

Volatility Modelling and Tail-Risk Measurement in Cryptocurrency Markets

Haotian Bai

University of London, Singapore, Singapore
hbai005@mymail.sim.edu.sg

Abstract. In global financial markets, cryptocurrencies have become an important subject of study in financial risk management due to their high volatility and uncertainty. This paper systematically assesses the volatility and tail risk of the cryptocurrency market based on daily price data for Bitcoin (BTC) and Ethereum (ETH) from 1 July 2020 to 1 July 2025. The research methodology combines conditional heteroskedasticity models (Conditional Heteroskedasticity Models), particularly the Generalised Autoregressive Conditional Heteroskedasticity (GARCH) model and its asymmetric extension, the Glosten-Jagannathan-Runkle Generalized Autoregressive Conditional Heteroskedasticity (GJR-GARCH) model, and further incorporates the Peak Over Threshold (POT) method from Extreme Value Theory (EVT). Empirical results indicate that at the 95% confidence level, both models effectively capture cryptocurrency price volatility. However, at the stricter 99% level, the GARCH-EVT hybrid model demonstrates greater robustness in tail risk prediction, significantly enhancing the precision of risk measurement. The study's conclusions provide robust empirical support for risk management and policy formulation in the digital asset sector.

Keywords: Cryptocurrency, Tail Risk, GARCH, Extreme Value Theory (EVT)

1. Introduction

Over the past decade, the cryptocurrency market has rapidly evolved from experimental projects into a trillion-dollar financial subsystem. Due to its 24/7 trading and highly speculative nature, major cryptocurrencies such as Bitcoin (BTC) and Ethereum (ETH) exhibit volatility far exceeding that of traditional assets like stocks and currencies. In high-frequency trading, extreme price fluctuations can result in significant losses in a short period and potentially trigger systemic risks. Accurately assessing the volatility and tail risks of cryptocurrencies has become a cutting-edge issue in financial regulation and risk management.

Several studies have provided quantitative evidence in papers. Zhong and Fu constructed a tail risk spillover network based on Quantile Vector Autoregression (QVAR). Their research found that tail risk spillovers between cryptocurrencies and the Chinese energy market were significantly higher in extreme conditions than in normal conditions with long-term spillovers dominating total spillovers. It indicating that risk propagation is cyclical and persistent [1]. Maghyereh and Ziadat used the Conditional Autoregressive Value at Risk (CAViAR) and Time-Varying Parameter Vector

Autoregression (TVP-VAR) models to calculate the tail risk connectivity index. At risk levels of 1% and 5%, the total connectivity index was approximately 65%–66%, it indicated that external innovation explained more than 60% of risk fluctuations. In March 2020, this index rose to over 95% due to the COVID-19 shock, before falling back to around 70% [2]. Trucíos and Taylor used the generalized autoregressive scoring (GAS) model to study four cryptocurrency portfolios and found that the GAS model captured strong dynamic risk dependencies, with VaR prediction accuracy on average about 5% higher than that of Dynamic Conditional Correlation Generalized Autoregressive Conditional Heteroskedasticity (DCC-GARCH) [3]. Song and Chen combined AR, EGARCH, and extreme value theory to propose an improved VaR model. Test results show that the new model significantly reduces the default rate, especially for high-risk assets such as Bitcoin and crude oil. More accurate VaR predictions are helpful for asset allocation and regulation [4].

However, existing studies mainly focus on data from 2017 and earlier or single methods, and lack research that uses the latest data and comprehensively evaluates GARCH and extreme value theory. To address this gap, this paper selects daily data on BTC and ETH from 1 July 2020 to 1 July 2025, utilises multiple GARCH structures combined with the peak over-threshold (POT) method to analyse value at risk (VaR) and expected shortfall (ES) at different confidence levels, and uses out-of-sample backtesting to evaluate the model's effectiveness, thereby providing new empirical evidence for digital asset risk management.

2. Method

2.1. Data

The data in this paper is sourced from Investing.com and covers the daily closing prices of BTC and ETH from 1 July 2020 to 1 July 2025, with a total of 1,826 observations [5]. The raw data includes opening prices, high prices, low prices, and trading volumes. This paper calculates the logarithmic returns using closing prices, with dates arranged in chronological order. Bitcoin and Ethereum historical price data are updated in real-time. The dataset employed in this research exhibits a high degree of reliability. Investing.com, an internationally prevalent financial information platform, provides real-time and historical market data consistent with that of mainstream data providers such as Bloomberg and Refinitiv. Furthermore, the dataset utilised in this study is publicly accessible, reproducible, and subject to continuous updating, thereby ensuring the transparency and verifiability of the empirical analysis.

