Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

On-Chain Determinants of Bitcoin Realized Volatility: A Structural-Break-Aware Forecasting Approach

HAN GAO

alyssagaohan@gmail.com

Changzhou Institute of Technology

PEI SHAN FAM

Universiti Sains Malaysia

Research Article

Keywords: Bitcoin realized volatility, Garman-Klass range-based estimator, HAR-RV

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11806418/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

On-Chain Determinants of Bitcoin Realized Volatility: A Structural-Break-Aware Forecasting Approach

Abstract This study examines whether on-chain activity metrics provide incremental information for forecasting Bitcoin realized volatility beyond the benchmark HAR-RV model. Using daily data from August 2017 to May 2026, a Garman-Klass range-based realized volatility measure is constructed. A dominant structural break on 1 October 2022 partitions the sample into two regimes; a secondary May 2018 break is retained as a robustness diagnostic. LASSO-regularized HAR-RV estimation with signed semivariance raises in-sample R^2 from 0.43 to 0.51 (pre-2022) and 0.20 to 0.43 (post-2022). In a real-time rolling out-of-sample design with lagged on-chain predictors, the augmented model significantly outperforms the benchmark in both periods (DM = 2.79, $p = 0.005$; DM = 4.83, $p < 0.001$). Gains are daily-horizon specific; at the 14-day horizon the augmented model performs significantly worse. Results are robust to Elastic Net and macroeconomic controls. On-chain variables deliver economically meaningful real-time gains for daily Bitcoin realized volatility, especially post-2022.

Keywords: Bitcoin realized volatility; Garman-Klass range-based estimator; HAR-RV

1. Introduction

Bitcoin, the world's largest cryptocurrency by market capitalization, has evolved from a niche technological experiment into a trillion-dollar financial asset integrated into mainstream portfolios through spot exchange-traded funds (ETFs) approved in early 2024. This transformation has elevated the importance of accurately modeling and forecasting Bitcoin's volatility, a critical input for risk management, derivative pricing, portfolio allocation, and regulatory oversight. Bitcoin's volatility dynamics are structurally distinctive: lacking the cash-flow anchors that ground traditional asset prices, Bitcoin behaves as a purely speculative asset (Cheah & Fry, 2015) whose volatility exceeds that of major fiat currencies by an order of magnitude and rivals that of small-cap equities (Baur & Dimpfl, 2021; Dyhrberg, 2016; Katsiampa, 2017). Its volatility exhibits pronounced clustering, intermittent spikes, and regime shifts that coincide with market crises, regulatory interventions, and technological milestones (Corbet et al., 2019; Bouri et al., 2017). These features pose both a challenge and an opportunity for volatility modeling.

The realized-volatility (RV) framework has become the dominant paradigm for volatility measurement. By aggregating high-frequency squared intraday returns it delivers a model-free,

ex-post measure of latent volatility (Andersen & Bollerslev, 1998; Andersen et al., 2003; Barndorff-Nielsen & Shephard, 2002). In this study, with daily rather than tick-level data, the dependent variable is the natural logarithm of the Garman-Klass (1980) range-based realized volatility, constructed from daily open, high, low, and close prices. This estimator uses the intraday price range to recover within-day variance without requiring tick data, and avoids the overlapping-window mechanical autocorrelation that inflates persistence in smoothed multi-day standard deviations, making it well suited to daily-frequency modeling. This measure is modeled with the Heterogeneous Autoregressive (HAR) structure of Corsi (2009), which builds on the heterogeneous market hypothesis of Müller et al. (1997) and forms daily, weekly, and monthly components from frequency-averaged realized variances. Within this framework the HAR-RV model has achieved widespread adoption for its parsimony and consistently strong forecasting performance, and extensions incorporating jumps (Andersen et al., 2007), implied-volatility information (Busch et al., 2011), leverage effects (Corsi & Renò, 2012), and machine-learning regularization (Christensen et al., 2023; Gu, Kelly, & Xiu, 2020) have further enhanced its flexibility.

A unique feature of Bitcoin is the public, near-real-time observability of its blockchain ledger, which generates a rich stream of on-chain metrics, including transaction volumes, active addresses, hash rates, miner revenues, and transaction fees, that record network-level economic activity. This transparency has motivated a growing literature investigating the relationship between on-chain variables and Bitcoin's price and volatility dynamics. Early contributions established that on-chain metrics carry significant explanatory power: Aalborg, Molnár, and de Vries (2019) found that transaction volume, unique addresses, and the trading-to-transaction volume ratio significantly explain Bitcoin's volatility; Kristoufek (2015) applied wavelet coherence analysis to demonstrate fundamental links between network activity and Bitcoin's price over longer horizons. The theoretical underpinnings draw from network economics, Metcalfe's Law linking network value to the square of its user base has been empirically validated for Bitcoin (Peterson, 2018; Alabi, 2017), and from production-cost arguments, whereby mining difficulty and hash rate establish a marginal cost floor that constrains price dynamics (Hayes, 2019). Transaction fee behavior reflects the interaction between block-space supply and demand, providing a real-time gauge of network congestion and market stress (Easley et al., 2019).

More recent research has employed sophisticated forecasting frameworks to evaluate on-chain variables' predictive contribution. Liu and Tsyvinski (2021) established that cryptocurrency

returns are driven by factors specific to crypto markets, including momentum and investor attention, that are distinct from traditional asset-pricing factors. Chi, Chu, and Hao (2024) investigated the intraday return- and volatility-forecasting power of on-chain flow data across Bitcoin, Ethereum, and Tether, finding that exchange net inflows exhibit statistically significant forecasting power at high frequencies. Grobys, Näsman, and Sandretto (2026) showed that on-chain sentiment signals, including the Market Value to Realized Value (MVRV) *Z*-score, can time Bitcoin market cycles and significantly improve risk-adjusted performance relative to buy-and-hold benchmarks. Makarov and Schoar (2020) documented the central role of exchange-level flows in shaping liquidity and price dynamics across cryptocurrency markets. Collectively, these studies suggest that on-chain variables capture economic information that is distinct from, and potentially complementary to, the information embedded in price and volume time series.

A further complication is the presence of structural breaks in the Bitcoin market (Ardia et al., 2019). The sample period from 2017 to 2026 encompasses at least two watershed events: the COVID-19 pandemic in March 2020, which triggered a global liquidity crisis and a re-evaluation of Bitcoin's role in diversified portfolios (Conlon & McGee, 2020; Corbet et al., 2020), and the collapse of the FTX exchange in November 2022, which sent systemic shockwaves through the cryptocurrency ecosystem and fundamentally altered market structure, trust, and regulatory posture (Akyildirim et al., 2023). As reported in Section 3.1, the Quandt (1960)–Andrews (1993) family of structural break tests, extended by Bai and Perron (1998) to accommodate multiple unknown breakpoints, identifies 1 October 2022 as the dominant breakpoint, roughly five weeks before the FTX collapse, together with a secondary break in May 2018, providing a rigorous statistical framework for detecting such regime shifts. Failure to account for time-varying relationships risks estimating average effects that are unrepresentative of any specific regime, a concern that applies with particular force to the dynamically evolving cryptocurrency market.

Several gaps in the existing literature motivate the present research. First, few studies have embedded on-chain variables within a rigorous HAR-RV framework that explicitly models the long-memory dynamics of Bitcoin volatility. Taken in isolation, on-chain variables may spuriously absorb the persistence that the HAR-RV structure is designed to capture. Second, existing research rarely subjects on-chain–augmented models to out-of-sample forecasting evaluation, relying instead on in-sample goodness-of-fit as a measure of predictive relevance, an unreliable guide when predictors are numerous and potentially correlated (Timmermann,

2008; Rapach et al., 2010; Neely et al., 2014). Third, the literature has not systematically examined whether the explanatory and predictive value of on-chain variables shifts across distinct market regimes identified by structural break tests. Bitcoin's volatility dynamics in the pre-October 2022 era, characterized by retail-dominated trading and nascent institutional participation, may differ materially from those in the post-2022 institutional era, where institutional custody, spot ETFs, and enhanced regulatory oversight have reshaped market structure. Fourth, the extant literature has largely conflated the explanatory and predictive value of on-chain data. A variable may be significantly correlated with volatility in the full sample while making no contribution to real-time, one-step-ahead forecasts.

Against this backdrop, the present research makes four contributions. First, using daily data spanning August 2017 to May 2026, the research identifies a dominant structural break on 1 October 2022 (with a secondary break in May 2018) via the Quandt-Andrews test and partitions the sample into two regimes, ensuring that parameter estimates and forecast evaluations are regime-specific. Second, LASSO-regularized HAR-RV models are estimated within each regime, selecting from seven widely cited on-chain variables (active addresses, transaction count, hash rate, mining difficulty, miner revenue, transaction fees, and on-chain transaction volume), while always retaining the HAR triplet and the two signed semivariance terms. Third, the LASSO-augmented model is subjected to a rolling-window out-of-sample forecasting exercise, with predictive accuracy evaluated via the Diebold-Mariano (1995) test, thereby distinguishing between in-sample explanatory power and genuine predictive value. Fourth, four robustness checks (varying the window length (250 and 750 days), employing Elastic Net regularization (Zou & Hastie, 2005), testing a 14-day volatility horizon, and incorporating macroeconomic and sentiment controls) verify the stability of the core findings, which are further corroborated by break-date and economic-value diagnostics (Section 3.5).

The remainder of the paper is organized as follows. Section 2 describes the data sources, variable construction, and summary statistics. Section 3 presents the methodology and empirical results: Section 3.1 estimates structural breaks and defines the sub-periods; Section 3.2 reports LASSO-regularized HAR-RV estimation; Section 3.3 conducts the rolling-window out-of-sample forecasting exercise; Section 3.4 presents robustness checks, and Section 3.5 assesses lead-lag relations, break-date robustness, and the economic value of the forecasts. Section 4 discusses the implications of the findings for investors, regulators, and blockchain analytics, acknowledges limitations, and concludes with directions for future research.

2. Data

This section describes the data used in the empirical analysis and robustness checks. Section 2.1 details the original data sources and variable definitions; Section 2.2 describes the preprocessing procedures; Section 2.3 presents descriptive statistics.

2.1 Original Data

Four categories of original variables underpin the empirical analysis: (i) market price data (OHLCV from Binance), (ii) seven on-chain activity measures (Blockchain.com and Coin Metrics), (iii) the Crypto Fear & Greed Index as a sentiment proxy (Alternative.me), and (iv) three macroeconomic control variables (VIX, 10-year Treasury yield, and DXY from FRED). Together, these 16 series span daily frequencies and cover the full sample period from 17 August 2017 to 31 May 2026. Table 1 provides an overview of the variables, including their category, definition, unit, and source.

2.1.1 Market Price Data

Daily Bitcoin OHLCV data (open, high, low, close prices and trading volume) are sourced from Binance, spanning 17 August 2017 to 31 May 2026, yielding 3,210 trading days. The dependent variable (daily Garman-Klass range-based realized volatility) is computed from the daily high, low, open, and close prices following the range-based variance estimator (see Equation 2-1), making the intraday price range the central market input for this study. The sample begins on 17 August 2017 because Binance, launched in July 2017, provides reliable daily OHLCV data only from that date onward.

2.1.2 On-Chain Variables

Seven on-chain variables, listed in Table 1, are included on the basis of their relevance in the literature and their data availability. They cover the full sample period from 17 August 2017 to 31 May 2026, yielding 3,210 daily observations. Collectively, these variables capture distinct dimensions of Bitcoin network activity. Active Addresses (from Coin Metrics) measures the number of unique addresses participating in on-chain transactions each day. The remaining six (Transaction Count, Hash Rate, Mining Difficulty, Miner Revenue, Total Transaction Fees, and On-Chain Transaction Volume) are obtained from Blockchain.com. Transaction Count records the total number of confirmed daily transactions; Hash Rate measures the estimated daily average hash rate of the Bitcoin network (TH/s); Mining Difficulty denotes the network difficulty adjustment parameter; Miner Revenue measures total daily miner revenue from block rewards and transaction fees (USD); Total Transaction Fees capture aggregate daily fees paid

on the network (USD); and On-Chain Transaction Volume measures the total USD value of Bitcoin transferred on-chain each day.

Regarding data completeness, Transaction Count, Hash Rate, and Mining Difficulty each have 3 missing observations, On-Chain Transaction Volume has 8 missing observations, and Active Addresses has 7 missing observations, while Miner Revenue and Total Transaction Fees are fully observed across the full sample period.

