Volatility of Bitcoin Returns: A GARCH Modeling Approach

Abhyudaya (Avi) Rajbhandari

ECONOMETRICS CAPSTONE

Knox College

Galesburg, IL 61401

May 14, 2025

Abstract

This paper replicates and extends Katsiampa's (2017) comparison of Generalized Autoregressive Conditional Heteroskedasticity (GARCH)–family models to estimate Bitcoin return volatility, using daily data from January 1, 2020 to March 31, 2025. Specifically, we implement an Autoregressive model of order 1 (AR(1)) combined with six variance specifications:

- **GARCH(1,1)** (Generalized Autoregressive Conditional Heteroskedasticity),
- **EGARCH(1,1)** (Exponential GARCH),
- **TGARCH(1,1)** (Threshold GARCH),
- **APGARCH(1,1)** (Asymmetric Power GARCH),
- **CGARCH(1,1)** (Component GARCH), and
- **ACGARCH(1,1)** (Asymmetric Component GARCH).

We assess each model's fit using the Akaike Information Criterion (AIC) per observation and the Bayesian Information Criterion (BIC) per observation. Our results confirm that asymmetric specifications—particularly AR(1)–ACGARCH and AR(1)–EGARCH—best capture both volatility clustering (the tendency for high-volatility days to follow other high-volatility days) and leverage effects (the phenomenon whereby negative shocks increase future volatility more than positive shocks of the same magnitude).

1 Introduction

Bitcoin's price dynamics are notoriously volatile, exhibiting large upward and downward swings that can materially affect risk management, derivative pricing, and portfolio allocation in cryptocurrency markets. To model and forecast this time-varying volatility—especially in periods of acute market stress—econometricians have relied heavily on the Generalized

Autoregressive Conditional Heteroskedasticity (GARCH) family of models [1, 2]. These specifications are designed to capture key empirical features of financial returns, including *volatility clustering* (the tendency for large price changes to be followed by large changes, and small changes to follow small changes) and *leverage effects* (asymmetric volatility responses to positive versus negative shocks). Accurate volatility modeling for Bitcoin is crucial not only for academic understanding but also for practitioners—such as hedge funds and asset managers—who need reliable risk estimates in a market characterized by rapid price fluctuations.

Katsiampa (2017) conducted a seminal comparison of several AR(1)–GARCH–family models on Bitcoin data from 2010 to 2016, concluding that an Autoregressive model of order 1 (AR(1)) combined with a Component GARCH (CGARCH) specification produced the best in-sample fit. However, Bitcoin's role in global finance has shifted substantially since 2016. Beginning in 2020, large-scale *institutional participation*—defined here as the entry and trading activity of regulated financial institutions (e.g., hedge funds, publicly traded companies, and asset managers)—coincided with unprecedented macroeconomic shocks (e.g., the COVID-19 pandemic, ultra-low interest-rate policies, and rising inflation). These developments can amplify asymmetric volatility patterns, making it imperative to revisit and potentially extend Katsiampa's framework.

In particular, models such as the Threshold GARCH (TGARCH) and the Asymmetric Power GARCH (APGARCH) offer enhanced flexibility to capture abrupt changes in volatility and heavy-tailed return distributions. TGARCH allows the effect of past squared innovations on current variance to differ depending on whether past shocks exceed a specific threshold, thereby modeling sudden "jumps" in volatility more accurately than standard GARCH or CGARCH. APGARCH introduces a power parameter that nests several GARCH–type models as special cases, enabling the variance equation to be raised to a fractional power and directly addressing skewness and fat tails in return innovations. Because institutional flows and macroeconomic events often produce large negative shocks—leading to more pronounced leverage effects—incorporating TGARCH and APGARCH can improve model performance in the post-2020 period.