2.2. Volatility modelling

This study selected three types of models: standard GARCH, GJR-GARCH with leverage effects, and exponential GARCH (EGARCH), and assumed that the innovation term followed a Student's *t*-distribution. The optimal structure was selected based on log-likelihood and information criteria. Compared with more complex models, the above three types of models can capture conditional heteroscedasticity and leverage effects and are easy to combine with extreme value theory for tail risk assessment.

To apply these models, this study first defines the return series. Let P_t be the closing price on day t . The logarithmic return is defined as $r_t = \ln(P_t) - \ln(P_{t-1})$. In this way, this paper specifies the GARCH family of models, whose standard form is as follows:

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

Where $\varepsilon_t = r_t - \mu_t$ is the residual, $\omega > 0$ is the long-term variance, α reflects the impact of recent shocks on volatility, and β measures the persistence of volatility. Since negative returns often have a stronger impact on future volatility than positive returns, this paper further adopts the GJR-GARCH model and adds a step function I_{t-1} to the conditional variance equation to construct the leverage effect term $\gamma I_{t-1} \varepsilon_{t-1}^2$. The exponential GARCH (EGARCH) model takes the logarithm of the conditional variance so that the parameters do not need to be positive. Its formula is

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \alpha \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - E|z_{t-1}| \right) + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \quad (2)$$

Where $z_t = \varepsilon_t / \sigma_t$ is the standardised residual. To reflect the fat tail distribution of returns, this study assumed that z_t follows a Student's t-distribution and estimate the degrees of freedom parameter Df. These heteroscedasticity models have been widely applied in cryptocurrency research.

2.3. Extreme value theory and tail risk

To estimate the tail risk of the return distribution, this study combines the POT method from extreme value theory. This method focuses on sample values $y = x - u$ that exceed a high threshold u , and assumes that the excess follows a generalized Pareto distribution (GPD), whose distribution function is

$$G_{\xi, \beta}(y) = 1 - (1 + \xi y / \beta)^{-1/\xi} \quad (3)$$

where ξ is the shape parameter controlling the tail thickness, and β is the scale parameter. When the sample size is N , the number of over-threshold observations is n , and the confidence level is $1 - p$, after estimating ξ and β using maximum likelihood estimation, the conditional VaR can be approximated as follows:

$$\text{VaR}_{1-p} = \mu_t + \sigma_t \left[u + \frac{\beta}{\xi} \left(\left(\frac{N}{n} p \right)^{-\xi} - 1 \right) \right] \quad (4)$$

Where μ_t and σ_t are the estimated conditional mean and standard deviation of the GARCH model, respectively. Compared with traditional historical simulation methods, the POT method uses extreme value distributions to improve the estimation accuracy of tail extreme events. This paper applies this method to the standardised residuals generated by the GJR-GARCH model and compares the VaR and ES performance of the pure GJR-GARCH model and the GARCH-EVT hybrid model at 95% and 99% confidence levels.

3. Research results

3.1. Model parameter estimation and diagnosis

To compare the fitting performance of different GARCH structures, this paper uses four models to fit the logarithmic return series of BTC and ETH: normal distribution (sGARCH-norm), Student's t-distribution (sGARCH-std), Student's t-distribution with leverage effect (GJR-GARCH-std), and exponential GARCH (eGARCH-std). Table 1 lists the parameter estimates and diagnostic indicators for each model.

Table 1. GARCH family model parameter estimation and diagnosis (BTC & ETH)

Model	BTC				ETH			
	sGARCH-norm	sGARCH-std	GJR-GARCH-std	eGARCH-std	sGARCH-norm	sGARCH-std	GJR-GARCH-std	eGARCH-std
Parameters								
μ	-0.00156	-0.00100	-0.00100	-0.00090	-0.00124	-0.00115	-0.00113	-0.00094
ω	0.00003	0.00001	0.00001	-0.02169	0.00002	0.00002	0.00002	-0.05766
α	0.0674	0.0571	0.0551	-0.0108	0.0569	0.0763	0.0838	-0.0071
β	0.900	0.942	0.943	0.997	0.931	0.923	0.919	0.991
Diagnostics								
BTC			ETH					
Model	sGARCH-norm	sGARCH-std	Model	sGARCH-norm	sGARCH-std	Model	sGARCH-norm	sGARCH-std
γ	NA	NA	0.00214	0.12316	NA	NA	-0.00786	0.14369
Df	NA	3.18	3.18	2.84	NA	3.65	3.63	3.53
LogLik	-4.17	-4.33	-4.33	-4.34	-3.65	-3.77	-3.77	-3.78
AIC	-4.15	-4.31	-4.30	-4.31	-3.63	-3.75	-3.75	-3.76
BIC	-4.17	-4.33	-4.33	-4.34	-3.65	-3.77	-3.77	-3.78

3.2. Out-of-sample VaR and ES tests

After determining that GJR-GARCH-std is the optimal volatility model, this paper further calculates VaR and ES at 95% and 99% confidence levels based on this model and the GARCH-EVT hybrid model and conducts backtesting on the external sample period. Table 2 lists the expected number of defaults, the actual number of defaults, and the statistical values of the Kupiec proportion error test and ES regression test.

