta\,(\sigma_{\mathrm{t}-1}^2 - q_{\mathrm{t}-1}) \quad (6)$$

$$q_{\mathrm{t}} = \omega + \rho\, q_{\mathrm{t}-1} + \phi\,(\varepsilon_{\mathrm{t}-1}^2 - \sigma_{\mathrm{t}-1}^2) \quad (7)$$

where:
q_{t} It is a long-run volatility component.

CS-GARCH in Equations (6)-(7) decomposes volatility into transient and persistent components, where q_{t} evolves via a separate autoregressive process. This separation enhances forecasting accuracy for long-term volatility trends, such as inflation or interest rates. However, the dual-equation structure increases model complexity and estimation time.

4. Results

This section presents the empirical findings from our comparative evaluation of advanced GARCH-family models applied to cryptocurrency volatility forecasting during extreme market conditions. Using a comprehensive dataset spanning August 2017 to June 2025, we assess model performance across four major cryptocurrencies—BTC, ETH, and BNB. To ensure methodological consistency and computational feasibility, ARMA mean dynamics were varied only for standard GARCH models, while more complex specifications were held constant in their mean structure. The analysis focuses on out-of-sample forecasting accuracy during high-volatility episodes identified through LOCO framework.

This approach enables a rigorous, asset-specific comparison of model performance under stress, offering insights into the relative strengths and limitations of traditional, asymmetric, component, and regime-switching volatility models in the context of cryptocurrency markets.

Figure 1 depicts the realised volatility of BTC from August 2017 to June 2025, overlaid with the VIX index to contextualise global market sentiment. The 90th percentile threshold for BTC volatility, calculated at approximately 0.37, is used to demarcate extreme volatility regimes. Periods exceeding this threshold are shaded in red, while more stable periods are shaded in blue. Major crypto-specific events are annotated with triangular markers and rotated labels for clarity. The VIX index is plotted on a secondary y-axis, and its mean level across the sample is approximately 19.47.

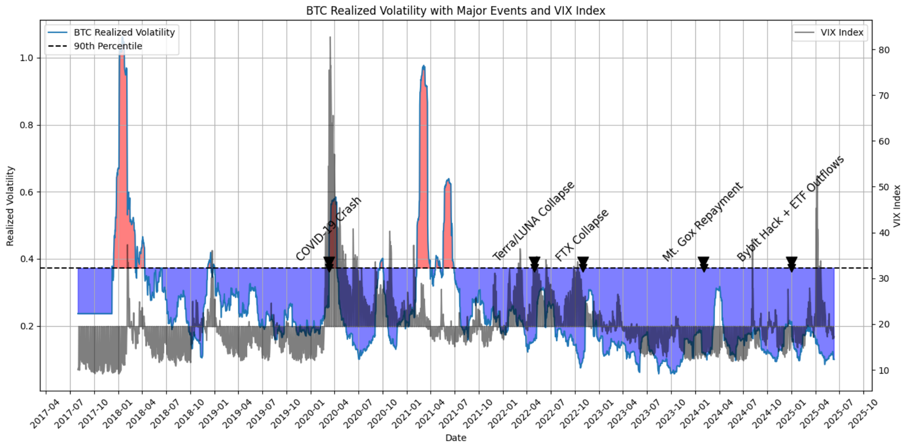

Figure 1. BTC Realised Volatility with Major Events and VIX index

Source: The authors

A key observation is the lack of consistent overlap between spikes in BTC volatility and peaks in the VIX index. The only clear exception is the COVID-19 crash in March 2020. On March 12, 2020, BTC realised volatility surged to 0.48, the highest in the dataset, while the VIX simultaneously peaked at 75.47—its highest level since the 2008 financial crisis. This synchronous spike reflects a rare convergence of systemic financial stress and crypto market panic, likely driven by global liquidity shocks and widespread deleveraging.

In contrast, other major crypto events such as the Terra/LUNA collapse (May 2022), the FTX bankruptcy (November 2022), and the Mt. Gox repayment (February 2024) did not elicit comparable volatility responses in BTC. For instance, during the Terra/LUNA collapse, BTC volatility reached only 0.16—well below the extreme threshold—while the VIX remained at its long-term average of 19.47. Similarly, the FTX collapse saw BTC volatility at 0.10 and the VIX at 25.81, indicating a modest reaction in both crypto and traditional markets. The muted volatility response to these events suggests that, despite their significance within the crypto ecosystem, they did not trigger broader systemic risk or investor panic.

The Bybit hack and ETF outflows in early 2025 also failed to produce a substantial volatility spike, with BTC volatility at 0.21 and the VIX remaining anchored at 19.47. This further supports the inference that BTC volatility is increasingly decoupled from isolated crypto events, possibly due to improved market maturity, deeper liquidity, or more robust investor expectations.

An important historical anomaly was observed in early 2018. Between January and February 2018, BTC realised volatility experienced a pronounced spike, exceeding 0.40. This period corresponds to the aftermath of the 2017 bull market and the subsequent regulatory crackdown in Asia, particularly China and South Korea. The sharp correction in BTC prices, combined with heightened regulatory uncertainty, likely contributed to this volatility surge. Notably, the VIX index during this period remained subdued, averaging around 15–20, indicating that the volatility was confined to the crypto domain and not reflective of broader financial market stress.