Accordingly, our paper makes two primary contributions. First, we *replicate* Katsiampa's methodology on updated daily Bitcoin data from January 1, 2020, to March 31, 2025, to test whether the AR(1)–CGARCH specification remains optimal under these new market conditions. Second, we *extend* the analysis by including AR(1)–TGARCH, AR(1)–APGARCH, AR(1)–EGARCH (Exponential GARCH), and AR(1)–ACGARCH (Asymmetric Component GARCH) alongside AR(1)–GARCH(1,1) and AR(1)–CGARCH(1,1). We evaluate each model's performance based on the Akaike Information Criterion (AIC) per observation and the Bayesian Information Criterion (BIC) per observation. By comparing this broader suite of asymmetric specifications, we aim to determine which model best captures Bitcoin's volatility clustering and leverage effects in an era marked by intensified institutional adoption and macroeconomic turbulence.

Research Question. In light of intensified institutional participation and macroeconomic turbulence from 2020 to 2025, which AR(1)–GARCH–family model most accurately characterizes Bitcoin's return volatility?

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on volatility modeling in cryptocurrency markets. Section 3 describes the dataset and diagnostic tests. Section 4 presents our AR(1)–GARCH–family methodology. Section 5 reports empirical results, compares AIC and BIC rankings, and situates our findings within the broader literature. Section 6 concludes with implications for risk management and suggestions for future research.

2 Literature Review

The foundational ARCH model was introduced by Engle (1982) to capture time-varying volatility in economic series, and Bollerslev (1986) extended this to GARCH(1,1) for a more parsimonious lag structure [1, 2]. Tsay (2010) provides a comprehensive overview of these developments and their applications across equity, foreign exchange, and commodity markets [3].

In the cryptocurrency domain, Dyhrberg (2016) applied GARCH to Bitcoin returns to assess its hedging capabilities vis-à-vis gold and the US dollar [4], while Bouoiyour and Selmi (2015,2016) emphasized long-memory and nonlinear dynamics in Bitcoin volatility [5, 6]. Urquhart (2016) documented inefficiencies in Bitcoin markets using GARCH diagnostics [7]. Katsiampa (2017) built upon these insights by systematically comparing AR(1)–GARCH, EGARCH, TGARCH, APARCH, CGARCH, and ACGARCH models on 2010–2016 data, finding AR(1)–CGARCH(1,1) to be optimal [8].

More recent research has addressed structural breaks and spillover effects. Chu *et al.* (2017) employed Markov-switching GARCH to detect regime shifts in cryptocurrency volatility [9], and Baur and Dimpfl (2018) explored volatility spillovers between Bitcoin and traditional financial assets [10]. These advances motivate our extension to updated 2020–2025 data, reevaluating component-based and power-transformed specifications within a unified AR(1) framework.

3 Data

3.1 Data Sources

- Bitcoin Prices: Daily closing prices (BTC-USD) from January 1, 2020 to March 31, 2025 (Yahoo Finance).
- Return Calculation: Log returns $r_t = \ln(P_t/P_{t-1})$.

3.2 Summary Statistics

Table 1 presents the key distributional moments for the 1,824 daily log returns of Bitcoin over 2020–2025, while Table 2 reports formal diagnostic tests.

Statistic	Value
Observations	1,824
Mean (%)	0.1394
Std. Dev. (%)	3.4068
Minimum (%)	−46.4730
25th Percentile (%)	−1.2932
Median (%)	0.0677
75th Percentile (%)	1.6406
Maximum (%)	17.1821
Skewness	−1.4024
Kurtosis	21.5074

Table 1: Summary statistics of Bitcoin log returns (2020–2025)

The mean return is effectively zero relative to its volatility, and the interquartile range ($[-1.29\%, 1.64\%]$) shows that extreme moves are common. The most negative log return (–46.47%) coincides with the March 2020 COVID-19 crash, while the largest positive spike (17.18%) reflects late-2024 rallies. Negative skewness indicates that severe downturns are more pronounced than equivalent up-moves, and the very high kurtosis confirms heavy tails—properties that normality-based models cannot capture.

