

MASTER

MATHEMATICAL FINANCE

Master's Final Work

Dissertation

Sentiment and Volatility in Cryptocurrency Markets:
An Empirical Investigation Using GARCH-X, EVT, and Regime-
Switching Models

HAOTIAN LUO

Jury:

President: Prof. João Paulo Vicente Janela

Supervisor: Prof. Carlos Manuel Jorge da Costa

Member: Prof. Flávio Alexandre Costa Romão

June - 2025



MASTER

MATHEMATICAL FINANCE

Master's Final Work

Dissertation

Sentiment and Volatility in Cryptocurrency Markets:
An Empirical Investigation Using GARCH-X, EVT, and Regime-
Switching Models

HAOTIAN LUO

Supervisor

Prof. Carlos J. Costa

This document was created solely for the pursuit of a master's degree

June - 2025

List of abbreviations

AR – Autoregressive.

AIC – Akaike Information Criterion.

AICc – Corrected Akaike Information Criterion.

BTC – Bitcoin.

DOGE – Dogecoin.

EGARCH – Exponential Generalized Autoregressive Conditional Heteroskedasticity.

ES – Expected Shortfall.

ETH – Ethereum.

EVT – Extreme Value Theory.

GARCH – Generalized Autoregressive Conditional Heteroskedasticity.

GARCH-X – GARCH model with exogenous variables (e.g., sentiment).

GPD – Generalized Pareto Distribution.

HMM – Hidden Markov Model.

LINK – Chainlink.

MFW - Master's Final Work

MS-GARCH – Markov Switching GARCH model.

POT – Peaks Over Threshold.

QQ-plot – Quantile-Quantile plot.

RMSE – Root Mean Squared Error.

σ^2 – Conditional variance.

VaR – Value at Risk.

ξ – GPD shape parameter (tail index).

β – GPD scale parameter.

ABSTRACT

Volatility modeling plays a key role in understanding and managing financial risk, particularly in high-frequency and sentiment-driven markets such as cryptocurrency. However, traditional models often struggle to capture extreme fluctuations caused by sudden shifts in investor behavior. This study investigates whether public sentiment data obtained from platforms like LunarCrush and Google Trends can improve the forecasting of volatility and tail risk in crypto assets. To verify this, we apply a set of advanced time-series models to hourly price data for four major cryptocurrencies (BTC, ETH, DOGE, LINK) for the period 2020 to 2025.

The modeling framework integrates multiple layers, including GARCH and EGARCH variants with external sentiment regressors, regime-switching volatility via Markov models, and tail modeling via Generalized Pareto Distribution. Model performance is assessed both in terms of volatility forecast accuracy and risk coverage metrics such as Value-at-Risk (VaR) and Expected Shortfall (ES). This article pays particular attention to changes in distribution behavior during panic and performs Monte Carlo simulations to assess forward looking tail risk.

The results show that for both ETH and DOGE, the sentiment-enhanced GARCH model outperforms the standard model, especially during periods of heightened sentiment volatility. The regime-switching model shows that negative sentiment significantly increases the probability of entering a high-risk state, while the tail model suggests that once in such a state, the distribution of returns becomes quite heavy. DOGE exhibits the most severe tail risk amplification, while BTC and ETH show a more stable, but still significant, variation. In sum, this work provides evidence that sentiment signals not only predict short-term volatility, but also effectively capture the structural changes that drive extreme downside risk in cryptocurrency markets.

LIST OF FIGURES

Figure 1 – Distribution of absolute returns — event vs control windows	20
Figure 2 - Regime Probability by Sentiment Deciles	24
Figure 3 - Regime-Specific 99% ES Under Normal and Panic Regimes	26
Figure 4 - Log–Log Tail Survival Curves in Normal vs. Panic Regimes	27
Figure 5 - Low-Volatility Regime Residuals: Time Series and Distributional Properties	29
Figure 6 - High-Volatility Regime Residuals and QQ-Plots	30
Figure 7 - Simulated vs. Realized Conditional Volatility	31
Figure 8 - Results under Simulated Panic	34
Figure 9 - Residual Diagnostics for LINK: Low- and High-Volatility Regimes	44
Figure 10 - Residual Diagnostics for ETH: Low- and High-Volatility Regimes	45
Figure 11 - Residual Diagnostics for DOGE: Low- and High-Volatility Regimes	47
Figure 12 - Residual Diagnostics for BTC: Low- and High-Volatility Regimes	48

LIST OF TABLES

Table I - Comparision Of Return Distributions Between Event And Control Windows During Sentiment Shocks	21
Table II - Estimated EGARCH-X MODEL Coefficients And Model Performance	22
Table III - VAR Violation Tests Before And After Inclusing Sentiment In GARCH-X.....	22
Table IV - Effect of Fear on Panic-Regime Probability	25
Table V - Panic vs. Calm: Probability of a Top-10% Tail Event.....	25
Table VI - Tail Risk Estimates Using GPD Under Normal And Panic Regimes.....	27
Table VII - Monte Carlo Risk Estimates under Panic Regime.....	35

Contents

1. Introduction.....	1
2. Literature Review.....	2
2.1 Classical and Asymmetric Volatility Models	2
2.2 GARCH-X Models: Integrating Exogenous Information	3
2.3 Regime-Switching Models and Volatility States	4
2.4 Extreme Value Theory (EVT) and Tail Risk Estimation	4
2.5 Local Volatility Models: A Functional Alternative to Stochastic Volatility (Future work) ...	4
2.6 Summary and Research Gap.....	5
3. Methodology	6
3.1 Sentiment Data.....	6
3.2 Asset Data.....	6
3.3 Data Preparation	6
3.4.1 Modeling - Overview	7
3.4.2 Modelling Strategy	8
3.4.3 Statistical Adequacy	10
3.4.3.1 Information Criteria and Likelihood Ratio	10
3.4.3.2 Residual Whiteness.....	10
3.4.3.3 Remaining Conditional Heteroskedasticity	11
3.4.3.4 Sign-Bias (Asymmetry) Test	11
3.4.4 Forecast Accuracy.....	11
3.4.4.1 Loss Functions	12
3.4.4.2 Diebold–Mariano Test.....	12
3.4.5 Risk Adequacy	12
3.4.5.1 VaR Back-testing.....	13
3.4.5.2 Independence and Conditional Coverage	13
3.4.5.3 McNemar Test (Model-vs-Model).....	13
3.4.6 Tail-Fit Diagnostics	14
3.4.6.1 Anderson–Darling Test for GPD	14
3.4.6.2 Likelihood-Ratio for Panic vs. Calm	14
3.5. Deployment and Operational Integration.....	15

3.5.1 Data Pipeline.....	15
3.5.2 Model Refresh Cycle	15
3.5.3 Dissemination of Forecasts	16
3.5.4 Model Governance and Oversight	17
3.5.5 Quantitative-Trading Applications	17
4. Empirical Results	19
4.1 How Return Distributions Change Around Sentiment Shocks	20
4.2 Baseline Models: GARCH.....	21
4.2.1 Return-and-Sentiment Modelling Performance	21
4.2.2 VaR Back-testing Results	22
4.3 Markov-Switching GARCH model	23
4.4 Tail Behavior under Regime-Specific GPD.....	26
4.5 Residual Diagnostics and Regime-Specific Goodness-of-Fit	28
4.5.1 Methodology and Theoretical Background	28
4.5.2 Low- σ Regime Residuals: Stable and Symmetric	29
4.5.3 High- σ Regime Residuals: Tail Deviations and Heavy Extremes	30
4.5.4 Volatility Simulation: Realized vs Predicted	31
4.5.5 Summary and Implications	31
4.6 Volatility Paths and Forward-Looking Tail Risk under Regime-Specific GPDs	32
4.7 Comparative Insights Across Models	35
5. Discussion.....	37
6. Conclusion.....	38
7. Future Research.....	39
REFERENCES	40
APPENDICES	43
Appendix A: Data and Results.....	43
Appendix B: Programming.....	48

1. Introduction

In the past decade, cryptocurrencies such as Bitcoin (BTC), Ethereum (ETH), and a growing number of alternative coins have gained a great deal of attention from investors around the world. Cryptocurrencies, which started as a technological experiment, have become a major player in the financial markets, with players ranging from retail investors to large institutions and even governments down the road. One thing that makes the cryptocurrency market stand out is how quickly its price changes, often for reasons not directly related to economic data or company performance. (Dyhrberg, 2016, Baur & Dimpfl, 2018, Costa, 2024).

Unlike traditional markets, cryptocurrency prices often react strongly to online activity, such as social media discussions, influencer posts, or trending news. In some cases, a single tweet can move the market. These price movements are not always easy to explain using standard financial models, which typically assume that volatility changes slowly and predictably (Kristoufek, 2013, Mai et al., 2018, Aparicio et al., 2022). However, in crypto markets, volatility tends to spike suddenly and behave very differently during times of fear or excitement (Corbet et al., 2020, Gkillas & Longin, 2020).

