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 the Limits of GARCH in Event-Driven Markets: A Null-Universe Perspective with Evidence from the S&P 500 and Bitcoin

Li-Yung Chen

chenlijon.tw@yahoo.com.tw

Gangshan Branch of Zuoying Armed Forces General Hospital

Research Article

Keywords: GARCH applicability, null-universe test, standardized innovations, Bitcoin volatility, structural tension, heavy tails, event-sensitive markets

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8972371/v1

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

Additional Declarations: No competing interests reported.

Abstract

This study examines whether the Bitcoin (BTC) return process can be adequately described by a noise-driven conditional variance framework within the GARCH family. Rather than proposing additional GARCH extensions for cryptocurrency markets, we adopt a null-universe perspective: if a market is compatible with a GARCH-type specification, the standardized innovations implied by the fitted model should resemble stable noise; if not, extreme standardized events may occur more frequently than predicted under the null.

Using the S&P 500 index as a benchmark market across two eras (1962–1985 and 1986–2026), and Bitcoin at four time scales (30-minute, 1-hour, 4-hour, daily), we evaluate tail-event exceedances under both normal and Student-t innovations. We find that while S&P 500 tail deviations are largely accommodated by moderate-tailed Student-t distributions ($\nu = 5.6–10.0$), Bitcoin exhibits substantially heavier standardized residual tails across frequencies. Estimated degrees of freedom for BTC ($\nu \approx 3.0–4.0$) approach the lower boundary of finite higher moments, and residual tail rejection persists at higher frequencies even under the Student-t specification.

These results suggest a meaningful distinction in the empirical compatibility of GARCH-type models across asset classes. We outline a diagnostic distinction between noise-dominant and event-sensitive market environments—as a diagnostic tool for assessing model applicability prior to specification refinement.

JEL Classification: C22; C58; G10; G17

1. Introduction

The GARCH framework, introduced by Engle (1982) and generalized by Bollerslev (1986), has become a central tool for modeling time-varying conditional variance in financial time series. Its empirical foundation lies in volatility clustering—the tendency for large price movements to be followed by large movements and calm periods by calm periods (Cont, 2001). Over the past four decades, numerous studies have shown that GARCH-type models effectively capture this feature for major equity indices, including the S&P 500 (Awartani & Corradi, 2005; Engle, 2001). This success has led to a broad family of extensions, such as EGARCH (Nelson, 1991), GJR-GARCH (Glosten et al., 1993), and FIGARCH (Baillie et al., 1996), designed to accommodate asymmetry, long memory, and heavy tails while preserving the core conditional variance structure.

With the emergence of Bitcoin and the broader cryptocurrency market, it was natural for researchers to apply the same modeling framework. A substantial literature has since developed. Katsiampa (2017) compared alternative GARCH specifications for Bitcoin; Chu et al. (2017) evaluated multiple GARCH variants across cryptocurrencies; Dyhrberg (2016) examined Bitcoin's relationship with gold and the U.S. dollar using GARCH-type models. Subsequent studies explored regime-switching specifications (Ardia et

al., 2019; Caporale & Zekokh, 2019; Tan et al., 2021), structural breaks (Abakah et al., 2020), realized GARCH approaches (Walther et al., 2019), and stochastic volatility alternatives (Chan et al., 2019).

Despite this extensive effort, the literature has not converged on a stable canonical specification for Bitcoin. Parameter sensitivity, regime dependence, and persistent tail deviations are frequently reported across samples and model variants (Ardia et al., 2019; Jiang et al., 2023; Silva & Maciel, 2025). These findings raise a broader question: beyond selecting among alternative extensions, how compatible is the underlying GARCH framework with the statistical structure of cryptocurrency returns?

This paper addresses that question from a diagnostic perspective. Rather than proposing an additional specification, we adopt a null-universe framework. If a market's return process is adequately described by a noise-driven conditional variance model, standardized innovations from a fitted GARCH specification should resemble stable noise within reasonable distributional assumptions. If systematic deviations persist in the tails after volatility filtering, this may indicate limitations of the framework under certain market conditions. We formalize this intuition through a tail-event null test calibrated against a benchmark market—the S&P 500—where GARCH models have historically demonstrated empirical usefulness. The benchmark serves both as a reference point and as a validation of the test's discriminatory capacity.

An important component of our design is a robustness analysis using Student-t innovations. A common interpretation of null rejection under normal innovations is distributional misspecification. Accordingly, we re-estimate all models under a GARCH-t specification with jointly estimated degrees-of-freedom parameter v. For the S&P 500, moderate values of v are sufficient to account for observed tail behavior, suggesting that deviations are largely distributional. For Bitcoin, estimated v values are substantially lower (approximately 3.0–4.0 across frequencies), and residual tail discrepancies persist at higher sampling frequencies. This two-step procedure allows us to distinguish between distributional adjustment within the GARCH structure and cases in which residual behavior remains difficult to reconcile under standard conditional variance dynamics.

The contributions of the paper are threefold. First, we introduce a null-universe diagnostic framework that can be applied prior to specification refinement, providing a structured way to assess model compatibility across markets. Second, by combining benchmark calibration with Student-t robustness checks, we offer empirical evidence that compatibility with GARCH-type dynamics may differ systematically across asset classes. Third, we provide an interpretation for why asymmetric and heavy-tailed extensions have historically improved equity volatility modeling, while similar refinements may yield more limited gains in certain cryptocurrency settings.