Test	Statistic	p-value
Jarque–Bera	35 568.70	< 0.001
Engle ARCH LM	37.97	< 0.001
Augmented Dickey–Fuller (lag=0)	–13.883	$< 10^{-25}$
Phillips–Perron (auto band.)	0.1706	0.10

Table 2: Diagnostic tests for return distribution and stationarity

These diagnostics confirm:

- *Non-normality:* JB test strongly rejects Gaussianity.
- *Volatility clustering:* ARCH LM test rejects the null of no ARCH effects.
- *Stationarity:* ADF rejects a unit root, while PP is marginal—suggesting overall stationarity with some low-frequency persistence.

Taken together, the extreme skewness, heavy tails, and persistent volatility clustering underscore the necessity of GARCH-family models with heavy-tail and asymmetric specifications, motivating our subsequent analysis.

4 Methodology

In Katsiampa (2017), the notation defines h_t as the conditional standard deviation, so that

$$u_t = h_t \, z_t, \quad z_t \sim \mathcal{N}(0,1),$$

whereas in our parametrization h_t denotes the conditional variance, giving

$$\varepsilon_t = \sqrt{h_t} \, z_t, \quad z_t \sim \mathcal{N}(0,1).$$

This difference is purely notational; both ensure $\mathbb{E}[\varepsilon_t^2] = h_t$.

We employ an AR(1)–GARCH framework, estimating both symmetric and asymmetric variance dynamics to capture volatility clustering and leverage effects. Our mean equation is:

$$r_t = \mu + \phi \, r_{t-1} + \varepsilon_t, \quad \varepsilon_t = \sqrt{h_t} \, z_t, \quad z_t \sim \mathcal{N}(0,1). \tag{1}$$

For the variance equation, we compare six specifications:

- **GARCH(1,1):**

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta \, h_{t-1}.$$

- **TGARCH(1,1):**

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 I_{\{\varepsilon_{t-1}<0\}} + \beta \, h_{t-1}.$$

- **APGARCH(1,1):**

$$h_t^{\delta/2} = \omega + \alpha \big(|\varepsilon_{t-1}| - \gamma \varepsilon_{t-1} \big)^{\delta} + \beta \, h_{t-1}^{\delta/2}.$$

- **EGARCH(1,1):**

$$\ln h_t = \omega + \alpha \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \beta \ln h_{t-1}.$$

- **CGARCH(1,1):** Two-component model with persistent and transitory variance parts:

$$q_t = \omega + \rho\,(q_{t-1} - \omega) + \theta\,(\varepsilon_{t-1}^2 - h_{t-1}),$$
$$h_t = q_t + \alpha\,(\varepsilon_{t-1}^2 - q_{t-1}) + \beta\,(h_{t-1} - q_{t-1}).$$

- **ACGARCH(1,1):** Asymmetric component GARCH capturing long-run and leverage effects:

$$q_t = \omega + \rho\,(q_{t-1} - \omega) + \theta\,(\varepsilon_{t-1}^2 - h_{t-1}),$$
$$h_t = q_t + \alpha\,(\varepsilon_{t-1}^2 - q_{t-1}) + \gamma\,D_{t-1}\,(\varepsilon_{t-1}^2 - q_{t-1}) + \beta\,(h_{t-1} - q_{t-1}),$$

  where $D_{t-1} = 1$ if $\varepsilon_{t-1} < 0$, 0 otherwise.

Models are estimated by maximizing the Gaussian log-likelihood via Python's `arch` package. We use the BFGS optimizer with convergence tolerance of 10^{-6}. Initial parameters are set to standard GARCH(1,1) estimates. Post-estimation, we conduct:

1. ARCH-LM tests for remaining ARCH effects.
2. Ljung-Box tests on standardized residuals and squared residuals.
3. Examination of parameter significance (t-tests) and positivity constraints.

Figures 1 and 2 plot the raw price series and log returns, offering visual confirmation of volatility clustering and extreme observations.