Table 2. Overall out-of-sample backtesting results for VaR and ES

Assets	Confidence level	Models	Expected violations	Actual number of violations	Kupiec p-value	ES gamma estimation	ES regression test p-value
BTC	95%	GJR-GARCH-std	66.30	81	<0.001	0.069	0.124
	95%	GARCH-EVT	66.30	81	<0.001	0.069	0.124
	99%	GJR-GARCH-std	13.26	14	<0.001	-0.014	0.544
	99%	GARCH-EVT	13.26	14	<0.001	-0.014	0.544
ETH	95%	GJR-GARCH-std	66.30	81	<0.001	0.071	0.094
	95%	GARCH-EVT	66.30	81	<0.001	0.071	0.094
	99%	GJR-GARCH-std	13.26	15	<0.001	0.012	0.586
	99%	GARCH-EVT	13.26	15	<0.001	0.012	0.586

The results show that, at a 95% confidence level, there should theoretically be 66.30 defaults, but the actual number of defaults for BTC and ETH reached 81 respectively. At the 99% level, the expected number of defaults for the two assets is 13.26, but the actual number of defaults is 14 and 15, respectively. The p-values of the Kupiec test are all significantly less than 0.001, indicating that the model generally underestimates tail risk. In terms of ES regression testing, only BTC has a γ estimate of -0.014 at the 99% confidence level, with a p-value of 0.544, failing to reject the unbiasedness hypothesis. None of the other cases passed the significance test. This indicates that, regardless of whether extreme value theory is combined, the single GARCH model still has a bias in its description of extreme losses. It is worth noting that, in the sample of this study, the POT method did not significantly improve the number of defaults and test statistics.

4. Discussion

Model estimates and out-of-sample backtesting results indicate that the volatility characteristics of the cryptocurrency market are complex and difficult to predict. First, the GJR-GARCH-std model performed best in the parameter estimation stage, with its α , β , and γ reflecting significant volatility clustering and leverage effects. However, the β of the eGARCH model being close to 1 indicates extremely strong volatility persistence. The degrees of freedom of the student t-distribution ranging from 2.84 to 3.65 indicate that the distribution of BTC and ETH returns exhibits severe fat tails, which is consistent with the findings of Theodossiou et al. who tested the fat tails and kurtosis of Bitcoin returns using the GJR-GARCH model [6]. Secondly, out-of-sample testing shows that both the pure GJR-GARCH and the GARCH-EVT model combined with POT severely underestimate tail risk. At the 95% confidence level, the expected number of defaults is 66.30, while the actual number of defaults is 81. At a 99% confidence level, the expected number of defaults was 13.26, whereas the actual number of defaults was 14. Although the Kupiec test indicated that the model failed the

default rate test, the ES regression test did not reject the hypothesis of unbiasedness for BTC at this level.

The model's prediction bias reversed direction at different stages, further illustrating that a single-parameter GARCH model is unable to adapt to structural changes. In contrast to the results of Song and Chen's improved model, which effectively improves the accuracy of VaR prediction, the model used in this paper lacks sufficient capture of extreme losses [4]. This difference can also be partly attributed to differences in data frequency and simulation methods. For example, Dashti and Lian used high-frequency data and block self-regression historical simulation to capture microstructural noise, thereby significantly improving tail risk estimation [7].

A comparative literature review reveals that the findings of this study are consistent with Zhong and Fu's tail risk spillover network analysis, which indicates that tail risk spillover significantly increases under extreme conditions and that long-term spillover dominates [1]. The sample period of this study includes the bull market of 2021 and the bear market of 2022. The number of defaults in the out-of-sample data differs significantly between the two phases, reflecting cyclical and persistent characteristics. Maghyreh and Ziadat calculated the tail risk connectivity using TVP-VAR, which rose above 95% during the COVID-19 pandemic and subsequently fell back to 70% [2], further indicating that macroeconomic shocks led to a sharp increase in tail risk.