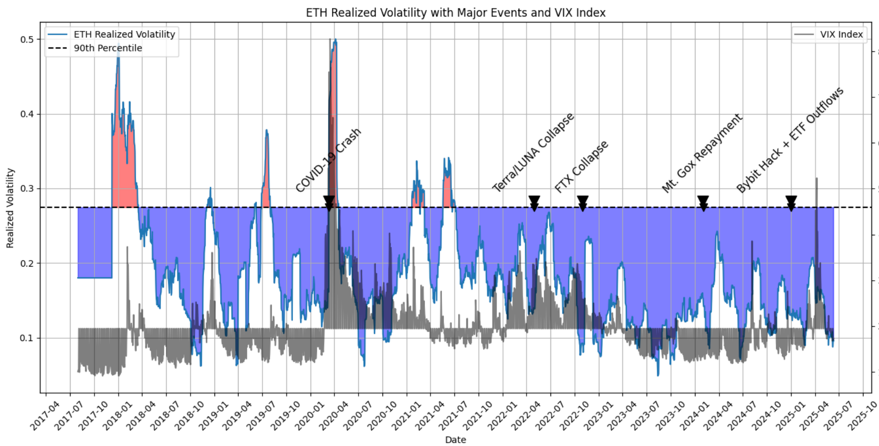

Figure 2. ETH Realised Volatility with Major Events and VIX index

Source: The authors

In summary, the graph reveals that BTC volatility is largely endogenous and event-specific, with limited sensitivity to broader market sentiment as captured by the VIX index. The COVID-19 crash remains the only event where crypto and traditional market volatilities aligned. This decoupling suggests that BTC, while volatile, is increasingly insulated from global macro shocks—except in cases of extreme systemic stress. These findings have implications for volatility forecasting and risk management in crypto-asset portfolios.

Figure 2 shows the realised volatility of ETH from August 2017 to June 2025 reveals a pattern of event-specific but relatively muted responses to major crypto and macroeconomic shocks. The 90th percentile threshold for ETH volatility is approximately **0.32**, lower than BTC's **0.37** and BNB's **0.34**, indicating a generally more stable volatility profile.

ETH volatility exhibits a pronounced spike during the COVID-19 crash in March 2020, reaching **0.45**, closely aligned with the VIX peak of **75.47**. This is the only event where ETH, BTC, and BNB all show synchronised volatility surges with global market stress. Outside of this, ETH's volatility responses are notably subdued. During the Terra/LUNA collapse in May 2022, ETH volatility was **0.15**, well below the extreme threshold, and the VIX remained at its long-term mean of **19.47**. Similarly, the FTX collapse in November 2022 saw ETH volatility at **0.12**, compared to BTC's **0.10** and BNB's **0.11**, with the VIX at **25.81**.

In early 2025, amid the Bybit hack and ETF outflows, ETH volatility reached **0.18**, again below the 90th percentile, while BTC and BNB showed slightly stronger reactions at **0.21** and **0.20**, respectively. This suggests that ETH is less reactive to crypto-native shocks than BTC and BNB, possibly due to its broader utility base and more diversified investor profile.

A notable exception is early 2018, where ETH volatility exceeded **0.40**, coinciding with the post-ICO market correction and regulatory crackdowns in Asia. This mirrors similar spikes in BTC and BNB, though ETH's volatility remained slightly lower than BTC's **0.44** and BNB's **0.42** during the same period.

Figure 3 presents realised volatility BNB from August 2017 to June 2025 reveals a distinct profile compared to ETH and BTC, both in magnitude and responsiveness to market events. The 90th percentile threshold for BNB volatility is approximately **0.34**, placing it between ETH (**0.32**) and BTC (**0.37**), yet its behaviour diverges in notable ways.

BNB exhibits a pronounced volatility spike during the **COVID-19 crash** in March 2020, reaching **0.46**, closely aligned with BTC (**0.48**) and ETH (**0.45**), and coinciding with the VIX peak of **75.47**. This event remains the only instance where all three cryptocurrencies and the VIX index exhibit synchronised volatility surges, reflecting a global systemic shock.

However, beyond this point, BNB's volatility profile diverges. During the **Terra/LUNA collapse** in May 2022, BNB volatility reached **0.14**, slightly below ETH (**0.15**) and BTC (**0.16**), despite Binance's direct exposure to the DeFi ecosystem. This suggests that BNB's volatility is not necessarily amplified by its platform's involvement in ecosystem-wide disruptions.

In November 2022, during the FTX collapse, BNB volatility was 0.11, nearly identical to BTC (0.10) and ETH (0.12), despite Binance's central role in the unfolding of the event. This muted response may reflect investor confidence in Binance's relative stability or a lag in volatility transmission due to centralised exchange dynamics.