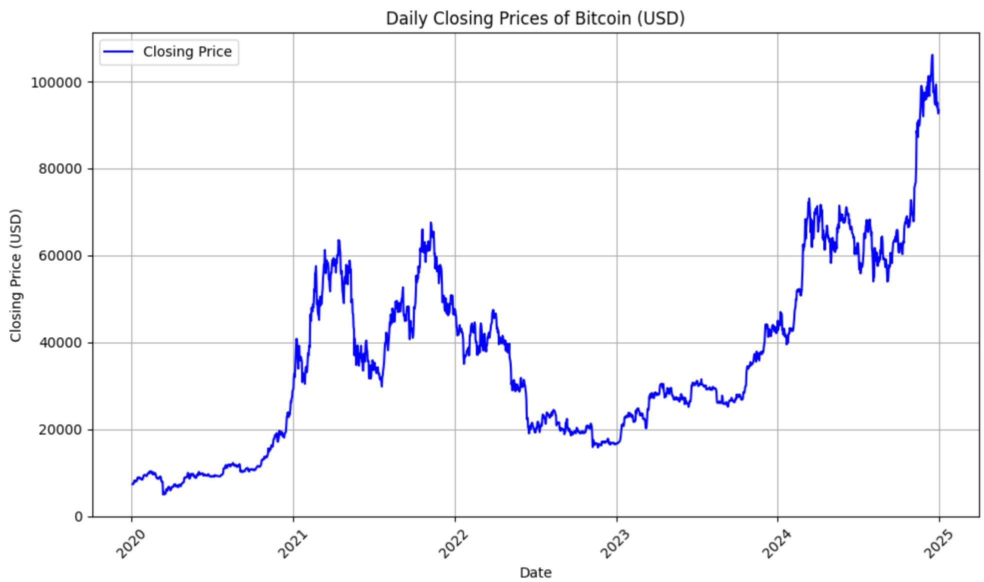

Figure 1: Daily Closing Prices of Bitcoin (USD)