The remainder of the paper is organized as follows. Section 2 reviews the evolution of GARCH models and related cryptocurrency applications. Section 3 describes the data. Section 4 outlines the methodology. Section 5 presents empirical results. Section 6 discusses implications, and Section 7 concludes.

2. Background and Literature

2.1. The GARCH Family and Its Equity Market Success

The ARCH model of Engle (1982) was motivated by evidence that the variance of macroeconomic residuals was time-varying and clustered. Bollerslev (1986) extended this framework to GARCH, allowing conditional variance to depend on its own lagged values, providing a parsimonious and empirically tractable representation. Cont (2001) documented a set of stylized facts in financial returns—heavy unconditional tails, volatility clustering, and weak linear autocorrelation—that GARCH-type models are designed to capture.

For major equity indices, particularly the S&P 500, GARCH(1,1) has demonstrated considerable empirical durability. Subsequent extensions were developed to address observed asymmetries and persistence. Nelson (1991) proposed EGARCH to model leverage effects, while Glosten et al. (1993) introduced GJR-GARCH to capture asymmetric volatility responses. Engle and Ng (1993) formalized diagnostic tools such as the news impact curve. Awartani and Corradi (2005) showed that asymmetric variants often improve out-of-sample performance relative to the baseline GARCH(1,1), though the benchmark specification remains competitive over shorter horizons.

Collectively, this literature indicates that equity markets, despite occasional tail episodes, are broadly compatible with conditional variance dynamics of the GARCH family when appropriate refinements are introduced.

2.2. GARCH Applications to Bitcoin: Model Proliferation and Instability

A substantial literature has applied GARCH-type models to Bitcoin and other cryptocurrencies. Early contributions evaluated alternative specifications within the standard framework. Katsiampa (2017) compared multiple GARCH variants for Bitcoin, while Chu et al. (2017) tested twelve specifications across cryptocurrencies. Dyhrberg (2016) examined Bitcoin using GARCH-type models in relation to gold and the U.S. dollar.

Subsequent studies introduced additional structural features. Regime-switching GARCH models (Ardia et al., 2019; Caporale & Zekokh, 2019; Tan et al., 2021), models incorporating structural breaks (Abakah et al., 2020), realized GARCH approaches (Walther et al., 2019), and stochastic volatility alternatives (Chan et al., 2019) have all been proposed. While these extensions often improve forecasting performance in specific samples, empirical results frequently depend on time period, asset selection, and model specification (Fung et al., 2022; Silva & Maciel, 2025). Jiang et al. (2023) document long-memory and structural break features that complicate standard short-memory GARCH dynamics.

Taken together, the literature reflects ongoing model refinement and adaptation rather than convergence toward a single stable specification. This motivates a complementary question: beyond selecting among

variants, how compatible is the core conditional variance framework with the statistical properties of cryptocurrency returns?

2.3. A Conditional View of Model Compatibility

To frame this question, we introduce a conceptual distinction between markets in which distributional refinements appear sufficient to reconcile tail behavior and those in which residual discrepancies persist despite such adjustments. In the former case, the underlying noise-driven conditional variance structure remains broadly effective, with extensions addressing specific empirical features. Equity markets, particularly the S&P 500, provide a reference example of this pattern.

Cryptocurrency markets exhibit structural characteristics that may influence volatility dynamics, including continuous trading, heterogeneous participant composition, and the presence of liquidation cascades and exchange-specific shocks. These features raise the possibility that tail concentration may reflect mechanisms not fully captured by standard conditional variance dynamics. Rather than presuming incompatibility, this paper evaluates that possibility empirically through a null-universe diagnostic framework.

3. Data

3.1. S&P 500 (Benchmark Market)

S&P 500 index data are obtained from TradingView using the CBOE SPX feed. A data-quality inspection indicates that prior to 1962-01-01, OHLC values are identical (Open = High = Low = Close), suggesting that intraday dispersion is not recorded in the available series. Accordingly, volatility modeling begins on 1962-01-03.

For analytical purposes, the sample is divided into two subperiods: 1962–1985 and 1986–2026. This partition avoids selecting breakpoints based on ex post volatility events and permits evaluation across distinct historical market environments, including episodes such as Black Monday (1987), the dot-com correction, the global financial crisis, and the COVID-19 shock.

3.2. Bitcoin (Cryptocurrency Market)

Bitcoin data consist of Coinbase BTCUSD OHLC series obtained from TradingView at four sampling frequencies: 30-minute, 1-hour, 4-hour, and daily. Due to platform export limits, sample lengths differ across frequencies.

Each frequency is evaluated within its own simulated null framework using the corresponding empirical sample length. As a result, inference is conducted separately at each frequency, and rejection decisions do not rely on cross-frequency comparisons of absolute exceedance rates.

3.3. Returns

For each series, we compute close-to-close log returns scaled to percent for numerical stability: r = 100 × (ln(Close_t) − ln(Close_{t − 1})).

4. Methodology

4.1. Baseline GARCH Specification

For each return series, we estimate a standard GARCH(1,1) model with a constant mean and conditional normal innovations. Parameters are obtained via maximum likelihood estimation under standard non-negativity constraints, with covariance stationarity imposed through the restriction α + β < 1. Multiple initial values are used to mitigate sensitivity to local optima.

4.2. Standardized Innovations as the Basis of the Null Test

Standardized innovations are computed as

$$Z_t = \epsilon_t / \sigma_t,$$

where $\epsilon_t$denotes the residual and $\sigma_t$the fitted conditional volatility.

Under the GARCH-Normal specification, the standardized innovations are expected to approximate standard normal noise. Accordingly, large absolute realizations (e.g., |Z_t| > 5) should occur with low probability under the maintained null.

