Proceedings of the 5th International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/2026.36621

Measuring Crypto Tail Risk under Extreme Shocks: An EVT-GARCH and CAViaR Comparative Study

Yucheng Song

School of Economics, University of Nottingham Ningbo China, Ningbo, China
hnyys25@nottingham.edu.cn

Abstract. Cryptocurrency markets face frequent extreme price swings, flash crashes and liquidity shocks. The expansion of decentralized finance, stablecoins and interconnected trading venues further complicates tail-risk measurement. This study examines extreme downside risk in Bitcoin, Ethereum, Uniswap, and DAI, treating the 2022 FTX collapse as the primary exogenous shock event. Daily closing prices from the Kaggle dataset are aligned by calendar date and transformed into logarithmic returns. Descriptive statistics and Jarque-Bera tests are first conducted, followed by an EVT-GARCH model incorporating an FTX event dummy and an Asymmetric Slope CAViaR model for tail-risk estimation. The results reveal significant deviations from normality, heavy tails, and substantial differences across asset types. The FTX window is associated with additional volatility in all four assets, including de-pegging-related risk in DAI. Backtesting at the 99% and 95% VaR levels indicates that both models provide reliable risk forecasts. EVT-GARCH generates smoother risk estimates, while CAViaR responds more quickly to recent negative shocks and records smaller violation-count deviations for the benchmark assets. These findings support the joint use of complementary models in cryptocurrency stress testing and risk monitoring.

Keywords: extreme value theory, CAViaR, cryptocurrency, tail risk, FTX collapse

1. Introduction

Cryptocurrency markets have become an established part of digital finance but exhibit risk characteristics distinct from conventional assets, combining volatility clustering, abrupt price jumps and rapid cross-protocol shock transmission. The November 2022 FTX collapse demonstrated how a centralized trading venue failure could disrupt confidence, liquidity and market structure across the broader crypto ecosystem [1]. The expansion of decentralized finance, stablecoins and governance tokens further complicates the issue, as functionally diverse assets transmit and absorb shocks heterogeneously [2]. Lower-tail dependence also intensifies during stressed conditions and varies across mainstream cryptocurrencies, DeFi tokens and stablecoins [3]. This raises the core question of whether a single tail-risk model can capture these responses and deliver reliable downside-risk forecasts during extreme events.

Existing research provides two main methodological foundations for this analysis. GARCH-type models are widely used to capture volatility persistence and have been applied to several

DOI: 10.54254/2754-1169/2026.36621

cryptocurrency return series [4]. The EVT-GARCH approach extends conditional volatility modelling by applying Extreme Value Theory to standardized residuals for separate tail estimation [5]. CAViaR directly models conditional VaR quantiles without specifying the full return distribution [6]. Recent studies show high-frequency crypto data contain complex tail behavior requiring flexible risk methods [7]. Broad comparisons find non-Gaussian assumptions suit crypto returns better, with model rankings varying across assets [8]. Residual dependence may remain after GARCH filtering, making robustness checks critical for tail estimation [9]. However, direct comparison of the two models across mixed asset types under the FTX event remains limited.

This study addresses this gap using Bitcoin, Ethereud number of violations is approximately 12 as the aligned sample contains 1,199 daily observations. Both models produce violm, Uniswap and Dai. It first examines distributional properties of aligned daily log returns, then estimates an EVT-GARCH model with an FTX event dummy, and finally compares its VaR forecasts with Asymmetric Slope CAViaR via 99% and 95% backtesting. The cross-asset design reveals risks obscured when the cryptocurrency market is treated as homogeneous. Empirical results confirm severe non-normality, FTX-related additional volatility, and acceptable but distinct forecasting performance across the two models. This study integrates volatility filtering, tail estimation and quantile modelling into a consistent crisis framework, supporting combined model use in stress testing and short-horizon risk monitoring.

2. Materials and methods

2.1. Data source and preprocessing

The empirical analysis uses daily OHLCV data from Top 50 Cryptocurrency Data, which contains core trading fields for major crypto assets [10]. The sample selects Bitcoin, Ethereum, Uniswap, and Dai to represent different market functional categories. Raw price series are cleaned and aligned by calendar date, with the Kaggle symbol UNI7083 renamed UNI for consistency. The common sample starts on September 18, 2020 due to UNI's shortest listing history, and ends on December 31, 2023, yielding 1,199 daily log return observations. Using returns improves cross-asset comparability but does not eliminate non-normality and volatility clustering, so descriptive statistics and Jarque-Bera tests are conducted before model estimation.

