



Regime-aware volatility forecasting for bitcoin: A hybrid EGARCH–hidden Markov model approach

Catherine Ngo^{1*}

Orson Chi²

Yeong Nain Chi³

^{1,2}Department of Agriculture, Food and Resource Sciences University of Maryland Eastern Shore Princess Anne, Maryland 21853, USA.

¹Email: ctngo@umes.edu

²Email: ychi@umes.edu

³Department of Computer Science and Engineering Technology University of Maryland Eastern Shore Princess Anne, Maryland 21853, USA.

²Email: ochi@umes.edu

Licensed:

This work is licensed under a Creative Commons Attribution 4.0 License.

Keywords:

Bitcoin,

Cryptocurrency risk management

EGARCH

Hidden Markov model

Regime switching

Value-at-risk

Volatility forecasting.

Abstract

This study develops a regime-aware hybrid volatility framework for daily Bitcoin returns from January 2020 to May 2026 by integrating EGARCH with a three-state Gaussian Hidden Markov Model (HMM). EGARCH was selected as the best single-regime model, while the HMM identified quiet (1.0% volatility; 43.2% of observations), normal (3.4%; 43.4%), and crisis (6.0%; 13.4%) regimes, with crisis periods lasting about 9.6 days. Bitcoin returns display pronounced non-normality, negative skewness, autocorrelation, and ARCH effects, supporting the use of fat-tailed and regime-switching methods. In a strict out-of-sample evaluation over 465 test days, the Hybrid EGARCH+HMM outperformed standalone EGARCH, reducing RMSE by 31.2%, MAE by 47.2%, and QLIKE by 19.6%, and consistently achieving better monthly forecasts. Although an MS-EGARCH+HMM model improved accuracy, it remained inferior to the simpler hybrid approach. Results were robust across alternative volatility measures, subsamples, and Monte Carlo simulations, supporting practical Bitcoin risk management and forecasting applications.

Received: 20 August 2026

Revised: 16 September 2026

Accepted: 30 September 2026

Published: 8 October 2026

(* Corresponding Author)

Funding: This study received no specific financial support.

Institutional Review Board Statement: Not applicable.

Transparency: The authors declare that the manuscript is honest, truthful and transparent, that no important aspects of the study have been omitted and that all deviations from the planned study have been made clear. This study followed all rules of writing ethics.

Data Availability Statement: The corresponding author may provide study data upon reasonable request.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the conception and design of the study. All authors have read and agreed to the published version of the manuscript.

1. Introduction

Bitcoin has emerged as one of the most important and controversial financial innovations of the past decade. Since its introduction in 2009, it has evolved from a niche digital currency into a globally traded financial asset with substantial participation from retail investors, institutional asset managers, hedge funds, exchanges, and derivative markets. Despite its increasing adoption, Bitcoin continues to exhibit levels of price volatility far exceeding those observed in traditional asset classes such as equities, bonds, and foreign exchange. Extreme price fluctuations, sudden market crashes, volatility clustering, heavy-tailed return distributions, and rapidly changing market conditions make Bitcoin particularly challenging to model and forecast. These characteristics have generated considerable interest among researchers and practitioners seeking robust methods to estimate and predict Bitcoin volatility, which remains a critical input for portfolio allocation, derivative pricing, risk

management, market-making, and regulatory capital determination (Bouri, Gupta, Tiwari, & Roubaud, 2017; Corbet, Lucey, Urquhart, & Yarovaya, 2019).

Volatility forecasting occupies a central role in modern financial economics. Since the seminal work of Engle (1982) who introduced the Autoregressive Conditional Heteroskedasticity (ARCH) model, and Bollerslev (1986) who generalized the framework through the Generalized ARCH (GARCH) model, conditional volatility modeling has become a standard tool for analyzing financial time series characterized by heteroskedasticity and volatility clustering. These models recognize that volatility is not constant through time but instead evolves dynamically according to past information. The widespread success of GARCH-family models has led to numerous extensions capable of capturing stylized facts such as persistence, leverage effects, asymmetry, and heavy tails, all of which are particularly relevant for cryptocurrency markets (Bollerslev, 1986; Engle, 1982).

Bitcoin returns exhibit many of the stylized facts commonly observed in financial markets, but often in amplified form. Numerous studies have documented that Bitcoin returns are characterized by excess kurtosis, skewness, volatility clustering, and time-varying conditional variance (Corbet et al., 2019; Katsiampa, 2017). However, unlike traditional financial assets, cryptocurrency markets operate continuously, are largely decentralized, and remain highly sensitive to technological developments, regulatory announcements, macroeconomic events, cybersecurity incidents, and market sentiment. Consequently, periods of relative calm can transition rapidly into episodes of extraordinary volatility. Such behavior presents significant challenges for conventional econometric models that assume a single volatility-generating process throughout the sample period.

Among the various GARCH extensions proposed in the literature, the Exponential GARCH (EGARCH) model developed by Nelson (1991) has become particularly influential. Unlike standard GARCH specifications, EGARCH models the logarithm of conditional variance, ensuring non-negativity without imposing parameter restrictions and allowing volatility to respond asymmetrically to positive and negative shocks. This feature is especially relevant for Bitcoin markets, where adverse news events, exchange failures, regulatory interventions, and periods of market panic may have disproportionately large effects on volatility. Previous research has shown that EGARCH and other asymmetric GARCH specifications frequently outperform symmetric alternatives when forecasting cryptocurrency volatility (Katsiampa, 2017; Nelson, 1991).

Although GARCH-family models have demonstrated considerable success in modeling conditional volatility, an important limitation remains. Traditional GARCH specifications implicitly assume that a single volatility process governs the entire sample period. In practice, however, financial markets often alternate between different volatility regimes characterized by distinct behaviors, persistence levels, and risk profiles. For example, Bitcoin experienced relatively tranquil consolidation periods during portions of 2023 and 2025, while the COVID-19 market dislocation in March 2020, the collapse of Terra-Luna in 2022, and the FTX bankruptcy generated periods of exceptional market stress. A single conditional variance equation may struggle to adequately capture both tranquil and crisis periods simultaneously, potentially leading to biased forecasts and reduced predictive accuracy.

Regime-switching methodologies provide a natural solution to this problem. Hamilton (1989) introduced the Markov-switching framework, demonstrating how economic and financial time series can transition among latent states governed by probabilistic state transitions. Closely related to this approach, Hidden Markov Models (HMMs) provide a flexible statistical framework for inferring unobservable regimes from observed data. The theoretical foundations and practical implementation of HMMs were comprehensively described by Rabiner (1989) whose tutorial remains the standard reference across numerous scientific disciplines. In financial applications, HMMs have been increasingly employed to identify latent market regimes, distinguish between high- and low-volatility environments, and improve forecasting performance by accounting for structural changes that are difficult to capture using conventional time-series models alone (Hamilton, 1989; Rabiner, 1989).

Recent advances in cryptocurrency forecasting research have further highlighted the importance of accommodating regime dependence and structural heterogeneity. Huang, Sangiorgi, and Urquhart (2024) demonstrate that machine learning models can outperform conventional GARCH specifications in forecasting Bitcoin volatility, suggesting that nonlinear dynamics remain incompletely captured by traditional approaches. Similarly, Chinazzo and Jeleskovic (2024) emphasize the persistent challenges associated with modeling Bitcoin's highly volatile and evolving risk structure. Recent studies employing regime-based methodologies indicate that market states contain valuable information for forecasting and investment decision-making, supporting the view that cryptocurrency markets operate under distinct latent conditions rather than a single homogeneous process (Agakishiev, Härdle, Becker, & Zuo, 2025; Karam, El Maalouf, & Dirani, 2026).

Motivated by these developments, the present study proposes a regime-aware hybrid volatility forecasting framework that combines the strengths of EGARCH and Hidden Markov Models. The proposed Hybrid EGARCH+HMM model first estimates conditional volatility using EGARCH and subsequently adjusts those forecasts based on probabilities derived from a three-state Gaussian HMM. The resulting framework allows volatility forecasts to adapt dynamically to changing market regimes while maintaining the parsimonious structure and interpretability of the EGARCH model. In addition, this study develops and evaluates a more

flexible Markov-Switching EGARCH (MS-EGARCH+HMM) specification that permits regime-specific volatility processes.

The study contributes to literature in three important ways. First, it provides a comprehensive examination of Bitcoin return dynamics during the post-2020 institutional adoption era, a period characterized by unprecedented growth, increased market participation, and multiple episodes of market stress. Second, it develops and empirically evaluates two regime-aware forecasting frameworks, Hybrid EGARCH+HMM and MS-EGARCH+HMM, using rigorous out-of-sample testing procedures and multiple forecast evaluation criteria. Third, it investigates whether forecasting gains arise from genuine latent regime structure rather than statistical overfitting through robustness analyses and extensive Monte Carlo simulations. By integrating traditional econometric volatility modeling with latent-state classification, the study seeks to advance both the methodological and practical understanding of cryptocurrency risk forecasting.