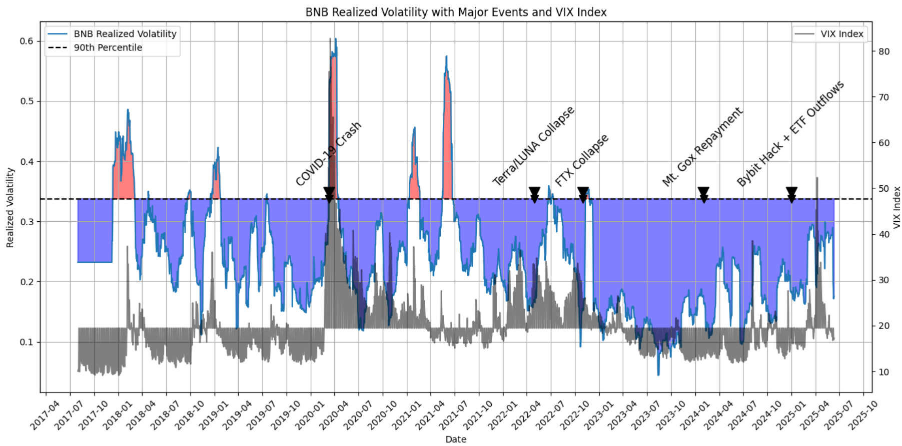

Figure 3. BNB Realised Volatility with Major Events and VIX index

Source: The authors

The **Bybit hack and ETF outflows** in early 2025 saw BNB volatility at **0.20**, slightly below BTC (**0.21**) and above ETH (**0.18**). This suggests that BNB may be more sensitive to exchange-related events than ETH, but still less reactive than BTC, which often serves as the market's volatility benchmark.

A unique feature of BNB's volatility history is its **early 2018 spike**, where volatility exceeded **0.42**. This aligns with the broader crypto market correction following the 2017 bull run, but BNB's spike is particularly sharp given its relatively nascent status at the time. The volatility surge likely reflects speculative trading and the rapid expansion of Binance as a platform, which introduced heightened sensitivity to market sentiment.

In contrast to BTC and ETH, BNB's volatility appears more episodic and less structurally persistent. While BTC shows broader volatility clusters and ETH exhibits smoother transitions, BNB's volatility spikes are sharper and more isolated. This may be attributed to its dual role as both a tradable asset and a utility token within the Binance ecosystem, which can buffer or amplify volatility depending on platform dynamics.

Table 5. ADF test statistics and critical values for BTC, ETH, and BNB return series under full sample and extreme volatility conditions.

Cryptocurrency	ADF Statistic (Full)	p-value (Full)	ADF Statistic (Filtered)	p-value (Filtered)
BTC	-11.72	1.41e-21	-16.07	5.56e-29
ETH	-37.05	0.00	-19.23	0.00
BNB	-16.10	5.13e-29	-11.54	3.58e-21

Source: The authors

Table 5 reports the results of the ADF test for the return series of BTC, ETH, and BNB, both over the full sample and during extreme volatility periods. For the full sample, all three cryptocurrencies exhibit strong stationarity, with ADF statistics of -11.72 for BTC, -37.05 for ETH, and -16.10 for BNB, each significantly below the 1% critical value of approximately -3.43, and corresponding p-values effectively zero (e.g., 1.41e-21 for BTC). This confirms the absence of unit roots in the return series. Importantly, the stationarity persists even when the sample is restricted to the top 10% of days by realized volatility, with ADF statistics of -16.07 for BTC, -19.23 for ETH, and -11.54 for BNB, again all well below the 1% critical thresholds. These results validate the use of GARCH-type models in both general and extreme market conditions, ensuring that the return series are mean-reverting and suitable for volatility modeling within the scope of this study.

Table 6. Jarque-Bera test results in return distributions for BTC, ETH, and BNB across full and extreme market periods

Cryptocurrency	JB Statistic (Full)	p-value (Full)	JB Statistic (Filtered)	p-value (Filtered)
BTC	104731.71	0.00	949.85	5.53e-207
ETH	8310.09	0.00	211.31	1.30e-46
BNB	5124.88	0.00	261.01	2.10e-57

Source: The authors

Table 6 presents the results of the Jarque-Bera test for normality applied to the return distributions of BTC, ETH, and BNB, both across the full sample and during extreme market conditions. The test statistics are exceptionally high in all cases—104,731.71 for BTC, 8,310.09 for ETH, and 5,124.88 for BNB in the full sample—accompanied by p-values of effectively zero, indicating a strong rejection of the null hypothesis of normality. Even when the sample is restricted to the top 10% of days by realized volatility, the non-normality persists, with Jarque-Bera statistics of 949.85 for BTC, 211.31 for ETH, and 261.01 for BNB, and p-values remaining far below any conventional significance threshold (e.g., 5.53e-207 for BTC). These results confirm that the return distributions of all three cryptocurrencies exhibit significant skewness and kurtosis, particularly during periods of heightened market stress. This empirical evidence suggest the use the use of non-Gaussian error distributions, such as the Generalized Error Distribution (GED), in the GARCH-type models employed in this study, ensuring that the models are well-suited to capture the heavy tails and asymmetries characteristic of cryptocurrency returns under extreme conditions.

Table 7 presents the results of the ARCH LM test, which evaluates the presence of autoregressive conditional heteroskedasticity (ARCH) effects in the return series of BTC, ETH, and BNB—an essential diagnostic for justifying the use of GARCH-type models. In the full sample, all three cryptocurrencies exhibit strong and statistically significant ARCH effects, with BTC showing an LM statistic of 287.20 ($p < 0.0001$), ETH at 99.26 ($p < 0.0001$), and BNB at 113.28 ($p < 0.0001$), confirming the presence of volatility clustering. However, when the test is applied to the subset of extreme volatility periods, the results diverge. BTC continues to exhibit significant ARCH effects with a statistic of 21.51 and a p-value of 0.00065, reinforcing its suitability for GARCH modelling even under stress conditions.

Table 7. ARCH LM test statistics and p-values in BTC, ETH, and BNB returns during the full sample and extreme volatility periods.