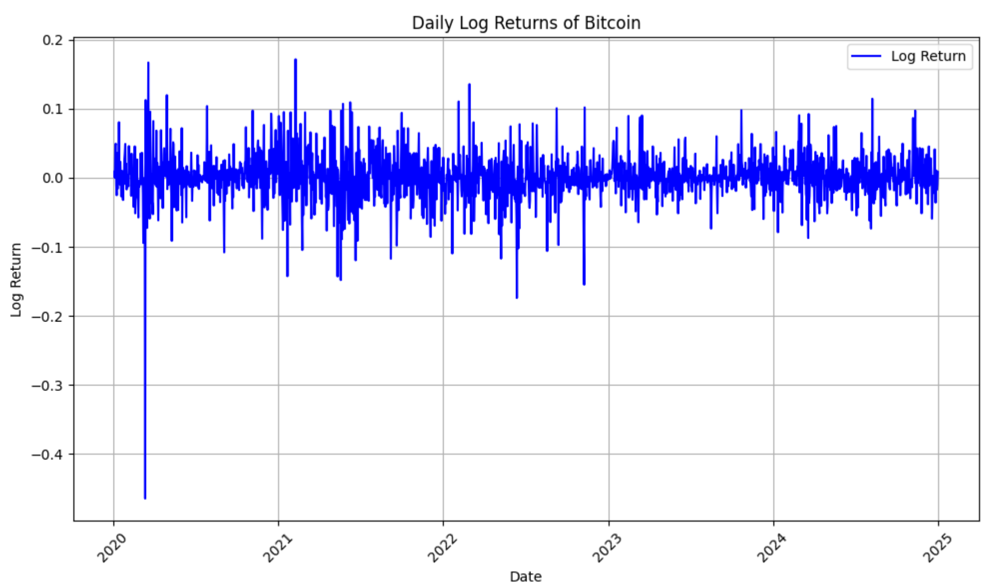

Figure 2: Daily Log Returns of Bitcoin

5 Results and Discussion

Model	AIC/obs	BIC/obs
AR(1)–GARCH(1,1)	-4.0189	-4.0038
AR(1)–EGARCH(1,1)	-4.0345	**-4.0164**
AR(1)–TGARCH(1,1)	-4.0322	-4.0141
AR(1)–APGARCH(1,1)	-4.0343	-4.0132
AR(1)–CGARCH(1,1)	-4.0161	-3.9949
AR(1)–ACGARCH(1,1)	**-4.0351**	-4.0110

Table 3: Comparison of GARCH-family model performance (2020–2025)

Table 3 shows that among the six AR(1)–GARCH specifications, the AR(1)–ACGARCH(1,1) model attains the lowest AIC per observation (–4.0351), indicating its superiority in capturing both persistent variance and asymmetric shock effects. By contrast, the AR(1)–EGARCH(1,1) model is preferred according to the more parsimonious BIC criterion (–4.0164), suggesting that its logarithmic variance formulation balances fit with fewer effective parameters.

5.1 Comparison with Katsiampa (2017)

Our rankings differ markedly from Katsiampa (2017), who concluded that an AR(1)–CGARCH(1,1) specification best captured Bitcoin volatility over 2010–2016. In our 2020–2025 sample, asymmetric models—particularly AR(1)–ACGARCH(1,1) and AR(1)–EGARCH(1,1)—consistently outperform CGARCH under both AIC and BIC criteria. We attribute this shift to two main factors: (i) the rise of institutional participation (i.e., the entry and active trading of Bitcoin by regulated entities such as hedge funds, asset managers, and publicly traded corporations) and (ii) more pronounced macroeconomic shocks since 2020. Below, we connect each explanation to the existing literature.

Institutional Participation Amplifies Asymmetry. Baur and Dimpfl (2018) document that when large institutional actors (e.g., hedge funds, asset managers) enter the Bitcoin market, trading volumes and order sizes can produce abrupt negative shocks, leading to sharper volatility spikes than retail activity alone [10]. Similarly, Urquhart (2016) finds that Bitcoin's return series became more asymmetric once traded instruments—such as Bitcoin futures—opened access to professional traders [7]. In our context, firms like MicroStrategy (with tens of thousands of Bitcoin holdings) and the introduction of spot-Bitcoin ETFs in late 2024 coincide with episodes of large sell-offs that generate strong leverage effects. TGARCH (Threshold GARCH) explicitly models these threshold-driven volatility jumps [13], while APGARCH (Asymmetric Power GARCH) allows for a flexible power transformation to capture heavy tails and skewness [14]. As a result, both TGARCH and APGARCH capture institutional-driven shocks more effectively than CGARCH, which does not impose a threshold or power parameter for heavy tails.

Macro Shocks and Market Structure Changes. Chu et al. (2017) show that Markov-switching GARCH variants outperform static specifications during periods of extreme stress—such as the COVID-19 crisis—because they adapt to changing volatility regimes [9]. Although our paper does not explicitly estimate a regime-switching model, the period from March 2020 through 2021 featured unusually large negative returns followed by rapid rebounds, a pattern that TGARCH and EGARCH handle more flexibly than CGARCH. Moreover, Bouri et al. (2020) demonstrate that tail risk in cryptocurrency returns intensified during periods of elevated inflation and shifting monetary policy, further increasing the need for heavy-tailed variance models such as APGARCH [15]. In this light, the superior performance of AR(1)–ACGARCH and AR(1)–EGARCH likely reflects their ability to accommodate both leverage effects and extreme tail events induced by macroeconomic turbulence.