4.3. Null-Universe Rejection Criterion

Tail exceedances are evaluated at thresholds $k \in \{4,5,6\}$. Under the normal-innovation null, the expected exceedance probabilities at these thresholds are small. For each series, we report empirical exceedance frequencies, expected counts under the fitted model, binomial p-values, and Monte Carlo p-values based on 5,000 simulated paths.

This framework provides a transparent way to assess whether observed tail realizations are consistent with model-implied behavior.

4.4. Robustness Check: Student-t Innovations

To examine whether rejection under the normal specification reflects distributional misspecification, each series is re-estimated under a GARCH(1,1) model with Student-t innovations, jointly estimating the degrees-of-freedom parameter ν. Expected exceedance counts and Monte Carlo p-values are recalculated using the fitted ν.

If the Student-t specification yields non-rejection with moderate degrees of freedom (e.g., $\nu > 5$), this suggests that tail discrepancies may be addressed through distributional refinement within the conditional variance framework. If substantially lower values of $\nu$are required and rejection persists, the results indicate increasing tension between model-implied and observed tail behavior.

5. Results

5.1. GARCH(1,1) Parameter Estimates and Innovation Tail Statistics

Table 1 reports GARCH(1,1) parameter estimates and standardized innovation tail statistics for all six series. For the S&P 500, parameter estimates are relatively stable across both subsamples, with $\alpha + \beta$ below unity (0.9965 and 0.9822). In contrast, the BTC 30-minute series approaches the IGARCH boundary ($\alpha + \beta = 1.0000$, marked *), indicating persistence near the non-stationary region under the standard parameterization.

The kurtosis of standardized innovations Z ranges from 3.82 (S&P 500, 1962–1985) to 19.14 (BTC 1-hour). Across all sampling frequencies, Bitcoin exhibits substantially higher residual kurtosis than the equity benchmark after volatility filtering.

Figure 1 provides a four-panel comparison of kurtosis and tail exceedance rates across the six series. The kurtosis panels indicate a pronounced separation in magnitude between the S&P 500 and Bitcoin series: the S&P 500 reaches a maximum of 6.56 (post-1986), whereas Bitcoin's minimum value is 11.18 (30-minute frequency).

At the $k = 6$ threshold, S&P 500 (1962–1985) records zero exceedances, while all four Bitcoin frequencies display positive counts. This contrast highlights the differing concentration of extreme standardized realizations across the two asset classes under the fitted GARCH(1,1) specification.

Table 1
GARCH(1,1) Parameter Estimates and Innovation Tail Statistics

Series	N	μ	ω	α	β	α + β	Kurt(Z)	P(|Z|>5)
S&P 500 (1962–1985)	6,030	0.0392	0.00439	0.0849	0.9116	0.9965	3.82	0.000332
S&P 500 (1986–2026)	10,112	0.0647	0.02378	0.1136	0.8686	0.9822	6.56	0.001187
BTC 30-min	20,041	0.0035	0.00357	0.2253	0.7747	**1.0000***	11.18	0.002595
BTC 1-hour	27,561	-0.0001	0.01744	0.2460	0.7306	0.9767	19.14	0.003302
BTC 4-hour	20,033	0.0186	0.03097	0.0825	0.9082	0.9908	14.19	0.002795
BTC Daily	4,067	0.1814	0.65640	0.1850	0.7937	0.9786	15.12	0.001967

*Notes: μ = mean return; ω = GARCH intercept; α = ARCH coefficient; β = GARCH coefficient; α + β = persistence. Kurt(Z) = excess kurtosis of standardized innovations; P(|Z|>5) = empirical tail rate. * BTC 30-min hits the IGARCH boundary. Blue rows: S&P 500 benchmark. Red rows: Bitcoin.*

5.2. Tail-Event Null Tests (GARCH-Normal)

Table 2 reports Monte Carlo results for tail exceedances at thresholds $k \in \{4,5,6\}$. Under the normal-innovation null, simulated exceedance counts at these thresholds are typically very small. In several cases, p-values below 10^{-300} indicate that none of the simulated paths approached the observed exceedance counts.

For the S&P 500 (1962–1985), the null cannot be rejected at $k = 6$(p = 1.00), indicating consistency with model-implied tail behavior at this threshold. The post-1986 subsample exhibits stronger deviations, consistent with major historical volatility episodes; however, exceedance frequencies remain substantially lower than those observed for Bitcoin.

Across all four Bitcoin frequencies, the null is rejected at each threshold under the GARCH-Normal specification. Observed exceedance counts exceed model-implied expectations by large margins, and Monte Carlo p-values are below 0.0001 in all cases. This pattern is observed consistently across sampling intervals from 30-minute to daily.

Figure 2 displays empirical distributions of standardized residuals Zoverlaid on the standard normal density. The S&P 500 series broadly track the normal benchmark, with moderate tail deviations. In contrast, the Bitcoin series exhibit higher central concentration and heavier tails relative to the normal

reference. This pattern persists after filtering by the fitted GARCH model, indicating that residual tail concentration remains under the normal-innovation specification.