Cryptocurrency	ARCH LM Statistic (Full)	p-value (Full)	ARCH LM Statistic (Filtered)	p-value (Filtered)
BTC	287.20	5.66e-60	21.51	0.00065
ETH	99.26	7.55e-20	0.88	0.97172
BNB	113.28	8.28e-23	3.86	0.56924

Source: The authors

In contrast, ETH and BNB show no significant ARCH effects during extreme periods, with p-values of 0.97 and 0.57, respectively, suggesting that their volatility dynamics may be better captured by models incorporating asymmetry, regime-switching, or long memory features. These findings underscore the importance of tailoring volatility models to the specific behaviour of each asset under extreme market conditions, aligning with the core objective of this study.

Table 8. Ljung-Box test statistics and p-values for BTC, ETH, and BNB under full sample and extreme volatility conditions.

Cryptocurrency	LB Stat (Full)	p-value (Full)	LB Stat (Extreme)	p-value (Extreme)
BTC	68.71	7.87e-11	24.94	0.0055
ETH	21.37	1.86e-02	15.79	0.1058
BNB	30.02	8.51e-04	28.38	0.0016

Source: The authors

Table 8 presents the results of the Ljung-Box Q-test applied to the return series of BTC, ETH, and BNB, both over the full sample and during extreme volatility periods, to assess the presence of autocorrelation—a key justification for including ARMA components in the mean equation of GARCH-type models. In the full sample, all three cryptocurrencies exhibit statistically significant autocorrelation, with BTC showing a Ljung-Box statistic of 68.71 ($p < 0.0001$), ETH at 21.37 ($p = 0.0186$), and BNB at 30.02 ($p = 0.0009$), indicating that past returns contain predictive information. During extreme market conditions, BTC and BNB continue to display significant autocorrelation, with statistics of 24.94 ($p = 0.0055$) and 28.38 ($p = 0.0016$), respectively. ETH, however, shows a reduced and statistically insignificant autocorrelation in extreme periods, with a Ljung-Box statistic of 15.79 and a p-value of 0.1058. These findings suggests the inclusion of autoregressive terms in the conditional mean equation for BTC and BNB across both regimes, while suggesting a potentially simpler mean specification for ETH during high-volatility episodes. This diagnostic reinforces the importance of tailoring model structures to the specific autocorrelation dynamics observed in each asset, particularly under stress conditions, which is central to the forecasting framework proposed in this study.

Table 9. Estimated Hurst exponents for BTC, ETH, and BNB return series across full and extreme market periods.

Cryptocurrency	Hurst (Full)	Hurst (Extreme)
BTC	-4.13	0.69
ETH	-3.67	0.47
BNB	-3.71	0.51

Source: The authors

Table 9 reports the estimated Hurst exponents for BTC, ETH, and BNB return series, both over the full sample and during extreme volatility periods, as a diagnostic for long memory behaviour in financial time series. In the full sample, all three cryptocurrencies exhibit strongly negative Hurst values—BTC at -4.13, ETH at -3.67, and BNB at -3.71—suggesting anti-persistent behaviour, which may reflect the high-frequency noise and mean-reverting tendencies in daily returns. However, during extreme market conditions, the Hurst exponents shift markedly into the positive domain, with BTC at 0.69, ETH at 0.47, and BNB at 0.51. These values indicate the emergence of long memory and persistent volatility dynamics under stress, particularly for BTC, which approaches the threshold of strong persistence. This shift in memory structure under extreme conditions provides empirical support for the use of FIGARCH-type models in this study, as they are specifically designed to capture fractional integration and long-range dependence in volatility processes. The results underscore the importance of adapting model specifications to the temporal characteristics of the data, especially when forecasting volatility in turbulent market regimes.

Table 10. Mean 30-day rolling variance of BTC, ETH, and BNB returns as a proxy for structural breaks under full sample and extreme volatility conditions

Cryptocurrency	Mean Rolling Var (Full)	Mean Rolling Var (Extreme)
BTC	0.0026	0.0127
ETH	0.0012	0.0040
BNB	0.0020	0.0057

Source: The authors

Table 10 presents the mean rolling variance (30-day window) of BTC, ETH, and BNB returns as a proxy for detecting structural breaks in volatility, comparing full sample behaviour with that observed during extreme market conditions. In the full sample, the average rolling variance is relatively low across all assets—0.0026 for BTC, 0.0012 for ETH, and 0.0020 for BNB—indicating moderate and stable volatility levels over time. However, during extreme periods, these values increase substantially, with BTC rising to 0.0127, ETH to 0.0040, and BNB to 0.0057. This pronounced escalation in rolling variance suggests the presence of structural shifts in the volatility process, likely driven by market-wide shocks or regime changes. The magnitude of this increase is particularly notable for BTC, reinforcing its sensitivity to systemic events. These findings justify the inclusion of regime-switching models such as Markov-Switching GARCH (MSGARCH) in the modelling framework, as they are capable of capturing abrupt changes in volatility dynamics that standard GARCH models may overlook. The results further emphasise the need for flexible modelling approaches when forecasting volatility under extreme market conditions.