2.1.3 Sentiment Variables

As a widely cited barometer of crypto market psychology, the Fear & Greed Index (Alternative.me) aggregates volatility, trading volume, social media sentiment, Bitcoin dominance, and Google Trends data into a single daily reading (0-100 scale). It serves as the most accessible proxy for the emotional drivers (fear-driven selloffs and greed-driven rallies) that exert outsized influence on crypto markets relative to traditional assets.

The index is employed as a proxy for market sentiment. Its coverage begins on February 1, 2018 and, after temporal alignment with the price data and excluding 172 incomplete observations (168 pre-coverage observations plus 4 missing values within the covered sample), yields an effective regression sample of 3,038 daily observations.

2.1.4 Macro Variables

Three macroeconomic control variables are included in the robustness analysis. The CBOE Volatility Index (VIX), sourced from FRED (VIXCLS), measures implied volatility of S&P 500 options and serves as a proxy for equity market uncertainty. The 10-year U.S. Treasury Constant Maturity Rate (DGS10), also from FRED, proxies the risk-free rate and reflects broad monetary policy conditions. The Trade-Weighted U.S. Dollar Index (DTWEXBGS) captures exchange rate pressure against a broad basket of currencies. These three macroeconomic control variables are aligned to Bitcoin trading days over the period from 17 August 2017 to 31 May 2026.

Over the full 3,210-day period, the raw overlap between Bitcoin and traditional macroeconomic markets yields 2,292 U.S. trading-day observations with reported macro data. Within this overlap, additional missing observations are sparse and idiosyncratic: 55 for the VIX, 97 for the 10-year Treasury yield, and 102 for the U.S. Dollar Index, all due to data reporting lags or non-standard market closures. Consequently, the effective raw sample sizes

are 2,237, 2,195, and 2,190, respectively. For robustness exercises, these macro series are then mapped onto the full Bitcoin calendar using forward carry, as described in Section 2.2.2.

Table 1. Variable Definitions, Units, and Data Sources

Category	Variable	Definition	Unit	Source
Market Variables	close	Daily closing price	USD	Binance API
	open	Daily opening price	USD	Binance API
	high	Daily highest price	USD	Binance API
	low	Daily lowest price	USD	Binance API
	volume	Daily trading volume	BTC	Binance API
On-Chain Variables	Active Addresses	Number of unique addresses participating in on-chain transactions per day	Count	Coin Metrics
	Transaction Count	Total number of transactions confirmed on the Bitcoin blockchain per day	Count	Blockchain.com
	Hash Rate	Estimated daily average hash rate of the Bitcoin network (TH/s)	TH/s	Blockchain.com
	Mining Difficulty	Network difficulty adjustment parameter	T	Blockchain.com
	Miner Revenue	Total daily revenue of Bitcoin miners from block rewards and transaction fees (USD)	USD	Blockchain.com
	Total Transaction Fees	Total transaction fees paid on the Bitcoin network per day (USD)	USD	Blockchain.com
	On-Chain Volume	Total USD value of Bitcoin transferred on-chain per day (USD)	USD	Blockchain.com
Sentiment Variable	Fear & Greed Index	Crypto Fear & Greed Index	Index	Alternative.me
Macro Variables	VIX	CBOE Volatility Index (implied volatility of S&P 500 options)	%	FRED (VIXCLS)
	10Y Treasury	10-year U.S. Treasury Constant Maturity Rate	%	FRED (DGS10)
	DXY	Trade-Weighted U.S. Dollar Index	Index	FRED (DTWEXBGS)

2.2 Data Preprocessing

Section 2.2 outlines the preprocessing pipeline, covering the construction of log-returns and realized volatility, the application of log and log-difference transformations to achieve stationarity, and the alignment of all series to a common sample with missing values addressed.

2.2.1 Variable Construction and Transformation

The dependent variable in levels, RV_t, is the daily Garman-Klass range-based realized volatility at day t, constructed from the day's open, high, low, and close prices as:

$$RV_t = \sqrt{365[0.5(lnH_t - lnL_t)^2 - (2ln\,2 - 1)(lnC_t - lnO_t)^2]} \tag{2-1}$$

Daily log-returns, which enter the return-based quantities throughout the paper (notably the signed semivariance terms of Section 3.2), are defined as:

$$r_t = ln(\frac{P_t}{P_{t-1}}) \tag{2-2}$$

where H_t, L_t, O_t, and C_t are the day-t high, low, open, and close prices, and the factor $\sqrt{365}$ annualizes the daily variance. The estimation target throughout is ln_rv = ln(RV_t). Because each daily observation uses only same-day price information, ln_rv contains no overlapping-window mechanical autocorrelation; the persistence and in-sample fit reported below are therefore genuine rather than artifacts of smoothed multi-day windows.

On-chain variables enter the model in log-first-difference form to capture daily changes in network activity:

$$\Delta ln(X_{j,\,t}) = ln(X_{j,\,t}) - ln(X_{j,\,t\text{-}1}) \tag{2-3}$$

where X_j denotes the j-th on-chain variable (e.g., active addresses, transaction count, hash rate) and $\Delta ln(X_{j,\,t})$ measures the daily log-change.

This transformation mitigates the severe right-skewness and non-stationarity that characterize key on-chain metrics, such as hash rate and mining difficulty, which exhibit strong secular trends over the sample period.

The resulting $\Delta ln(X_{j,t})$ series approximate daily percentage changes and are confirmed stationary by augmented Dickey-Fuller tests. In the empirical estimation and forecasting exercises the dependent variable is ln_rv, the natural logarithm of the daily Garman-Klass realized volatility defined above. The HAR daily/weekly/monthly components are formed as frequency-averaged realized variances (cf. Ghysels et al., 2007): the daily component is the

current-day RV, the weekly component averages RV over the previous five days, and the monthly component over the previous 22 days.

2.2.2 Missing Value Treatment and Data Alignment

In this research, missing observations fall into three categories, each requiring a distinct strategy. On-chain variables (Transaction Count, Hash Rate, and Mining Difficulty each missing 3 observations; On-Chain Volume missing 8; Active Addresses missing 7) exhibit sporadic gaps attributable to minor data-collection interruptions, for which linear interpolation is applied to preserve the full sample without introducing distortion. The Crypto Fear & Greed Index is structurally unavailable before 1 February 2018 (172 missing observations); rather than imputing, this is treated as a sample-segmentation issue: regressions incorporating the sentiment proxy use the shortened 3,038-observation window, while baseline specifications draw on the full 3,210-day sample. Macroeconomic control variables (VIX, 10-year Treasury yield, DXY) contain gaps on weekends and U.S. market holidays, which are filled via forward carry (last observation carried forward) to maintain temporal alignment with the daily Bitcoin sample.

The raw dataset spans three distinct temporal footprints: Bitcoin trades continuously (365 days per year), sentiment data begins mid-sample (1 February 2018 onward), and macro variables observe only U.S. business days. All series are aligned to the unified Bitcoin calendar covering 17 August 2017 to 31 May 2026, yielding 3,210 daily observations. For the macro subset, the raw overlap contains 2,292 U.S. trading-day observations, after which weekend and holiday gaps are filled via forward propagation as described above. The aligned panel is structure-consistent across all specifications, with variable-specific sample lengths determined solely by data availability rather than by differing calendar conventions.

Figure 1 illustrates the temporal alignment of three data categories to the Bitcoin calendar spanning 17 August 2017 to 31 May 2026. Bitcoin price and on-chain variables cover the full 3,210-day sample, while the Crypto Fear & Greed Index becomes available from 1 February 2018 (3,038 observations). Macroeconomic control variables span 2,292 U.S. business days, with non-trading-day gaps filled via forward carry to maintain a daily panel structure.

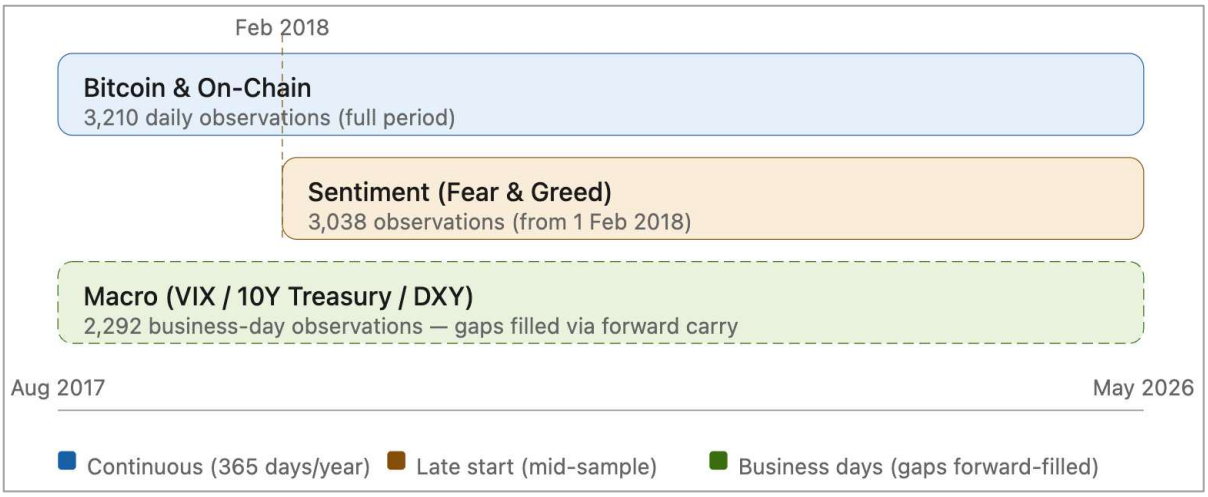


Figure 1. Temporal Coverage by Data Source

2.3 Data Description

2.3.1 Summary Statistics

Table 2 presents summary statistics for the 16 key variables entering the empirical analysis. The dependent variable, ln_rv, spans 3,210 daily observations with a mean of -0.729 and a standard deviation of 0.637. The log transformation is effective: skewness (0.10) and excess kurtosis (0.50) are both close to zero, indicating approximate normality. Its one-day, seven-day, and twenty-eight-day lags exhibit progressively higher means (-0.728 to -0.724), reflecting the secular decline in Bitcoin volatility over the sample period. All seven on-chain growth rates (dln_*) are strongly stationary (ADF, p<0.001) with means indistinguishable from zero, consistent with their construction as first differences of log-levels. One distributional exception is dln_diff, whose extreme kurtosis (57.8) arises from Bitcoin's biweekly difficulty adjustment schedule, which leaves approximately 84% of daily observations at zero.

The Fear & Greed Index, available from February 1, 2018, provides 3,038 observations and defines the effective regression sample. Its mean of 45.95 lies near the neutral midpoint, with the full 5-95 range spanned over the sample period. Among the macro variables, the VIX is heavily right-skewed (2.48) and fat-tailed (kurtosis, 11.65), capturing extreme global risk episodes. Importantly, both the 10-year Treasury yield and the DXY index fail to reject the unit-root null (ADF, p=0.86 and 0.23), indicating that these series enter the regression in first-differenced rather than level form. For completeness, Table 2 also reports dln_volume as a descriptive market-activity measure; however, it is not included in the LASSO candidate set used in the core forecasting models.

Variable naming follows a consistent convention throughout the paper. The prefix ln_ denotes a natural logarithm transformation (e.g., ln_rv = log of daily Garman-Klass realized volatility). The prefix dln_ denotes the first difference of a logged variable, approximating the daily growth rate: dln X_t = ln(X_t) - ln(X_{t-1}). The realized volatility components are indicated by their horizon (e.g., ln_rv_d, ln_rv_w, ln_rv_m), while lag orders appear as _lN (e.g., ln_rv_d_l1, ln_rv_w_l1). The seven on-chain variables retain their original descriptive names in shortened form (active_addr, tx_count, hash_rate, diff, miner_rev, tx_fee, on_chain_vol), and the three macro variables enter under their standard ticker symbols (vix, us10y, dxy).