Table 2
Monte Carlo Null-Universe Rejection: Tail Exceedance Tests (GARCH-Normal)

Series	k	Obs	Exp	Obs Rate	Exp Rate	p-Binom	p-MC
S&P 500 (1962–1985)	4	7	0.38	0.001161	$6.33e^{-05}$	$1.68e^{-07}$	< 0.0001
	5	2	0	0.000332	$5.73e^{-07}$	$5.96e^{-06}$	< 0.0001
	6	0	0	0.000000	$1.97e^{-09}$	**1.0000**	**1.0000**
S&P 500 (1986–2026)	4	28	0.64	0.002769	$6.33e^{-05}$	$< 1e^{-300}$	< 0.0001
	5	12	0.01	0.001187	$5.73e^{-07}$	$< 1e^{-300}$	< 0.0001
	6	5	0	0.000494	$1.97e^{-09}$	$< 1e^{-300}$	< 0.0001
BTC 30-minute	4	126	1.27	0.006287	$6.33e^{-05}$	$< 1e^{-300}$	< 0.0001
	5	52	0.01	0.002595	$5.73e^{-07}$	$< 1e^{-300}$	< 0.0001
	6	27	0	0.001347	$1.97e^{-09}$	$< 1e^{-300}$	< 0.0001
BTC 1-hour	4	204	1.75	0.007402	$6.33e^{-05}$	$< 1e^{-300}$	< 0.0001
	5	91	0.02	0.003302	$5.73e^{-07}$	$< 1e^{-300}$	< 0.0001
	6	46	0	0.001669	$1.97e^{-09}$	$< 1e^{-300}$	< 0.0001
BTC 4-hour	4	131	1.27	0.006539	$6.33e^{-05}$	$< 1e^{-300}$	< 0.0001
	5	56	0.01	0.002795	$5.73e^{-07}$	$< 1e^{-300}$	< 0.0001
	6	29	0	0.001448	$1.97e^{-09}$	$< 1e^{-300}$	< 0.0001
BTC Daily	4	25	0.26	0.006147	$6.33e^{-05}$	$< 1e^{-300}$	< 0.0001
	5	8	0	0.001967	$5.73e^{-07}$	$< 1e^{-300}$	< 0.0001
	6	5	0	0.001229	$1.97e^{-09}$	$< 1e^{-300}$	< 0.0001

Notes: Obs = observed exceedance count; Exp = expected count under GARCH-Normal null; p-Binom = one-sided binomial p-value; p-MC = Monte Carlo p-value (N = 5,000 simulations). Gray row (S&P 500 1962–85, k = 6): null not rejected (p = 1.00). Blue rows: S&P 500. Red rows: Bitcoin.

Figure 3 plots observed-to-expected (Obs/Exp) ratios on a log scale across all series and thresholds. Under the null hypothesis, this ratio is expected to be close to unity. The S&P 500 (1962–1985) at k = 6 lies at or below 1, consistent with behavior expected under the GARCH-Normal benchmark.

In contrast, BTC exhibits substantially elevated ratios at k = 6 across all sampling frequencies: approximately 13,700× (30-minute), 23,400× (1-hour), 14,700× (4-hour), and 2,500× (daily). These magnitudes indicate that extreme standardized events occur far more frequently than predicted by the GARCH-Normal specification. Such discrepancies suggest that tail risk in Bitcoin may not be adequately characterized by a conditional variance structure combined with normal innovations, particularly at higher frequencies. Consequently, risk measures derived under this specification should be interpreted with caution in event-sensitive market environments.

5.3. Robustness Check: GARCH-Normal vs. GARCH-t

A natural interpretation of rejection under the GARCH-Normal specification is potential misspecification of the innovation distribution rather than limitations of the conditional variance structure itself. Table 3 therefore compares results under Normal and Student-t innovations across all six series.

For the S&P 500, allowing for Student-t innovations largely reconciles the observed tail exceedances. Across both eras and all three thresholds, p-values increase to the range 0.22–1.00, and the null is no longer rejected. The estimated degrees of freedom are moderate ($\nu = 10.0$ for 1962–1985 and $\nu = 5.6$ for 1986–2026), consistent with the well-documented leptokurtosis of equity returns (Cont, 2001). These results suggest that, in the benchmark market, tail deviations can be accommodated within the GARCH framework through distributional refinement.

For Bitcoin, the estimated degrees of freedom are substantially lower, with $\nu \approx 3.0–4.0$ across frequencies. At such values, higher-order moments become undefined or highly sensitive to sampling variability: kurtosis is undefined at $\nu \leq 4$ and variance at $\nu \leq 2$. While the Student-t specification substantially reduces exceedance discrepancies relative to the Normal case, residual rejection persists at higher frequencies (e.g., BTC 30-minute and 1-hour at k = 4, p = 0.0012 and 0.0016).

Taken together, these findings indicate that for Bitcoin, reconciliation within the GARCH framework requires very heavy-tailed innovations and may remain incomplete at shorter sampling intervals. This pattern suggests that distributional adjustment alone may not fully account for the observed tail behavior in high-frequency cryptocurrency returns.