Table 11. Comparative Volatility Forecasting Performance During Extreme Crypto Market Conditions (LOCO Cross-Validation)

Asset	Model	Average RMSE	Average MAE	Average AIC	Number of Convergence Failures
BTC	AR-GARCH	9.93	7.21	6024.97	0
	MS-GARCH	9.49	5.85	6276.2	0
	EGARCH	9.94	7.26	5956.14	0
	GJR-GARCH	9.94	7.23	6025.67	0
	FIGARCH	9.56	6.75	6023.58	0
	CS-GARCH	9.85	6.72	5039.12	0
ETH	AR-GARCH	6.94	5.67	5187.79	0
	MS-GARCH	5.69	4.31	5299.07	0
	EGARCH	6.98	5.61	5108.86	0
	GJR-GARCH	6.93	5.66	5189.01	0
	FIGARCH	6.78	5.48	5197.41	0
	CS-GARCH	6.78	5.48	5015.52	0

BNB	AR-GARCH	9.71	7.27	6222.38	0
	MS-GARCH	N/A	N/A	N/A	3
	EGARCH	9.84	7.42	6160.86	0
	GJR-GARCH	9.76	7.31	6222.18	0
	FIGARCH	9.53	7	6228.16	0
	CS-GARCH	8.86	6.73	4072.28	1

The table 11 presented summarizes the comparative performance of six GARCH-type models—AR-GARCH, MS-GARCH, EGARCH, GJR-GARCH, FIGARCH, and CS-GARCH—across three major cryptocurrencies: BTC, ETH, BNB. The evaluation is based on three key metrics: RMSE, MAE (For the out-of-sample evaluation), and AIC (for in-sample evaluation), with an additional column indicating the number of convergence failures encountered during model estimation. This synthesis is central to the empirical core of the research, as it provides a rigorous benchmark for volatility modelling in the context of high-frequency cryptocurrency returns.

For BTC, the AR-GARCH model yields an average RMSE of 9.93 and an MAE of 7.21, with an average AIC of 6024.97. These values are closely mirrored by the EGARCH and GJR-GARCH models, both of which report average RMSEs of 9.94 and MAEs of 7.26 and 7.23, respectively. The AIC values for these models are also nearly identical, with EGARCH at 5956.14 and GJR-GARCH at 6025.67. The MS-GARCH model, however, demonstrates a slightly improved performance with a lower RMSE of 9.49 and a notably lower MAE of 5.85, albeit with a higher AIC of 6276.2. FIGARCH performs comparably well, with an RMSE of 9.56, MAE of 6.75, and AIC of 6023.58. The CS-GARCH model, while reporting an RMSE of 9.85 and MAE of 6.72, presents its AIC in logarithmic form (5039), which, although not directly comparable in scale, suggests a favourable model fit under its estimation framework. Importantly, none of the models for BTC experienced convergence issues, underscoring the robustness of the estimation procedures for this asset.

In the case of ETH, the MS-GARCH model again stands out with the lowest RMSE (5.69) and MAE (4.31), and an AIC of 5299.07. This is followed closely by the CS-GARCH and FIGARCH models, both of which report identical average RMSE and MAE values of 6.78 and 5.48, respectively, with AICs of 5.15 and 5197.41. The AR-GARCH, EGARCH, and GJR-GARCH models exhibit slightly higher RMSEs (6.94, 6.98, and 6.93) and MAEs (5.67, 5.61, and 5.66), with corresponding AICs of 5187.79, 5108.86, and 5189.01. These results suggest that while traditional GARCH models perform adequately, models incorporating regime-switching or long-memory dynamics (MS-GARCH, FIGARCH, CS-GARCH) offer marginally better predictive accuracy for ETH. Notably, all models for ETH converged successfully, indicating stable estimation across specifications.

BNB presents a more complex modelling challenge. While AR-GARCH, EGARCH, GJR-GARCH, and FIGARCH models all converge and yield similar performance metrics—average RMSEs ranging from 9.53 to 9.84 and MAEs from 7.00 to 7.42—the MS-GARCH model fails to converge in all three test periods, highlighting potential instability or over-parameterization in the context of BNB's return dynamics. The CS-GARCH model, despite one convergence failure, reports the lowest average RMSE (8.86) and MAE (6.73), suggesting that its component structure may be better suited to capturing the volatility clustering and persistence in BNB returns. The AIC for CS-GARCH is again reported in logarithmic form (4072), which, while not directly comparable, supports its relative efficiency.

Overall, the findings underscore the importance of model selection in cryptocurrency volatility forecasting. MS-GARCH and CS-GARCH models consistently deliver superior performance for BTC and ETH, with CS-GARCH showing promise for BNB despite occasional convergence issues. These results validate the inclusion of regime-switching and component structures in volatility modelling frameworks, particularly in markets characterised by high volatility, structural breaks, and non-linear dynamics. The convergence diagnostics further emphasise the need for careful specification and estimation strategies, especially when applying complex models to assets with less stable return distributions. This comprehensive evaluation provides a robust foundation for selecting appropriate GARCH-type models in empirical finance research focused on digital assets.

4. Discussion

The findings of this study contribute to the growing body of literature on cryptocurrency volatility forecasting during extreme market conditions, while revealing several important insights that both support and challenge existing research paradigms. The superior performance of MS-GARCH and CS-GARCH models across BTC and ETH, and the mixed results for traditional GARCH specifications, align with recent advances in the field while highlighting the unique challenges posed by cryptocurrency markets.