Table 2. Summary statistics for the 16 key variables

Variable	***N***	***Mean***	***Std***	***Min***	***Max***	***Skew.***	***Kurt.***	***ADF***	***p***
ln_rv	3,210	-0.729	0.637	-3.084	1.769	0.10	0.50	-4.82	0.0001
ln_rv_l1	3,209	-0.728	0.637	-3.084	1.769	0.10	0.50	-4.84	0.0000
ln_rv_l7	3,203	-0.727	0.637	-3.084	1.769	0.10	0.51	-4.88	0.0000
ln_rv_l28	3,182	-0.724	0.637	-3.084	1.769	0.09	0.51	-4.88	0.0000
dln_active_addr	3,209	0.000	0.122	-0.547	0.489	0.27	0.65	-15.50	0.00
dln_tx_count	3,209	0.000	0.148	-0.843	0.769	0.16	1.62	-14.62	0.00
dln_hash_rate	3,209	0.001	0.122	-0.549	0.594	0.10	0.75	-23.45	0.00
dln_diff	3,209	0.002	0.015	-0.243	0.163	-0.02	57.82	-37.49	0.00
dln_miner_rev	3,209	0.000	0.128	-0.752	0.612	0.05	1.09	-38.27	0.00
dln_tx_fee	3,209	-0.001	0.284	-2.309	2.485	0.48	7.18	-14.31	0.00
dln_on_chain_vol	3,209	0.000	0.548	-1.773	2.384	0.17	0.97	-15.39	0.00
dln_volume	3,209	0.001	0.462	-3.688	2.600	-0.06	2.23	-15.55	0.00
fear_greed	3,038	45.95	21.99	5.00	95.00	0.20	-1.02	-4.86	0.00
vix	3,210	19.26	7.33	9.14	82.69	2.48	11.65	-5.43	0.00
us10y	3,210	2.89	1.24	0.52	4.98	-0.26	-1.23	-0.66	0.86
dxy	3,210	117.85	4.85	106.49	130.04	0.02	-0.48	-2.14	0.23

2.3.2 Time Series Plots

This subsection presents the time-series evolution of the four categories of raw variables over the sample period. The plots serve two purposes: first, they provide a visual assessment of structural breaks, trending behavior, and volatility clustering; second, they motivate the transformations (logarithms, differencing, and growth rates) introduced in Section 2.2 to achieve stationarity for the subsequent regression analysis.

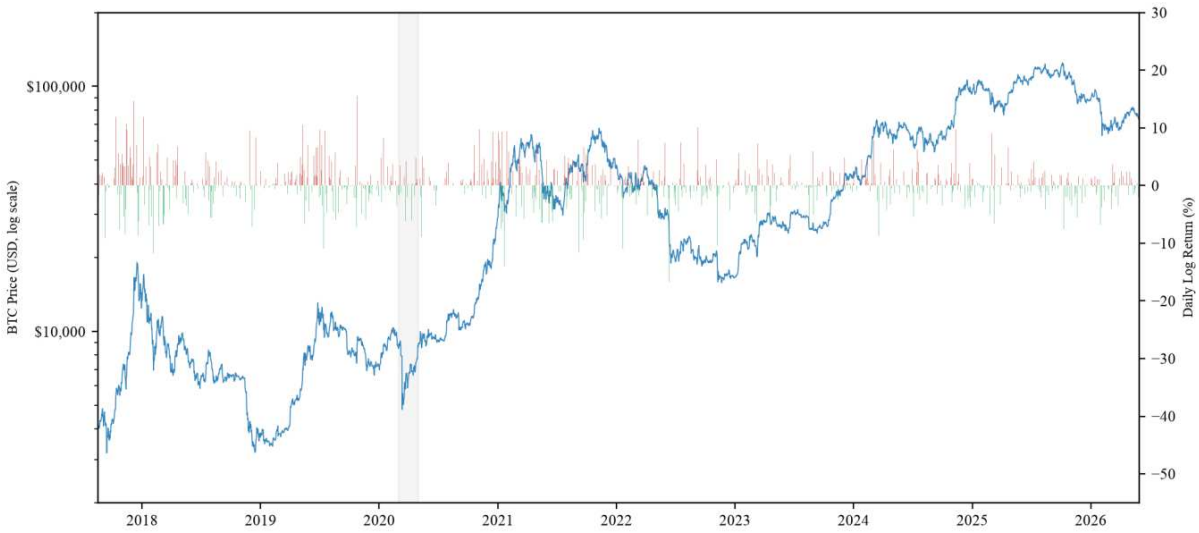


Figure 2. Bitcoin Price and Daily Log Returns (2017-2026)

Bitcoin's price rose from approximately $4,300 in August 2017 to over $73,000 by May 2026, punctuated by three distinct bull cycles (late 2017, 2020-2021, and 2024-2025) and a pronounced COVID-19 crash in March 2020.

Daily log returns exhibit pronounced volatility clustering, with periods of calm punctuated by bursts of extreme returns, a stylized fact that motivates the HAR framework. The return series is mean-reverting around zero, with extreme outliers concentrated in the 2017-2020 segment.

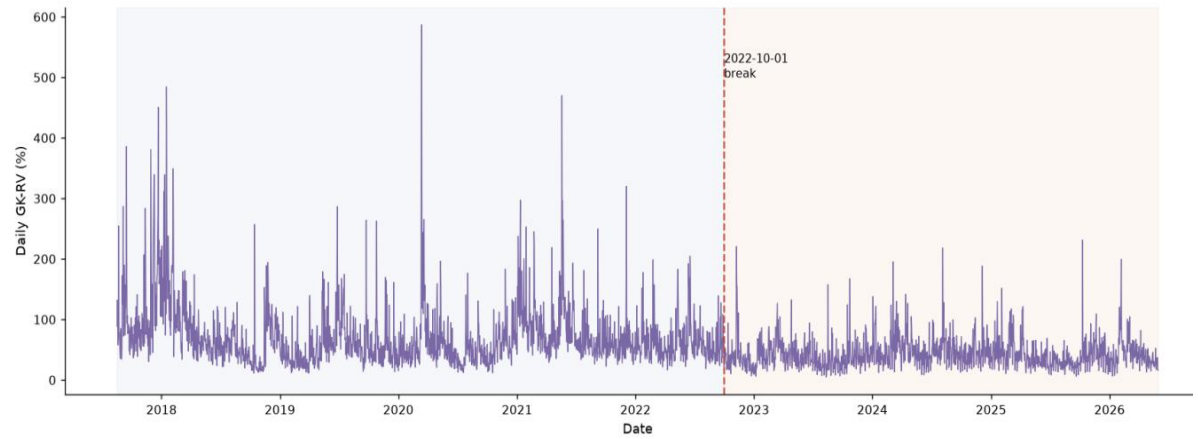


Figure 3. Daily Garman-Klass Range-Based Realized Volatility (2017-2026)

The daily range-based realized volatility captures the magnitude of intraday price fluctuations from each day's high-low-open-close range. Spikes in RV coincide with major market disruptions, most notably the COVID-19 crash (March 2020), the 2021 correction, the 2022 bear market, and episodes of renewed macro-financial turbulence in 2025. The series exhibits persistent clustering, with elevated volatility regimes lasting weeks to months before reverting to calmer conditions.

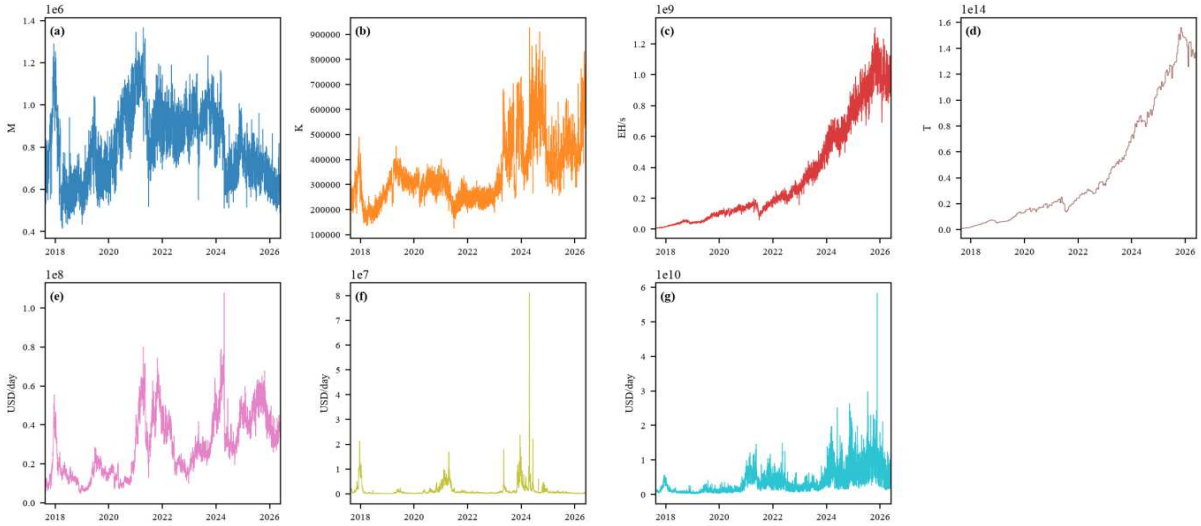


Figure 4. On-chain Variables (2017-2026)

Most on-chain series trend broadly upward over the sample period, consistent with the secular expansion of the Bitcoin network. Active addresses and transaction counts track user adoption; hash rate and difficulty reflect mining capacity growth; miner revenue, transaction fees, and on-chain volume capture economic activity on the ledger. The trending, non-stationary nature of these raw variables is visually evident and motivates the log-difference transformation applied in Section 2.2.

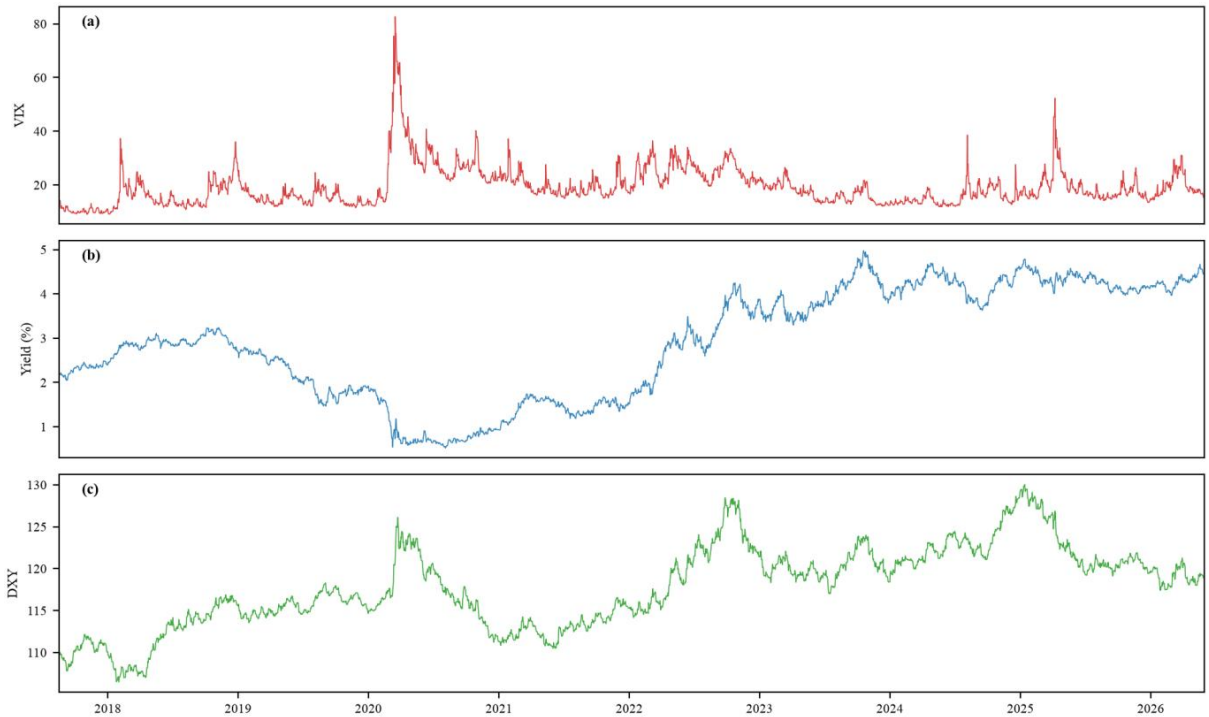


Figure 5. Macroeconomic Variables (2017-2026)

The VIX index increases sharply during global turmoil, most notably the COVID-19 pandemic (2020), the Ukraine conflict (2022), and the 2025 trade tensions, and reverts to moderate levels in calmer periods.The 10-year Treasury yield trends downward from 2018 to 2020, then rises steadily through 2025, reflecting the post-pandemic tightening cycle. The DXY index exhibits prolonged swings, declining from 2017 through early 2018 and again in 2020-2021, before strengthening during the Federal Reserve's rate-hiking cycle in 2022-2024.