Table 3
GARCH-Normal vs. GARCH-t Null Test Comparison

Series	v (df)	k	Obs	Exp (Normal)	p-MC (Normal)	Exp (t)	p-MC (t)	Verdict
S&P 500 (1962–1985)	v = 10.0	4	7	0.38	< 0.0001	7.23	0.5934	**Cannot reject**
		5	2	~ 0	< 0.0001	1.4	0.3966	**Cannot reject**
		6	0	~ 0	1.0000	0.32	1.0000	**Cannot reject**
S&P 500 (1986–2026)	v = 5.6	4	28	0.64	< 0.0001	30.13	0.4670	**Cannot reject**
		5	12	0.01	< 0.0001	10.13	0.3018	**Cannot reject**
		6	5	~ 0	< 0.0001	3.99	0.2196	**Cannot reject**
BTC 30-min	v = 4.0	4	126	1.27	< 0.0001	96.59	0.0012	**REJECT**
		5	52	0.01	< 0.0001	42.39	0.0608	**Marginal**
		6	27	~ 0	< 0.0001	21.26	0.1278	**Marginal**
BTC 1-hour	v = 3.2	4	203	1.75	< 0.0001	164.87	0.0016	**REJECT**
		5	91	0.02	< 0.0001	83.7	0.1466	**Marginal**
		6	45	~ 0	< 0.0001	47.69	0.6654	**Cannot reject**
BTC 4-hour	v = 3.0	4	131	1.27	< 0.0001	123.04	0.1482	**Marginal**
		5	56	0.01	< 0.0001	64.29	0.6178	**Cannot reject**
		6	29	~ 0	< 0.0001	37.56	0.9098	**Cannot reject**
BTC Daily	v = 3.1	4	25	0.26	< 0.0001	24.79	0.3462	**Cannot reject**

Notes: v = estimated Student-t degrees of freedom. Exp(Normal) = expected exceedances under Normal null. Exp(t) = expected under t-null with fitted v. p-MC = Monte Carlo p-value (N = 5,000). Green cells: cannot reject null. Yellow: marginal (0.05 < p < 0.15). Red: reject. Blue rows: S&P 500. Red rows: Bitcoin.

Series	v (df)	k	Obs	Exp (Normal)	p-MC (Normal)	Exp (t)	p-MC (t)	Verdict
		5	8	~ 0	< 0.0001	12.83	0.9376	**Cannot reject**
		6	5	~ 0	< 0.0001	7.43	0.7492	**Cannot reject**

Notes: v = estimated Student-t degrees of freedom. Exp(Normal) = expected exceedances under Normal null. Exp(t) = expected under t-null with fitted v. p-MC = Monte Carlo p-value (N = 5,000). Green cells: cannot reject null. Yellow: marginal (0.05 < p < 0.15). Red: reject. Blue rows: S&P 500. Red rows: Bitcoin.

6. Discussion

The empirical results suggest a meaningful distinction in the degree of compatibility between GARCH-type models and different market environments. In the S&P 500 benchmark, extreme standardized events are infrequent relative to model-implied expectations once moderate heavy-tailed innovations are introduced. In this setting, distributional refinements such as Student-t innovations and asymmetric extensions (e.g., EGARCH, GJR-GARCH) appear sufficient to accommodate observed tail behavior. The pre-1986 subsample, in particular, illustrates that under appropriate distributional assumptions the null cannot be rejected even at k = 6, supporting the view that GARCH-type conditional variance dynamics provide a reasonable approximation in noise-dominant equity markets.

For Bitcoin, the empirical pattern differs. Tail exceedances remain elevated across sampling frequencies, and reconciliation within the GARCH framework requires substantially lower degrees of freedom. Although the Student-t specification reduces discrepancies relative to the Normal case, residual rejection persists at shorter intervals. These findings indicate that heavy-tailed distributional adjustment alone may not fully account for the observed tail concentration in high-frequency cryptocurrency returns. One possible interpretation is that discrete event-driven mechanisms—such as liquidation cascades, regulatory announcements, or exchange-specific shocks—play a more prominent role in shaping volatility dynamics than in traditional equity markets.

This conditional interpretation also provides context for the historical evolution of GARCH variants. In equity markets, where volatility clustering is dominant and extreme events are comparatively infrequent, incremental refinements to the conditional variance equation have proven effective. In contrast, in environments characterized by more concentrated tail activity, additional structural features beyond standard conditional variance dynamics may warrant consideration. This perspective is consistent with prior findings documenting regime dependence, structural breaks, and model instability in cryptocurrency volatility modeling (Ardia et al., 2019; Abakah et al., 2020; Jiang et al., 2023; Silva & Maciel, 2025).

The proposed null-universe framework is not limited to the specific comparison examined here. Rather, it can be interpreted as a diagnostic tool for assessing model compatibility across markets. In settings

where the null is strongly rejected even after distributional adjustment, researchers may consider whether alternative modeling approaches—such as jump-diffusion, regime-switching, or event-based specifications—provide a more suitable representation of the underlying dynamics. The aim is not to dismiss the GARCH framework, but to clarify the conditions under which it offers an adequate approximation.

7. Conclusions

Using a null-universe diagnostic perspective supplemented by a Student-t robustness analysis, this study documents systematic differences in the empirical compatibility of GARCH-type models across asset classes. For the S&P 500, volatility clustering is effectively captured within the GARCH framework, and tail deviations are largely reconciled through moderate heavy-tailed innovations ($\nu = 5.6–10.0$). These findings are consistent with the long-standing empirical usefulness of GARCH-type models in equity markets.

For Bitcoin, null rejection under the GARCH-Normal specification is substantial across time scales, and reconciliation within a GARCH-t framework requires considerably lower degrees of freedom ($\nu \approx 3.0–4.0$). Although heavy-tailed innovations mitigate part of the discrepancy, residual tail concentration remains at higher frequencies. These results suggest that the adequacy of standard conditional variance dynamics may vary across market environments, particularly in settings characterized by concentrated tail events.

Rather than positioning the findings as a rejection of the GARCH framework, we interpret them as evidence that model applicability may be conditional on market structure. The null-universe approach proposed here can serve as a preliminary diagnostic prior to specification refinement. Future research may explore whether alternative parametric or semi-parametric frameworks—such as jump-based or event-sensitive models—better satisfy the null-universe criterion in markets with pronounced tail activity. This broader question remains open.