Some models like Black-Scholes (Black & Scholes, 1973) or the basic GARCH model (Cont, 2001, Taleb, 2020) are commonly used to estimate risk and price volatility. But in the crypto world, these models often underestimate how large and frequent extreme changes can be. To address this, more advanced approaches—such as GARCH extensions, regime-switching models, and jump models—have been developed over time (Duffie et al., 2000, Haas et al., 2004, Heston, 1993). These allow for more flexible reactions to market shocks. Still, most of these models rely only on historical price data, and don't consider how public mood or investor sentiment might play a role.

Thanks to platforms like LunarCrush and Santiment, it is now possible to track and quantify how people feel about cryptocurrencies in real time. These platforms use data from Twitter, Reddit, and other sources to produce sentiment indicators, such as levels of fear or hype (García & Schweitzer, 2015, Smales, 2022). In this work, I made use of Python in notebook to collect and process high frequency sentiment data, focusing specifically on panic indicators. This data was then combined with traditional financial models to better understand how market emotions affect volatility and extreme risk.

This MFW aims to study how sentiment data collected from social media can improve the way we model volatility and risk in crypto markets. I choose four cryptocurrencies for this analysis which are Bitcoin (BTC), Ethereum (ETH), Dogecoin (DOGE), and Chainlink (LINK). I use Google Colab as the main analysis platform and the process are performed in my github(<https://github.com/LuoToby/Markov-Switch-Garch-in-sentiment-effect>). I applied a series of time series models such as GARCH-X, EGARCH-X, and Markov switching GARCH and tested whether adding sentiment variables helped improve the accuracy of risk forecasts. In addition, I used Extreme Value Theory (EVT) to focus on the behavior of extreme returns and assess how well the models captured tail risks.

The main goal of this study is not to predict prices, but to understand whether changes in public sentiment can help explain changes in volatility and the risk of large losses. By combining emotional data with traditional financial techniques, this research tries to give a clearer picture of how crypto markets behave, especially during periods of high stress.

2. Literature Review

2.1 Classical and Asymmetric Volatility Models

Some early models of financial risk, like the Black-Scholes-Merton framework (Black & Scholes, 1973) were built on the idea that market volatility stays constant over time. But in practice, especially when looking at equity markets or high-frequency financial data, this assumption doesn't match reality.

Empirical studies (Cont, 2001) show that market returns often display sharp spikes, heavy tails, and periods where volatility becomes clustered. These patterns suggest that markets behave much more unpredictably than early models assumed.

To address this, Engle introduced the concept of Autoregressive Conditional Heteroskedasticity (ARCH) (Engle, 1982), which made it possible for volatility to change over time depending on past price movements. This idea was later extended by Bollerslev into the more flexible GARCH (p, q) model (Bollerslev, 1986), and the conditional variance σ_t^2 is defined as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 \quad (2.1)$$

where: ϵ_{t-i}^2 are past squared residuals ("news shocks"), σ_{t-j}^2 are past conditional variances, α_i, β_j are model parameters, with constraints $\alpha_i, \beta_j \geq 0, \alpha_0 > 0$.

While GARCH models have been widely used for capturing volatility clustering, one of their main limitations is that they treat positive and negative shocks in the same way. Markets often react more strongly to bad news than to good news. To better capture this asymmetry, the Exponential GARCH (EGARCH) model was introduced (Nelson, 1991). Unlike standard GARCH, EGARCH models the logarithm of the conditional variance, which allows it to reflect different volatility responses depending on the sign of the return. After using a log transformation, formula (2.1) becomes:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \beta_i \log \sigma_{t-i}^2 + \sum_{j=1}^q \gamma_j \left(\frac{\epsilon_{t-j}}{\sigma_{t-j}} \right) + \theta_j \left(\left| \frac{\epsilon_{t-j}}{\sigma_{t-j}} \right| - E \left| \frac{\epsilon_{t-j}}{\sigma_{t-j}} \right| \right) \quad (2.1.1)$$

This formulation allows the model to account for leverage effects, where negative shocks increase future volatility more than positive shocks of the same magnitude.

2.2 GARCH-X Models: Integrating Exogenous Information

In cryptocurrency markets, investor sentiment is not only an abstract idea but also often generated directly from online platforms, which make it both measurable and highly relevant. Some websites like LunarCrush, Santiment, and The TIE collect and summarize social activity from sources like Reddit, Twitter, Google Trends, and GitHub. From this data, they produce real time sentiment metrics like Galaxy Score, AltRank, and volume-weighted sentiment indexes (Kristoufek, 2013, García & Schweitzer, 2015, Mai et al., 2018). Several studies have found that these indicators are closely linked to short-term price movements and sudden spikes in volatility (Smales, 2022, Corbet et al., 2020).

To incorporate external variables such as sentiment into volatility modeling, we can extend standard GARCH models into GARCH-X frameworks. The GARCH-X model introduces an exogenous regressor x_t for example like sentiment's data, which modifying the conditional variance equation as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \delta x_{t-1} \quad (2.2)$$

where δ captures the direct influence of the sentiment variable x_t on volatility. Conrad and Loch applied this framework using macroeconomic uncertainty indices as exogenous inputs (Conrad & Loch, 2015).

In crypto markets, (Dyhrberg, 2016) finds that including Reddit or tweet-based sentiment signals improves both in sample fit and out of sample Value-at-Risk (VaR) forecasting accuracy. When combined with EGARCH, we obtain the EGARCH-X model:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \beta_i \log \sigma_{t-i}^2 + \gamma \left(\frac{\epsilon_{t-1}}{\sigma_{t-1}} \right) + \theta \left(\left| \frac{\epsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \delta x_{t-1} \quad (2.2.2)$$

2.3 Regime-Switching Models and Volatility States

To capture the fact that markets may alternate between high and low volatility regimes, Markov switching GARCH (MS-GARCH) models were developed (Hamilton, 1989). These models assume that the parameters of the GARCH process vary depending on a hidden, discrete regime variable $S_t \in \{1, 2, \dots, K\}$, which evolves according to a first order Markov chain.

In the simplest two-regime MS-GARCH (1,1) model:

$$\sigma_t^2 = \alpha_0^{(S_t)} + \alpha_1^{(S_t)} \epsilon_{t-1}^2 + \beta_1^{(S_t)} \sigma_{t-1}^2 \quad (2.3)$$

The transition probabilities $P_{ij} = P(S_t = j | S_{t-1} = i)$ define the likelihood of moving between regimes. These models are especially useful in capturing sentiment-triggered regime switches, where panic or euphoria shifts market participants between "normal" and "crisis" states (Baruník & Křehlík, 2018).

2.4 Extreme Value Theory (EVT) and Tail Risk Estimation

While volatility models explain average risk, extreme value theory (EVT) is essential for quantifying tail risks, for example like those associated with crashes, large losses, or VaR exceedances (McNeil & Frey, 2000, Gkillas & Longin, 2020). The Peaks over Threshold (POT) method fits a Generalized Pareto Distribution (GPD) to the excesses over a high threshold u :

$$P(X > x | X > u) = \left(1 + \xi \frac{x-u}{\beta} \right)^{-1/\xi}, \quad x > u \quad (2.4)$$

where: ξ is the shape parameter which controls the heaviness of the tail, β is the scale parameter. The *Higher* ξ implies fatter tails and higher risk. In this study, EVT is used to model extreme losses in crypto returns, especially during periods of extreme sentiment (Gkillas & Longin, 2020).

2.5 Local Volatility Models: A Functional Alternative to Stochastic Volatility

(Future work)

This thesis mostly treats volatility as a random, time-varying process, like in GARCH or regime-switching models. But another common approach sees volatility as a fixed function of price and time. This is called local volatility, where variance is written as $\sigma(S, t)$ and modeled with functional approximations.

Although this method is most applied in option pricing contexts, its structure provides useful insight into how volatility might depend systematically on both asset price levels and time horizons. One of the early and widely discussed formulations, introduced by (Hull & White, 1987), and (Daglish et al., 2007), and these models' local volatility are described as:

$$\sigma(S, t) = a_0 + a_1 \ln(S/S_0) + a_2 \ln^2(S/S_0) + a_3(T - t) + a_4(T - t)^2 + a_5 \ln(S/S_0)(T - t) \quad (2.5)$$

This equation captures how volatility can shift with changes in log-moneyness (price relative to a base level S_0) and time to maturity. Alternative forms normalize the price-time interaction, such as:

$$\sigma(S, t) = b_0 + b_1 \frac{\ln(S/S_0)}{\sqrt{T - t}} + b_2 \left(\frac{\ln(S/S_0)}{\sqrt{T - t}} \right)^2 + b_3 \left(\frac{\ln(S/S_0)}{\sqrt{T - t}} \right)^3 \quad (2.6)$$

More generalized versions allow even more flexible fitting across market conditions, especially for modeling volatility skew or smile:

$$\sigma(S, t) = c_0 + c_1 \left(\frac{\ln(S/S_0)}{(T - t)^{d_0}} \right) + c_2 \left(\frac{\ln(S/S_0)}{(T - t)^{d_0}} \right)^2 + c_3 \left(\frac{\ln(S/S_0)}{(T - t)^{d_0}} \right)^3 \quad (2.7)$$

Finally, for skew and smile modeling, nonlinear functional approximations are used:

$$\sigma(S, t) = \sigma_{\text{ATM}}(t) + \sigma_{\text{skew}}(t) \tanh(\gamma_{\text{skew}}(t) \ln(S/S_0) - \theta_{\text{skew}}(t)) \quad (2.8)$$

$$\sigma(S, t) = \sigma_{\text{smile}}(t) \left(1 - \text{sech}(\gamma_{\text{smile}}(t) \ln(S/S_0) - \theta_{\text{smile}}(t)) \right) \quad (2.9)$$

Although this thesis does not implement local volatility models directly, understanding these formulations helps frame the broader landscape of volatility modeling. They highlight how volatility can be estimated using price and time-based functions an approach that differs from the stochastic and sentiment driven models used in this study.