Similarly, the VaR and ES default counts in this study showed a significant increase during the peak of the pandemic. Trucíos and Taylor confirmed that the DCC-GARCH model improves VaR prediction performance by 5% compared to the GAS model [3], suggesting that introducing a dynamic scoring mechanism or a more flexible heteroskedastic structure may improve predictions. This also consistent with Sözen's findings on the differences in applicability between the Threshold Generalized Autoregressive Conditional Heteroskedasticity (TGARCH) and EGARCH models across different currencies. Additionally, Wu and Yueh employed the Lévy-GJR-GARCH model to examine jump and skewness effects in return distributions, it concluded that the choice of innovation distribution significantly impacts the predictive accuracy of VaR and ES [8]. Building upon this foundation, Ahelegbey and Giudici introduced the Extreme Downside Hedging (EDH) and Extreme Downside Correlation (EDC) metrics. Their findings revealed that BTC acts as a risk 'exporter' while Ethereum functions as a risk 'absorber'. Thus illuminating the directionality of risk contagion from a network perspective [9]. Furthermore, Zhou et al. constructed a multi-scale graph neural network framework for predicting cryptocurrency volatility and had discovered that deep learning models could significantly enhance forecasting accuracy [10]. Overall, these studies demonstrate that both advanced econometric models and emerging machine learning approaches play a significant role in improving cryptocurrency volatility modelling and tail risk assessment.

5. Conclusion

This study aims to assess the effectiveness of the GJR-GARCH-std and GJR-GARCH-EVT models in predicting tail risk for Bitcoin and Ethereum by comparing them.

Contrary to theoretical expectations, the empirical findings of this study indicated that the advanced risk models examined exhibit significant limitations in cryptocurrency markets. The core conclusion of the study is that, during the backtesting period, neither model provided reliable risk forecasting, primarily manifested as a systematic underestimation of Value at Risk (VaR). Furthermore, research has found that even when incorporating Extreme Value Theory (EVT) to adjust for tail risks, the predictive performance of the model did not see any substantial improvement. Subsequent phased backtesting further revealed that the model's failure mode exhibited instability. Its predictive bias oscillating between underestimating and overestimating risk

across different periods. The findings reveal the difficulties in directly applying mainstream risk models, such as GARCH-EVT, to cryptocurrency risk management. The contribution of this study lies in empirically delineating the failure boundaries of such models in emerging markets. This cautionary conclusion underscores that rigorous, multi-dimensional backtesting validation is indispensable before deploying any risk model in practical application. Future research directions should include models capable of capturing structural breaks in the market, such as Markov-switching GARCH models, as well as exploring dynamic threshold selection methods.

References

- [1] Zhong, R., & Fu, X. (2025). Tail risk spillover effect between the cryptocurrency and China's energy market. *Entropy*, 27(7), 704.
- [2] Maghyreh, A., & Ziadat, S. A. (2024). Pattern and determinants of tailrisk transmission between cryptocurrency markets: New evidence. *Financial Innovation*, 10, 77.
- [3] Trucíos, C., & Taylor, J. W. (2023). Modelling and forecasting risk dependence and portfolio VaR for cryptocurrencies. *Empirical Economics*, 65, 1011–1043.
- [4] Song, Y., & Chen, R. (2025). Volatility modelling and tail risk estimation of financial assets: Evidence from gold, oil, bitcoin, and stocks. *Risks*, 13(3), 57.
- [5] Investing.com. (2025, July 1). Bitcoin and Ethereum historical data. Retrieved July 21, 2025, from <https://www.investing.com/crypto/bitcoin/historical-data> and <https://www.investing.com/crypto/ethereum/historical-data>
- [6] Theodossiou, P., Ellina, P., & Savva, C. S. (2022). Stochastic properties and pricing of Bitcoin using a GJRGARCH model with conditional skewness and kurtosis components. *Review of Quantitative Finance and Accounting*, 59(2), 695–716.
- [7] Dashti, Z., & Lian, B. (2024). Estimating tail risk in ultrahighfrequency cryptocurrency data. *International Journal of Financial Studies*, 12, 37.
- [8] Wu, H. T., & Yueh, M. L. (2025). Cryptocurrency risk management using Lévy processes and timevarying volatility. *Review of Quantitative Finance and Accounting*. Advance online publication.
- [9] Ahelegbey, D. F., & Giudici, P. (2020). Tail risk measurement in cryptoasset markets. *SSRN Electronic Journal*.
- [10] Zhou, Y., Xie, C., Wang, G. J., Gong, J., & Zhu, Y. (2025). Forecasting cryptocurrency volatility: A novel framework based on the evolving multiscale graph neural network. *Financial Innovation*, 11, 87.