Abbreviations

The following abbreviations are used in this manuscript:

ARCH Autoregressive Conditional Heteroskedasticity

BTC Bitcoin

EGARCH Exponential GARCH

GARCH Generalized Autoregressive Conditional Heteroskedasticity

GJR-GARCH Glosten-Jagannathan-Runkle GARCH

IGARCH Integrated GARCH

MLE Maximum Likelihood Estimation

S&P 500 Standard and Poor's 500 Index

VaR Value-at-Risk

OHLC Open, High, Low, Close (price data)

Declarations

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Ethical Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Author Contribution

Conceptualization, L.-Y.C.; Methodology, L.-Y.C.; Software, L.-Y.C.; Formal analysis, L.-Y.C.; Data curation, L.-Y.C.; Writing—original draft, L.-Y.C.; Writing—review and editing, L.-Y.C. The author has read and agreed to the published version of the manuscript.

Data Availability

S&P 500 data are sourced from TradingView (CBOE SPX feed). Bitcoin OHLC data are sourced from TradingView (Coinbase BTCUSD at 30-minute, 1-hour, 4-hour, and daily frequencies). All data are publicly available through TradingView.

References

1. Abakah, E. J. A., Gil-Alana, L. A., Madigu, G., & Romero-Rojo, F. (2020). Volatility persistence in cryptocurrency markets under structural breaks. *Research in International Business and Finance, 52*, 101205.
2. Ardia, D., Bluteau, K., & Rüede, M. (2019). Regime changes in Bitcoin GARCH volatility dynamics. *Finance Research Letters, 29*, 266–271.
3. Awartani, B. M. A., & Corradi, V. (2005). Predicting the volatility of the S&P-500 stock index via GARCH models: The role of asymmetries. *International Journal of Forecasting, 21*(1), 167–183.
4. Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 74*(1), 3–30.
5. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*(3), 307–327.
6. Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics, 69*(3), 542–547.
7. Caporale, G. M., & Zekokh, T. (2019). Modelling volatility of cryptocurrencies using Markov-switching GARCH models. *Research in International Business and Finance, 48*, 143–155.
8. Chan, S., Chu, J., Nadarajah, S., & Osterrieder, J. (2019). A statistical analysis of cryptocurrencies. *Journal of Risk and Financial Management, 10*(4), 17.
9. Chu, J., Chan, S., Nadarajah, S., & Osterrieder, J. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management, 10*(4), 17.
10. Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance, 1*(2), 223–236.
11. Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar—A GARCH volatility analysis. *Finance Research Letters, 16*, 85–92.
12. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50*(4), 987–1007.
13. Engle, R. F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives, 15*(4), 157–168.
14. Engle, R. F., & Ng, V. K. (1993). Measuring and testing the impact of news on volatility. *The Journal of Finance, 48*(5), 1749–1778.
15. Fung, K., Jeong, J., & Pereira, J. (2022). More to cryptos than bitcoin: A GARCH modelling of heterogeneous cryptocurrencies. *Finance Research Letters*, 47, 102544.
16. 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.
17. Jiang, Z., Mensi, W., & Yoon, S. M. (2023). Risks in major cryptocurrency markets: Modeling the dual long memory property and structural breaks. *Sustainability, 15*(3), 2193.

18. Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters, 158*, 3–6.
19. Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica, 59*(2), 347–370.
20. Silva, L., & Maciel, L. (2025). Cryptocurrency price returns volatility modeling and forecasting with GARCH models. *RAUSP Management Journal, 60*(1), 225–239.
21. Tan, C. Y., Koh, Y. B., Ng, K. H., & Ng, K. H. (2021). Dynamic volatility modelling of Bitcoin using time-varying transition probability Markov-switching GARCH model. *The North American Journal of Economics and Finance, 56*, 101377.
22. Walther, T., Klein, T., & Bouri, E. (2019). Exogenous drivers of Bitcoin and cryptocurrency volatility—A mixed data sampling approach to forecasting. *Journal of International Financial Markets, Institutions and Money, 63*, 101133.