2.6 Summary and Research Gap

The existing literature shows significant progress in modeling volatility via GARCH extensions, incorporating behavioral sentiment variables and modeling regime switches. However, a unified framework that integrates sentiment, volatility states, and tail risk modeling under one system is rare, especially in the crypto context. This thesis contributes by jointly estimating GARCH-X, EVT, and MS-GARCH models with real-time sentiment data, tested across four major cryptocurrencies.

3. Methodology

3.1 Sentiment Data

To explore the link between investor sentiment and market volatility, this study focuses on high-frequency behavioral data paired with daily log returns from January 1, 2020, to June 3, 2025. While platforms like Twitter, Reddit, or Google Trends provide a wealth of raw information, working directly with this data can be quite difficult due to API restrictions, messy formats, and inconsistent timing. Instead of manually collecting and cleaning data from multiple platforms, we used LunarCrush, a platform that aggregates crypto-related sentiment into clean, ready-to-use time series.

Through LunarCrush’s API, we collected hourly data on a range of indicators: Galaxy Score (which combines social engagement and market activity), Bullish/Bearish sentiment scores (counting positive and negative posts), Reddit Scores, Social Volume, and OHLCV market data. Using Python, we pulled and cleaned the data, converted timestamps to UTC, and structured the results into pandas DataFrames.

For most modeling purposes, especially GARCH and EVT, we used daily aggregated values. But for more detailed analysis, such as regime-switching or intraday forecasting, we preserved the original hourly frequency.

3.2 Asset Data

This study focuses on four cryptocurrencies, chosen not just for their popularity but also for how differently they tend to react to market sentiment. Bitcoin (BTC), as the most widely recognized and institutionally adopted crypto asset, serves as a natural benchmark. Ethereum (ETH), on the other hand, plays a major role in decentralized applications and smart contracts, often reacting to technical events like protocol upgrades. Dogecoin (DOGE) represents a unique case. Its meme driven and extremely sensitive to social media trends and viral moments, making it ideal for studying sentiment driven price swings. Lastly, Chainlink (LINK) is included as a more neutral asset with a smaller retail presence, helping us test how models behave when sentiment has less visible influence.

By working with this mix of assets ranging from highly institutional (BTC) to highly speculative and retail driven (DOGE) we’re able to look at sentimental effects from different angles. This setup allows for a more balanced and meaningful comparison when applying the same volatility and risk models across all four assets.

3.3 Data Preparation

Once the data was retrieved, several transformation steps were applied to prepare

it for modeling. First, we constructed key input features for each asset. Then we collect continuous sentiment variables which include Galaxy Score, Bullish, Bearish, and Reddit metrics. After that we standardized this data by using z-scores to remove scale differences and ensure stationarity. This transformation helps avoid bias when combining features with different magnitudes:

$$z_{i,t} = \frac{x_{i,t} - \bar{x}_i}{\sigma_i}, \quad \text{for } i \in \{\text{galaxy, bullish, bearish, reddit}\} \quad (3.1)$$

To account for short-term temporal effects, we added one hour, and two hour lagged versions of the sentiment indicators. These help the models detect possible causal relationships between past sentiment and present-day volatility:

$$\{\text{Bullish}_{t-1}, \text{Bearish}_{t-1}, \text{Galaxy}_{t-1}\} \rightarrow \sigma_t^2 \quad (3.2)$$

In addition, a binary panic dummy variable was created to flag unusually negative sentiment events. Specifically, a day was flagged as "panic" if the standardized bearish sentiment exceeded 2 or Reddit sentiment dropped below -1.5. This simple rule allowed us to construct regime indicators that were later used in both Markov-switching models and conditional volatility equations:

$$Dt_{\text{panic}} = \{1, \text{if } z_{\text{bearish}, t} > 2 \text{ or } 0, z_{\text{reddit}, t} < -1.5 \text{ otherwise}\} \quad (3.3)$$

Since the raw LunarCrush data was recorded at an hourly frequency, we needed to resample and aggregate it to daily frequency for most of the GARCH and EVT models. This was done using either the daily maximum (for peak sentiment signals) or volume-weighted averages (for smoother metrics). For example, the daily bullish sentiment was computed as the maximum hourly value within each trading day, while the Galaxy score was averaged across the day:

$$\text{Bullish}_{\text{daily}} = \max_{h \in H_t}(\text{Bullish}_h), \quad \text{Galaxy}_{\text{daily}} = \frac{1}{|H_t|} \sum_{h \in H_t} \text{Galaxy}_h \quad (3.4)$$

Where H_t denotes all hours within trading day t

Finally, all features including sentiment, market returns, and lagged indicators were merged into a single master DataFrame. This panel was indexed by using a multilevel structure which is based on asset symbols and timestamp (UTC), meanwhile, the missing values were conservatively forward filled to ensure continuity. The resulting dataset provided a clean, synchronized time series for each asset which are ready for modeling across GARCH-X, regime-switching, and tail risk estimation frameworks.

3.4.1 Modeling - Overview

This chapter follows a layered modeling strategy designed to gradually capture the

complexity of volatility and tail risk in cryptocurrency markets. Comparing with assuming a single model from the outset, we start with baseline volatility models GARCH and EGARCH to account for standard volatility clustering and asymmetry. When diagnostic checks reveal persistent autocorrelation, asymmetry, or unexplained conditional variance, we introduce additional complexity in stages. First, sentiment-augmented models such as GARCH-X and EGARCH-X incorporate lagged indicators like social volume or bearish signals. If volatility appears to switch between distinct states, we then apply a two regime Markov switching GARCH (MS-GARCH) framework, where the transition probabilities are influenced by observed sentiment patterns. Finally, to address residual fat tails that conventional models fail to capture, we employ Extreme Value Theory (EVT) on standardized residuals which allow us to model tail risk under both calm and panic regimes.

Each modeling layer is evaluated through a consistent selection process. Statistical adequacy is first tested using the Ljung-Box Q-statistic (Equation 4.23), ARCH-LM test (Equation 4.1.2), and sign-bias tests (Equation 4.2.5) to detect remaining structure in residuals. Models that pass these checks are compared using the Bayesian Information Criterion (BIC, Equation 4.2.1), and out-of-sample forecasting performance is assessed with Root Mean Squared Error RMSE (Equation 4.26) and the QLIKE loss function (Equation 4.27). For model comparison, Diebold–Mariano (Equation 4.28) is used to formally determine whether one model significantly outperforms another. When evaluating Value-at-Risk (VaR) performance, we apply the Kupiec LR test for coverage accuracy (Equation 4.29), rejecting models with poor risk coverage. If EVT is required, tail behavior is assessed using log–log survival diagnostics and Generalized Pareto tail fitting (Equation 4.31–Equation 4.32). This evidence-driven approach ensures that each additional layer of complexity is only introduced when clearly warranted by the data.

3.4.2 Modelling Strategy

Before estimating volatility models, we first filter the raw return series to remove any potential autocorrelation. This helps avoid spurious ARCH effects that might distort the volatility estimates (Box & Jenkins, 1976). To do this, we apply an ARMA(p,q) model to the return series r_t :

$$r_t = \mu + \sum_{i=1}^p \phi_i r_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + \hat{a}_t, \quad \hat{a}_t \sim \text{i.i.d. } (0, \sigma_a^2) \quad (4.10)$$

The order pair $(p, q) \in \{0, 1, 2\}^2$ is selected by the minimum Akaike information criterion (AIC). Residuals $\hat{a}_t$ are accepted only if the Ljung–Box statistic $Q(20)$ fails to reject the null of no serial correlation at the 5 % level. We henceforth set $\varepsilon_t = \hat{a}_t$ and estimate the standard GARCH (P, Q) process to capture volatility clustering (Engle, 1982, Bollerslev, 1986):

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^P \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^Q \beta_j \sigma_{t-j}^2, \quad \alpha_0 > 0, \alpha_i, \beta_j \geq 0 \quad (4.11)$$

For the model to be weakly stationary, the condition $\sum_i \alpha_i + \sum_j \beta_j < 1$ must be satisfied. We estimate the model using BFGS optimization and allow the innovations ε_t to follow Gaussian, Student-t, or skewed-t distributions. The corresponding log-likelihood function is:

$$\ell(\theta) = \sum_{t=1}^T [\log f(\varepsilon_t/\sigma_t; \nu) - \log \sigma_t] \quad (4.12)$$