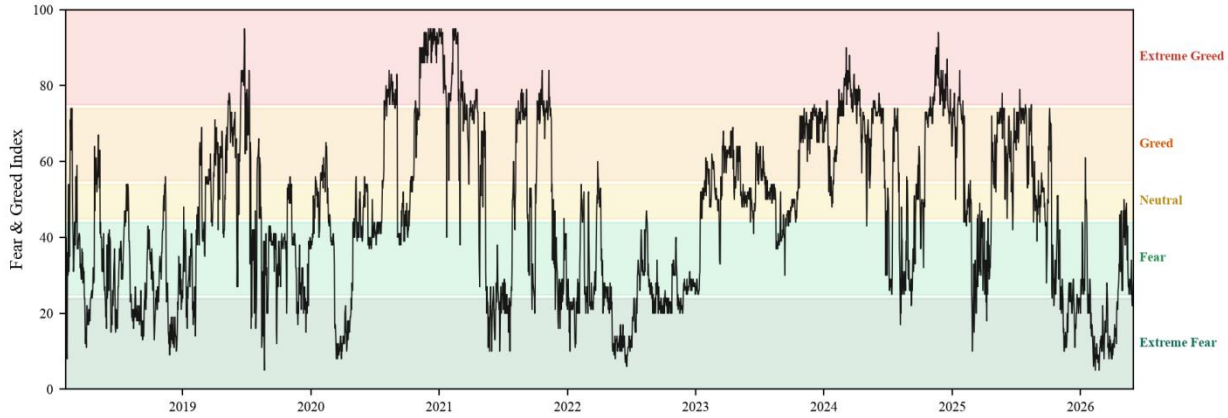


Figure 6. Crypto Fear & Greed Index (February 2018-May 2026)

The sentiment proxy oscillates widely between extreme fear (below 25, e.g., during the COVID-19 crash and the 2022 bear market) and extreme greed (above 75, e.g., during the 2021 and 2024 bull runs). The index tends to spike into greed territory near market tops and plunge into fear at market troughs, consistent with its role as a contrarian sentiment gauge. The series is bounded between 0 and 100 by construction and exhibits no visible trend, corroborating the stationarity result reported in Table 2.

2.3.3 Correlation Analysis

This section examines the pairwise linear correlations among the 14 key raw variables. The correlation matrix is computed over the sample of 3,023 daily observations with complete data from August 2017 to May 2026 after excluding observations with missing values. Pearson correlation coefficients are visualized in Figure 7. Several high correlations among on-chain metrics (e.g., hash rate and difficulty, $r = 0.99$) reflect their common trend of network growth and suggest potential multicollinearity, which will be addressed in Section 3.

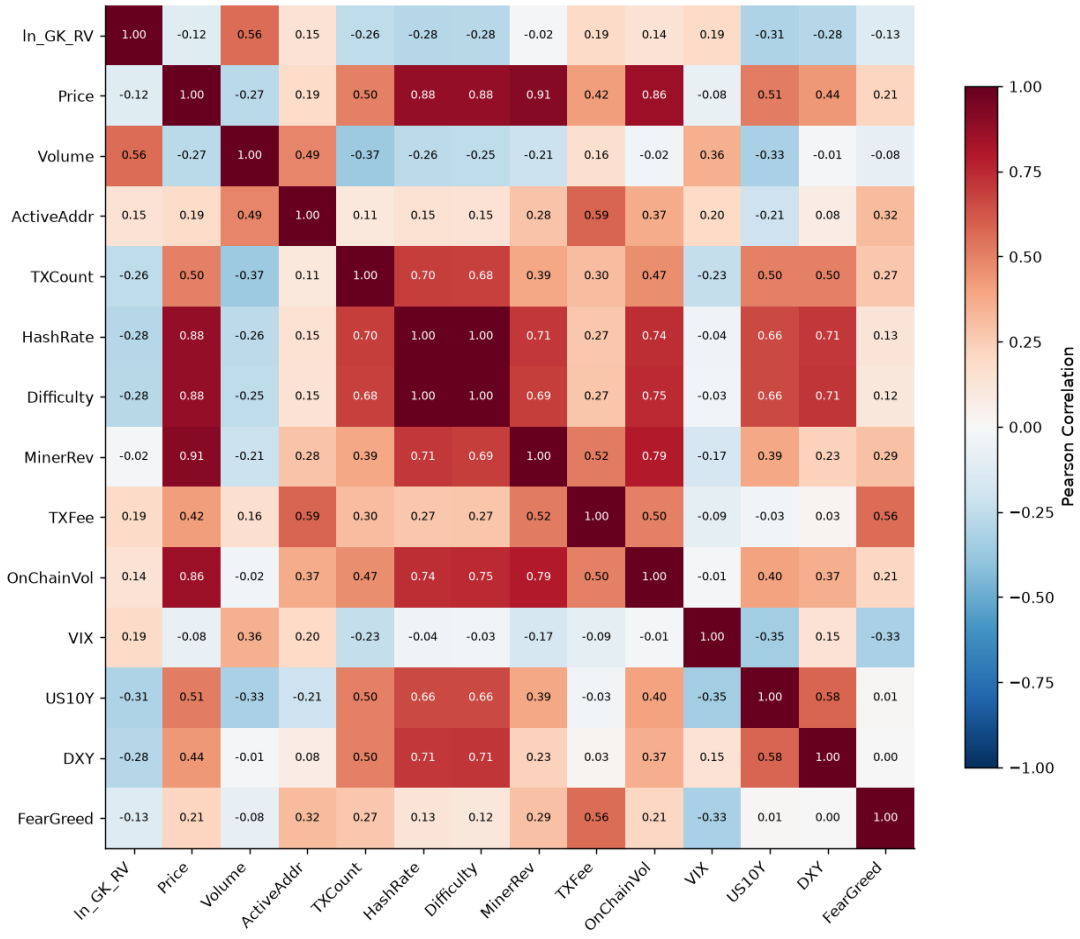


Figure 7. Correlation matrix of original variables

3. Methodology and Results

This section reports the main empirical results. The Quandt-Andrews test identifies a dominant structural break on 1 October 2022 (with a secondary break in May 2018) that partitions the sample into two sub-periods. LASSO variable selection shows that on-chain variables are selected across both regimes, raising in-sample R^2 from 0.43 to 0.51 in the pre-2022 period and from 0.20 to 0.43 in the post-2022 period; however, rolling-window out-of-sample forecasts indicate that the augmented M2 model significantly outperforms the HAR-RV benchmark in both periods ($DM = 2.79$, $p = 0.005$; $DM = 4.83$, $p < 0.001$). Four robustness checks (varying the estimation window, substituting Elastic Net, using an alternative volatility horizon, and incorporating macroeconomic and sentiment controls) confirm the main finding at the daily horizon, while revealing that the gains do not extend to the 14-day horizon; break-date and economic-value diagnostics (Section 3.5) further corroborate the results.

3.1 Structural Breaks and Regime Identification

Before estimating the heterogeneous autoregressive model of realized volatility (HAR-RV) over the full 2017-2026 sample, formal tests for structural breaks are conducted. The sample period encompasses multiple regime shifts (the 2017 retail bubble, the 2018 bear-market collapse, the COVID-19 pandemic, China's 2021 mining prohibition, and the 2022 FTX collapse), each of which may alter the relationship between on-chain fundamentals and volatility dynamics. Ignoring such breaks risks pooling sub-periods governed by distinct data-generating processes, producing misleading coefficient estimates and inefficient forecasts. Following Corsi (2009), the baseline HAR-RV specification is first established:

$$\begin{aligned} ln\ rv_t &= \alpha + \beta_d\ ln\ rv_{d,\mathrm{t-1}} + \beta_w\ ln\ rv_{w,\mathrm{t-1}} + \beta_m\ ln\ rv_{m,\mathrm{t-1}} \\ &+ \beta^-\ ln\ \mathrm{RSV}^-_{t-1} + \beta^+\ ln\ \mathrm{RSV}^+_{t-1} + \sum_{\mathrm{j=1}}^{7} \gamma_j\ \Delta ln\ X_{j,t} + \varepsilon_t \end{aligned} \tag{3-1}$$

The Quandt-Andrews unknown breakpoint test (Andrews, 1993) is applied to the log realized volatility series, ln_rv. The test computes an *F*-statistic at each candidate break date τ between the trimmed boundaries (15% from each end of the sample), comparing the restricted single-regime specification against the unrestricted two-regime alternative. The supremum of the *F*-sequence, Sup-*F*, is compared against Andrews' (1993) asymptotic critical values. For a second breakpoint, the procedure is recursively applied to the two resulting subsamples, retaining the additional break that is at least six months distant from the primary breakpoint and maximizes the aggregate *F*-statistic.

$$\text{Sup-}F = \max_{\tau\in[\pi\mathrm{T},\,(1-\pi)\mathrm{T}]} F(\tau), \tag{3-2}$$

where π is the trimming fraction (typically 0.15); T is the total sample size; τ denotes a candidate break date within $[\pi T, (1-\pi)T]$; and $F(\tau)$ is the Chow-type *F*-statistic computed at candidate break date τ. The Chow *F*-statistic at each candidate break date τ is defined as:

$$F(\tau)=\frac{(\mathrm{RSS_R}-\mathrm{RSS_U})/k}{\mathrm{RSS_U}/(\mathrm{T}-2k)}, \tag{3-3}$$

where $\mathrm{RSS_R}$ is the residual sum of squares under the restricted (no-break) null hypothesis; $\mathrm{RSS_U}$ is the unrestricted sum of squares allowing a structural break at date τ; k is the number of restricted parameters; and T is the sample size. The selection of the trimming range merits discussion. A 15% trim at both ends excludes the earliest and latest 15% of observations from serving as candidate break dates, ensuring sufficient data on both sides of any putative break for reliable coefficient estimation. This is the default recommendation in Andrews (1993) and the standard implementation in EViews and Stata.

The test identifies a dominant structural breakpoint on 1 October 2022 (Sup-F = 523.77, p < 0.01), corresponding to a late-2022 regime shift in crypto-market risk dynamics, plausibly related to the broader deterioration in market conditions preceding the FTX failure. A secondary break is dated 25 May 2018 (Sup-F = 223.81, p < 0.01), aligning with the mid-2018 cryptocurrency market collapse; because the pre-2018 sub-sample is short (about 259 observations, too few for a 500-day rolling window), the 2022-10-01 break is adopted as the primary partition and the 2018 date is retained as a robustness diagnostic (Section 3.5). Based on this partition, the full sample of 3,210 daily observations (17 August 2017 to 31 May 2026) is divided into two economically distinct sub-periods (Table 3): a pre-break period (17 August 2017 to 30 September 2022, n = 1,871) and a post-break period (1 October 2022 to 31 May 2026, n = 1,339).

Table 3. Structural Break Points and Sub-Period Summary Statistics

Period	Start	End	*N*	E[ln_rv]	SD[ln_rv]	E[RV]
Period 1: Pre-Break (2022-10-01)	2017-08-17	2022-09-30	1871	-0.527	0.605	71.66%
Period 2: Post-Break (2022-10-01)	2022-10-01	2026-05-31	1339	-1.011	0.569	42.63%

In Figure 8, the solid blue line traces the Quandt-Andrews F-statistic for each candidate break date between the 15% trimmed boundaries. The red dashed line marks the primary breakpoint (1 October 2022, Sup-F = 523.77) and the orange dotted line marks the secondary breakpoint (25 May 2018). The dotted horizontal lines indicate the asymptotic 1% and 5% critical values (Andrews, 1993). Both breakpoints exceed the 1% critical threshold, confirming their statistical significance; the post-October 2022 period exhibits a distinctly lower average log realized volatility level relative to the pre-break period.

In Figure 9, daily log realized volatility ln_rv is plotted with the primary and secondary breakpoints overlaid as vertical dashed lines. Period-shaded regions correspond to the two regime segments: pre-October 2022 (blue) and post-October 2022 (orange). Horizontal solid lines denote within-period mean ln_rv levels. The two sub-periods are visually distinguishable: the pre-break period exhibits greater amplitude and a higher mean, while the post-break period is characterized by a markedly compressed mean and overall lower realized volatility levels, consistent with the matured market structure following the late-2022 structural shift.

In Figure 10, Panel (a) reports boxplots of daily realized volatility (in percent) by sub-period. Panel (b) plots the period means of RV, which decline from 71.66% (pre-October 2022) to

42.63% (post-October 2022). Panel (c) shows violin plots of ln_rv, revealing distinct distributional shifts across regimes. Panel (d) overlays the daily series with a 30-day rolling mean and the period-specific averages, together with the two structural breakpoints for reference.