Figures

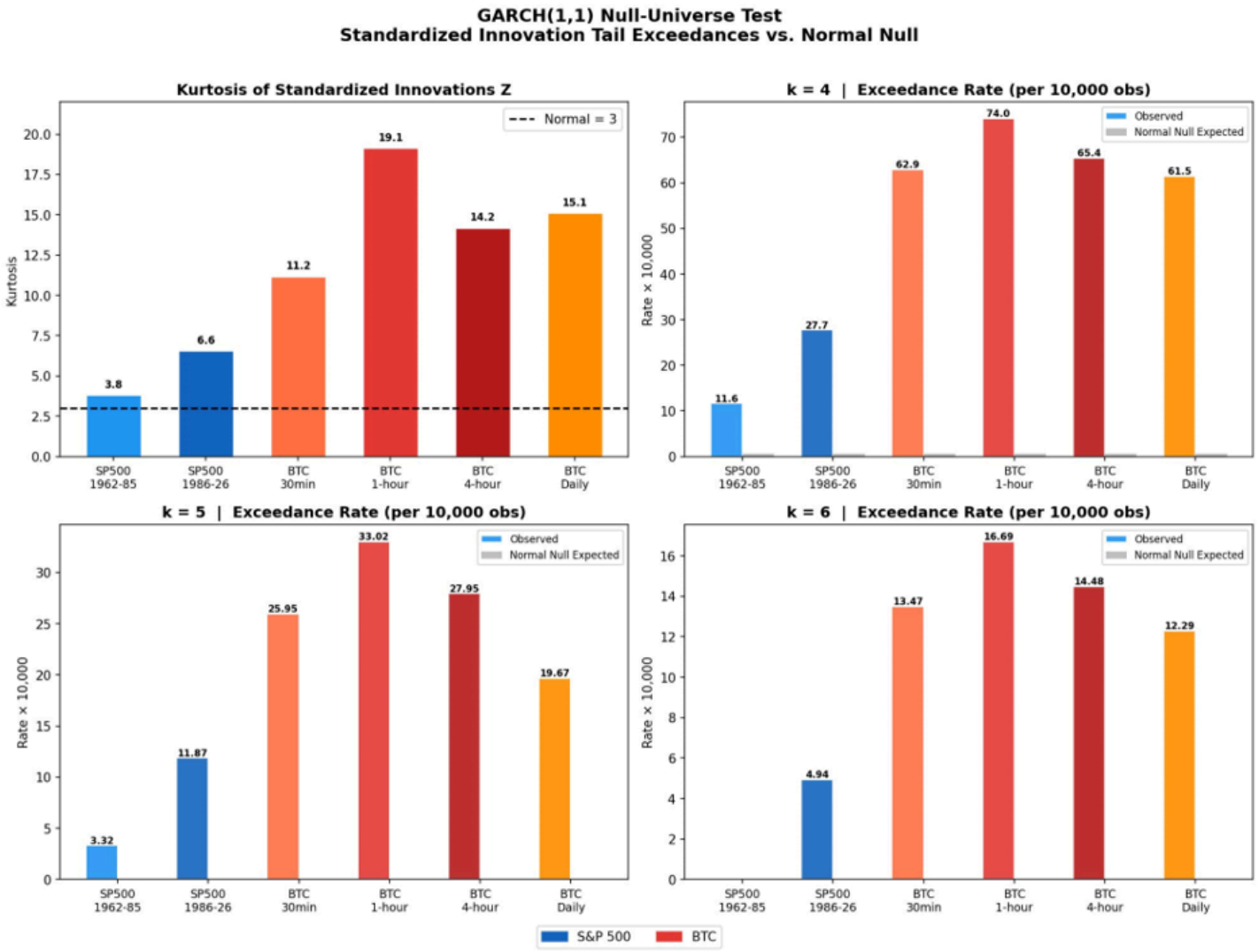


Figure 1

GARCH(1,1) Null-Universe Test: Kurtosis and Tail Exceedance Rates Across All Series. Four-panel comparison of standardized innovation kurtosis (upper left) and exceedance rates at k = 4, 5, 6 (remaining panels). Blue bars: S&P 500. Red/orange bars: Bitcoin. Dashed line: Normal null benchmark (kurtosis = 3; rate = expected under N(0,1)).