We assess the model fit using Q(20) and ARCH-LM tests on the standardized residuals $\hat{z}_t = \varepsilon_t/\hat{\sigma}_t$. If the basic GARCH model fits well statistically but still fails to capture important variations driven by sentiment, we extend the model to a GARCH-X form by adding lagged sentiment variables to the variance equation (Conrad & Loch, 2015)

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \delta x_{t-1} \quad (4.13)$$

where x_{t-1} is the lagged z-scored sentiment indicator as defined in (Equation (3.1)). The joint null hypothesis $H_0: \delta = 0$ is tested by Wald statistics. If the hypothesis is rejected at 5 %, sentiment is retained; otherwise, we fall back to the baseline GARCH. Next, we account for asymmetric volatility effects, where negative shocks often increase volatility more than positive ones of the same magnitude (Nelson, 1991, Glosten et al., 1993). Then we can apply for EGARCH-X model:

$$\log \sigma_t^2 = \omega + \sum_{i=1}^p \beta_i \log \sigma_{t-i}^2 + \gamma \left(\frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right) + \theta \left(\left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right) + \delta x_{t-1} \quad (4.14)$$

In here $\gamma < 0$ indicates a leverage effect. If γ is insignificant, but the Engle–Ng sign-bias test still rejects symmetry, we use a GJR-GARCH-X model instead:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \gamma_1 \varepsilon_{t-1}^2 1_{\varepsilon_{t-1} < 0} + \beta_1 \sigma_{t-1}^2 + \delta x_{t-1} \quad (4.15)$$

If the residuals from these models still show signs of regime shifts or bimodal distributions, we switch to a Markov-Switching GARCH-X (MS-GARCH-X) model. This allows volatility parameters to depend on a hidden state $S_t \in \{1, 2\}$:

$$\sigma_t^{2, (S_t)} = \sigma_0^{(S_t)} + \sigma_1^{(S_t)} \varepsilon_{t-1}^2 + \beta_1^{(S_t)} \sigma_{t-1}^2 + \delta x_{t-1} \quad (4.16)$$

The latent state S_t follows a first-order Markov chain with transition matrix $[p_{ij}]_{i,j=1}^2$ and $p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)$, meanwhile this model is estimated via EM algorithm using Hamilton filtering. We perform a likelihood-ratio test against the GARCH-X benchmark; if the test statistic exceeds $\chi_{4, 0.95}^2 = 9.49$, we retain the regime-switching specification.

To estimate extreme risk, we apply Extreme Value Theory (EVT) to the standardized residuals $\hat{z}_t = \varepsilon_t / \hat{\sigma}_t$. We extract the tail values above a high threshold u (97th percentile):

$$X_t = |\hat{z}_t| - u > 0 \quad (4.17)$$

By the Pickands–Balkema–de Haan theorem, X_t follows a Generalised Pareto distribution $\text{GPD}(\xi, \beta)$:

$$F_X(x) = 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, \quad x > 0 \quad (4.18)$$

And the Parameters $(\hat{\xi}, \hat{\beta})$ are estimated by maximum likelihood

$$\ell(\xi, \beta) = -N \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^N \log \left(1 + \xi \frac{x_i}{\beta}\right) \quad (4.19)$$

From this we compute Tail risk measures derived as:

$$\text{VaR}_\alpha = u + \frac{\hat{\beta}}{\hat{\xi}} \left[(N/n(1 - \alpha))^{\hat{\xi}} - 1 \right], \quad \text{ES}_\alpha = \frac{\text{VaR}_\alpha}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi}u}{1 - \hat{\xi}} \quad (4.20)$$

We further test whether sentiment affects the heaviness of tails by splitting the sample using the panic dummy variable D_t^{panic} , and re-estimating equations (4.18)–(4.20). A likelihood-ratio test with 2 degrees of freedom is then used to test for significant differences in (ξ, β) across sentiment states.

3.4.3 Statistical Adequacy

3.4.3.1 Information Criteria and Likelihood Ratio

As introduced in [Equation \(4.12\)](#) all models are estimated by maximum likelihood. The fitness quality is assessed by:

$$\text{AIC} = -2l_{\max} + 2k, \quad \text{BIC} = -2l_{\max} + k \ln T \quad (4.21)$$

where $l_{\max}$ is the log-likelihood, k is the number of estimated parameters, and T is the sample size. In nested models, our improvements are tested by:

$$\text{LR} = 2(l_{\text{full}} - l_{\text{restricted}}) \sim \chi_{k_{\text{diff}}}^2 \quad (4.22)$$

3.4.3.2 Residual Whiteness

If the ARIMA pre-filter in [Equation \(4.10\)](#) was correctly specified, and the

conditional variance captured appropriately (via Eq. 4.11 or 4.13). standardized residuals should exhibit no serial correlation. We apply the Ljung–Box Q-test to standardized residual residuals $\hat{z}_t = \varepsilon_t / \hat{\sigma}_t$, and the test statistics is:

$$Q(m) = T(T+2) \sum_{h=1}^m \frac{\hat{\rho}_h^2}{T-h}, \quad \hat{\rho}_h = \text{ACF}(\hat{z}_t, h) \quad (4.23)$$

Which is Failing to reject $H_0(p > 0.05)$ at $m=20$ confirms no linear dependence remains. (absence of autocorrelation)

3.4.3.3 Remaining Conditional Heteroskedasticity

The GARCH framework [Equation \(4.11\)](#) assumes that variance depends on past squared shocks. If such dependence remains unexplained in residuals, then either misspecification or asymmetry is present: By using ARCH–LM test on $\hat{z}_t^2$:

$$\hat{z}_t^2 = c + \sum_{i=1}^m \eta_i \hat{z}_{t-i}^2 + u_t \quad (4.24)$$

Under H_0 (no remaining ARCH), $TR^2 \sim \chi_m^2$. A significant result indicates model misspecification and prompts estimation of an asymmetry-aware model.

3.4.3.4 Sign-Bias (Asymmetry) Test

If [Equation \(4.11\)](#) or [\(4.13\)](#) fails to account for asymmetrical volatility response, we expect sign-dependent residual patterns. we apply the Engle–Ng (1993) test (Engle & Ng, 1993):

$$\hat{z}_t^2 = c + \gamma_S S_{t-1} \hat{z}_{t-1} + \gamma_Q Q_{t-1} \hat{z}_{t-1}^2 + v_t \quad (4.25)$$

where $S_{t-1} = 1_{\{\hat{z}_{t-1} < 0\}}$, and $Q_{t-1} = 1_{\{\hat{z}_{t-1} < 0\}} \hat{z}_{t-1}^2$. Joint significance of γ_S and γ_Q and $(\gamma_S, \gamma_Q \neq 0)$ justifies the use of EGARCH or GJR-type models.

3.4.4 Forecast Accuracy

Even if residuals pass diagnostic tests under Section 3.5.1, it remains essential to assess how well the model predicts out-of-sample volatility. Forecast accuracy serves as a forward-looking test of model utility, particularly for financial decision-making tasks such as dynamic hedging or risk allocation.

Using a rolling window of 1,000 observations, each model (from [Equation 4.11](#) to

[4.15](#)) is re-estimated daily to issue one-step-ahead forecasts $\widehat{\sigma_{t-1|t}}^2$. These are compared against realized volatility proxies to quantify predictive loss.

3.4.4.1 Loss Functions

We use two widely accepted loss functions: RMSE captures scale-sensitive errors between predicted and realized conditional variances:

$$\text{RMSE} = \sqrt{\frac{1}{T_{\text{test}}} \sum_t (\widehat{\sigma_t^2} - \sigma_{t,\text{real}}^2)^2} \quad (4.26)$$

QLIKE (Hansen & Lunde, 2005), robust to variance scale distortion, measures relative forecast efficiency:

$$\text{QLIKE} = \frac{1}{T_{\text{test}}} \sum_t \left[\frac{\sigma_{t,\text{real}}^2}{\widehat{\sigma_t^2}} - \log \left(\frac{\sigma_{t,\text{real}}^2}{\widehat{\sigma_t^2}} \right) - 1 \right] \quad (4.27)$$

In the end, Models with lower RMSE and QLIKE are retained for tail risk evaluation

3.4.4.2 Diebold–Mariano Test

To formally test whether including exogenous sentiment variables (as in Equation [4.13–4.15](#)) improves volatility prediction, we apply the Diebold–Mariano (DM) test to compare loss differentials between models: For two competing models A and B with loss $L_{A,t}, L_{B,t}$ define $d_t = L_{A,t} - L_{B,t}$ Under where

$$H_0: E(d_t) = 0 ; \text{ DM} = \frac{\bar{d}}{\sqrt{(2\pi\widehat{f_d}(0))/T_{\text{test}}}} \sim \mathcal{N}(0,1) \quad (4.28)$$

where $\widehat{f_d}(0)$ is the Newey–West estimate of the spectral density at zero. Rejection of H_0 supports the statistical significance of the sentiment-augmented GARCH-X or EGARCH-X layers.

3.4.5 Risk Adequacy

Having a volatility model that fits well in sample is a good start, but for real-world risk management, what matters more is how well it performs at the extremes. In practice, financial institutions care less about minor prediction errors and more about whether the model can correctly anticipate large losses. That’s where Value-at-Risk (VaR) and

Expected Shortfall (ES) come in. In this section, I examine whether the final standardized residuals, defined as $\hat{z}_t = \varepsilon_t / \hat{\sigma}_t$, from the best-performing models, accurately capture the empirical distribution of extreme losses.