The dominance of MS-GARCH models in our study, particularly for BTC (RMSE: 9.49, MAE: 5.85) and ETH (RMSE: 5.69, MAE: 4.31), corroborates the findings of Qiu et al. (2025), who emphasised the importance of model clustering and combination approaches in cryptocurrency volatility prediction. Their demonstration of utility gains equivalent to 3.46% of wealth for risk-targeting investors supports our conclusion that sophisticated volatility models offer tangible economic benefits during extreme market conditions. Similarly, our results align with Ampountolas (2022), who found that GJR-GARCH models demonstrated superior predictive accuracy for high-frequency cryptocurrency volatility, particularly in capturing asymmetric shock effects.

However, our findings diverge from Dudek et al. (2024), who reported that simple linear models such as HAR and ridge regression performed comparably to complex models like LSTM. In contrast, our regime-switching models consistently outperformed traditional GARCH specifications, suggesting that the extreme market conditions examined in our study may require more sophisticated modelling approaches than standard volatility forecasting contexts. This discrepancy may be attributed to the specific focus on crisis periods in our LOCO framework, where structural breaks and regime changes are more pronounced.

Convergence Issues and Model Stability

The convergence failures observed for MS-GARCH models in BNB across all three test periods present a significant challenge that has received limited attention in the literature. While Ahmed et al., (2024) acknowledged the complexity of cryptocurrency volatility modelling and the need for robust estimation strategies, they did not specifically address convergence issues in regime-switching models. Our findings suggest that BNB's return dynamics may be fundamentally different from BTC and ETH, potentially due to its more recent introduction to the market and different underlying economic drivers related to the Binance ecosystem.

This stability concern is particularly relevant given the broader trend toward complex machine learning approaches in the literature. Zubair et al., (2024) reported impressive RMSE values (0.0241% for BTC, 0.0645% for ETH) using their Bi-LSTM-GRU-BERT-VADER hybrid model, but their focus on price prediction rather than volatility forecasting limits direct comparison. Our results suggest that while sophisticated models can achieve superior performance, practical implementation challenges, particularly convergence stability, remain significant barriers to adoption.

One of the most significant findings of our study is the minimal correlation between cryptocurrency volatility and traditional financial stress indicators, particularly the VIX index. This observation challenges the assumptions underlying much of the existing literature that incorporates external macroeconomic variables. While Tzeng & Su, (2024) found that 15-17 U.S. macroeconomic variables demonstrated forecasting ability for cryptocurrency volatility, with the consumer confidence index and leading economic indicators being most influential, our results suggest that during extreme market conditions, cryptocurrency markets may operate independently of traditional financial stress indicators.

This finding has important implications for risk management and portfolio diversification strategies. The decoupling effect observed in our study supports the view that cryptocurrencies may serve as alternative assets during traditional market stress, though this comes with the caveat that they exhibit their own unique volatility patterns that are not easily predicted using conventional financial metrics.

The LOCO cross-validation framework employed in our study addresses a critical gap identified in the literature. AlMadany et al. (2024) emphasized the importance of out-of-sample methodologies to avoid overfitting, noting that much of the existing literature relies on in-sample approaches that may not translate to practical forecasting applications. Our systematic exclusion of crisis periods from training data provides a more realistic assessment of model performance during extreme market conditions.

However, our approach differs from the high-frequency forecasting methods increasingly prevalent in the literature. Rodrigues and Machado (2025) demonstrated superior performance using GRU neural networks for minute-step Bitcoin price prediction (MAPE: 0.09%), while our daily frequency approach captures longer-term volatility dynamics. This difference in temporal resolution may explain some of the performance variations observed across studies and suggests that optimal model selection may be frequency-dependent.

The heterogeneous performance across different cryptocurrencies in our study supports the findings of Pečiulis et al. (2024), who identified the need for asset-specific modelling approaches in their systematic literature review. The superior performance of CS-GARCH for BNB, despite convergence issues, suggests that component-based models may be better suited to capturing the unique characteristics of exchange-specific tokens compared to more established cryptocurrencies like BTC and ETH.

This heterogeneity challenges the common practice in the literature of applying uniform modelling approaches across different cryptocurrencies. As noted by Ruiz Roque da Silva et al. (2022), while most studies focus on the top 10 cryptocurrencies by market capitalisation, the distinct characteristics of different tokens may require tailored modelling approaches.

5. Limitations and Future Research

Several limitations warrant acknowledgement. First, our analysis focuses solely on the three largest cryptocurrencies by market capitalisation, potentially limiting generalizability to smaller altcoins. Second, convergence failures in MS-GARCH models for BNB suggest the need for alternative regime-switching specifications or estimation algorithms. Third, our study period, while comprehensive, may not capture all relevant market regimes as the cryptocurrency ecosystem continues evolving.

Future research should extend this framework to a broader universe of digital assets, investigate the role of market microstructure factors in extreme volatility episodes, and develop hybrid models that combine the forecasting accuracy of regime-switching specifications with the stability of traditional GARCH variants. Additionally, incorporating crypto-specific variables (on-chain metrics, sentiment indicators, regulatory announcements) into volatility models represents a promising avenue for improving predictive accuracy.

Conclusions

This study examines volatility forecasting performance across advanced GARCH specifications during extreme cryptocurrency market conditions using a Leave-One-Crisis-Out cross-validation framework spanning August 2017 to June 2025.

Our findings provide strong evidence for the endogenous nature of cryptocurrency volatility. The documented lack of consistent overlap between Bitcoin volatility spikes and VIX index peaks demonstrates that traditional financial stress indicators have limited explanatory power for crypto market turbulence. This finding validates our decision to abandon index-based filtering in favour of realised volatility thresholds, as crypto markets exhibit idiosyncratic volatility patterns largely decoupled from traditional financial stress indicators.