The remainder of the paper is organized as follows. Section 2 reviews the relevant literature on cryptocurrency volatility modeling, GARCH-family models, and regime-switching approaches. Section 3 describes the data and presents preliminary diagnostics. Section 4 develops econometric methodology. Section 5 presents estimation results. Section 6 evaluates out-of-sample forecasting performance. Section 7 examines Value-at-Risk forecasting and backtesting results. Section 8 provides robustness checks and Monte Carlo evidence. Section 9 discusses the findings, and Section 10 concludes.

2. Literature Review

2.1. Volatility Modeling of Bitcoin Returns

The growing importance of Bitcoin as a financial asset has stimulated extensive research into its volatility dynamics and forecasting. Unlike traditional assets, Bitcoin exhibits extreme price fluctuations, pronounced volatility clustering, heavy tails, and frequent structural shifts, making volatility forecasting particularly challenging. Early studies established that cryptocurrency returns possess statistical properties that differ substantially from those of equities, bonds, and foreign exchange markets, thereby requiring specialized modeling approaches (Corbet et al., 2019; Urquhart, 2016).

The application of GARCH-family models to Bitcoin volatility was pioneered by Katsiampa (2017) who compared several ARCH and GARCH specifications and demonstrated that no single model consistently dominates across forecasting criteria. This finding highlighted the complexity of Bitcoin's volatility process and motivated subsequent research into more flexible volatility frameworks. Baur and Dimpfl (2018) further documented asymmetric volatility behavior in Bitcoin markets, while Bouri, Gupta, et al. (2017) emphasized the role of uncertainty and extreme market movements in cryptocurrency price dynamics. Together, these studies established volatility clustering and time-varying conditional variance as central features of Bitcoin returns.

Among GARCH-family extensions, the Exponential GARCH (EGARCH) model introduced by Nelson (1991) has attracted particular attention in cryptocurrency research. The EGARCH framework models the logarithm of conditional variance, ensuring positivity without imposing parameter restrictions and allowing volatility to respond asymmetrically to market shocks. These properties are especially valuable in cryptocurrency markets, where large price movements occur frequently and volatility persistence is often near unity. Several studies have found EGARCH models to outperform conventional GARCH specifications when forecasting Bitcoin volatility because they better capture nonlinear responses to information arrivals and periods of elevated market stress (Conrad, Custovic, & Ghysels, 2018; Gkillas & Katsiampa, 2018).

More recent studies have continued to investigate volatility forecasting in cryptocurrency markets using both traditional econometric and machine-learning methods. Huang et al. (2024) have reported that neural network architectures such as Long Short-Term Memory (LSTM) and hybrid CNN-LSTM models often outperformed conventional GARCH models for short- and medium-horizon Bitcoin volatility forecasts. Similarly, Chinazzo and Jeleskovic (2024) have showed that Bitcoin exhibited persistently high expected volatility regardless of whether volatility that is measured using historical, implied, or GARCH-based approaches. These findings suggest that volatility forecasting remains an active research area in which traditional statistical approaches and newer machine-learning methods continue to compete.

2.2. Distributional Characteristics and Tail Risk

A central challenge in modeling Bitcoin volatility is the presence of severe departures from normality. Numerous empirical studies have documented that Bitcoin returns exhibit excess kurtosis, heavy tails, skewness, and extreme downside risk. These features violate the assumptions underlying many conventional econometric models and necessitate the use of alternative error distributions.

Gkillas and Katsiampa (2018) have employed extreme value theory to examine the tail behavior of major cryptocurrencies and concluded that Bitcoin returns exhibit exceptionally heavy tails, implying that extreme price movements occur much more frequently than predicted by Gaussian models. Borri (2019) similarly has demonstrated that Bitcoin exhibits substantially higher conditional tail risk than traditional financial assets and that risk measures such as Value-at-Risk (VaR) and Expected Shortfall are particularly sensitive to distributional assumptions.

Consequently, Student's t and generalized error distributions have become standard components of cryptocurrency volatility models. These distributions allow for greater probability mass in the tails and provide more realistic estimates of risk during market crises. The use of fat-tailed error specifications has consistently improved model fit and forecasting performance across numerous cryptocurrency studies (Borri, 2019; Bouri, Molnár, Azzi, Roubaud, & Hagfors, 2017).

Another notable feature of financial volatility is the leverage effect; whereby negative returns increase future volatility more than positive returns of equal magnitude. While leverage effects are well documented in equity markets (Black, 1976; Christie, 1982) empirical evidence for Bitcoin remains mixed. Baur and Dimpfl (2018) have found limited support for asymmetric volatility responses in Bitcoin, suggesting that volatility dynamics in cryptocurrency markets may differ fundamentally from those observed in conventional financial assets. This distinction further motivates the search for alternative mechanisms, such as regime switching, to explain Bitcoin's volatility behavior.

2.3. Regime-Switching Models and Financial Volatility

Traditional GARCH models assume that a single set of parameters governs volatility throughout the sample period. However, financial markets often experience structural breaks and shifts between low-volatility and high-volatility environments. To address this limitation, researchers have increasingly employed regime-switching methodologies.

The theoretical foundation for regime-switching analysis was established by Hamilton (1989) who introduced the Markov-switching framework for modeling economic regimes. In this approach, observed time-series behavior is driven by an unobservable state variable that evolves according to a Markov process. Hamilton's framework has since become widely used in macroeconomics and finance for modeling structural change.

The integration of conditional heteroskedasticity with regime-switching dynamics was undertaken by Hamilton and Susmel (1994) who developed the Markov-Switching ARCH model. Subsequent contributions by Cai (1994) and Gray (1996) extended the approach to GARCH processes, enabling volatility persistence to vary across regimes. However, Markov-switching GARCH models face significant computational challenges because the conditional variance depends on the entire history of regime transitions. Klaassen (2002) and Haas, Mittnik, and Paoletta (2004) developed tractable estimation frameworks that significantly improved the feasibility of Markov-switching GARCH models in empirical finance.

Recent research continues to demonstrate the value of regime-switching methods in financial forecasting. Agakishiev et al. (2025) show that cryptocurrency returns exhibit meaningful regime behavior that can be exploited for improved portfolio management and forecasting. Similarly, Karam et al. (2026) demonstrate that regime-dependent machine-learning models frequently outperform single-regime approaches when market dynamics vary substantially across states. These findings provide growing evidence that cryptocurrency markets are characterized by latent regimes that influence return and volatility behavior.

2.4. Hidden Markov Models and Cryptocurrency Markets

An alternative approach to explicit Markov-switching volatility models is provided by Hidden Markov Models (HMMs). The HMM framework introduced by Rabiner (1989) models observations as arising from a set of unobservable states that evolve according to a Markov chain. Unlike MS-GARCH models, which specify separate variance equations for each regime, HMMs directly model the probability distribution of returns conditional on latent states.

The popularity of HMMs in finance stems from their computational efficiency and their ability to identify changing market conditions in real time. Guidolin and Timmermann (2008) demonstrate that regime identification through HMMs substantially improves portfolio allocation decisions in international equity markets. More generally, HMMs have proven effective for detecting transitions between bull and bear markets, low- and high-volatility regimes, and crisis periods.

Applications of HMMs to cryptocurrency markets remain comparatively limited but have produced promising results. Osterrieder and Lorenz (2017) identify distinct phases of Bitcoin market behavior using state-dependent models, while Borri (2019) finds evidence of substantial variation in risk characteristics across different market environments. Nevertheless, relatively few studies have combined HMM-based regime classification with GARCH-type volatility dynamics in a single forecasting framework. The integration of HMMs with EGARCH models therefore represents a natural extension of the existing literature, allowing conditional volatility forecasts to adapt dynamically to evolving market regimes.

2.5. Cryptocurrency-Specific Sources of Regime Behavior

Bitcoin's volatility dynamics are influenced by a unique set of institutional and technological factors that are largely absent from traditional financial markets. One of the most widely discussed mechanisms is the Bitcoin halving cycle, in which miner rewards are reduced by 50% approximately every four years. Hayes (2017) and Hårdle, Harvey, and Reule (2020) argue that these supply shocks contribute to long-term boom-and-bust cycles and may generate distinct market regimes.

Regulatory decisions also play a major role in shaping Bitcoin volatility. Government restrictions on cryptocurrency trading and mining, taxation policies, central bank announcements, and the approval of exchange-traded funds (ETFs) often produce abrupt shifts in market sentiment. Exchange failures and cybersecurity incidents represent another important source of structural change. Events such as the collapse of the FTX exchange in November 2022 generated extreme market turbulence and illustrate how cryptocurrency-specific shocks can trigger rapid transitions between volatility states.