3.4.5.1 VaR Back-testing

The first step is to test whether the number of VaR breaches matches what we would expect statistically. For instance, at a 1% VaR level, we should see roughly 1% of observations fall below the predicted threshold. Let N be the number of observed violations over a test sample of T_{test} days. The observed breach rate is: $\hat{p} = N/T_{\text{test}}$, and the unconditional coverage test is:

$$\text{LR}_{\text{UC}} = -2 \ln[(1 - \alpha)^{T-N} \alpha^N] + 2 \ln[(1 - \hat{p})^{T-N} \hat{p}^N] \sim \chi_1^2 \quad (4.29)$$

If this test fails (for example, like too many or too few violations), the model may be underestimating or overestimating tail risk.

3.4.5.2 Independence and Conditional Coverage

Even if the overall number of breaches looks acceptable, we also want to know whether those violations are randomly distributed. If they tend to cluster, it suggests that the model is missing some form of regime switching or time-dependence (e.g., failure of Equation 4.15 or 4.16):

We construct a 2x2 transition matrix of the hit sequence (I_{t-1}, I_t) with n_{ij} which implies the number of transitions from state i to state j . The independence test is then:

$$\text{LR}_{\text{IND}} = -2 \ln \left[\frac{(1-\hat{p})^{n_{00}} \hat{p}^{n_{01}} (1-\hat{p})^{n_{10}} \hat{p}^{n_{11}}}{(1-\hat{p}_0)^{n_{00}} \hat{p}_0^{n_{01}} (1-\hat{p}_1)^{n_{10}} \hat{p}_1^{n_{11}}} \right] \sim \chi_1^2 \quad \text{with} \quad \hat{p}_j = \frac{n_{0j}}{n_{0j} + n_{1j}} \quad \text{and the total}$$

conditional coverage test is:

$$\text{LR}_{\text{CC}} = \text{LR}_{\text{UC}} + \text{LR}_{\text{IND}} \quad (4.30)$$

If LR_{CC} is significant, it means the model fails both in breach frequency and clustering, and further structure—such as regime-switching volatility—might be necessary.

3.4.5.3 McNemar Test (Model-vs-Model)

Sometimes two models may pass the same back testing metrics but behave differently in terms of which specific days they get wrong. To test whether two competing models produce statistically different VaR violation patterns, we use the McNemar test:

$$\chi_{\text{McN}}^2 = \frac{(n_{01} - n_{10})^2}{n_{01} + n_{10}} \sim \chi_1^2 \quad (4.31)$$

where n_{01} and n_{10} represent the number of times only one of the two models produces a breach. A significant result indicates that the models behave differently in practice—even if they seem similar on the surface.

3.4.6 Tail-Fit Diagnostics

The EVT layer introduced in Equation (4.17–4.20) aims to model standardized residual exceedances using Generalized Pareto Distributions (GPD). We now assess whether these tail assumptions are statistically valid.

3.4.6.1 Anderson–Darling Test for GPD

To test the goodness-of-fit of the GPD, we apply the Anderson–Darling test to the standardized exceedances $X_i = |\hat{z}_i| - u$ and the test statistic is:

$$A^2 = -N - \frac{1}{N} \sum_{i=1}^N \left[(2i-1) \left(\ln \hat{F}(X_{(i)}) + \ln \left(1 - \hat{F}(X_{(N+1-i)}) \right) \right) \right] \quad (4.32)$$

A high A^2 statistic suggests the GPD doesn't fit the data well—either because the threshold is set too low or the data's tail structure is not stationary.

3.4.6.2 Likelihood-Ratio for Panic vs. Calm

The results indicate that the GPD may not provide a good fit to the data in certain cases. This could be due to the chosen threshold being too low, or because the tail behavior of the data is not stable across time. To examine how sentiment-driven states affect the distribution of extremes, we use the panic indicator D_t^{panic} from Equation (3.3) to split the standardized residuals by regime and re-estimate the GPD parameters (ξ, β) within each state. To formally test whether the tail structure differs between regimes, we compute a likelihood-ratio test comparing a model with pooled tails (log-likelihood l_{joint}) to a regime-specific model with separate likelihoods l_{panic} and l_{calm} , so the test statistic is given by:

$$\Lambda = -2(l_{\text{joint}} - l_{\text{panic}} - l_{\text{calm}}) \sim \chi_2^2 \quad (4.33)$$

A significant test result suggests that tail risk is regime-dependent—implying that sentiment shifts not only influence volatility but also reshape the heaviness of the distribution tails. This provides further justification for including panic indicators in

both the volatility and risk modeling components.

3.5. Deployment and Operational Integration

While the previous sections focused on statistical validity, in practice a model is only as good as its ability to function reliably and consistently in a live environment. This chapter describes how the final volatility and tail risk model is developed and updated in real time. Following the "Deployment" phase of the CRISP-DM framework, the section shows how sentiment, price, and Google Trends data are piped into the system, how the model is reconstructed daily, and how the outputs can be used in risk monitoring. The entire setup is broken into six key components.

3.5.1 Data Pipeline

To enable real time forecasting, we built an automated data pipeline that brings together three key inputs: hourly price data, sentiment scores from LunarCrush, and search interest levels from Google Trends. These sources are collected through scheduled Python scripts that retrieve the latest values, then we format them consistently, after that we store everything in UTC. In the code, prices are saved as floats, count-type variables as integers, and all timestamps follow the ISO 8601 standard. All the daily versioned snapshots are also stored to ensure that any past forecast can be fully reconstructed using the exact inputs available at the time.

Occasionally, we observe irregular spikes or drops in the data, which are usually during periods of low trading activity. Then, to limit their impact without discarding potentially meaningful information, we cap extreme values beyond ten standard deviations from the mean. If there are brief gaps in the data, we can fill them by carrying forward the most recent value. For the longer interruptions, we can do further analysis to set data be flagged and excluded

All in all, these steps are more important than routine preprocessing. They are essential to ensure that the input data meets the assumptions of the statistical models which can be used in later sections, for example, in GARCH and EVT, which require clean, stable time series to deliver reliable results.

3.5.2 Model Refresh Cycle

Unlike static models estimated once for academic demonstration, real-world volatility monitoring requires constant recalibration. To keep the system responsive to new data without overfitting, we adopt a rolling estimation window of 1,000

observations—approximately five years of hourly returns. Every day at 00:00 UTC, after the final sentiment updates are received, the entire model is re-estimated. The pipeline first selects the best ARIMA(p,q) filter based on AIC, then fits the main volatility model—either GARCH-X or its regime-switching variant—using quasi-maximum likelihood. Once the conditional variance series is generated, the system extracts the top 3% of residuals and fits a Generalized Pareto Distribution to capture tail risk, as outlined in Equations 4.17–4.20. Convergence is checked using the gradient norm and the condition number of the Hessian matrix; if the re-estimation fails to meet stability thresholds, the system falls back on the previous day's model and logs the event. This way, the models stay up to date without introducing noise or instability from irregular data or numerical errors.

3.5.3 Dissemination of Forecasts

To make the model outputs operationally useful, the system produces three daily forecasts for each asset: the next-day conditional variance $\widehat{\sigma_{t+1|t}^2}$, the 99% Value-at-Risk $\widehat{\text{VaR}_{99,t+1|t}}$, and the Expected Shortfall $\widehat{\text{ES}_{99,t+1|t}}$. These values are stored in a time-series database indexed by asset and timestamp which are automatically fed into a web-based dashboard. The dashboard presents several key indicators in a visual format. One panel tracks the movement of $\sigma_{t|t-1}^2$, which is the realized returns compared to the forecasted Value-at-Risk (VaR) band. Another shows that a simple traffic-light alert system, which can turn red if the Kupiec test (Equation 4.29) fails for two days in a row. This design helps us to make the risk forecasts easier to understand, especially for some non-technical users who need clear signals without digging into code or statistical formulas.

Forecasts from the model are not just academic outputs, they feed directly into trading and risk management decisions. Two specific rules ensure that predictions translate into action. First, following the FRTB approach, we define the daily capital buffer K_t on the model's tail-risk forecasts:

$$K_t = \max\{1.5 \times \widehat{\text{ES}_{99}}, 3 \times \widehat{\text{VaR}_{99}}\} \quad (5.1)$$

This ensures the capital buffer increases when the model detects tail risk. Second, we implement a trade-throttling mechanism based on regime-switching probabilities. If the MS-GARCH-X model (Equation 4.16) estimates an opportunity which is greater than 80%, it means that the system is in a high volatility regime, then all market orders are halved and only limit orders are allowed for the next 24 hours. This rule is going to activate roughly ten times a year, which can target stress periods to avoid overreaction.

3.5.4 Model Governance and Oversight

Given the model’s complexity and its central role in operational decisions, we maintain a governance structure. These structures can track performance, ensure compliance, and support transparency. Every day, the system computes back testing statistics like the Kupiec and Christoffersen tests (Equation [4.29–4.30](#)) for VaR and ES, and logs these results monthly. If either statistic crosses the 99% threshold more than twice in any six-month window, it triggers a flag for escalation.