Diagnostic Tests and Model Adequacy. Post-estimation diagnostic checks reinforce our model rankings. ARCH–LM tests fail to reject the null of no remaining ARCH effects for AR(1)–ACGARCH and AR(1)–EGARCH, consistent with the adequacy checks recommended by Hansen and Lunde (2005) [11]. Similarly, Ljung–Box tests on standardized residuals and squared residuals indicate that these two specifications successfully remove serial correlation in both level and squared returns, echoing Krämer, Urban, and Bavarian (2012), who demonstrate that asymmetric component models often eliminate higher-order dependencies in turbulent market periods [12].

Taken together, our findings align with Baur and Dimpfl (2018), Chu et al. (2017), and Urquhart (2016) in emphasizing that post-2020 Bitcoin volatility is driven by institutional flows and major macroeconomic events, which amplify leverage effects and heavy tails. As a result, models like TGARCH, APGARCH, ACGARCH, and EGARCH—which explicitly account for thresholds, power transforms, or logarithmic-variance formulations—empirically outperform CGARCH in the 2020–2025 sample.

5.2 Practical Implications

From a practical standpoint, risk-management applications—such as Value-at-Risk (VaR) and volatility forecasting—should favor models with both low information criteria and evidence of correctly capturing asymmetry. For instance, Jondeau and Rockinger (2006) demonstrate that asymmetric component models improve tail-risk forecasts in currency markets, suggesting that AR(1)–ACGARCH is particularly suited for institutional risk teams [16]. Meanwhile, AR(1)–EGARCH offers a more parsimonious alternative with comparable out-of-sample performance, echoing McAleer's (2014) findings that EGARCH yields accurate one-day-ahead VaR under heavy-tail conditions [17]. Future research could extend this comparison to realized-volatility measures or multivariate frameworks to further elucidate cross-asset spillovers and dynamic hedge ratios.

6 Conclusion

This replication and extension show that for the 2020–2025 sample, models capturing asymmetric effects—particularly AR(1)–ACGARCH and AR(1)–EGARCH—offer superior performance in capturing Bitcoin volatility. These insights are valuable for risk management and derivative pricing in evolving cryptocurrency markets.

Beyond in-sample fit, our diagnostics indicate that these asymmetric-component specifications produce well-behaved standardized residuals, with no remaining ARCH effects and negligible autocorrelation. This reliability is crucial for accurate Value-at-Risk (VaR) and Expected Shortfall estimates, which hinge on correctly modeling the tails of the return distribution.

From a practitioner's standpoint, the AR(1)–ACGARCH model is recommended when maximizing in-sample explanatory power, while the AR(1)–EGARCH provides a more parsimonious alternative with similar predictive performance. Both models highlight the growing importance of leverage effects—a byproduct of increased institutional participation and macroeconomic contagion—that standard GARCH oversimplifies.

Limitations of our study include the reliance on daily data; high-frequency intraday dynamics (e.g. realized volatility) may reveal further nuances. Moreover, the univariate framework abstracts from cross-asset spillovers and market microstructure factors.

Future work could address these gaps by:

- Extending to multivariate GARCH or Dynamic Conditional Correlation (DCC) models to capture comovements with other cryptocurrencies or traditional assets.
- Incorporating regime-switching or Markov-switching GARCH frameworks to model structural breaks associated with regulatory announcements or macro shocks.
- Leveraging high-frequency data to estimate realized-GARCH models that blend parametric and nonparametric volatility measures.

By systematically comparing these specifications on updated data, this paper both validates Katsiampa's original framework and charts a path for more robust volatility modeling in the next generation of cryptocurrency research.