Unlike equity markets, Bitcoin trades continuously on a 24-hour, seven-day-a-week basis. Consequently, information produced during nights, weekends, and holidays is immediately incorporated into prices rather than accumulating until the next trading session. This continuous trading environment contributes to elevated volatility and strengthens the case for identifying multiple latent market states characterized by different levels of risk.

2.6. Evaluation of Volatility Forecasting Models

Accurate evaluation of volatility forecasts requires appropriate loss functions and statistical testing procedures. Hansen and Lunde (2005) demonstrate that simple GARCH models are remarkably difficult to outperform in traditional equity markets, establishing a benchmark against which more complex models are often compared. However, the distinctive characteristics of Bitcoin volatility suggest that additional model flexibility may be more valuable in cryptocurrency markets.

Patton (2011) develops a theoretical framework for evaluating volatility forecasts and shows that loss functions such as Mean Squared Error (MSE) and the Quasi-Likelihood (QLIKE) loss are consistent measures of forecast accuracy. As a result, these metrics have become standard tools for comparing competing volatility forecasting models.

Formal statistical inference regarding forecast superiority is typically conducted using the Diebold and Mariano (1995) which evaluates whether differences in forecast errors are statistically significant. When forecast errors exhibit serial correlation, Newey-West heteroskedasticity and autocorrelation consistent (HAC) corrections are commonly applied to obtain robust inference. These techniques have become standard practice in volatility forecasting studies and provide a rigorous framework for assessing the predictive value of regime-aware models relative to conventional alternatives.

3. Materials

3.1. Data Description

The dataset comprises daily closing prices for Bitcoin quoted in US dollars, sourced from Yahoo Finance (ticker: BTC-USD), spanning 2 January 2020 to 14 May 2026 (Figure 1). The sample yields $N = 2,325$ daily log-returns, computed as $r_t = 100 \times \ln(P_t / P_{t-1})$, expressed as percentage per day. The sample was selected to capture the post-institutional-adoption era of Bitcoin, which includes five distinct market episodes: (i) the COVID-19 pandemic dislocation (March 2020), (ii) the 2021 institutional adoption bull market, (iii) the 2022 bear market and FTX collapse, (iv) the 2023–2024 recovery and spot ETF approval cycle, and (v) the 2024 halving, 2025 all-time high (\$124,752.53 on 6 October 2025), and subsequent 2026 correction. Bitcoin trades continuously on a 24/7 basis; no market closures or missing observations exist in the dataset.



Figure 1. Time Series Plot of Bitcoin Daily Closing Prices from 1/2/2020 to 5/14/2026 in US dollars.

3.2. Descriptive Statistics

Bitcoin's price rose from \$7,202.55 at the opening of January 2020 to \$81,051.25 at the close of the sample on 14 May 2026, a total return of 1,060.28% and a compound annual growth rate (CAGR) of 47.00%. The coefficient of variation of 0.64 reflects the extreme dispersion of price levels relative to the mean, consistent with an asset that traversed price ranges from under \$5,000 to over \$124,000 within the sample period. The all-time high of \$124,752.53 was reached on 6 October 2025; the sample low of \$4,970.79 occurred on 12 March 2020 during the COVID-19 market dislocation (Table 1).

Table 1. Bitcoin Price-Level Summary Statistics.

Statistic	Close (\$)	High (\$)	Low (\$)
Mean	48,887.94	50,239.45	47,472.58
Median	42,250.55	43,419.20	40,961.07
Std dev	31,315.21	32,010.83	30,569.48
Minimum	4,970.79	5,280.79	4,748.63
Maximum	124,752.53	126,040.00	122,700.00
Coeff. of variation	0.6406	0.6373	0.6440

Note: Statistics computed on daily closing prices. High and Low columns represent daily intraday highs and lows. All figures in US dollars.

Table 2 presents comprehensive summary statistics. The return distribution departs dramatically from normality on every dimension tested. The Jarque–Bera statistic of 46,599 ($p \approx 0$) reflects the combined effect of negative skewness (−1.319) and extreme excess kurtosis (21.773), a level approximately eight times that observed in major equity index returns. The single worst daily return of −46.47% (12 March 2020) would be assigned probability $\approx 10^{-270}$ under a Gaussian model with the same mean and standard deviation. Returns beyond five standard deviations occur four times in the sample (0.17% of days), roughly 3,000 times the Gaussian prediction.

Table 2. Summary statistics for Bitcoin daily log-returns (2 January 2020 – 14 May 2026).

Statistic	Value	Statistic	Value
Sample period	2 Jan 2020 - 14 May 2026	Excess kurtosis	21.773
Observations (N)	2,325	Jarque–Bera stat.	46,599 ($p \approx 0$)
Mean return (%/day)	0.1041	ADF statistic (8 lags)	−15.806 ($p < 0.001$)
Median (%/day)	0.0415	KPSS statistic	0.182 ($p = 0.100$)
Std deviation (%/day)	3.2113	ARCH-LM(10)	51.71 ($p < 0.001$)
Minimum (%)	−46.473 (12 Mar 2020)	LB $Q(10)$ - returns	$p = 0.00046$
Maximum (%)	17.182 (8 Feb 2021)	LB $Q(10)$ - sq. returns	$p < 0.001$
Skewness	−1.319	ACF lag-1	−0.060
Ann. mean return	26.24%	Ann. volatility	50.98%
Sharpe ratio	0.515	Max drawdown	−76.63%
Total return	+1,060.28%	CAGR	47.00%

Note: N = 2,325 trading days. Log-returns defined as $r_t = 100 \times \ln(P_t / P_{t-1})$. ADF = Augmented Dickey–Fuller with lag order selected by AIC; critical values at 1% and 5% are −3.433 and −2.863. KPSS H_0 : Stationary; $p \geq 0.10$ cannot reject stationarity. ARCH-LM = Engle (1982) LM test with 10 lags. LB = Ljung–Box test. ACF lag-1 = −0.060 (Statistically significant). Green = Stationarity/Pass; Red = Non-Normality/ARCH alert. All figures in US dollars.

Table 3 disaggregates the distributional properties by calendar year, revealing substantial heterogeneity. The year 2020 exhibits the highest excess kurtosis (51.44), entirely attributable to the March 2020 crash. The year 2021 is nearest to normality (kurtosis = 1.52), consistent with a broadly trending regime. The bear market year 2022 shows the lowest mean return (−0.28%/day). The years 2023 and 2025 display the lowest realized volatility (standard deviations of 2.27% and 2.19% respectively), underscoring the time-variation that single-regime models must accommodate.

Table 3. Annual breakdown of return distribution statistics.

Year	n	Mean (%/d)	Std (%/d)	Skewness	Ex. kurtosis	Min (%)	Max (%)
2020	365	+0.382	4.013	−4.085	51.44	−46.47	+16.71
2021	365	+0.128	4.204	−0.107	1.52	−14.81	+17.18
2022	365	−0.282	3.361	−0.600	4.73	−17.41	+13.58
2023	365	+0.257	2.265	+0.694	3.02	−7.36	+9.81
2024	366	+0.217	2.781	+0.338	1.74	−8.71	+11.46
2025	365	−0.018	2.193	−0.004	2.31	−9.08	+9.12
2026*	134	−0.057	2.771	−0.602	8.19	−15.23	+11.80

Note: * 2026 covers 1 January – 14 May only (n = 134). Mean and Std expressed as %/day. Skewness and excess kurtosis are Fisher moments. Green (Mean) = positive mean; Red = Negative. Red (Ex. kurtosis) = value exceeds 6. All figures in US dollars.

3.3. Pre-Estimation Diagnostic Tests

Three diagnostic tests establish the statistical foundation for the modelling choices. First, the Augmented Dickey–Fuller test (ADF; [Dickey and Fuller \(1979\)](#)) with AIC-selected lag order of 8 strongly rejects the unit root null for log-returns (ADF statistic = −15.806, $p < 0.001$), confirming stationarity. The complementary KPSS test ([Kwiatkowski, Phillips, Schmidt, & Shin, 1992](#)) cannot reject the stationarity null (statistic = 0.182, $p \geq 0.10$). Price levels, by contrast, exhibit a clear unit root (ADF $p = 0.640$). Together these results confirm the standard result that log-returns, not prices, should be the object of modelling.

Second, [Engle \(1982\)](#) ARCH–LM test with 10 lags yields a statistic of 51.71 ($p < 0.001$), providing overwhelming evidence of conditional heteroscedasticity and directly motivating the GARCH framework. The Ljung–Box $Q(10)$ test on squared returns rejects the null of no autocorrelation ($p < 0.001$), confirming persistent volatility clustering. These properties establish that the conditional variance is a time-varying process, not a constant.