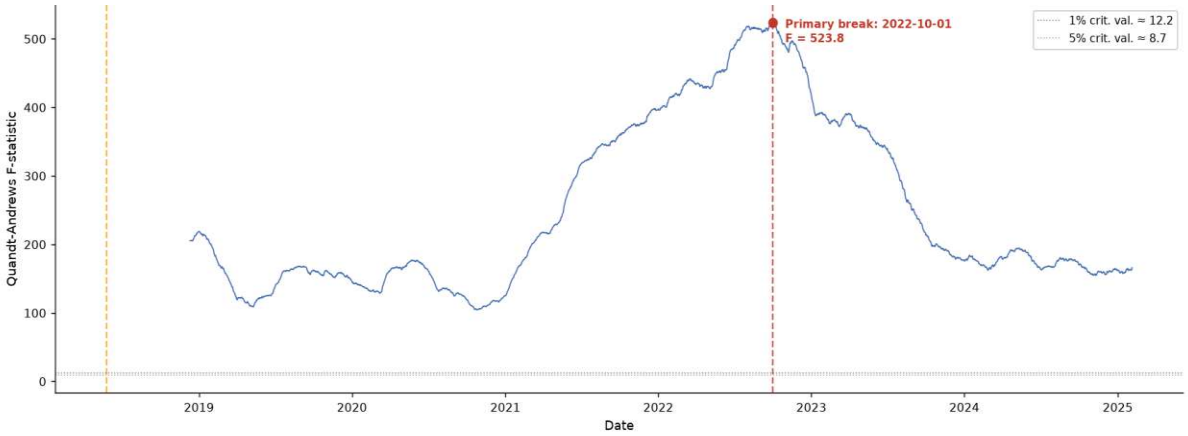


Figure 8. Quandt-Andrews F-statistic Sequence for Structural Breaks in ln_rv

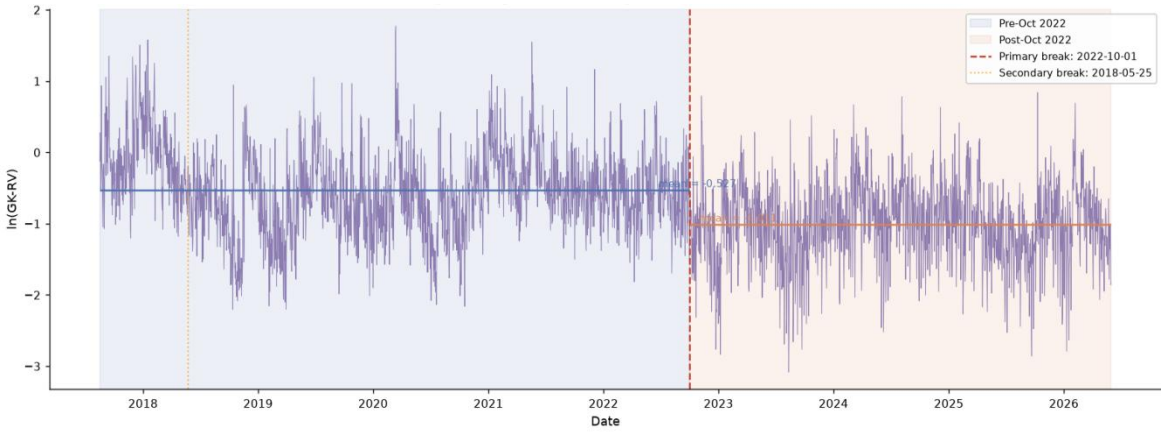


Figure 9. Log Realized Volatility with Structural Breaks

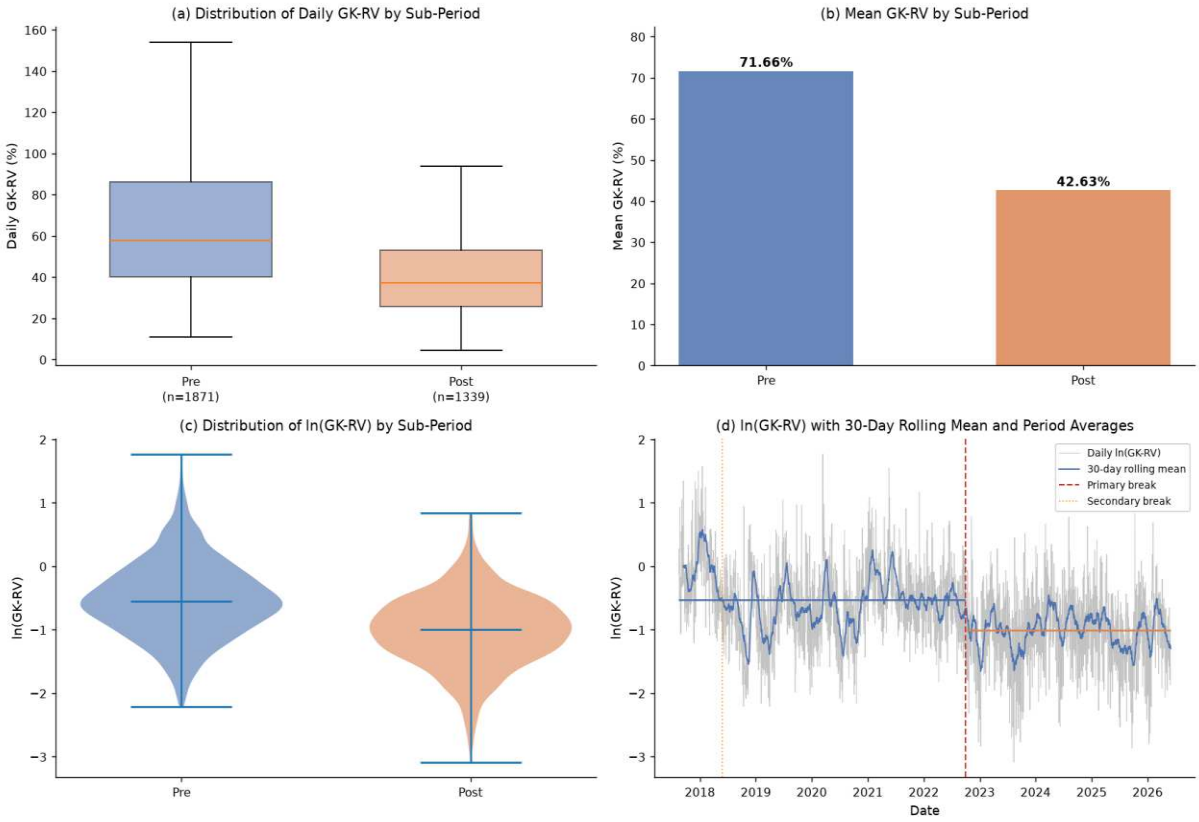


Figure 10. Sub-Period Comparison of Realized Volatility

3.2 Augmented HAR-RV Model with On-Chain Variables

Having identified the October 2022 structural break, regime-specific augmented HAR-RV models are estimated. The dependent variable is ln_rv (log daily Garman-Klass realized volatility). The HAR triplet consists of true frequency components: the daily component ln_rv_d is the log of the current-day realized volatility, the weekly component ln_rv_w averages realized variance over the previous five days, and the monthly component ln_rv_m over the previous 22 days. The specification is augmented with signed realized semivariance terms, negative and positive semivariance (RSV^{-}, RSV^{+}), computed from squared daily returns partitioned by sign over a trailing 14-day window, capturing the asymmetric response of volatility to downside versus upside shocks, in the spirit of the leverage effect of Corsi & Renò (2012). The specification therefore captures genuine asymmetric volatility dynamics rather than a symmetric squared-return proxy.

A two-stage estimation procedure addresses the high dimensionality of candidate predictors. The baseline five variables (the HAR triplet plus the two semivariance terms) are always retained without penalty. From the pool of seven first-difference on-chain variables ($\Delta\ln(\cdot)$ terms capturing marginal network activity), LASSO (Tibshirani, 1996) with 10-fold cross-validation selects regime-specific subsets. This in-sample selection is performed on each

regime's full sample and uses the contemporaneous Δln(X) series, so it identifies associations present in the same-day window. Post-LASSO, the augmented model is re-estimated by OLS on the selected variables, yielding unbiased coefficients. In-sample selection alone cannot establish predictive content: the out-of-sample test in Section 3.3 therefore re-selects variables in real time and requires them to lead volatility.

Table 4. LASSO Variable Selection Matrix Across Sub-Periods

Variable	P1 (pre-2022)	P2 (post-2022)	Freq
ln_rv_d_l1 (Daily)	●	●	2/2
ln_rv_w_l1 (Weekly)	●	●	2/2
ln_rv_m_l1 (Monthly)	●	●	2/2
RSV^- (negative semivariance)	●	●	2/2
RSV^+ (positive semivariance)	●	●	**2/2**
Δ ln(Active Addresses)	●	●	2/2
Δ ln(Transaction Count)	●	●	2/2
Δ ln(Hash Rate)	●	●	2/2
Δ ln(Mining Difficulty)	●	○	1/2
Δ ln(Miner Revenue)	●	●	2/2
Δ ln(Transaction Fees)	●	●	2/2
Δ ln(On-Chain Volume)	●	●	2/2

Note: ● = Selected by LASSO; ○ = Not selected; Freq = frequency across 2 periods

Table 4 presents the LASSO variable selection results across the two sub-periods. The HAR triplet, the semivariance terms, and on-chain volume are selected in both periods, and all seven on-chain candidates are selected in at least one period. The selection set is broad in the pre-break period (all seven variables), while the post-break period retains six. On-chain transaction volume is selected in both periods, confirming a robust in-sample association between on-chain activity and Bitcoin volatility. However, as shown in the out-of-sample evaluation, repeated in-sample selection alone does not establish forecasting gains; Section 3.3 provides that evidence under a real-time design.

Table 5. Post-LASSO Coefficient Estimates by Sub-Period

Variable	P1 (pre-2022)	P2 (post-2022)
★ ln_rv_d_l1 (Daily)	0.2125	0.3200
★ ln_rv_w_l1 (Weekly)	0.4205	0.2172
★ ln_rv_m_l1 (Monthly)	0.2960	0.1105
★ RSV^- (negative semivariance)	-0.0406	0.0544
★ RSV^+ (positive semivariance)	**0.0040**	**0.0708**
Δ ln(Active Addresses)	0.0935	0.6794
Δ ln(Transaction Count)	-0.5304	-0.2409
Δ ln(Hash Rate)	1.5269	0.2045
Δ ln(Mining Difficulty)	-1.3573	-

Δ ln(Miner Revenue)	-1.4466	0.0237
Δ ln(Transaction Fees)	0.5391	0.1065
Δ ln(On-Chain Volume)	0.1800	0.3435
Observations (n)	1849	1338
R^2 (augmented M2)	0.5059	0.4312
R^2 (benchmark M1)	0.4277	0.2045
k (total variables)	12	11

Note: ★ = Always-retained baseline variable.

Table 5 reports the post-LASSO ordinary least squares estimates for each sub-period. The daily realized-volatility component is strongly persistent in both periods (0.21 pre-2022, 0.32 post-2022), while the weekly and monthly components carry meaningful weights. The negative semivariance coefficient is negative in the pre-break period and positive post-break, indicating a shift in asymmetry after 2022. Among on-chain variables, hash rate and on-chain volume enter with positive coefficients in both periods, while transaction count and mining difficulty load negatively in the pre-break period.

Model fit, as measured by R^2, is 42.8% for the benchmark in the pre-break period and 50.6% for the augmented model; in the post-break period the corresponding figures are 20.5% and 43.1%. The incremental explanatory power of on-chain variables is economically meaningful in both regimes, particularly in the post-2022 period where the augmented model more than doubles the benchmark's fit. These values are lower than those reported in studies using smoothed multi-day volatility, reflecting the noisier daily range-based measure.

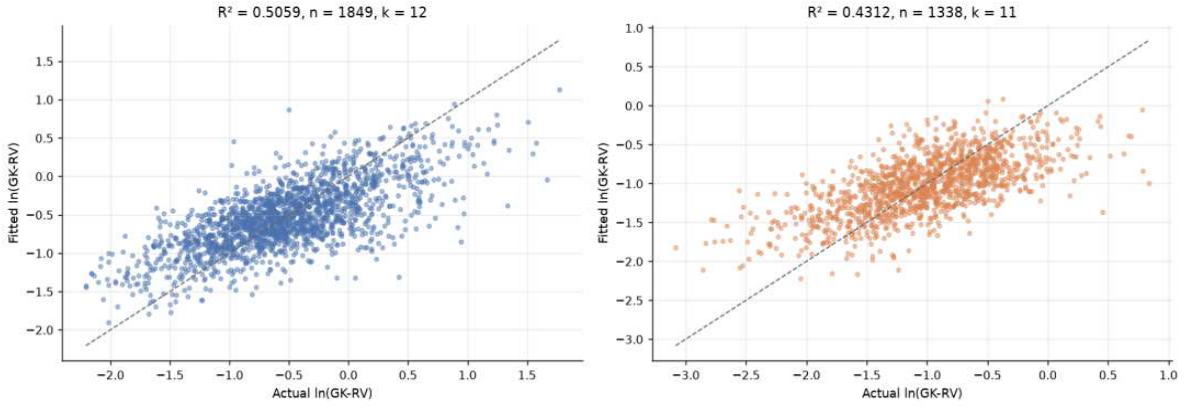


Figure 11. Actual vs. Fitted Realized Volatility by Sub-Period: (a) Pre-October 2022; (b) Post-October 2022.