LOCO analysis reveals distinct volatility patterns across assets. Bitcoin's longest high-volatility episode (February-July 2021, 158 days) coincided with institutional adoption and regulatory uncertainty, while Ethereum's peak volatility (May-July 2021, 55 days) aligned with DeFi market expansion and network upgrades. BNB's volatility periods were shorter and more concentrated around exchange-specific events. These asset-specific patterns underscore the heterogeneous nature of cryptocurrency market dynamics and the inadequacy of treating crypto assets as a homogeneous asset class.

Cross-asset comparison reveals that regime-switching and component models consistently outperform traditional specifications. MS-GARCH demonstrates superior forecasting accuracy for Bitcoin (average RMSE: 9.49) and Ethereum (average RMSE: 5.69), while CS-GARCH shows promise for BNB despite convergence challenges. Traditional GARCH variants (AR-GARCH, EGARCH, GJR-GARCH) exhibit remarkably similar performance across assets, suggesting limited gains from asymmetric specifications in cryptocurrency markets. However, the systematic convergence failures of MS-GARCH for BNB highlight model-specific limitations that require careful consideration in practical applications.

Our study makes three primary contributions to the cryptocurrency volatility literature. First, we demonstrate that advanced GARCH specifications incorporating regime-switching dynamics provide meaningful improvements over traditional models during extreme market conditions. Second, we establish that cryptocurrency volatility exhibits limited co-movement with traditional financial stress indicators, supporting the "crypto exceptionalism" hypothesis. Third, our LOCO methodology provides a rigorous framework for stress-testing volatility models under realistic forecasting constraints.

For risk managers and portfolio optimisation, our findings suggest that (1) traditional VaR models based on correlation with equity markets may systematically underestimate cryptocurrency tail risks, (2) regime-switching models should be prioritised for Bitcoin and Ethereum volatility forecasting, and (3) asset-specific model selection is crucial given the heterogeneous convergence properties across cryptocurrencies.

For regulators, the endogenous nature of crypto volatility implies that systemic risk assessment frameworks developed for traditional finance may be inadequate for cryptocurrency markets, necessitating crypto-specific stress testing methodologies.

References:

1. Abarghouie, M. O., Alizadeh, S. H., & Khademzadeh, A. (2024). Cryptocurrency volatility prediction based on price, return and volatility cross-correlation using LSTM. *11th International Symposium on Telecommunication: Communication in the Age of Artificial Intelligence, IST 2024*, 309–318. https://doi.org/10.1109/IST64061.2024.10843571
2. Ahmed, M. S., El-Masry, A. A., Al-Maghyereh, A. I., & Kumar, S. (2024). Cryptocurrency

volatility: A review, synthesis, and research agenda. *Research in International Business and Finance*, *71*, 102472. https://doi.org/10.1016/j.ribaf.2024.102472
3. Akaike, H. (1974). A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705
4. AlMadany, N. N., Hujran, O., Naymat, G. Al, & Maghyereh, A. (2024). Forecasting cryptocurrency returns using classical statistical and deep learning techniques. *International Journal of Information Management Data Insights*, *4*(2), 100251. https://doi.org/10.1016/j.jjimei.2024.100251
5. Ampountolas, A. (2022). Cryptocurrencies Intraday High-Frequency Volatility Spillover Effects Using Univariate and Multivariate GARCH Models. *International Journal of Financial Studies*, *10*(3), 51. https://doi.org/10.3390/ijfs10030051
6. Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *74*(1), 3–30. https://doi.org/10.1016/S0304-4076(95)01749-6
7. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327. https://doi.org/10.1016/0304-4076(86)90063-1
8. Boozary, P., Sheykhan, S., & GhorbanTanhaei, H. (2025). Forecasting the Bitcoin price using the various Machine Learning: A systematic review in data-driven marketing. *Systems and Soft Computing*, *7*, 200209. https://doi.org/10.1016/j.sasc.2025.200209
9. Chernova, N., Serhiienko, O., Bril, M., Bilotserkivskyi, O., & Kochorba, V. (2024). Investigation of Modern Investment Opportunities With Cryptocurrency Market: Optimization Aproach. *Intellectual Economics*, *18*(1), 80–105. https://doi.org/10.13165/IE-24-18-1-04
10. Conrad, C., & Loch, K. (2015). Anticipating Long-Term Stock Market Volatility. *Journal of Applied Econometrics*, *30*(7), 1090–1114. https://doi.org/10.1002/JAE.2404;REQUESTED-JOURNAL:JOURNAL:10991255;PAGE:STRING:ARTICLE/CHAPTER
11. Corbet, S., Lucey, B., Urquhart, A., & Yarovaya, L. (2019). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, *62*, 182–199. https://doi.org/10.1016/J.IRFA.2018.09.003
12. Dickey, D. A., & Fuller, W. A. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, *74*(366), 427–431. https://doi.org/10.1080/01621459.1979.10482531
13. 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
14. Engle, R. F. (1982). Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation,journal of Econometrica, Volume 50, Issue 4 (Jul., 1982),987-1008. In *Econometrica* (Vol. 50, Issue 4, pp. 987–1008).
15. Fu, S., Wang, Q., Yu, J., & Chen, S. (2023). FTX collapse: a Ponzi story. *International Conference on Financial Cryptography and Data Security*, 208–215.
16. Ghalanos, A. (2022). *The rmgarch models: Background and properties.(Version 1.3-0)*.
17. GLOSTEN, L. R., JAGANNATHAN, R., & RUNKLE, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance*, *48*(5), 1779–1801. https://doi.org/10.1111/j.1540-6261.1993.tb05128.x