Third, and distinctively for the post-2020 sample, the Ljung–Box $Q(10)$ test on raw returns also rejects the null ($p < 0.001$, lag-1 ACF = −0.060). This statistically significant short-run mean reversion is absent from the full 2014–2026 history (where LB(10) $p = 0.025$), suggesting a structural change in Bitcoin microstructure associated with the growth of algorithmic market-making and cross-exchange arbitrage in the post-institutional-adoption era. While economically small and insufficient to generate profitable trading strategies after transaction costs, this autocorrelation motivates the inclusion of a non-zero conditional mean in all estimated models.

4. Methods

4.1. GARCH-Family Specifications

Let r_t denote the daily log-return, μ the conditional mean, and σ_t^2 the conditional variance. All specifications share the mean equation:

$$r_t = \mu + \sigma_t \varepsilon_t, \varepsilon_t \sim D(0, 1) \quad (1)$$

where $D(0,1)$ denotes a zero-mean, unit-variance distribution. Throughout, $a_t \equiv r_t - \mu = \sigma_t \varepsilon_t$ denotes the mean-corrected return (the model innovation) and $z_t \equiv \varepsilon_t = a_t / \sigma_t$ the standardized residual. The baseline GARCH(1,1) specification of [Bollerslev \(1986\)](#) parameterizes the conditional variance as:

$$\sigma_t^2 = \omega + \alpha a_{t-1}^2 + \beta \sigma_{t-1}^2, \omega > 0, \alpha \geq 0, \beta \geq 0 \quad (2)$$

with stationarity requiring $\alpha + \beta < 1$. The GJR–GARCH(1,1) of [Glosten, Jagannathan, and Runkle \(1993\)](#) augments [Equation 2](#) with an asymmetric term:

$$\sigma_t^2 = \omega + (\alpha + \gamma 1_{\{a_{t-1} < 0\}}) a_{t-1}^2 + \beta \sigma_{t-1}^2 \quad (3)$$

where γ captures the leverage effect. The preferred specification, the EGARCH(1,1) of [Nelson \(1991\)](#) models the log-conditional variance:

$$\ln(\sigma_t^2) = \omega + \alpha(|z_{t-1}| - E|z_{t-1}|) + \gamma z_{t-1} + \beta \ln(\sigma_{t-1}^2), z_t = \varepsilon_t = \frac{a_t}{\sigma_t} \quad (4)$$

The signed term γz_{t-1} captures the leverage effect, the asymmetric response of volatility to negative versus positive return shocks, while the centered magnitude term $\alpha(|z_{t-1}| - E|z_{t-1}|)$ governs the symmetric ARCH response.

The logarithmic transformation in [Equation 4](#) ensures conditional variance positivity without imposing sign restrictions on ω , α , or β , an important advantage when parameters take near-boundary values as they do for Bitcoin. The FIGARCH(1,d,1) model of [Baillie, Bollerslev, and Mikkelsen \(1996\)](#) replaces the lag polynomial in [Equation 2](#) with a fractionally integrated operator, parameterized by $d \in (0,1)$. All specifications are estimated by maximum likelihood using robust ([Bollerslev & Wooldridge, 1992](#)) standard errors. The study tests three error distributions D : standard Normal, Student- t with degrees of freedom ν , and skewed Student- t . Model selection uses $AIC = -2\mathcal{L} + 2k$ and $BIC = -2\mathcal{L} + k \ln(N)$.

4.2. Hidden Markov Model

Let $S_t \in \{0, 1, 2\}$ denote a discrete latent state following a first-order Markov chain. The transition probability matrix is:

$$P = [p_{ij}] \text{ where } p_{ij} = P(S_t = j | S_{t-1} = i), \sum_j p_{ij} = 1 \quad (5)$$

Conditionally on $S_t = s$, returns follow a Gaussian distribution:

$$r_t | S_t = s \sim N(\mu_s, \sigma_s^2) \quad (6)$$

Parameters are estimated by the Baum–Welch expectation-maximization algorithm ([Baum, Petrie, Soules, & Weiss, 1970](#)). State assignment uses the [Viterbi \(1967\)](#) for the most-likely sequence and the forward algorithm for the real-time filtered state probabilities used in all out-of-sample forecasts (the forward–backward smoother is reserved for full-sample regime characterization):

$$\hat{p}_t(s) = P(S_t = s | r_{1:t}) \quad (7)$$

The number of states $K = 3$ is selected by BIC, which achieves its minimum at $K = 3$ (BIC = 11,392.5 vs. 11,457.5 for $K = 2$ and 11,420.7 for $K = 4$). States are ordered by estimated conditional volatility σ_s : quiet ($S = 0$), normal ($S = 1$), crisis ($S = 2$).

4.3. Hybrid EGARCH+HMM Model

The Hybrid EGARCH+HMM model addresses the key limitation of standalone EGARCH, its regime-invariant dynamics, by scaling the EGARCH conditional volatility forecast using HMM-derived regime probabilities. Define the regime-specific scaling multiplier:

$$m_s = \frac{E[|r_t| | S_t = s]}{E[\hat{\sigma}_{t,EGARCH} | S_t = s]} \quad (8)$$

estimated from the full in-sample data. The hybrid conditional volatility forecast is then:

$$\hat{\sigma}_{t|Hybrid} = \hat{\sigma}_{t|EGARCH} \times \sum_{s=0}^{K-1} \hat{p}_t(s) m_s \quad (9)$$

This formulation preserves the EGARCH within-regime dynamics (the fast-updating α and slow-decaying β) while using soft-weighted regime probabilities to adjust the volatility level continuously as the HMM infers the current market state. The multiplicative structure has a natural interpretation: $m_0 \approx 0.28$ implies that EGARCH over-predicts volatility in the quiet regime by a factor of approximately 3.6, consistent with the persistent-memory parameter $\beta = 0.986$ retaining a long shadow of past crises.

All quantities in Equation 9 are computable in real time: $\hat{p}_t(s)$ conditions only on the history $r_{1:t}$, using the forward algorithm without look-ahead. The transition matrix and emission parameters are estimated on the full sample.

4.4. Markov-Switching EGARCH (MS-EGARCH+HMM) Model

The MS-EGARCH+HMM specification extends the hybrid framework by fitting a structurally distinct conditional variance process in each regime. Per-regime model selection is guided by the statistical properties of each regime's return subseries: (i) the quiet-regime EGARCH yields $\beta \approx 0$ and $v \approx 170$ (near-Gaussian, near-constant variance), so a constant-volatility process is used: $\hat{\sigma}_t^{(0)} = \bar{\sigma}_0$; (ii) the normal regime uses EGARCH(1,1)-t with stable positive parameters; (iii) the crisis-regime EGARCH yields $\alpha = -0.233$ (a numerical pathology from 311 asymmetric crash observations), so GARCH(1,1)-t is used for the crisis regime. Let $\hat{\sigma}_t^{(N)}$ and $\hat{\sigma}_t^{(C)}$ denote the normal-regime EGARCH and crisis-regime GARCH conditional volatilities, computed by applying regime-specific parameters to the full return series with regime unconditional variance initialization. The MS-EGARCH conditional volatility is:

$$\hat{\sigma}_{t|MS-EGARCH} = \hat{p}_t(0) \bar{\sigma}_0 + \hat{p}_t(1) \hat{\sigma}_t^{(N)} + \hat{p}_t(2) \hat{\sigma}_t^{(C)} \quad (10)$$

This specification captures the qualitative insight that the volatility-generating mechanism is structurally different across regimes: memoryless i.i.d. in the quiet state, strongly persistent EGARCH dynamics in the normal state, and slowly-decaying pure GARCH turbulence in the crisis state.

4.5. Forecast Evaluation

One-step-ahead volatility forecasts are evaluated on the 465-observation test period (February 2025 – May 2026) using three loss functions. Let $\hat{r}_t^v$ denote the realized-volatility proxy and $\hat{\sigma}_t$ the model forecast.

Root Mean Squared Error:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{r}_t^v - \hat{\sigma}_t)^2} \quad (11)$$

Mean Absolute Error:

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{r}_t^v - \hat{\sigma}_t| \quad (12)$$

QLIKE (Patton, 2011):

$$QLIKE = \frac{1}{n} \sum_{t=1}^n \left[\frac{(\hat{r}_t^v)^2}{\hat{\sigma}_t^2} + \ln(\hat{\sigma}_t^2) \right] \quad (13)$$

QLIKE is the only loss function that is consistent, i.e., ranks models correctly in population, under standard regularity conditions regardless of the realized-volatility proxy used (Patton, 2011). The primary proxy is $|r_t|$; robustness is assessed using the Parkinson (1980) and Garman and Klass (1980) range-based estimators. Statistical significance of pairwise forecast differences is assessed using the Diebold and Mariano (1995) test with Newey–West HAC variance correction (5 lags).