Traditional normal GARCH-VaR serves as the conceptual benchmark in this study. It estimates time-varying conditional volatility and derives VaR values based on normal distribution quantiles, featuring transparent calculation logic and effective capture of volatility clustering. However, this approach tends to underestimate downside losses for return series with skewness and heavy-tailed properties. The empirical comparison focuses on EVT-GARCH and CAViaR, as both relax the restrictive distributional assumptions of the normal GARCH-VaR framework while maintaining dynamic risk measurement capabilities.

2.2. EVT-GARCH measurement framework

Cryptocurrency returns often contain persistent volatility and a small number of unusually large observations. EVT-GARCH addresses both features in two stages. A GARCH filter first estimates conditional volatility, after which Extreme Value Theory is applied to the lower tail of the standardized residuals. Separating the volatility process from the residual tail reduces dependence on a normal distribution and allows the final VaR estimate to respond to both current market conditions and extreme-loss behavior. The GARCH variance specification is written as Model (1):

DOI: 10.54254/2754-1169/2026.36621

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma D_t^{FTX} \quad (1)$$

σ_t^2 denotes conditional variance, ε_{t-1}^2 is the previous squared shock, and σ_{t-1}^2 is lagged conditional variance. D_t with the superscript FTX equals one from November 2 to November 15, 2022, and zero otherwise. The coefficient γ measures the additional variance associated with the event window after accounting for normal GARCH persistence. Since the output does not include standard errors, the coefficient is interpreted through its sign and economic magnitude rather than formal statistical significance. After GARCH estimation, the Peak Over Threshold method fits extreme standardized losses to a Generalized Pareto Distribution, whose shape and scale parameters describe tail behavior and exceedance dispersion. The estimated tail quantile is combined with conditional volatility to derive dynamic VaR, granting EVT-GARCH an edge in severe-loss estimation, though fixed threshold use may cause result sensitivity.

2.3. CAViaR model specification

CAViaR directly models the conditional VaR quantile and does not require a complete distribution for returns. This feature is useful when cryptocurrency returns are non-normal and when recent shocks alter downside risk more quickly than a full volatility model can adjust. CAViaR also retains persistence through lagged VaR, while its asymmetric form allows negative and positive returns to affect risk differently. Two classical specifications are considered. The Symmetric Absolute Value model updates current VaR using lagged VaR and the absolute size of the previous return. The Asymmetric Slope model separates positive and negative lagged returns and is used in the final comparison because negative news, forced selling, and liquidity withdrawal can produce stronger downside effects. EVT-GARCH therefore estimates tail risk through volatility filtering and residual-tail modeling, while CAViaR estimates the risk quantile directly. Their forecasting performance is evaluated at the 99% confidence level and checked again at 95%.

VaR backtesting compares predicted risk limits with realized returns. At the 99% level, the expected violation rate is 1%. A model with too many violations understates downside risk, while a model with very few violations may be excessively conservative. The comparison reports expected and actual violations for BTC and ETH, the two benchmark assets with formal backtesting output. The Kupiec unconditional coverage test assesses whether the total violation frequency is consistent with the theoretical rate. The Christoffersen conditional coverage test also considers whether violations are independent over time. Both dimensions matter because a model may record an acceptable total number of violations while still allowing them to cluster during periods of market stress.

3. Empirical results

3.1. Descriptive statistics and normality tests

The distributional analysis establishes whether tail-sensitive models are justified. As shown in Table 1, all four-return series have kurtosis above the normal benchmark of 3, and every Jarque-Bera p-value is below 0.001. The evidence rejects normality and indicates that average volatility alone cannot describe the probability of extreme losses. UNI has the largest standard deviation at 0.0607, while BTC and ETH record 0.0335 and 0.0437. DAI appears stable in ordinary periods because its standard deviation is only 0.0018, yet its kurtosis reaches 62.06, and its skewness is negative. This combination reflects a distribution concentrated near zero with a small number of unusually large

DOI: 10.54254/2754-1169/2026.36621

negative observations. During the FTX collapse window, liquidity pressure and temporary de-pegging movements produced returns that were large relative to DAI's normal range. The result shows why a low average volatility measure should not be interpreted as an absence of tail risk.

Table 1. Descriptive statistics of daily log returns

Assets	Observations	Mean	Std.Dev	Skewness	Kurtosis	Jarque-Bera (J-B) Stat	p-value
BTC	1199	0.0011	0.0335	-0.2000	6.49	617.43	<0.001
ETH	1199	0.0015	0.0437	-0.3998	8.31	1438.95	<0.001
UNI	1199	0.0000	0.0607	0.2831	8.61	1588.53	<0.001
DAI	1199	0.0000	0.0018	-0.8105	62.06	174377.0	<0.001

Figure 1 provides a visual check of the same argument by comparing empirical kernel density curves with a normal benchmark. The empirical distributions are more sharply concentrated around zero and retain more probability in the tails than the normal curve. It explains why normal GARCH-VaR may be insufficient. EVT and CAViaR are retained because both are designed to reduce the effect of restrictive distributional assumptions on downside-risk forecasts.