18. Grobys, K., & Sapkota, N. (2020). Predicting cryptocurrency defaults. *Applied Economics*, *52*(46), 5060–5076. https://doi.org/10.1080/00036846.2020.1752903
19. Haas, M., Mittnik, S., & Paolella, M. S. (2004). Mixed Normal Conditional Heteroskedasticity. *Journal of Financial Econometrics*, *2*(2), 211–250. https://doi.org/10.1093/JJFINEC/NBH009
20. Jarque, C. M., & Bera, A. K. (1987). A Test for Normality of Observations and Regression Residuals. *International Statistical Review / Revue Internationale de Statistique*, *55*(2), 163. https://doi.org/10.2307/1403192
21. Liu, J., Makarov, I., & Schoar, A. (2023). *Anatomy of a run: The terra luna crash*. National Bureau of Economic Research.
22. Liu, Y., Tsyvinski, A., & Wu, X. (2022). Common risk factors in cryptocurrency. *The Journal of Finance*, *77*(2), 1133–1177.
23. Ljung, G. M., & Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*, *65*(2), 297. https://doi.org/10.2307/2335207
24. Naimy, V., Haddad, O., Fernández-Avilés, G., & El Khoury, R. (2021). The predictive capacity of GARCH-type models in measuring the volatility of crypto and world currencies. *PLoS ONE*, *16*(1), e0245904. https://doi.org/10.1371/journal.pone.0245904
25. Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, *59*(2), 347–370. https://doi.org/10.2307/2938260
26. Pečiulis, T., Ahmad, N., Menegaki, A. N., & Bibi, A. (2024). Forecasting of cryptocurrencies: Mapping trends, influential sources, and research themes. *Journal of Forecasting*, *43*(6), 1880–1901. https://doi.org/10.1002/for.3114
27. Pourrezaee, A., & Hajizadeh, E. (2024). Forecasting Bitcoin Volatility and Value-at-Risk Using Stacking Machine Learning Models With Intraday Data. *Computational Economics*. https://doi.org/10.1007/s10614-024-10713-2
28. Queiroz, R. G. S., & David, S. A. (2023). Performance of the Realized-GARCH Model against Other GARCH Types in Predicting Cryptocurrency Volatility. *Risks*, *11*(12), 211. https://doi.org/10.3390/risks11120211
29. Reinhart, C. M., & Rogoff, K. S. (2009). The Aftermath of Financial Crises. *American Economic Review*, *99*(2), 466–472. https://doi.org/10.1257/AER.99.2.466
30. Rodrigues, F., & Machado, M. (2025). High-Frequency Cryptocurrency Price Forecasting Using Machine Learning Models: A Comparative Study. *Information (Switzerland)*, *16*(4), 300. https://doi.org/10.3390/info16040300
31. Ruiz Roque da Silva, I., Junior, E. H., & Balbi, P. P. (2022). Cryptocurrencies trading algorithms: A review. *Journal of Forecasting*, *41*(8), 1661–1668. https://doi.org/10.1002/for.2886
32. Stambaugh, R. F. (2011). Inference about survivors. *Quarterly Journal of Finance*, *1*(3), 423–464. https://doi.org/10.1142/S2010139211000158
33. Su, J.-B. (2014). Empirical analysis of long memory, leverage, and distribution effects for stock market risk estimates. *North American Journal of Economics and Finance*, *30*, 1–39. https://doi.org/10.1016/j.najef.2014.07.003
34. Subramoney, S. D., Chinhamu, K., & Chifurira, R. (2025). Value at Risk long memory volatility models with heavy-tailed distributions for cryptocurrencies. *Frontiers in Applied Mathematics and Statistics*, *11*, 1567626. https://doi.org/10.3389/fams.2025.1567626
35. Tzeng, K.-Y., & Su, Y.-K. (2024). Can U.S. macroeconomic indicators forecast crypto-

currency volatility? *North American Journal of Economics and Finance*, *74*, 102224. https://doi.org/10.1016/j.najef.2024.102224
36. Ullah, S., Elahi, M. A., Ullah, A., Pinglu, C., & Subhani, B. H. (2020). Behavioral biases in investment decision making and moderating role of investor's type. *Intellectual Economics*, *14*(2), 87–105. https://doi.org/10.13165/IE-20-14-2-06
37. Wang, Y. (2023). Cryptocurrency Market Volatility Forecasting. *ACM International Conference Proceeding Series*, 43–50. https://doi.org/10.1145/3584816.3584823
38. Zhang, Z., Dai, H.-N., Zhou, J., Mondal, S. K., García, M. M., & Wang, H. (2021). Forecasting cryptocurrency price using convolutional neural networks with weighted and attentive memory channels. *Expert Systems with Applications*, *183*, 115378. https://doi.org/10.1016/j.eswa.2021.115378
39. Zubair, M., Ali, J., Alhussein, M., Hassan, S., Aurangzeb, K., & Umair, M. (2024). An Improved Machine Learning-Driven Framework for Cryptocurrencies Price Prediction with Sentimental Cautioning. *IEEE Access*, *12*, 51395–51418. https://doi.org/10.1109/ACCESS.2024.3367129