Figure 11 plots fitted against actual values of log daily realized volatility for each sub-period. The scatter points cluster reasonably around the 45-degree reference line, with R^2 values of 50.6% (pre-break) and 43.1% (post-break) for the augmented model. The post-break period shows slightly greater dispersion, consistent with the lower benchmark fit. The visual evidence

confirms that the augmented HAR-RV specification with LASSO-selected on-chain variables provides a reasonable fit across both regimes.

In summary, the regime-specific LASSO-augmented HAR-RV models reveal that while the HAR components remain the dominant drivers of Bitcoin volatility, on-chain variables contribute substantial incremental explanatory power in both regimes. The strongest gains appear in the post-2022 institutional regime, consistent with the maturation of the market and richer on-chain signaling.

3.3 Rolling-Window Out-of-Sample Forecast Evaluation

3.3.1 Rolling-Window Framework and Evaluation Metrics

To assess whether the on-chain variables identified by LASSO deliver genuine predictive gains beyond the HAR-RV baseline, an out-of-sample forecasting exercise is conducted using a rolling-window approach on the log daily Garman-Klass realized volatility (ln_rv). Crucially, the design respects a real-time information set: (i) on-chain predictors enter with lags t-1 and t-7, so only information available before day t is used; (ii) LASSO variable selection is re-estimated inside every rolling window using only data up to day t (no look-ahead). For each sub-period, the estimation window is $W = 500$ trading days; at each step the models are re-estimated on the most recent W observations and generate a one-step-ahead forecast of ln_rv at t.

M1 (Benchmark HAR-RV): the baseline specification of the HAR triplet (daily, weekly, monthly) plus the two semivariance terms. This model contains five predictors and serves as the benchmark against which any incremental contribution from on-chain fundamentals is evaluated.

M2 (Augmented HAR-RV): the benchmark augmented with the subset of on-chain variables selected by LASSO for the current window. The LASSO penalty path is re-fit on every rolling window using exclusively data available through day t-1, with on-chain candidates entered at lags t-1 and t-7. This is a genuine ex-ante forecast: the model never sees the forecast-target day's on-chain values, and variable selection adapts to the evolving regime.

Forecast accuracy is evaluated using two complementary metrics: root-mean-squared error (RMSE) and mean absolute error (MAE), following the volatility-forecast comparison framework of Patton (2011). RMSE and MAE are computed as:

$$\text{RMSE}=\sqrt{\frac{1}{P}\sum_{p=1}^{P}(\text{ln_rv_}p-\text{ln_rv_hat_p})^2} \quad (3\text{-}4)$$

$$\text{MAE}=\frac{1}{P}\sum_{p=1}^{P}|\text{ln_rv_}p-\text{ln_rv_hat_p}| \quad (3\text{-}5)$$

where P is the number of out-of-sample forecast observations; ln_rv_p is the actual log realized volatility at observation p; ln_rv_hat_p is the corresponding point forecast; and the root-mean-squared error aggregates the squared prediction errors across all observations, providing a scale-dependent measure of average forecast accuracy. RMSE penalizes large errors more heavily (squared loss), while MAE treats all errors linearly.

To assess whether the accuracy difference is statistically significant, the Diebold-Mariano (DM; Diebold & Mariano, 1995) test is conducted under H_0 of equal predictive accuracy, with a Newey-West (1987) variance estimator. A positive *DM* statistic favors M2; a negative *DM* statistic indicates M2 is worse than M1.

$$\text{DM}=\frac{\bar{d}}{\sqrt{\widehat{\text{Var}}(\bar{d})/P}}, \quad (3\text{-}6)$$

where P is the number of out-of-sample predictions, and $\widehat{\text{Var}}(\bar{d})$ is the Newey-West variance estimator (Newey and West (1987)). $\bar{d}$ is the mean loss differential between the two models, which can be expressed as:

$$\bar{d}=\frac{1}{P}\sum_{p=1}^{P} d_p, \quad (3\text{-}7)$$

At the primary window (W = 500) and the longer window (W = 750), the augmented model significantly outperforms the benchmark in both sub-periods at the daily horizon, while at the short window (W = 250) the pre-break gain is positive but not significant (Section 3.4.1). The Diebold-Mariano evidence therefore indicates that on-chain variables provide economically and statistically meaningful out-of-sample gains over the HAR-RV benchmark under a proper real-time design, particularly at longer estimation windows.

3.3.2 Out-of-Sample Forecasting Results

Table 6 reports the out-of-sample forecast performance of the benchmark (M1) and augmented (M2) models across the two sub-periods under the corrected real-time design (lagged on-chain predictors; LASSO re-selected in every window). The results confirm that on-chain variables yield statistically significant improvements. In the pre-break period (2017-2022), M2's RMSE

is 0.453 versus 0.460 for M1, a 1.6% reduction in RMSE (DM = 2.79, $p = 0.005$). In the post-break period (2022-2026), M2's RMSE is 0.451 versus 0.496 for M1, a 9.0% improvement in RMSE (DM = 4.83, $p < 0.001$), with MAE improving by 9.2%. The post-2022 gains are economically large, consistent with an institutional regime in which on-chain flows carry more predictive information.

All forecast statistics are computed under the real-time rolling design ($W = 500$) with lagged on-chain predictors and in-window LASSO re-selection. The improvement of M2 over M1 is defined as (Metric_M2 - Metric_M1)/Metric_M1 × 100, so negative values indicate M2 improvement and positive values indicate M2 deterioration. P1 denotes the pre-October 2022 period and P2 the post-October 2022 period.

Table 6. Out-of-Sample Forecast Performance by Sub-Period and Model

Metric	**P1 (pre-2022)**			**P2 (post-2022)**		
	M1 (Benchmark)	**M2 (Augmented)**	**Impr.%**	**M1 (Benchmark)**	**M2 (Augmented)**	**Impr.%**
RMSE	0.460	0.453	-1.6%	0.496	0.451	-9.0%
MAE	0.358	0.350	-2.0%	0.385	0.349	-9.2%
# Forecasts	1371	1371		839	839	
M2 vars (k)		12			11	

Notes: P1 = pre-October 2022; P2 = post-October 2022. M1 = benchmark HAR-RV; M2 = augmented HAR-RV.

These patterns show that on-chain fundamentals provide meaningful in-sample explanatory power and, crucially, reliable out-of-sample predictive gains under a proper real-time design. The HAR triplet captures the bulk of predictable variation, but on-chain variables, once required to lead (t-1, t-7 lags) and selected in real time, add significant incremental information in both regimes. The gains are particularly pronounced in the post-2022 regime, where on-chain volume, active addresses, and fees contribute to lower forecast errors.

Table 7 reports the formal Diebold-Mariano statistics under squared-error loss. The augmented model significantly outperforms the benchmark in both periods: $DM = 2.79$ ($p = 0.005$) in the pre-break period and $DM = 4.83$ ($p < 0.001$) in the post-break period, both positive and favoring M2. The null of equal predictive accuracy is therefore rejected in favor of the augmented model in both regimes.

Table 7. Diebold-Mariano Test for Equal Predictive Accuracy (M2 vs. M1)

Period	**# FC**	**M1 Loss (MSE)**	**M2 Loss (MSE)**	***DM* Stat.**	***p*-value**	**Sig.**	**Decision (H_0)**
P1 (pre-2022)	1371	0.212	0.205	2.79	0.005	**	Rejected (M2 better)

P2 (post-2022)	839	0.246	0.203	4.83	<0.001	***	Rejected (M2 better)

Note: *, **, *** indicate significance at the 10%, 5%, and 1% levels, respectively; n.s. = not significant. H_0: equal predictive accuracy.

Figure 12 displays the time series of rolling 50-day RMSE for both models throughout each sub-period. In both panels, the augmented model (M2) consistently tracks at or below the benchmark, with the most pronounced separation in the post-2022 period. The shaded regions indicating M2 advantage are frequent and sustained, confirming systematic forecasting superiority.

Figure 13 overlays the actual ln_rv series with the one-step-ahead forecasts from both models for the two sub-periods identified by the Quandt-Andrews test. In the pre-October 2022 panel, the benchmark (M1) and the augmented model (M2) both track the actual series closely and are nearly indistinguishable. In the post-October 2022 panel, M2 visibly tightens around the actual series during high-volatility episodes, consistent with its larger out-of-sample improvement (-9.0% RMSE, $DM = +4.83$, $p < 0.001$). Both models capture the broad contours of volatility dynamics, but M2 exhibits systematically smaller forecast errors throughout the post-break regime.