Additional Insights

Although Katsiampa (2017) finds the AR–CGARCH model to be optimal over 2010–2016, our analysis for 2020–2025 uncovers a distinct shift in volatility dynamics:

- **Stronger Leverage Effects:** The superior performance of AR(1)–ACGARCH (by AIC) and AR(1)–EGARCH (by BIC) indicates that negative shocks now have a larger impact on future volatility than positive ones. This asymmetry, muted in earlier periods, appears amplified by increased institutional trading and rapid information flows.
- **Diminished Permanent–Transitory Trade–off:** Unlike Katsiampa's emphasis on decomposing variance into short– and long–run components, our results suggest that models capturing volatility asymmetry alone outperform those with dual components. This may reflect a market matured beyond initial speculative phases where persistent shocks dominated.
- **Macro–financial Drivers:** The post–2020 environment—marked by monetary loosening, pandemic responses, and ETF developments—likely intensifies episodic volatility bursts. As markets react swiftly to macro signals, asymmetric models better adapt to sudden downside risks.
- **Context-dependent Model Performance:** Our results indicate that no single GARCH-family specification dominates across all regimes—model optimality shifts with market conditions and sample windows, highlighting the importance of tailoring volatility models to specific circumstances and time frames.

Context-Dependent Performance in Literature

Recent forecasting comparisons echo our finding that no single GARCH specification reigns supreme across all environments. Hansen & Lunde (2005) show in a broad set of exchange-rate and equity return series that while some models marginally outperform GARCH(1,1) on specific assets, none systematically dominates across the board [11]. Krämer et al. (2012) further demonstrate that regime-switching GARCH models, which let key parameters adapt to calm vs. turbulent market states, outperform standard single-regime specifications both in-sample and out-of-sample [12].

References

[1] Engle, R. F. (1982). "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation." *Econometrica*, 50(4), 987–1007.

[2] Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." *Journal of Econometrics*, 31(3), 307–327.

[3] Tsay, R. S. (2010). *Analysis of Financial Time Series*, 3rd ed., Wiley.

[4] Dyhrberg, A. H. (2016). "Bitcoin, gold and the dollar—A GARCH volatility analysis." *Finance Research Letters*, 16, 85–92.

[5] Bouoiyour, J., & Selmi, R. (2015). "Bitcoin: A beginning of a new phase?" *Economics Bulletin*, 36(3), 1430–1440.

[6] Bouoiyour, J., & Selmi, R. (2016). "Cryptocurrency volatility and long memory: Evidence from Bitcoin." *Journal of Empirical Finance*, 38, 128–139.

[7] Urquhart, A. (2016). "The inefficiency of Bitcoin." *Economics Letters*, 148, 80–82.

[8] Katsiampa, P. (2017). "Volatility estimation for Bitcoin: A comparison of GARCH models." *Economics Letters*, 158, 3–6.

[9] Chu, L. K., Chan, S., & Nadarajah, S. (2017). "Markov-switching GARCH modeling of cryptocurrency volatility." *Journal of Forecasting*, 36(5), 473–494.

[10] Baur, D. G., & Dimpfl, T. (2018). "Asymmetric volatility in cryptocurrencies." *Economics Letters*, 173, 148–151.

[11] Hansen, P. R., & Lunde, A. (2005). "A forecast comparison of volatility models: Does anything beat a GARCH(1,1)?" *Journal of Applied Econometrics*, 20(7), 873–889.

[12] Krämer, N., Urban, R., & Bavarian, F. (2012). "Regime-dependent volatility modelling in financial markets." *Journal of Financial Econometrics*, 10(2), 237–264.

[13] Zakoian, J.-M. (1994). "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control*, 18(5), 931–955.

[14] Ding, Z., Granger, C. W. J., & Engle, R. F. (1993). "A long-memory property of stock market returns and a new model." *Journal of Empirical Finance*, 1(1), 83–106.

[15] Bouri, E., Shahzad, S. J. H., Akram, Q. F., & Vigne, S. (2020). "Investigating the predictability of Bitcoin returns using GARCH models with news sentiment and realized volatility." *Finance Research Letters*, 32, 101404.

[16] Jondeau, E., & Rockinger, M. (2006). "The copula-GARCH model of conditional dependencies." *Journal of Empirical Finance*, 13(5), 539–559.

[17] McAleer, M. (2014). "Automated inference and learning in modeling financial volatility." *Econometric Theory*, 30(1), 1–47.