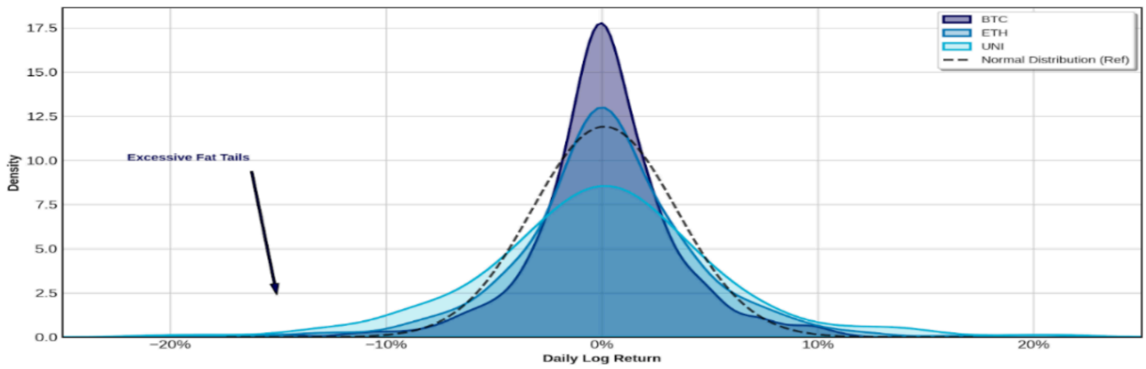

Figure 1. Empirical density of log returns vs. normal distribution (picture credit: original)

3.2. Return dynamics and parameter estimates around the FTX crisis

Figure 2 plots 30-day rolling volatility to show how the four assets behaved around the FTX event window. It contextualizes the event dummy coefficients within the long-term volatility trajectory instead of isolating the crisis as an independent indicator. BTC, ETH, and UNI experienced clear volatility increases around the crisis, while DAI recorded smaller but abnormal movements relative to its usual range. UNI remained the most volatile asset for much of the sample. The common movement around the event window is consistent with cross-asset stress, but the scale of the response differs substantially across asset types.

DOI: 10.54254/2754-1169/2026.36621

Figure 2. 30-Day rolling volatility (picture credit: original)

Table 2 reports the EVT-GARCH parameters and the GPD estimates. For POT estimation, the threshold is defined as the 95th percentile of standardized loss distribution, which corresponds to the lower 5th percentile of standardized returns. This rule retains the most severe downside observations while preserving enough exceedances for estimation, and the same threshold is used for all assets to maintain comparability. All FTX dummy coefficients are positive, indicating additional variance during the event window after lagged shocks and volatility are considered. The coefficient is largest for UNI at 0.000348, followed by ETH at 0.000308 and BTC at 0.000145. DAI has a much smaller coefficient because its return scale is narrower, but the positive sign remains consistent with event-related stress. The GPD results reveal different tail structures. DAI has the largest shape parameter at 0.4868, which is consistent with rare but severe deviations from its usual price range. BTC also has a positive shape parameter, while ETH is close to zero and UNI is slightly negative.

Table 2. Parameter estimates of the EVT-GARCH model

Parameters	BTC	ETH	UNI	DAI
ω (constant)	0.00023	0.000023	0.000040	0.000001
α (ARCH term)	0.0613	0.0748	0.0740	0.4791
β (GARCH term)	0.9183	0.9136	0.9166	0.5200
γ (FTX Dummy)	0.000145	0.000308	0.000348	0.000008
ε (GPD Shape)	0.1051	0.0047	-0.0063	0.4868
β (GPD Scale)	0.6781	0.7287	0.7146	0.3531

3.3. Cross-evaluation of tail risk measurement

Table 3 compares 99% VaR backtesting results for EVT-GARCH and Asymmetric Slope CAViaR. The expected number of violations is approximately 12 as the aligned sample contains 1,199 daily observations. Both models produce violation counts close to the theoretical value, but CAViaR exhibits the smaller full-sample bias. For BTC, EVT-GARCH records 14 violations and CAViaR records 13. For ETH, the corresponding violation figures are 15 and 12, respectively. CAViaR differs from the expected count by one violation for BTC and none for ETH, whereas EVT-GARCH exceeds the expected count by two and three violations. A smoother VaR path does not guarantee better coverage: the higher EVT-GARCH counts indicate that realized losses crossed its risk limit

DOI: 10.54254/2754-1169/2026.36621

more often than expected. Nevertheless, all Kupiec and Christoffersen p-values are above 0.05, so neither correct coverage nor violation independence is rejected for either model.