4.6. Value-at-Risk Specification and Backtesting

One-day-ahead VaR at confidence level $1 - \alpha$ is defined as the loss threshold not exceeded with probability α :

$$VaR_t^{(1-\alpha)} = -(\mu + \hat{\sigma}_t F^{-1}(\alpha)) \quad (14)$$

where $F^{-1}(\alpha)$ is the α -quantile of the assumed error distribution. Under the Normal distribution, $F^{-1}(\alpha) = \Phi^{-1}(\alpha)$; under the Student-t with ν degrees of freedom, $F^{-1}(\alpha) = t^{-1}\nu(\alpha)$, understood as the standardized quantile $\sqrt{(\nu-2)/\nu} t_{\nu}^{-1}(\alpha)$ so that the innovation has unit variance. The Expected Shortfall (Conditional VaR) at confidence $1 - \alpha$ is:

$$ES_t^{(1-\alpha)} = \hat{\sigma}_t \frac{f_t(t_{\nu}^{-1}(\alpha))}{\alpha} \frac{\nu + [t_{\nu}^{-1}(\alpha)]^2}{\nu-1} \quad (15)$$

where f_t is the Student-t density. The Kupiec (1995) unconditional coverage likelihood-ratio test evaluates whether the observed Value-at-Risk (VaR) violation frequency $\hat{\alpha} = \frac{1}{n} \sum_{t=1}^n 1\{r_t < -VaR_t\}$ is statistically consistent with the nominal tail probability α :

$$LR_{uc} = -2[n_1 \ln(\alpha) + n_0 \ln(1 - \alpha) - n_1 \ln(\hat{\alpha}) - n_0 \ln(1 - \hat{\alpha})] \sim \chi^2 \quad (16)$$

where n_1 = number of violations and $n_0 = n - n_1$. The model passes if p-value > 0.05 (null: observed rate equals nominal rate).

Because unconditional coverage alone ignores the temporal clustering of exceptions, the Christoffersen (1998) conditional-coverage test is also reported. It combines the Kupiec statistic with a first-order Markov independence statistic LR_{ind} into a joint test of correct conditional coverage:

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi^2 \quad (17)$$

A model is rejected if it generates either the wrong number of violations or violations that arrive in clusters.

5. In-Sample Estimation Results

5.1. GARCH Model Selection

Table 4 reports maximum likelihood estimates for ten GARCH specifications. The EGARCH(1,1)-t achieves the lowest AIC (11,200.0) and BIC (11,228.7), establishing it as the preferred specification. All Student-t or skewed-t models substantially outperform the Gaussian benchmark ($\Delta AIC = 521.9$), quantifying the catastrophic cost of normality misspecification. The EGARCH(1,1)-skew-t comes within 0.3 AIC points but requires an additional parameter; we select the simpler EGARCH(1,1)-t as the parsimonious preferred model. The GJR-GARCH leverage term, which would capture asymmetric volatility (larger response to negative than positive shocks), is statistically insignificant ($p = 0.647$), confirming that Bitcoin does not exhibit the equity-style leverage effect.

Table 4. In-sample GARCH model comparison.

Model	Distribution	k	Log-Lik.	AIC	BIC	ΔAIC
GARCH(1,1)	Normal	4	-5,857.0	11,721.9	11,744.9	+521.9
GARCH(1,1)	Student-t	5	-5,600.6	11,211.2	11,240.0	+11.2
GARCH(1,1)	Skew-t	6	-5,599.9	11,211.7	11,246.2	+11.7
GJR-GARCH(1,1)	Student-t	6	-5,600.5	11,212.9	11,247.4	+12.9
FIGARCH(1,1)	Student-t	6	-5,600.6	11,213.3	11,247.8	+13.3
EGARCH(1,1) — skew-t	Skew-t	6	-5,594.1	11,200.3	11,234.8	+0.3
EGARCH(1,1) ★	Student-t	5	-5,595.0	11,200.0	11,228.7	0.0

Note: k = number of estimated parameters. ★ = selected preferred model. AIC = $-2LL + 2k$; BIC = $-2LL + k \cdot \ln(N)$. ΔAIC measured relative to the best (EGARCH-t). Green = best model row; red = misspecified (Gaussian). All models estimated by maximum likelihood with Bollerslev-Wooldridge robust standard errors.

5.2. EGARCH(1,1)-t Parameter Estimates

Table 5 presents the full parameter estimates for the preferred EGARCH(1,1)-t model. The shock parameter $\alpha = 0.1526$ ($t = 5.17$, $p < 0.001$) captures the log-variance sensitivity to standardized return magnitude. The persistence parameter $\beta = 0.9858$ ($t = 146.5$) is close to unity, implying a half-life of $\ln(0.5)/\ln(0.9858) \approx 48.6$ days for volatility shocks. In the EGARCH formulation, the stationarity condition is $\beta < 1$ (not $\alpha + \beta < 1$), which is satisfied. The sum $|\alpha| + \beta = 1.138$ exceeds unity but this is not an integrated variance process, it reflects the logarithmic parameterization.

Table 5. EGARCH(1,1)-t maximum-likelihood parameter estimates.

Parameter	Symbol	Estimate	Std. Error	t-Statistic	p-Value
Conditional mean	μ	0.0784	0.0420	1.869	0.062†
Log-variance intercept	ω	0.0567	0.0203	2.797	0.005**
Shock coefficient	α	0.1526	0.0295	5.168	<0.001***
Persistence	β	0.9858	0.0067	146.54	<0.001***
Degrees of freedom	ν	3.1056	0.2228	13.94	<0.001***

Note: Log-likelihood = -5,595.00. AIC = 11,199.99. BIC = 11,228.74. Robust (Bollerslev–Wooldridge) standard errors. Residual diagnostics: LB(10) on standardized residuals $p = 0.1612$ (no serial correlation); LB(10) on squared standardized residuals $p = 0.8640$ (No remaining ARCH effects). Both diagnostics confirm adequate model fit. † $p < 0.10$; ** $p < 0.01$; *** $p < 0.001$.

5.3. HMM Regime Characterization

Table 6 presents the three-state HMM regime properties. The quiet regime ($S = 0$) characterizes 43.2% of the sample, periods of near-Gaussian, near-constant volatility (0.988%/day, approximately 15.7% annualized) comparable to equity index volatility. The normal regime ($S = 1$, 43.4% of days) corresponds to the typical Bitcoin trending environment, with volatility of 3.40%/day (54% annualized). The crisis regime ($S = 2$, 13.4% of days) captures sustained turbulence: 6.01%/day (95% annualized), six times the quiet-regime level, with a mean sojourn of 9.6 days. The EGARCH scale multipliers m_s show that EGARCH systematically over-predicts volatility in the quiet regime ($m_0 = 0.28$, a 72% over-prediction) while tracking more accurately in the normal and crisis regimes.

Table 6. HMM 3-state regime properties.

Property	Quiet ($S = 0$)	Normal ($S = 1$)	Crisis ($S = 2$)
Observations	1,004 (43.2%)	1,010 (43.4%)	311 (13.4%)
Mean return (%/day)	0.032	0.314	-0.346
Volatility (%/day)	0.988	3.398	6.007
Annualised volatility	15.7%	53.9%	95.4%
Mean sojourn (days)	2.9	3.7	9.6
EGARCH scale mult. (m_s)	0.28	0.82	0.88
Regime-specific model	Constant $\tilde{\sigma}_0$	EGARCH(1,1)-t	GARCH(1,1)-t

Note: States ordered by conditional volatility. Mean sojourn = $1/(1 - p_{\{ii\}})$. Scale multiplier $m_s = E[|r_{t,t}| \mid S = s] / E[\hat{\sigma}_{\{EGARCH\}} \mid S = s]$. Crisis regime ν (GARCH) = 5.17, substantially lower than global $\nu = 3.11$. Quiet-regime EGARCH: $\beta \approx 0$, $\nu \approx 170$ (effectively Gaussian constant-variance process).

5.4. HMM Transition Dynamics

Table 7 presents the estimated transition matrix, which reveals three structural features with direct implications for risk management. First, the crisis regime is highly self-persistent ($p_{22} = 0.896$, mean duration 9.6 days), reflecting the sustained nature of Bitcoin deleveraging episodes. Second, a direct quiet-to-crisis transition is essentially zero ($p_{02} \approx 0$), indicating that market turbulence escalates through the normal regime rather than in sudden jumps from calm to crisis. Third, the crisis regime never transitions directly to quiet ($p_{20} \approx 0$), requiring passage through the normal regime during de-escalation, a recovery path that typically requires two stages over several weeks.

Table 7. HMM transition probability matrix.