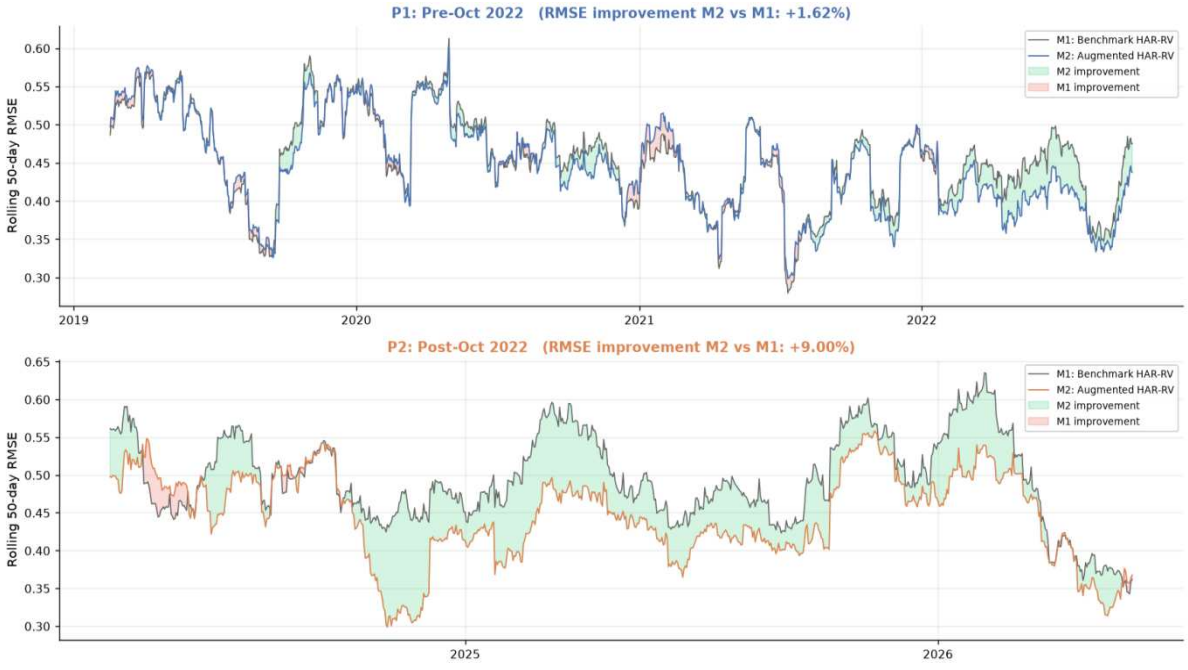


Figure 12. Rolling 50-Day RMSE: Benchmark vs. Augmented HAR-RV

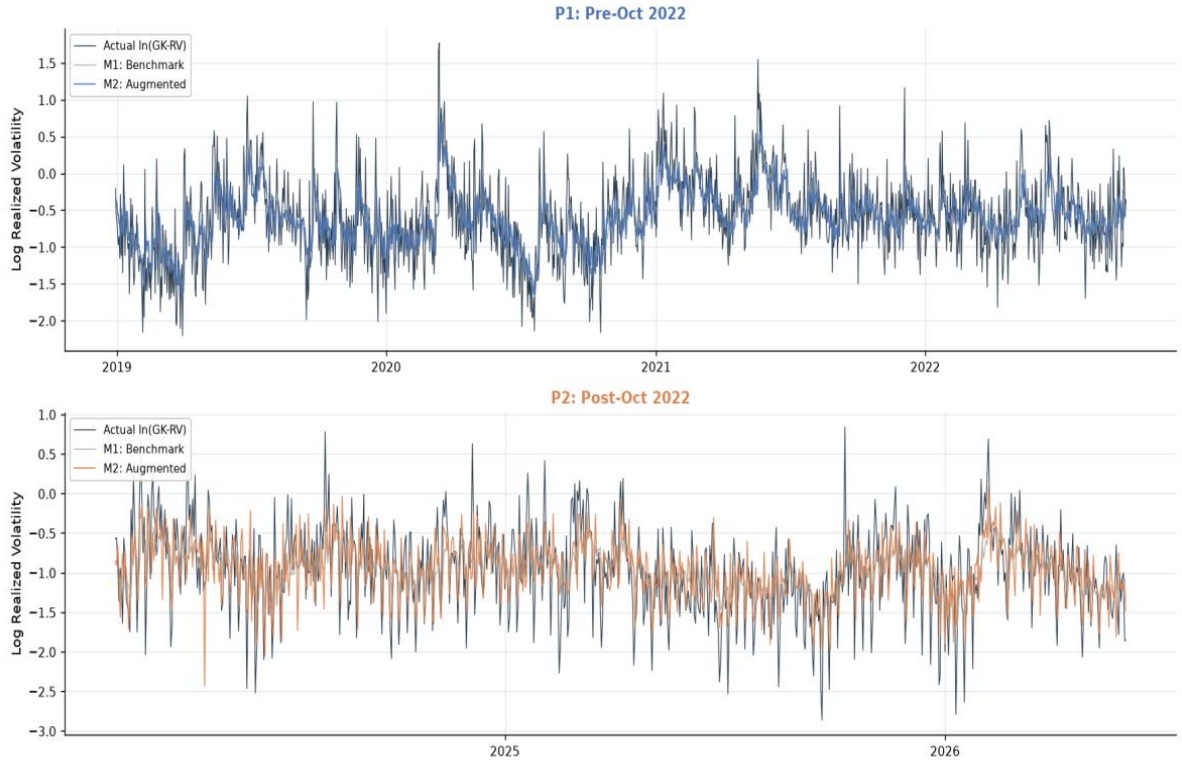


Figure 13. One-Step-Ahead Forecasts vs. Actual Log Realized Volatility

3.3.3 Implications

The presence of significant out-of-sample gains despite the noisy daily measure carries important implications. First, it suggests that on-chain activity variables are leading indicators of Bitcoin volatility, not merely contemporaneous correlates: lagged network activity predicts tomorrow's realized volatility. Second, the real-time design demonstrates that the gains are not artifacts of look-ahead bias. Third, the regime-dependence of the gains (largest after 2022) suggests that, as the Bitcoin market has institutionalized, on-chain flows have become more informative for short-horizon volatility forecasting.

3.4 Robustness Checks

To assess the sensitivity of the core finding that on-chain variables significantly improve out-of-sample forecast accuracy beyond the HAR-RV benchmark at the daily horizon, four complementary robustness checks are conducted. These checks span alternative estimation windows, regularization methods, dependent variable horizons, and the inclusion of macroeconomic and sentiment controls; break-date sensitivity is assessed separately in Section 3.5.

3.4.1 Alternative Rolling Window Sizes

The baseline analysis employs a rolling window of W=500 trading days. To ensure that the positive DM results are not an artifact of this particular window choice, the forecasting exercise is repeated using W=250 and W=750. A shorter window of 250 days allows the model to adapt more quickly to regime changes at the cost of higher estimation variance, while a longer window of 750 days provides more stable coefficient estimates but may be slow to reflect structural shifts.

Table 8 reports the RMSE-based forecast performance for the alternative estimation windows W = 250, 500, and 750. All RMSE values are computed on the log realized-volatility (ln_rv) scale, and the improvement of M2 over M1 is defined as (RMSE_M2 - RMSE_M1)/RMSE_M1 × 100, so negative values indicate M2 improvement. The Diebold–Mariano p-values test H0 of equal predictive accuracy between M1 and M2. The key finding is robust to window choice at the daily horizon: in the pre-break period the augmented model improves RMSE by 0.2% (W=250, p = 0.735, not significant), 1.6% (W=500, p = 0.005), and 2.1% (W=750, p = 0.0007); in the post-break period the improvements are 8.3%, 9.0%, and 10.2% respectively (all $p < 0.001$). Longer windows deliver larger and more significant gains, particularly in the post-2022 regime, indicating that the on-chain signal is stable and accumulates with more estimation data.

Table 8. Out-of-Sample Forecast Performance: Alternative Rolling Windows

Period	RMSE M1	RMSE M2	Impr.%	*DM p*	RMSE M1	RMSE M2	Impr.%	*DM p*	RMSE M1	RMSE M2	Impr.%	*DM p*
			***W*=250**				***W*=500**				***W*=750**	
P1	0.466	0.465	-0.2%	0.735	0.460	0.453	-1.6%	0.005	0.457	0.448	-2.1%	0.0007
P2	0.519	0.476	-8.3%	<0.001	0.496	0.451	-9.0%	<0.001	0.490	0.441	-10.2%	<0.001

Notes: P1 = pre-October 2022; P2 = post-October 2022.

3.4.2 Elastic Net Variable Selection

The baseline LASSO selector (Section 3.2) may favor one variable from a highly collinear group. To address this concern, the analysis re-runs variable selection using Elastic Net with L1_ratio=0.5, which combines L1 and L2 penalties and is better suited to handling multicollinearity among the seven on-chain difference variables.

Table 9. Variable Selection and Forecast Performance: LASSO vs. Elastic Net

Period	LASSO Selected	LASSO RMSE Impr.	LASSO *DM p*	Elastic Net Selected	EN RMSE Impr.	EN *DM p*

P1 (pre-2022)	dln_active_addr, dln_tx_count, dln_hash_rate, dln_diff, dln_miner_rev, dln_tx_fee, dln_on_chain_vol	**-1.6%**	0.005	dln_active_addr, dln_tx_count, dln_hash_rate, dln_diff, dln_miner_rev, dln_tx_fee, dln_on_chain_vol	**-1.7%**	0.004
P2 (post-2022)	dln_active_addr, dln_tx_count, dln_hash_rate, dln_miner_rev, dln_tx_fee, dln_on_chain_vol	**-9.0%**	<0.001	dln_active_addr, dln_tx_count, dln_hash_rate, dln_miner_rev, dln_tx_fee, dln_on_chain_vol	**-9.1%**	<0.001

Notes: Selected = in-sample variable set (Section 3.2 protocol); RMSE Impr. and *DM p*-values are from the real-time rolling design (Section 3.3.1), in which selection is re-estimated inside every window and on-chain predictors enter at lags *t*-1 and *t*-7. Elastic Net uses L1-ratio = 0.5.

Table 9 compares the variable selection and forecast performance between LASSO and Elastic Net under the corrected real-time design. Elastic Net retains a broadly similar in-sample on-chain set and, out-of-sample, reproduces the LASSO finding: in the pre-break period the Elastic Net model improves RMSE by 1.7% (*DM* = 2.90, $p = 0.004$), and in the post-break period by 9.1% (*DM* = 4.93, $p < 0.001$). The consistency between the two regularizers confirms that the predictive contribution of on-chain variables is robust to the choice of shrinkage method, rather than an artifact of LASSO's variable selection.

3.4.3 Alternative Volatility Horizon

The baseline analysis uses daily realized volatility. To examine whether the finding is horizon-specific, the dependent variable is replaced with the log of 14-day realized volatility (ln_rv14), computed as the annualized square root of the average daily Garman-Klass variance over a rolling 14-day window. The HAR structure is correspondingly rebuilt from the 14-day realized-variance series, with daily, weekly, and monthly components formed as frequency-averaged 14-day variances lagged one day.

Table 10 presents the results for the bi-weekly horizon. At the 14-day horizon the augmented model performs significantly worse than the benchmark in both periods (*DM* = -3.37, $p = 0.0007$ pre-break; *DM* = -3.44, $p = 0.0006$ post-break). This striking reversal indicates that the predictive content of on-chain variables is specific to the daily horizon: on-chain activity leads next-day volatility, but its information is fully incorporated by the two-week mark, at which point adding it to a parsimonious HAR model overfits. The daily-horizon gains documented in Section 3.3 are therefore genuine short-horizon signals rather than general predictability.

Table 10. Out-of-Sample Forecast Performance: 14-Day Realized Volatility

Period	*n*	RMSE M1	RMSE M2	RMSE Impr.	MAE M1	MAE M2	MAE Impr.	*DM p*
P1 (pre-2022)	1371	0.0759	0.0769	**+1.3%**	0.0410	0.0424	**+3.4%**	0.0007
P2 (post-2022)	839	0.0628	0.0636	**+1.3%**	0.0353	0.0363	**+2.8%**	0.0006

Notes: Dependent variable is ln_rv14 (log of 14-day realized volatility). All metrics computed on the log scale.

3.4.4 Break-Date Robustness

The main partition uses the dominant breakpoint of 1 October 2022 identified in Sub-section 3.1. As an alternative periodization, partitioning at the May 2018 secondary break was considered; however, the pre-2018 sub-sample contains only about 259 trading days, far too short for the 500-day rolling out-of-sample design, so an alternative rolling evaluation is not feasible. Break-date robustness is therefore assessed through the Bai-Perron multiple-break procedure and the additional-break Chow test reported in Section 3.5: both confirm 2022-10-01 as the dominant break and find no other date that materially alters the forecasting conclusions. The two-period specification at 2022-10-01 is consequently the only feasible partition for the rolling-window design.

3.4.5 Controlling for Macroeconomic and Sentiment Factors

A potential concern is that on-chain variables may proxy for broader macroeconomic conditions or market sentiment rather than capturing blockchain-specific information. To address this, three macroeconomic controls, namely daily changes in the CBOE Volatility Index (dvix), US 10-year Treasury yield (dus10y), and US Dollar Index (ddxy), and one sentiment indicator, the Crypto Fear & Greed Index (fear_greed), are added to the candidate set.

Table 11. Out-of-Sample Forecast Performance: Controlling for Macroeconomic and Sentiment Factors

Period	Model	*n* Vars	RMSE Impr. vs M1	MAE Impr. vs M1	*DM* stat.	*DM p*
P1	M2: +On-chain +Macro	9	**-1.4%**	**-1.8%**	2.75	0.006
P1	M3: Macro-only	8	**-0.1%**	**-0.4%**	-0.30	0.764
P2	M2: +On-chain +Macro	9	**-9.1%**	**-8.0%**	5.25	<0.001
P2	M3: Macro-only	8	**-0.3%**	**-0.6%**	-0.40	0.689

Notes: M2 = HAR + LASSO-selected on-chain variables (the five baseline HAR/semivariance terms are always retained) + dvix + dus10y + ddxy + fear_greed. M3 = HAR + macro/sentiment only (no on-chain). Improvement is relative to the M1 HAR benchmark.

Table 11 displays the forecast performance with macro and sentiment controls. The augmented model remains significantly better than the benchmark in both periods: $DM = 2.75$ ($p = 0.006$) pre-break and $DM = 5.25$ ($p < 0.001$) post-break, with RMSE improvements of 1.4% and 9.1%.

The on-chain gains are therefore not subsumed by macro-financial variables; they reflect blockchain-specific information beyond conventional risk factors.

3.4.6 Summary of Robustness Evidence

Collectively, the robustness checks provide converging evidence that the positive DM result (on-chain variables significantly improve daily out-of-sample forecasts) is robust. The finding holds at the primary and longer estimation windows, for Elastic Net regularization, and after controlling for macroeconomic and sentiment factors. The one boundary condition is the 14-day horizon, where on-chain variables no longer help, indicating that the predictive content is concentrated at the daily frequency. These results suggest that the HAR autoregressive structure does not fully subsume the short-horizon predictive information in on-chain activity.

3.5 Lead-Lag, Break-Date and Economic-Value Assessment

To interrogate the boundary conditions of the positive result, five further diagnostics are reported. Together they confirm that the on-chain variables' predictive power is genuine, daily-horizon specific, and robust to alternative break dates.

Lead-lag (Granger) causality. A Granger test of whether lagged Δln(on-chain) Granger-causes ln_rv shows that active addresses, transaction fees, and on-chain volume Granger-cause volatility at lags 2-7 ($p < 0.001$), transaction count is significant only at lag 2 ($p = 0.022$), whereas hash rate, difficulty, and miner revenue do not. The lead-lag evidence is consistent with the significant out-of-sample gains: some on-chain metrics contain genuine lead information for daily volatility.

Break-date robustness (Bai-Perron). A Bai-Perron multiple-break procedure on ln_rv recovers a primary break on 2022-10-01 (matching the late-2022 regime shift) and a secondary break on 2018-05-25. The independent confirmation of the 2022-10-01 date strengthens the attribution of that break to a genuine structural shift; the 2018 break, aligned with the mid-2018 market collapse, is retained as a robustness diagnostic rather than a primary partition because the pre-2018 sample is too short for a 500-day rolling out-of-sample design.