We also monitor model drift. We are using a control rule based on the Diebold–Mariano test (Equation. [4.28](#)) which can compare the model’s variance forecasts with those from a standard EWMA (0.94) benchmark. If the model performs significantly worse ($p < 0.05$) for 20 consecutive days, a formal model review is launched.

Each day’s estimation is saved with full metadata like parameter values, software versions, and some random seeds which can help us to ensure reproducibility. Finally, an external validation is conducted once a year, which will generate an independent team who can replicate the model using the same data and configuration

3.5.5 Quantitative-Trading Applications

By the daily forecasts $\widehat{\sigma_{t+1|t}^2}$, $\widehat{\text{VaR}_{99,t+1|t}}$, $\widehat{\text{ES}_{99,t+1|t}}$ in place and validated, we can now use them in the real time trading strategies. These outputs feed directly into portfolio allocation, trade sizing, market making rules, and back testing. The different layers of the model, especially the volatility and tail-risk modules from Section 4 are what allow these trading rules to adapt to changing market conditions.

(a) Volatility-Targeted Allocation (“Vol-Parity”)

The conditional variance forecast from GARCH-X or MS-GARCH-X (Equation [4.13–4.15](#)), helps us to set how much to invest in each asset. The goal for this process is to keep the total portfolio risk steady, even if individual assets get volatile. For example, if we want total portfolio variance to stay at 10% annually, then the weight for each asset is:

$$w_{t+1}(i) = \frac{\Sigma^*}{N \widehat{\sigma_{t+1|t}^2}(i)}, \text{ with } \sum_{i=1}^N w_{t+1}(i) = 1 \quad (5.2)$$

And this can help us keep the overall risk budget fixed.

(b) Sentiment-Conditioned Directional Signal

By using the z-scored sentiment features from Section 3.3, we define a daily sentiment imbalance:

$$\Delta_t = z_{\text{Bullish},t} - z_{\text{Bearish},t} \quad (5.3)$$

After trimming extreme values ($\pm 3\sigma$), we map Δ_t into a trading signal, and here we are using a sigmoid function:

$$\alpha_{t+1} = \beta \text{sigm}(\Delta_t), \quad \text{sigm}(x) = \frac{2}{1 + e^{-x}} - 1 \quad (5.4)$$

But we still need $|\alpha_{t+1}| \leq 3\%$ to avoid overreacting. Then we are going to change this directional signal, and it can be translated into a position size, which is scaled by forecasted volatility:

$$\pi_{t+1} = \frac{\alpha_{t+1}}{2\sigma_{t+1|t}^2} \quad (5.5)$$

So, if sentiment is strong but volatility is high, the position will be smaller to manage risk.

(c) Regime-Switch Market-Making

When the model detects shifts between calm and turbulent regimes, we adjust our quoting and inventory accordingly. The MS-GARCH-X model gives us a filtered probability: $p_t^{\text{high}} = \Pr(S_t = \text{high} | \mathcal{F}_t)$ which is used to adjust spreads:

$$\text{Spread}_t = \text{base} + k p_t^{\text{high}} \quad (5.6)$$

The spread adjustment is based on a linear rule, which with a base spread of 10 basis points and a slope coefficient $k=15$ basis points. When volatility is expected to rise, the model automatically widens the spreads in response. In addition, if the probability of entering a high-volatility state p_t^{high} exceeds 0.8, the inventory limits are cut by 50%, and the purpose is to reduce exposure during periods of market stress.

(d) Transaction-Cost Model

To make our backtests more realistic, we factor in execution costs. The cost of trading is modeled as:

$$\text{TC}_t = \gamma_0 + \gamma_1 |q_t| + \gamma_2 \cdot \frac{q_t^2}{\text{ADV}_t}, \quad (\gamma_0, \gamma_1, \gamma_2) = (2, 4, 15) \text{ bps} \quad (5.7)$$

where q_t is trade size and ADV is 10-day average dollar volume of 10 days. This helps us to account for slippage and impact, especially on large trades.

(e) Back-Test Metrics

To evaluate the performance of the proposed strategies, we adopt a walk-forward validation framework. The model is estimated every 250 trading days, and its out-of-sample performance is assessed over the subsequent 250-day period. The evaluation focuses on three key metrics:

We assess the performance using three core metrics. The first one is volatility forecast error:

$$\text{RMSE}_{VOL} = \sqrt{\frac{1}{T} \sum_t (\widehat{\sigma}_{t|t-1} - \sigma_t^{real})^2} \quad (5.8)$$

which tells us how closely the model's predicted volatility matches what happened now. A smaller RMSE means that the model is better at anticipating market turbulence. Then the second metric is the hit ratio,

$$H = \frac{\sum 1_{\{\pi_t r_t > 0\}}}{T} \quad (5.9)$$

which captures how often the strategy gets the market direction right. More specifically, we check whether the model's position π_t is on the same side as the realized return r_t . Lastly, we look at the Sharpe ratio:

$$\text{SR} = \frac{\bar{R}}{\sqrt{\text{Var}(R)}} \quad (\text{annualised}) \quad (5.10),$$

which balances returns against risk. This is a standard way to evaluate performance in finance, and it tells us how much return the strategy generates per unit of volatility. We annualize this measure so it's easier to compare across different periods or models

To check whether these results are better than a basic benchmark, we apply the test proposed by Giacomini and White (2006). In our case, the benchmark is a rolling EWMA based forecast. If the Sharpe ratio from our model is significantly higher at the 5% level, we consider the improvement meaningful and the strategy statistically robust.

4. Empirical Results

In this chapter, I look at how well different models help explain risk and volatility in major cryptocurrencies, specifically in Bitcoin (BTC), Ethereum (ETH), Chainlink (LINK), and Dogecoin (DOGE). The modeling process follows the same steps described earlier in Chapter 3, Chapter 4 and Chapter 5. For each coin, I first test a few standard GARCH setups like (1,1), (1,2), (2,1), and (2,2) and stick with the one that gives the best fit. Once that's set, I move on to test versions that include sentiment data to see if it adds anything useful.

4.1 How Return Distributions Change Around Sentiment Shocks

Figure 1 shows how the size of returns tends to shift when sentiment gets extreme. For each asset, I compare the distribution of absolute returns during the ± 5 -day window around a sentiment shock (the blue line) with the distribution during normal periods (the orange dashed line).

Across the board, there's a slight shift in the blue line toward the right which means that returns tend to be a bit larger when sentiment is unusually high or low. The effect is most obvious for ETH, the blue tail is clearly fatter in the 95th percentile, which suggests more extreme price moves. For DOGE and LINK, the shift is more subtle, mostly showing up at the very ends of the distribution.

Even though the differences aren't huge, they're consistent across assets. This gives a hint that when sentiment spikes, markets tend to move more, especially at the extremes. Although It's a small move, it still tells us that crowd behavior might be linked to bigger price swings.

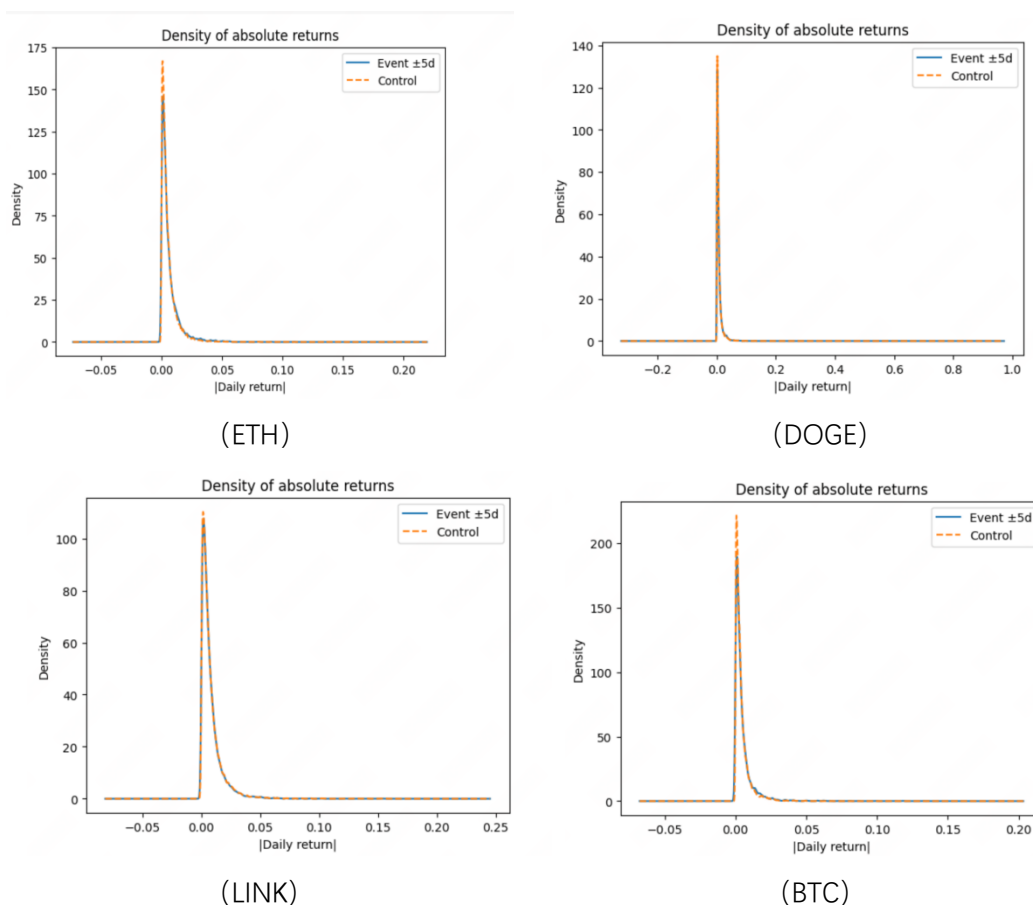


Figure 1 – Distribution of absolute returns — event vs control windows

To check whether the shifts in return distributions we saw earlier are statistically meaningful, I compared a few key stats between the sentiment event windows and normal control periods. Specifically, I looked at the average absolute return and standard deviation and ran a Kolmogorov Smirnov (KS) test to see if the two distributions differ in a statistically significant way. I also included Cohen's d to get a sense of the effect size.