From \ To	→ Quiet ($S = 0$)	→ Normal ($S = 1$)	→ Crisis ($S = 2$)
Quiet ($S = 0$)	0.6561	0.3439	≈ 0
Normal ($S = 1$)	0.2514	0.7281	0.0205
Crisis ($S = 2$)	≈ 0	0.1044	0.8956

Note: Diagonal entries (Bold) = Self-persistence probability. Off-diagonal shows transition probabilities. Zero entries confirmed at < 0.001 . The asymmetric entry/Exit structure implies that crisis detection is largely reactive (the market must already be in crisis to assign high $p(S=2)$), but the 9.6-day duration provides sufficient time for risk-management response.

6. Out-of-Sample Forecast Evaluation

6.1. Evaluation Design and Benchmarks

The evaluation design employs a strict 80/20 temporal split: the first 1,860 observations (January 2020 – February 4, 2025) form the training sample; the remaining 465 observations (February 5, 2025 – May 14, 2026) constitute the test period. All model parameters are estimated on the training sample and fixed throughout the test period. Benchmarks include the historical volatility estimator ($\tilde{\sigma}$ = sample standard deviation of training returns) and the rolling 21-day historical volatility (RW21), commonly used in practice.

6.2. Aggregate Out-of-Sample Performance

Table 8 presents the central out-of-sample results. The Hybrid EGARCH+HMM achieves the lowest RMSE (1.393), MAE (0.891), and QLIKE (2.150) of all specifications, representing improvements of 31.2%,

47.2%, and 19.6% respectively over standalone EGARCH. The MS-EGARCH+HMM is the second-best specification (RMSE = 1.523, -24.7% vs. EGARCH). Notably, the rolling 21-day historical volatility (RMSE = 1.858) outperforms both GARCH and EGARCH alone, reflecting the dominance of quiet-regime days in the test period (59.1% quiet, 33.8% normal, 7.1% crisis). This result, a simple historical estimator beating sophisticated GARCH, is predicted by theory: when the true variance is near-constant (quiet regime), GARCH's high-persistence structure creates unnecessary error by carrying forward the shadow of prior volatility. The hybrid solves exactly this problem.

Table 8. Out-of-sample volatility forecast evaluation (n = 465, Feb 2025 – May 2026).

Model	RMSE	MAE	QLIKE	Improvement vs EGARCH
Historical vol (Baseline)	2.441	2.183	2.926	—
Rolling 21-day vol	1.858	1.386	2.733	-8.2% RMSE
GARCH(1,1)-t	2.012	1.672	2.660	+0.6% (n.s.)
EGARCH(1,1)-t	2.024	1.687	2.675	Baseline
MS-EGARCH + HMM	1.523	1.287	2.238	-24.7% RMSE ***
Hybrid EGARCH + HMM ★	1.393	0.891	2.150	-31.2% RMSE ***

Note: RV proxy = Absolute daily log-return $|r_t|$. ★ = Best model overall. *** Diebold–Mariano $p < 0.001$ (Newey–West HAC, 5 lags). QLIKE = Patton (2011) consistent loss function. The Hybrid wins all 16 calendar months of the test period. GARCH and EGARCH are statistically indistinguishable (DM = 0.578, $p = 0.563$).

6.3. Statistical Significance: Diebold–Mariano Tests

Pairwise Diebold–Mariano tests under squared-error loss confirm the forecast ranking. The test statistic for EGARCH versus Hybrid is $DM = 11.37$ ($p < 0.001$), strongly rejecting equal accuracy in favor of the hybrid. GARCH versus Hybrid yields $DM = 9.71$ ($p < 0.001$). The comparison between the two hybrid specifications, MS-EGARCH versus Hybrid, yields $DM = -1.96$ ($p = 0.050$): the multiplicative hybrid is marginally but statistically significantly superior to the structurally richer MS-EGARCH. This result that the simpler model wins is important and is further interpreted in Section 8.

6.4. Robustness: Alternative Realized Volatility Proxies

Table 9 evaluates the hybrid against the Parkinson (1980) high–low range estimator and the Garman and Klass (1980) OHLC estimator, which are both less noisy than absolute returns and provide ex-post benchmarks with smaller idiosyncratic error. The hybrid improvement over EGARCH persists across all three proxies, 31.2% ($|r|$), 24.6% (Parkinson), and 19.0% (Garman & Klass), with no reversal in model rankings. The attenuation for smoother proxies is expected: range-based estimators down-weight extreme-return observations that are disproportionately represented in the hybrid's regime adjustments.

Table 9. Robustness: alternative realized volatility proxies.

RV Proxy	GARCH RMSE	EGARCH RMSE	Hybrid RMSE	Hybrid vs EGARCH
$ r $ (primary)	2.012	2.024	1.393	-31.2%
Parkinson (1980) H–L range	1.405	1.416	1.068	-24.6%
Garman and Klass (1980) OHLC	1.323	1.332	1.080	-19.0%

Note: Parkinson (1980): $\sigma^2 = [4 \ln 2]^{-1} \times [\ln(H_t/L_t)]^2$. Garman and Klass (1980): $\sigma^2 = 0.5[\ln(H/L)]^2 - (2\ln 2 - 1)[\ln(C/O)]^2$. Both in %/day. *** DM $p < 0.001$ for all Hybrid vs. EGARCH comparisons across all three proxies.

7. Value-at-Risk Backtesting and Risk Management

7.1. Kupiec Backtest Results

Table 10 presents VaR backtest results across all model-distribution combinations at confidence levels 95%, 99%, and 99.5%. The results reveal a clear pattern: Normal-distribution specifications are better calibrated than Student-t specifications for this dataset. The GARCH–Normal and EGARCH–Normal pass the Kupiec test at all three confidence levels, producing violation rates that are statistically consistent with their nominal values (3.44%, 1.29%, and 0.65% vs. expected 5.0%, 1.0%, and 0.5% respectively).

The Student-t distribution with $\nu = 3.1$ is systematically over-conservative: at the 95% confidence level, EGARCH–t produces only 1.29% violations vs. 5.0% expected (Kupiec $p < 0.001$, FAIL). This occurs because $\nu \approx 3$ implies near-Cauchy tails, the 5th percentile of the standardized t distribution is at -4.54σ rather than the -1.64σ of the Normal, meaning virtually no actual Bitcoin returns fall below the Student-t VaR threshold during a predominantly calm test period. While labelled "FAIL" in the Kupiec framework, this represents excess risk capital reservation, a conservative error in the direction preferred for regulatory purposes.

The Hybrid+Normal passes at the 95% level (7.10% violations, $p = 0.050$) but fails at 99% and 99.5%, reflecting the tighter volatility estimates the hybrid produces during quiet-regime periods. The MS-EGARCH+Normal passes at 99% and 99.5%, where the crisis-regime GARCH with $\nu = 5.17$ appropriately widens the tail estimate.

Table 10. VaR backtest — Kupiec unconditional-coverage test ($n = 465$).

Model	Distribution	Level	Vio. rate	Expected	Kupiec p	Result
GARCH(1,1)	Normal	95%	3.44%	5.00%	0.103	PASS
GARCH(1,1)	Normal	99%	1.29%	1.00%	0.547	PASS
GARCH(1,1)	Normal	99.5%	0.65%	0.50%	0.671	PASS
EGARCH(1,1)	Normal	95%	3.23%	5.00%	0.061	PASS
EGARCH(1,1)	Normal	99%	1.29%	1.00%	0.547	PASS
EGARCH(1,1)	Normal	99.5%	0.86%	0.50%	0.318	PASS
EGARCH(1,1)	Student-t	95%	1.29%	5.00%	<0.001	FAIL†
MS-EGARCH+HMM	Normal	99%	1.29%	1.00%	0.547	PASS
MS-EGARCH+HMM	Normal	99.5%	0.22%	0.50%	0.326	PASS
Hybrid + HMM	Normal	95%	7.10%	5.00%	0.050	PASS
Hybrid + HMM	Normal	99%	2.15%	1.00%	0.031	FAIL
Hybrid + HMM	Normal	99.5%	1.94%	0.50%	0.001	FAIL

Note: PASS = Kupiec LR test $p > 0.05$ (null: violation rate = nominal level). FAIL† = too conservative (violation rate < nominal — excess risk capital). FAIL (red) = too liberal (violation rate > nominal — under-reserved). Student-t with $\nu = 3.1$ systematically too conservative. Practical recommendation: EGARCH–Normal and GARCH–Normal are best calibrated across all confidence levels.

7.2. Current Market Risk Estimates

Table 11 provides risk estimates as of the final sample date, 14 May 2026. The HMM classifies the current market in the normal regime with 68.6% posterior probability (quiet 30.3%, crisis 1.0%), consistent with a period of moderate market activity following the 2026 sell-off recovery. The EGARCH conditional volatility of 2.228%/day (35.4% annualized) is elevated relative to the hybrid estimate of 1.466%/day (23.3% annualized), reflecting the EGARCH's residual memory from the February 2026 turbulence. We recommend a tiered approach: Hybrid–Normal for day-to-day position sizing and stop-loss calibration; EGARCH–t for stress testing and regulatory capital ($\text{VaR}(99\%)\text{--t} = 9.847\%$); and MS-EGARCH for worst-case scenario reserves.