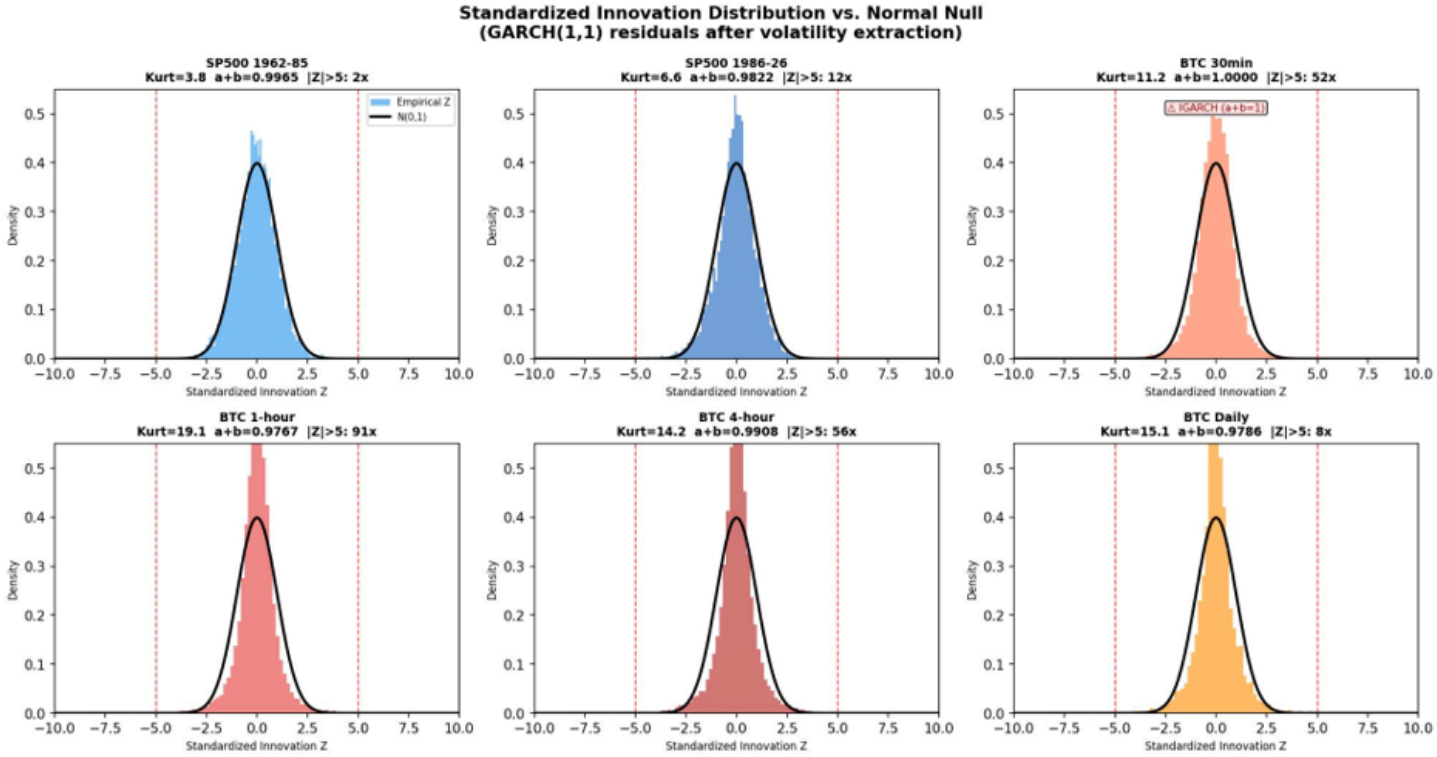
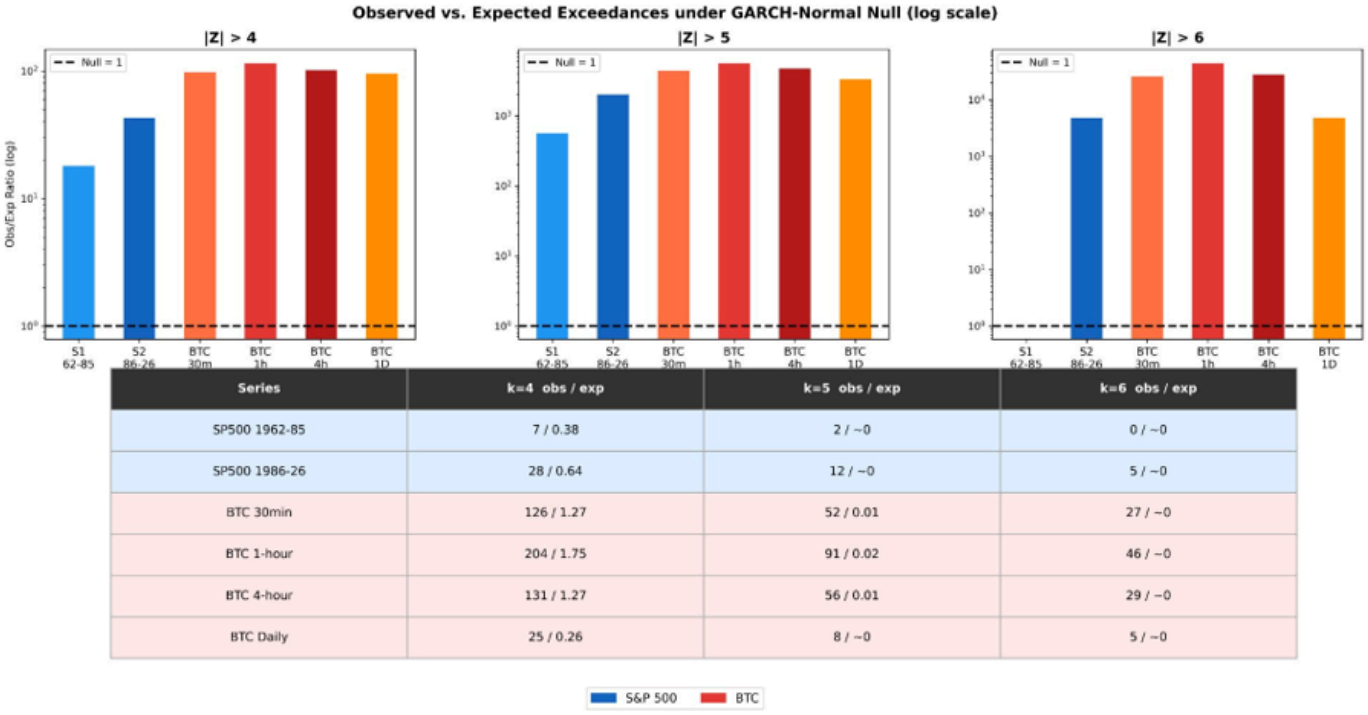


Figure 2

Standardized Innovation Distributions vs. N(0,1) Benchmark. Empirical histograms of GARCH(1,1) standardized residuals Z overlaid on the standard normal density (black curve). Red dashed lines mark ±5 standard deviations. Top row: S&P 500 series. Bottom row: Bitcoin series. BTC 30-min carries an IGARCH warning (α+β = 1.0000).



Series	k=4 obs / exp	k=5 obs / exp	k=6 obs / exp
SP500 1962-85	7 / 0.38	2 / ~0	0 / ~0
SP500 1986-26	28 / 0.64	12 / ~0	5 / ~0
BTC 30min	126 / 1.27	52 / 0.01	27 / ~0
BTC 1-hour	204 / 1.75	91 / 0.02	46 / ~0
BTC 4-hour	131 / 1.27	56 / 0.01	29 / ~0
BTC Daily	25 / 0.26	8 / ~0	5 / ~0

Figure 3

Observed vs. Expected Tail Exceedances under the GARCH-Normal Null (Log Scale). Bar charts show Obs/Exp ratios at k = 4, 5, 6 for all six series. Dashed line marks the null expectation (ratio = 1). Blue bars: S&P 500. Red/orange bars: Bitcoin. The embedded table provides raw counts. BTC 1-hour at k = 6 reaches 23,400×, highlighting the magnitude of deviation from model-implied expectations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendixs.docx