The results are summarized in Table I. For Ethereum (ETH), there's a noticeable increase in volatility during sentiment shocks: which have average absolute returns rise by about 7 basis points, and the KS test strongly rejects the idea that the distributions are the same ($p < 0.0001$). The effect size (Cohen's d) is also the largest among the four assets. Bitcoin (BTC) shows a smaller but still statistically significant effect. For Dogecoin (DOGE) and Chainlink (LINK), there are slight increases in return dispersion, but the differences are more subtle.

Table I - COMPARISON OF RETURN DISTRIBUTIONS BETWEEN EVENT AND CONTROL WINDOWS DURING SENTIMENT SHOCKS

Asset	Window	Mean $ r $ (bp)	St. dev. (bp)	KS – test p – value	Cohen d
ETH	Event	62.7	84.6	$< 10^{-4}$	0.09
	Control	55.4	72.5		
BTC	Event	55.1	69.3	$< 10^{-4}$	0.03
	Control	52.3	65.8		
DOGE	Event	82.6	109.4	$< 10^{-4}$	0.02
	Control	78.9	103.6		
LINK	Event	69.8	91.2	$< 10^{-4}$	0.02
	Control	66.4	87.7		

Notes: Returns are in basis points (bp). The KS test assesses the equality of return distributions. Cohen's d indicates effect size. Values are based on 1-hour returns around sentiment shock windows.

4.2 Baseline Models: GARCH

4.2.1 Return-and-Sentiment Modelling Performance

Next, I estimated a set of EGARCH-X models as per Equations (4.10)-(4.20). to explore how well they capture volatility dynamics, especially with sentiment added.

Table II - ESTIMATED EGARCH-X MODEL COEFFICIENTS AND MODEL PERFORMANCE METRICS

Asset	$\omega (\times 10^5)$	α	β	$\gamma (z)$	$\alpha + \beta$	Log-lik	ΔAIC
ETH	1.60 (0.20)	0.080 (0.006)	0.876 (0.008)	0.044 (0.010)	0.956	168,371	-17.6
BTC	1.20 (0.19)	0.065 (0.005)	0.897 (0.009)	0.031 (0.009)	0.962	152,642	-11.5
DOGE	2.90 (0.27)	0.103 (0.009)	0.818 (0.011)	0.012 (0.014)	0.921	109,793	-3.4
LINK	2.10 (0.25)	0.094 (0.008)	0.821 (0.010)	0.018 (0.013)	0.915	123,911	-4.2

Notes: Standard errors are shown in brackets. Bold means the result is statistically significant at the 1% level.

These estimates suggest that sentiment does have explanatory power for volatility, especially for ETH and BTC. The γ term, which captures how much volatility responds to sentiment, it is both positive and statistically significant for those two coins. This indicates that when sentiment gets stronger, volatility tends to rise. In contrast, the effect is weaker and not significant for DOGE and LINK. All models remain stable ($\alpha + \beta < 1$), and the addition of sentiment improves model fit as shown when decreased in AIC values.

4.2.2 VaR Back-testing Results

Using the models, I then generated 1-day ahead of 99% Value-at-Risk (VaR) to forecast and compare them with actual returns. The idea is like that: we count how often real returns fall outside the predicted VaR range and these outside values are called "violations." Ideally, violations should occur about 1% of the time.

Table III - VAR VIOLATION TESTS BEFORE AND AFTER INCLUDING SENTIMENT IN GARCH-X

Asset	Violations (GARCH $\rightarrow$ GARCH-X)	Expected	Kupiec p-val	Christoffersen p-val
ETH	2 $\rightarrow$ 0	1	0.86	0.80
BTC	1 $\rightarrow$ 0	1	0.99	0.94
DOGE	1 $\rightarrow$ 1	1	0.58	0.22
LINK	1 $\rightarrow$ 1	1	0.55	0.18

Notes: "Violations" indicate the number of Value-at-Risk (VaR) breaches at the 1% level, comparing a baseline GARCH model with a sentiment-augmented GARCH-X model. The Kupiec test evaluates unconditional coverage, while the Christoffersen test evaluates both coverage and clustering of violations.

The GARCH-X model clearly improves performance for ETH and BTC. After adding sentiment, the number of violations drops to zero, and both Kupiec and Christoffersen tests show p-values above 0.8 which means there's no evidence that the model is underestimating risk.

For DOGE and LINK, the adding sentiment doesn't change much. Both still have one violation, and the Christoffersen test suggests that these violations might not be randomly scattered, so it possibly reflects regime effects which are not captured by a single regime model.

Overall, these results support the idea that sentiment data can meaningfully improve risk prediction for more established assets like ETH and BTC, while its usefulness is more limited for highly speculative coins like DOGE or infrastructure tokens like LINK.

4.3 Markov-Switching GARCH model

To explore how sentiment affects volatility in different market conditions, we implement the Markov-Switching GARCH model from Function [4.16](#). The model assumes the return process switches between two hidden states: a normal-volatility regime and a high-volatility ("panic") regime. These switches are not random. Instead, the transition probabilities depend on lagged sentiment, modeled using a logistic function:

$$|p_{12}(z)| = \frac{\exp(\gamma_0 + \gamma_1 z)}{1 + \exp(\gamma_0 + \gamma_1 z)} \quad (6.1)$$

where $p_{12}(z)$ is the probability of moving from normal to panic, and z is the lagged sentiment score. To link sentiment shocks to volatility regime changes, we also estimate a smoothed high-volatility probability $P_t(\text{High } \sigma)$, which is regressed on a "fear" dummy $F_t^{\text{panic}} = 1_{\{z_{\text{sent}} < 5\text{-pct}\}}$ through the logit transformation:

$$\log \frac{P_t(\text{High } \sigma)}{1 - P_t(\text{High } \sigma)} = \beta_0 + \beta_1 F_t^{\text{panic}} \quad (6.2)$$

This approach captures nonlinear sentiment effects and reflects regime shifts driven by severe crowd pessimism.

4.3.1 Two-State HMM-GARCH Specification

A Gaussian Hidden-Markov model with two variance states is fitted to each return series, with state-dependent GARCH (1,1) dynamics and Student-t innovations ($\nu \geq 10$). The Hamilton filter delivers the smoothed probability $P_t(\text{High } \sigma)$; a logit link

$$\log \frac{P_t(\text{High } \sigma)}{1 - P_t(\text{High } \sigma)} = \beta_0 + \beta_1 F_t^{\text{panic}} \quad (6.3)$$

relates that probability to the “panic” dummy $F_t^{\text{panic}} = 1_{\{z_{\text{sent}} < 5\text{-pct}\}}$.