Table 11. Current risk estimates — 14 May 2026.

Risk Metric	EGARCH(1,1)–t	Hybrid EGARCH+HMM	Recommended use
Daily volatility (%/day)	2.228	1.466	Position sizing
Annualised volatility	35.4%	23.3%	Risk budgeting
VaR(95%) — Normal	3.665%	2.412%	Day-to-day stop-loss calibration
VaR(99%) — Normal	5.183%	3.411%	Position-level risk limits
VaR(95%) — Student-t	5.172%	3.403%	Stress testing
VaR(99%) — Student-t	9.847%	6.480%	Regulatory capital
ES/CVaR(95%) — Student-t	8.399%	5.527%	Expected loss beyond VaR
ES/CVaR(99%) — Student-t	14.961%	9.845%	Tail capital reserve
HMM state (14 May 2026)	Normal (68.6%)	Normal (68.6%)	Regime-conditional interpretation

Note: Risk metrics as of final sample date. HMM state probabilities: quiet = 30.3%, normal = 68.6%, crisis = 1.0%. All VaR and ES are 1-day ahead. $\text{ES}(99\%)\text{--t} = 14.961\%$ for EGARCH means the expected loss in the worst 1% of days is approximately 14.96% of position value. Tiered framework: (1) Hybrid–Normal for daily sizing; (2) EGARCH–t for stress test; (3) MS-EGARCH for capital reserves.

8. Robustness Analysis and Monte Carlo Validation

8.1. EGARCH Parameter Stability

Table 12 presents EGARCH(1,1)–t estimates across four expanding training samples ending at 50%, 60%, 70%, and 80% of the full sample. The shock parameter α ranges narrowly from 0.145 to 0.166 across all training windows; the persistence β from 0.984 to 0.991; and ν from 2.81 to 3.11. The stability of these parameter estimates across very different market periods from the COVID-shocked 2020 sample through the 2022 bear market and into the 2025 recovery confirms that the EGARCH(1,1)–t characterizes a genuine structural feature of Bitcoin return dynamics and is not an artefact of the specific training window chosen.

Table 12. EGARCH(1,1)–t parameter stability — expanding training samples.

Training end	n	α (Shock)	β (Persistence)	ν (DoF)
Mar 2023 (50%)	1,162	0.1619	0.9835	3.016
Oct 2023 (60%)	1,395	0.1658	0.9911	2.807
Jun 2024 (70%)	1,627	0.1599	0.9900	2.864
Feb 2025 (80%)	1,860	0.1449	0.9883	3.013
May 2026 (full)	2,325	0.1526	0.9858	3.106

Note: All subsamples estimated by maximum likelihood with Bollerslev–Wooldridge robust standard errors. Green = Full-sample estimates. Parameter ranges: $\alpha \in [0.145, 0.166]$, $\beta \in [0.983, 0.991]$, $\nu \in [2.807, 3.106]$ — all stable across subsamples.

8.2. Monte Carlo Simulation Under the EGARCH Null

A key concern for hybrid models is whether the apparent out-of-sample gains reflect genuine regime structure or spurious finite-sample flexibility of the HMM layer. We address this with a parametric Monte Carlo experiment. Under the null hypothesis that the true data-generating process is EGARCH(1,1)-t with the full-sample estimated parameters ($\mu = 0.0784$, $\omega = 0.0567$, $\alpha = 0.1526$, $\beta = 0.9858$, $v = 3.106$) and a latent three-state Markov chain with the estimated transition matrix and regime-specific volatility levels, we simulate 500 independent paths of length $N = 2,325$ and apply the identical 80/20 split and hybrid construction procedure to each simulation.

Under the null, the Monte Carlo mean RMSE improvement is -0.311 (standard deviation = 0.052), meaning that when the DGP is purely EGARCH, the hybrid systematically *loses* to EGARCH by 0.311 RMSE units in finite samples a predictable consequence of estimation noise in the HMM layer. The empirical RMSE improvement of $+0.631$ exceeds the Monte Carlo 99.9th percentile by 18.22 standard deviations ($p \approx 0$). An oracle variant, in which the true regime label is known at every time step, produces an identical mean of -0.311 , confirming that the limitation is not HMM classification accuracy but the fundamental inadequacy of the EGARCH DGP: if the true process is pure EGARCH, no regime adjustment helps regardless of classification perfection. The empirical data, by generating a $+0.631$ improvement against a null that would produce -0.311 , provide decisive evidence that Bitcoin volatility contains genuine latent regime structure beyond what EGARCH alone captures.

8.3. Why the Multiplicative Hybrid Outperforms MS-EGARCH

The result that the simpler multiplicative Hybrid marginally outperforms the structurally richer MS-EGARCH ($DM = -1.96$, $p = 0.050$) deserves interpretation. We identify three contributing factors. First, the per-regime EGARCH estimations are statistically fragile in the quiet ($n = 1,004$, $\beta \approx 0$, $v \approx 170$) and crisis ($n = 311$, negative α) regimes, the small-sample parameters introduce estimation noise that the multiplicative hybrid avoids by using a single globally-estimated EGARCH. Second, the MS-EGARCH model is not estimated by a full joint likelihood (which would require solving the Haas et al. (2004) path-dependence problem with Student-t innovations), so parameter estimates from separate regime subsamples are not mutually consistent. Third, during regime transitions, particularly quiet-to-normal, occur with probability of 0.344 per day, the MS-EGARCH crisis-regime GARCH path produces elevated forecasts even at low crisis probability, creating a mild upward bias absent from the multiplicative hybrid's smoother scaling. The practical implication is that the multiplicative hybrid is the preferred specification for day-to-day volatility forecasting, while the MS-EGARCH is valuable for its regime-conditional distributional interpretability in tail risk applications.

9. Discussion

The central finding that Bitcoin volatility exhibits genuine, quantifiable regime structure that materially improves volatility forecasting has several important implications for theory and practice. From a theoretical perspective, the near-unit-root EGARCH persistence ($\beta = 0.986$) is consistent with the integrated GARCH (IGARCH) result documented for equity markets (Engle & Bollerslev, 1986) but our regime evidence suggests an alternative interpretation: the high β may partly reflect the averaging of very different within-regime processes (near-zero β in quiet, $\beta = 0.966$ in normal) under a single-regime parameterization. This "spurious persistence" interpretation, formalized for regime-switching GARCH by Mikosch and Stărică (2004) and Lamoureux and Lastrapes (1990) is consistent with the MS-EGARCH finding that the normal-regime-specific $\beta = 0.966$ is lower than the global estimate.

The three-regime structure maps cleanly onto the Bitcoin market cycle. The quiet regime (15.7% annualized vol) corresponds to the consolidation and accumulation phases well-documented in Bitcoin's four-yearly halving cycle. The normal regime (54% ann.) corresponds to trending bull or bear conditions with regular volatility. The crisis regime (95% ann., 9.6-day sojourn) corresponds to the acute deleveraging episodes that punctuate Bitcoin's history, the COVID crash, the Terra/Luna implosion, and the FTX collapse. The asymmetric exit structure (crisis $\rightarrow$ normal $\rightarrow$ quiet, never direct crisis $\rightarrow$ quiet) reflects the two-stage recovery typical of leveraged asset markets: first, stop the bleeding (transition to normal); second, rebuild confidence (transition to quiet).

The degrees-of-freedom estimate $v = 3.106$ has direct practical consequences. With $v < 4$, the theoretical fourth moment is infinite, volatility is not just large but fundamentally unbounded. This is not an artefact of the post-2020 sample: the full 2014–2026 history yields $v = 2.81$, even lower. For VaR applications, this implies that Student-t-based VaR thresholds will be systematically over-conservative for most practical test periods, which are dominated by non-crisis regimes. The Kupiec test's FAIL labels for Student-t specifications in our backtest do not signal a risk underestimation problem, they signal excess conservatism. Practitioners must distinguish between these failure modes.

The 47.00% CAGR and 0.515 Sharpe ratio over the sample period, while impressive, are accompanied by a maximum drawdown of -76.63% , a figure that would be unacceptable in most institutional risk frameworks regardless of the return. The regime model contributes directly to managing this tension: by identifying the

current volatility state in real time, it allows position sizes to be reduced during elevated-regime periods, reducing drawdown risk without fully exiting the asset class. The hybrid's current estimate (23.3% annualized, normal regime, 68.6% probability) provides an actionable basis for this calibration.