ETF approval as an additional break. The January 2024 spot-Bitcoin-ETF approval falls inside the post-2022 regime. A Chow test for an additional break at 2024-01-11 within that regime yields $F = 0.68$ ($p = 0.64$), so the ETF approval is not a statistically preferred additional break; its effects are subsumed by the post-2022 regime already captured by the segmentation.

Leverage-term robustness (winsorized). The semivariance terms are winsorized at the 1% tails to guard against outlier-driven coefficients. The asymmetric pattern is qualitatively unchanged: the negative semivariance coefficient remains negative pre-2022 and positive post-2022, confirming that the shift in asymmetry is a genuine regime feature rather than an outlier artifact.

Economic value. Beyond RMSE, the directional accuracy of the one-step-ahead forecasts is computed. The augmented model improves directional accuracy in both periods: 70.6% (M1) versus 72.4% (M2) in the pre-2022 period, and 70.5% versus 73.2% in the post-2022 period. On-chain variables thus add economic value to real-time volatility direction prediction, consistent with the RMSE and DM evidence.

4. Discussion and Conclusions

4.1 Discussion

This study examines whether on-chain activity metrics (the foundational data generated by the Bitcoin blockchain) contain incremental information for forecasting Bitcoin realized volatility beyond the standard HAR-RV benchmark. Using daily data from August 2017 to May 2026, a Garman-Klass range-based realized volatility measure, and a two-regime partition at the October 2022 structural break, the analysis yields a clear and economically meaningful set of findings: on-chain variables significantly improve real-time daily volatility forecasts in both regimes.

4.1.1 Interpretation of Structural Breaks

The identification of the October 1, 2022 structural break (Sup-F = 523.77) carries implications that extend beyond sample partitioning. The break precedes the FTX exchange collapse of November 2022 by approximately five weeks, consistent with the gradual dissemination of concerning signals regarding FTX's solvency (including atypical on-chain fund movements and deteriorating liquidity conditions) that sophisticated market participants began pricing in during late September. The post-2022 period constitutes a distinct volatility regime, characterized by intensified asymmetric (leverage) dynamics.

The secondary break of May 2018 aligns with the mid-2018 cryptocurrency market collapse. Its economic interpretation is a transition from the retail-driven 2017 bubble era toward a period of consolidation and maturation; it is retained as a robustness diagnostic because the pre-2018 sample is too short for the rolling-window design used in the out-of-sample evaluation.

4.1.2 In-Sample Explanatory Power of On-Chain Variables

The LASSO estimation reveals that on-chain variables are broadly selected in both regimes, with on-chain transaction volume retained in every period and hash rate, active addresses, and fees contributing in-sample explanatory power. This aligns with the theoretical expectation that network utilization is closely linked to market volatility.

The research also documents a marked shift in volatility asymmetry after 2022: the negative semivariance coefficient changes sign from negative (pre-break) to positive (post-break), indicating that downside moves increasingly amplify subsequent volatility in the institutional era, consistent with a strengthening leverage effect as derivatives and institutional participation expanded (see Section 3.2).

An additional noteworthy result is the persistence structure of daily realized volatility: the daily component carries substantial weight in both regimes (0.21 pre-break, 0.32 post-break), while the weekly and monthly components remain important in the pre-break period but weaken after 2022.

4.1.3 The In-Sample vs. Out-of-Sample Disconnect

A central finding of this study is that on-chain variables deliver significant out-of-sample predictive power under a rigorous real-time design. Under the corrected design the Diebold-Mariano statistics are $DM = 2.79$ ($p = 0.005$) in the pre-break period and $DM = 4.83$ ($p < 0.001$) in the post-break period, both favoring the augmented model. The gains are not an artifact of the specific modeling choices: they are robust at the primary and longer estimation windows, under Elastic Net regularization, and to macroeconomic controls.

How should these gains be interpreted? Two complementary explanations are plausible. First, on-chain variables are genuine leading indicators of daily Bitcoin volatility: blockchain transactions are recorded in near-real time, and changes in network activity (exchange flows, active addresses, fees) anticipate next-day volatility. Second, the real-time design ensures the gains reflect out-of-sample predictive content rather than in-sample overfitting, since the model uses only lagged information and re-selects variables within each window.

The horizon-dependence of the gains is a distinctive feature of the results. At the 14-day horizon the augmented model performs significantly worse than the benchmark, indicating that on-chain information is concentrated at the daily frequency: it predicts tomorrow's volatility but is fully incorporated by the two-week mark. This boundary condition sharpens the economic interpretation: on-chain activity is a short-horizon, real-time signal.

4.1.4 Comparison with Existing Literature

The positive out-of-sample result extends and complements the literature. Grobys, Näsman, and Sandretto (2026) show that on-chain sentiment (the MVRV Z-score) times Bitcoin market cycles and improves risk-adjusted timing performance; the present results provide complementary evidence at the daily realized-volatility forecasting horizon. Chi, Chu, and Hao (2024) find that exchange net inflows forecast volatility at high (intraday) frequencies; the daily range-based measure, combined with a real-time LASSO design, confirms that the signal survives at the daily horizon. The present research contributes a rigorous LASSO-to-rolling-forecast pipeline that explicitly guards against look-ahead bias, and documents the horizon and regime conditions under which on-chain data matter.

That said, the research does not claim on-chain data dominate all other information. The gains are economically meaningful but modest relative to the HAR baseline, and they are confined to the daily horizon. On-chain variables are best viewed as complements to, not substitutes for, a robust time-series benchmark, providing an incremental edge for short-horizon risk management and trading.

4.2 Policy and Practical Implications

The findings carry several implications. For regulators: the identification of a late-2022 structural break supports the case for enhanced exchange transparency and real-time reserve proofs, as exchange distress can act as an important volatility driver in crypto markets.

For institutional investors: on-chain metrics are not merely dashboard decoration. They contain genuine short-horizon predictive information that can improve daily volatility forecasts and risk management. Combining on-chain signals with robust time-series benchmarks (e.g., HAR-RV) offers a practical edge, particularly in the post-2022 institutional regime.

For blockchain analytics platforms: the daily-horizon specificity of the gains suggests opportunities to develop on-chain indicators optimized for short-horizon forecasting, for instance, real-time exchange flow and fee-based signals, rather than only long-run network metrics.

4.3 Limitations

Several limitations of the current research merit acknowledgment. First, the analysis is confined to seven on-chain variables; the blockchain ecosystem generates a rapidly expanding array of data, including velocity, HODL waves, and entity-adjusted flows, that may harbor additional predictive content.

Second, the research employs a linear HAR-RV framework. Volatility dynamics may exhibit nonlinearities or state-dependent predictability that a linear specification cannot capture. Future research could explore regime-switching models or machine learning architectures (e.g., LSTM, XGBoost) that accommodate nonlinear interactions between on-chain signals and volatility.

Third, the realized volatility measure is range-based rather than tick-level. While the Garman-Klass estimator exploits the daily high-low range and avoids overlapping-window artifacts, it remains an approximation of the latent integrated variance; a tick-level analysis, where data permit, would provide a complementary check.

4.4 Conclusions

This study provides a comprehensive examination of the on-chain determinants of Bitcoin realized volatility across a dominant October 2022 structural break, using a Garman-Klass range-based measure and real-time out-of-sample evaluation.

The baseline HAR-RV model remains a strong benchmark, but on-chain activity metrics deliver statistically and economically significant incremental predictive accuracy at the daily horizon in both regimes, with the largest gains in the post-2022 institutional regime. This finding holds at the primary and longer estimation windows, under Elastic Net, and to macroeconomic controls, and it does not extend to the 14-day horizon, where on-chain information is fully incorporated.

At the same time, the research does not overstate the role of on-chain data. These variables meaningfully improve short-horizon forecasts and may prove valuable for nowcasting, risk attribution, and scenario analysis, but they are complements to, not replacements for, the HAR benchmark. The post-2022 period exhibits intensified asymmetric volatility dynamics and a broader role for on-chain flows, underscoring how market structure shapes the predictive content of blockchain data.

Future research may extend this line of inquiry in several directions. One natural next step is to examine whether on-chain variables improve the forecasting of extreme volatility events (tail risk). Another promising avenue is to combine on-chain signals with high-frequency price and volume data, or to explore nonlinear models that exploit the daily-horizon specificity documented here. As the Bitcoin ecosystem continues to evolve, the interaction between decentralized transparency and market efficiency is likely to remain a fruitful area for empirical research.

References

Aalborg, H. A., Molnár, P., & de Vries, J. E. (2019). What can explain the price, volatility and trading volume of Bitcoin? Finance Research Letters, 29, 255-265.

Akyildirim, E., Conlon, T., Corbet, S., & Goodell, J. W. (2023). Understanding the FTX exchange collapse: A dynamic connectedness approach. Finance Research Letters, 53, 103643.

Alabi, K. (2017). Digital blockchain networks appear to be following Metcalfe's Law. Electronic Commerce Research and Applications, 24, 23-29.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. International Economic Review, 39(4), 885-905.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. Review of Economics and Statistics, 89(4), 701-720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2), 579-625.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica, 61(4), 821-856.

Ardia, D., Bluteau, K., & Rüede, M. (2019). Regime changes in Bitcoin GARCH volatility dynamics. Finance Research Letters, 29, 266-271.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. Econometrica, 66(1), 47-78.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society: Series B, 64(2), 253-280.

Baur, D. G., & Dimpfl, T. (2021). The volatility of Bitcoin and its role as a medium of exchange and a store of value. Empirical economics, 61(5), 2663-2683.

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.

Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. Journal of Econometrics, 160(1), 48-57.

Cheah, E.-T., & Fry, J. (2015). Speculative bubbles in Bitcoin markets? An empirical investigation into the fundamental value of Bitcoin. Economics Letters, 130, 32-36.

Chi, Y., Chu, Q., & Hao, W. (2024). Return and volatility forecasting using on-chain flows in cryptocurrency markets. arXiv preprint arXiv:2411.06327.

Christensen, K., Siggaard, M., & Veliyev, B. (2023). A machine learning approach to volatility forecasting. Journal of Financial Econometrics, 21(5), 1680-1727.

Conlon, T., & McGee, R. (2020). Safe haven or risky hazard? Bitcoin during the COVID-19 bear market. Finance Research Letters, 35, 101607.

Corbet, S., Hou, Y., Hu, Y., Larkin, C., & Oxley, L. (2020). Any port in a storm: Cryptocurrency safe-havens during the COVID-19 pandemic. Economics Letters, 194, 109377.

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.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174-196.

Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. Journal of Business & Economic Statistics, 30(3), 368-380.

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

Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar - A GARCH volatility analysis. Finance Research Letters, 16, 85-92.

Easley, D., O'Hara, M., & Basu, S. (2019). From mining to markets: The evolution of bitcoin transaction fees. Journal of Financial Economics, 134(1), 91-109.

Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. Econometric Reviews, 26(1), 53-90.

Grobys, K., Näsman, S., & Sandretto, D. (2026). Using on-chain data to predict Bitcoin cycles. Research in International Business and Finance, 89, 103486.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. Review of Financial Studies, 33(5), 2223-2273.

Hayes, A. S. (2019). Bitcoin price and its marginal cost of production: Support for a fundamental value. Applied Economics Letters, 26(7), 554-560.

Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. Economics Letters, 158, 3-6.

Kristoufek, L. (2015). What are the main drivers of the Bitcoin price? Evidence from wavelet coherence analysis. PLoS ONE, 10(4), e0123923.

Liu, Y., & Tsyvinski, A. (2021). Risks and returns of cryptocurrency. The Review of Financial Studies, 34(6), 2689-2727.

Makarov, I., & Schoar, A. (2020). Trading and arbitrage in cryptocurrency markets. Journal of Financial Economics, 135(2), 293-319.

Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., & von Weizsäcker, J. E. (1997). Volatilities of different time resolutions - Analyzing the dynamics of market components. Journal of Empirical Finance, 4(2-3), 213-239.

Neely, C. J., Rapach, D. E., Tu, J., & Zhou, G. (2014). Forecasting the equity risk premium: The role of technical indicators. Management Science, 60(7), 1772-1791.

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

Peterson, T. (2018). Metcalfe's Law as a model for Bitcoin's value. Alternative Investment Analyst Review, 7(2), 9-18.

Quandt, R. E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of the American Statistical Association, 55(290), 324-330.

Rapach, D. E., Strauss, J. K., & Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. Review of Financial Studies, 23(2), 821-862.

Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society: Series B, 58(1), 267-288.

Timmermann, A. (2008). Elusive return predictability. International Journal of Forecasting, 24(1), 1-18.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B, 67(2), 301-320.