Table 3. VaR backtesting results at 99% confidence level

Asset	Model	Expected Violations	Actual Violations	Kupiec POF (p-value)	Christofferson (p-value)
BTC	EVT-GARCH	12	14	0.548	0.682
	CAViaR (AS)	12	13	0.765	0.810
ETH	EVT-GARCH	12	15	0.352	0.514
	CAViaR (AS)	12	12	0.998	0.855

Two robustness checks are conducted. First, the FTX dummy window is extended from the primary 14-day period to the full month of November 2022. The coefficients for BTC, ETH, and UNI remain positive but fall from 0.000145, 0.000308, and 0.000348 to 0.000112, 0.000215, and 0.000280. The decline suggests that the strongest volatility effect was concentrated around the core collapse period and became diluted when calmer days were included. Second, backtesting is repeated at the 95% VaR level, where the expected number of violations is approximately 60. EVT-GARCH and CAViaR record 63 and 58 violations for BTC, and 65 and 61 for ETH. All Kupiec and Christoffersen p-values remain above 0.05. The two models therefore retain acceptable coverage at a less extreme threshold, while CAViaR continues to show the smaller violation-count deviation.

4. Conclusion

This study finds that Bitcoin, Ethereum, Uniswap, and Dai returns are strongly non-normal, with distinct combinations of volatility, skewness, and tail risk. The FTX collapse window is associated with additional volatility across all four assets, but the economic scale and tail structure differ by asset type. Dai illustrates why stablecoins require separate attention: ordinary volatility is low, yet rare de-pegging movements produce extreme kurtosis and a large GPD shape estimate. Both EVT-GARCH and Asymmetric Slope CAViaR provide acceptable VaR forecasts at the 99% and 95% levels. EVT-GARCH produces smoother risk paths, while CAViaR reacts faster to recent negative returns and records smaller violation-count bias for BTC and ETH. No single model is consistently superior, and complementary application is recommended for different risk management scenarios.

The main contribution is a consistent comparison of volatility-based tail modelling and direct quantile modelling across multiple crypto asset types during the same crisis event, addressing gaps in studies that examine asset groups or risk models separately. The results help regulators, exchanges, institutional investors and researchers distinguish persistent risk from rapidly changing quantile risk. Limitations include daily observation frequency, formal backtesting restricted to BTC and ETH, and a fixed 95th percentile POT threshold, with no identified violation dates during the FTX window. Future research can adopt intraday data, extend backtesting coverage, test dynamic thresholds, compare violation timing, and incorporate on-chain indicators to improve early-warning performance.

References

[1] Jalan, A., & Matkovskyy, R. (2023). Systemic risks in the cryptocurrency market: Evidence from the FTX collapse. *Finance Research Letters, 53*, 103670.

DOI: 10.54254/2754-1169/2026.36621

[2] Vidal-Tomás, D., Briola, A., & Aste, T. (2023). FTX's downfall and Binance's consolidation: The fragility of centralised digital finance. *Physica A: Statistical Mechanics and Its Applications, 625*, 129044.
[3] Naifar, N. (2025). Mapping systemic tail risk in crypto markets: DeFi, stablecoins, and infrastructure tokens. *Journal of Risk and Financial Management, 18*(6), 329.
[4] Chu, J., Chan, S., Nadarajah, S., et al. (2017). GARCH modelling of cryptocurrencies. *Journal of Risk and Financial Management, 10*(4), 17.
[5] McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance, 7*(3-4), 271-300.
[6] Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics, 22*(4), 367–381.
[7] Giannopoulos, K., Nekhili, R., & Christodoulou-Volos, C. (2024). Estimating tail risk in ultra-high-frequency cryptocurrency data. *International Journal of Financial Studies, 12*(4), 99.
[8] Barson, Z., & Owusu Junior, P. (2024). Tail risk modelling of cryptocurrencies, gold, non-fungible tokens, and stocks. *Research in Globalization, 8*, 100229.
[9] Christodoulou-Volos, C. (2025). Robust tail risk estimation in cryptocurrency markets: Addressing GARCH misspecification with block bootstrapping. *Risks, 13*(9), 166.
[10] Park, S., & Park, H. W. (2020). Diffusion of cryptocurrencies: Web traffic and social network attributes as indicators of cryptocurrency performance. *Quality & Quantity, 54*(1), 297–314.