Several limitations should guide future research. First, the HMM transition matrix is estimated on the full sample and applied recursively, a mild in-sample optimism bias that would be eliminated by a genuine expanding-window HMM re-estimation, though at substantial computational cost. Second, the model is univariate; extending the HMM to a multivariate setting capturing cross-asset crypto co-volatility (e.g., BTC–ETH co-movements) would be theoretically richer and practically relevant for portfolio managers. Third, the MS-EGARCH model uses separate per-regime estimations rather than a full joint likelihood; implementing the Haas et al. (2004) path-dependence resolution with near-Cauchy Student-t innovations remains an open computational challenge. Fourth, high-frequency (intraday) data would enable the construction of realized volatility measures that are substantially less noisy than daily absolute returns, potentially yielding larger measured improvements.

10. Conclusion

This paper has proposed and rigorously validated a hybrid EGARCH–HMM framework for Bitcoin volatility forecasting and risk management. Estimated on 2,325 daily observations from January 2020 to May 2026, a period spanning five distinct market regimes, the framework delivers consistent, substantial improvements over every standalone benchmark.

The core empirical findings are: (i) the EGARCH(1,1) with Student-t errors ($v = 3.106$) is the best single-regime specification by AIC and BIC, with near-unit-root persistence ($\beta = 0.986$) and a half-life of 48.6 days; (ii) a three-state HMM identifies economically interpretable volatility regimes, quiet ($\sigma \approx 1\%/day$), normal ($\sigma \approx 3.4\%/day$), and crisis ($\sigma \approx 6\%/day$), with asymmetric transition dynamics (crisis sojourn ≈ 10 days, no direct crisis-to-quiet exit); (iii) the Hybrid EGARCH+HMM reduces out-of-sample RMSE by 31.2% and MAE by 47.2% relative to standalone EGARCH, winning every calendar month of the 16-month test period; (iv) a parametric Monte Carlo simulation with 500 paths under the EGARCH null places the empirical improvement 18.22 standard deviations above chance, confirming genuine regime structure; and (v) EGARCH–Normal provides the best-calibrated VaR at all three confidence levels in Kupiec backtesting.

The practical implication is clear: Bitcoin volatility is not well-served by any single model. The most effective approach is a regime-conditional hierarchy: the Hybrid EGARCH+HMM for day-to-day position sizing and stop-loss calibration, EGARCH–t for stress testing and regulatory capital, and MS-EGARCH for worst-case scenario reserve estimation. As of 14 May 2026, the market is classified in the normal regime with 68.6% probability, implying a current hybrid volatility estimate of 23.3% annualized and a 1-day VaR(99%) of 3.4% under the Normal distribution.

Future directions include extending the framework to multivariate crypto portfolios, incorporating order-book and on-chain data as regime-informative covariates, applying the hybrid structure to other highly non-normal asset classes such as energy commodities, and developing a fully Bayesian MS-EGARCH estimator that resolves the path-dependence problem for near-Cauchy Student-t innovations.

References

- Agakishiev, I., Härdle, W. K., Becker, D., & Zuo, X. (2025). Regime switching forecasting for cryptocurrencies. *Digital Finance*, 7(1), 107–131. <https://doi.org/10.1007/s42521-024-00123-2>
- 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](https://doi.org/10.1016/S0304-4076(95)01749-6)
- Baum, L. E., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1), 164–171. <https://doi.org/10.1214/aoms/1177697196>
- Baur, D. G., & Dimpfl, T. (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters*, 173, 148–151. <https://doi.org/10.1016/j.econlet.2018.10.008>
- Black, F. (1976). *Studies of stock price volatility changes*. Paper presented at the Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, American Statistical Association (pp. 177–181). Washington, DC: American Statistical Association.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), 307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- Bollerslev, T., & Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 11(2), 143–172. <https://doi.org/10.1080/07474939208800229>
- Borri, N. (2019). Conditional tail-risk in cryptocurrency markets. *Journal of Empirical Finance*, 50, 1–19. <https://doi.org/10.1016/j.jempfin.2018.11.002>
- Bouri, E., Gupta, R., Tiwari, A. K., & Roubaud, D. (2017). Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters*, 23, 87–95. <https://doi.org/10.1016/j.frl.2017.02.009>
- Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier? *Finance Research Letters*, 20, 192–198. <https://doi.org/10.1016/j.frl.2016.09.025>

- Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics*, 12(3), 309-316. <https://doi.org/10.1080/07350015.1994.10524546>
- Chinazzo, C., & Jeleskovic, V. (2024). Forecasting Bitcoin volatility: A comparative analysis of volatility approaches. *arXiv preprint arXiv:2401.02049*. <https://doi.org/10.48550/arXiv.2401.02049>
- Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics*, 10(4), 407-432. [https://doi.org/10.1016/0304-405X\(82\)90018-6](https://doi.org/10.1016/0304-405X(82)90018-6)
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841-862. <https://doi.org/10.2307/2527341>
- Conrad, C., Custovic, A., & Ghysels, E. (2018). Long-and short-term cryptocurrency volatility components: A GARCH-MIDAS analysis. *Journal of Risk and Financial Management*, 11(2), 23. <https://doi.org/10.3390/jrfm11020023>
- 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>
- 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(366a), 427-431. <https://doi.org/10.1080/01621459.1979.10482531>
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253-263. <https://doi.org/10.1080/07350015.1995.10524599>
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 50(4), 987-1007. <https://doi.org/10.2307/1912773>
- Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, 5(1), 1-50. <https://doi.org/10.1080/07474938608800095>
- Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, 53(1), 67-78. <https://doi.org/10.1086/296072>
- Gkillas, K., & Katsiampa, P. (2018). An application of extreme value theory to cryptocurrencies. *Economics Letters*, 164, 109-111. <https://doi.org/10.1016/j.econlet.2018.01.020>
- 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>
- Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42(1), 27-62. [https://doi.org/10.1016/0304-405X\(96\)00875-6](https://doi.org/10.1016/0304-405X(96)00875-6)
- Guidolin, M., & Timmermann, A. (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *The Review of Financial Studies*, 21(2), 889-935. <https://doi.org/10.1093/rfs/hhn006>
- Haas, M., Mittnik, S., & Paolella, M. S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4), 493-530. <https://doi.org/10.1093/jjfinec/nbh020>
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 57(2), 357-384. <https://doi.org/10.2307/1912559>
- Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64(1-2), 307-333. [https://doi.org/10.1016/0304-4076\(94\)90067-1](https://doi.org/10.1016/0304-4076(94)90067-1)
- Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: Does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7), 873-889. <https://doi.org/10.1002/jae.800>
- Härdle, W. K., Harvey, C. R., & Reule, R. C. G. (2020). Understanding cryptocurrencies. *Journal of Financial Econometrics*, 18(2), 181-208. <https://doi.org/10.1093/jjfinec/nbz033>
- Hayes, A. S. (2017). Cryptocurrency value formation: An empirical study leading to a cost of production model for valuing bitcoin. *Telematics and Informatics*, 34(7), 1308-1321. <https://doi.org/10.1016/j.tele.2016.05.005>
- Huang, Z.-C., Sangiorgi, I., & Urquhart, A. (2024). Forecasting Bitcoin volatility using machine learning techniques. *Journal of International Financial Markets, Institutions and Money*, 97, 102064. <https://doi.org/10.1016/j.intfin.2024.102064>
- Karam, S., El Maalouf, J., & Dirani, N. (2026). Hybrid regime-switching models for cryptocurrency prices: An asset-dependent performance analysis using Markov chains and random forests. *Stats*, 9(4), 71. <https://doi.org/10.3390/stats9040071>
- Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, 158, 3-6. <https://doi.org/10.1016/j.econlet.2017.06.023>
- Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, 27(2), 363-394. <https://doi.org/10.1007/s001810100100>
- Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2), 73-84. <https://doi.org/10.3905/jod.1995.407942>
- Kwiatkowski, D., Phillips, P. C., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3), 159-178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, 8(2), 225-234. <https://doi.org/10.1080/07350015.1990.10509794>
- Mikosch, T., & Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86(1), 378-390. <https://doi.org/10.1162/003465304323023886>
- Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 59(2), 347-370. <https://doi.org/10.2307/2938260>
- Osterrieder, J., & Lorenz, J. (2017). A statistical risk assessment of Bitcoin and its extreme tail behavior. *Annals of Financial Economics*, 12(01), 1750003. <https://doi.org/10.1142/S2010495217500038>
- Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, 53(1), 61-65. <https://doi.org/10.1086/296071>

- Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1), 246-256. <https://doi.org/10.1016/j.jeconom.2010.03.034>
- Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286. <https://doi.org/10.1109/5.18626>
- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, 148, 80-82. <https://doi.org/10.1016/j.econlet.2016.09.019>
- Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2), 260-269. <https://doi.org/10.1109/TIT.1967.1054010>