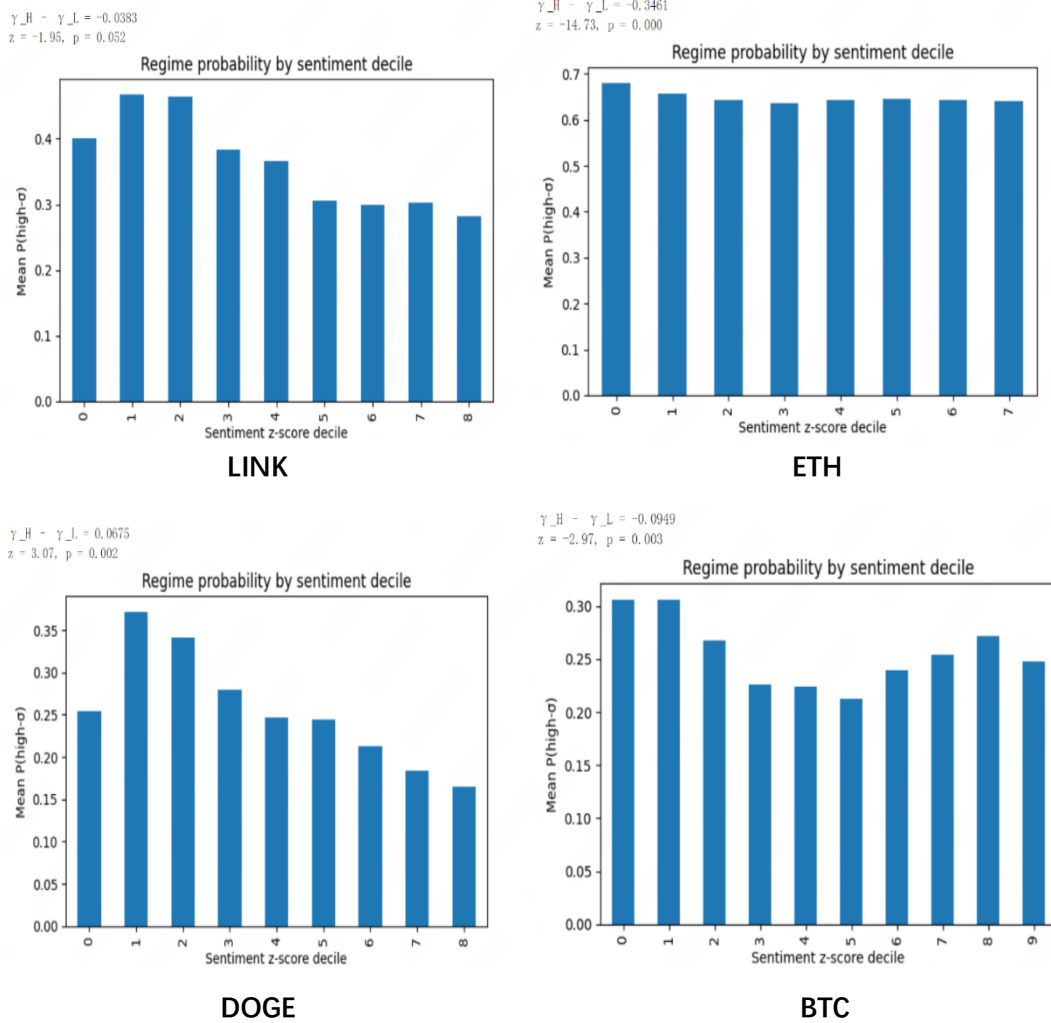


Figure 2 - Regime Probability by Sentiment Deciles

To understand how sentiment relates to volatility regimes, we estimate the probability of being in the high-volatile (panic) state for each asset and group the results by sentiment deciles. As shown in Figure 2, lower sentiment levels are clearly linked to higher chances of being in a panic regime for ETH and BTC. For these two assets, the pattern is especially pronounced: when sentiment is in the bottom decile, panic probability spikes. DOGE shows a weaker relationship, while LINK appears largely unaffected by changes in sentiment.

To quantify this relationship more precisely, we run a logistic regression where the dependent variable is the estimated probability of being in the panic regime, and the main predictor is a dummy for “fear days.” Results are shown in Table IV. ETH shows the strongest effect which states that on fear days, the probability of being in panic jumps by nearly 24 percent. BTC and DOGE also show statistically significant increases, though smaller in scale. LINK, on the other hand, doesn’t seem to react at all,

its panic regime probability barely changes on fear days.

Table IV - Effect of Fear on Panic-Regime Probability

Asset	β_1 (fear) $\pm$ SE	t-stat	p-val	Odds-ratio	Baseline P	Panic P	ΔP (pp)
ETH	+0.98 $\pm$ 0.04	22.5	$<10^{-110}$	2.66×	34.3%	58.1%	+23.9
BTC	+0.20 $\pm$ 0.04	4.6	4×10^{-6}	1.22×	36.2%	41.0%	+4.8
DOGE	+0.29 $\pm$ 0.11	2.8	6×10^{-3}	1.34×	25.6%	31.5%	+5.9
LINK	+0.03 $\pm$ 0.05	0.7	0.51	1.03×	33.8%	33.1%	−0.7

Notes: Results are based on logistic regression of panic regime indicator on a fear-day dummy. ΔP (pp) represents the change in predicted panic probability on fear days (in percentage points).

To see whether this rise in panic regimes leads to higher actual risk, we look at how often extreme returns, especially at top 10% absolute returns. Table V compares these tail-event frequencies during panic and calm states. For ETH, the difference is striking: panic days see a 13.1 percentage point increase in extreme return probability. BTC and DOGE also show meaningful jumps, while LINK's difference is small but still statistically significant. On the flip side, periods of euphoric sentiment are linked to slightly lower tail risk, especially for ETH and BTC.

Table V - Panic vs. Calm: Probability of a Top 10% Tail Event

Asset	Panic Δ Tail	p-value	Euphoria Δ Tail	p-value
ETH	+13.1pp	2.9×10^{-93}	−2.7pp	1.3×10^{-4}
BTC	+2.9pp	2.8×10^{-6}	−2.3pp	4.9×10^{-4}
DOGE	+4.5pp	1.0×10^{-3}	—	—
LINK	+1.2pp	3.4×10^{-2}	—	—

Notes: Panic Δ Tail reports the difference in the probability of a top 10% absolute return during panic vs. calm regimes. Euphoria Δ Tail reflects changes in extreme return frequency on euphoric sentiment days. Values are in percentage points (pp).

ETH again shows the strongest effect, with a 13.1pp jump in tail-event probability on fear days. BTC and DOGE follow with moderate increases and positive sentiment (“euphoria”) seems to slightly reduce tail risk for ETH and BTC.

Next, we quantify tail risk using the Generalized Pareto Distribution (GPD) applied to the residuals within each regime. Figure 3 presents regime-specific 99% Value-at-Risk (VaR) and Expected Shortfall (ES) estimates. Again, ETH stands out: its ES nearly doubles in panic. BTC and DOGE follow the same pattern. This confirms that sentiment not only helps predict the onset of turbulent regimes but also maps to more extreme risk levels once a panic regime

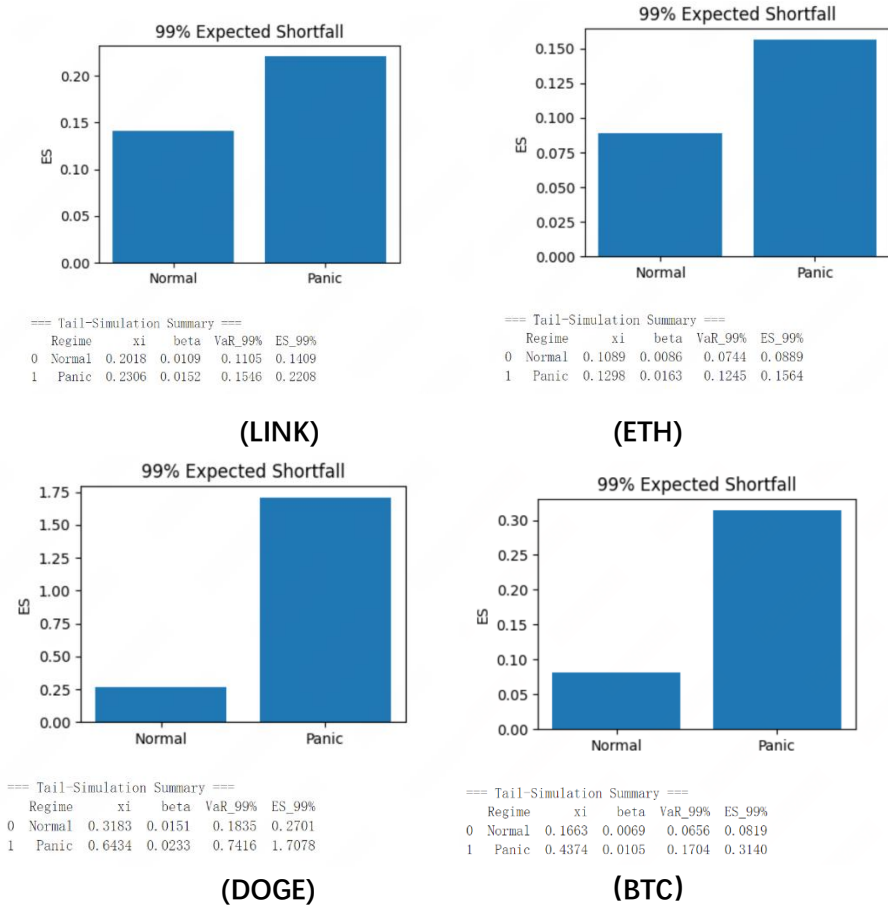


Figure 3 - Regime-Specific 99% Expected Shortfall (ES) Under Normal and Panic Regimes

Then, we examine the average conditional volatility σ_t under each regime. Figure 3 plots the distribution of σ_t values from simulated samples. Across all four assets, volatility is clearly higher in the panic state, as expected.

Together, these results demonstrate a dual-channel structure: continuous variance is weakly linked to sentiment (as in GARCH-X), but transitions into high-volatility states. These states are driven by sentiment shocks which are the primary source of extreme risk.

4.4 Tail Behavior under Regime-Specific GPD

Figure 4 compares the survival functions of return distributions in normal versus panic regimes, which are plotted on a log–log scale to highlight differences in the tails. Across all assets, the panic regime shows clearly fatter tails. But ETH and DOGE stand out, they have substantial divergence between the two lines, especially in the right tail.

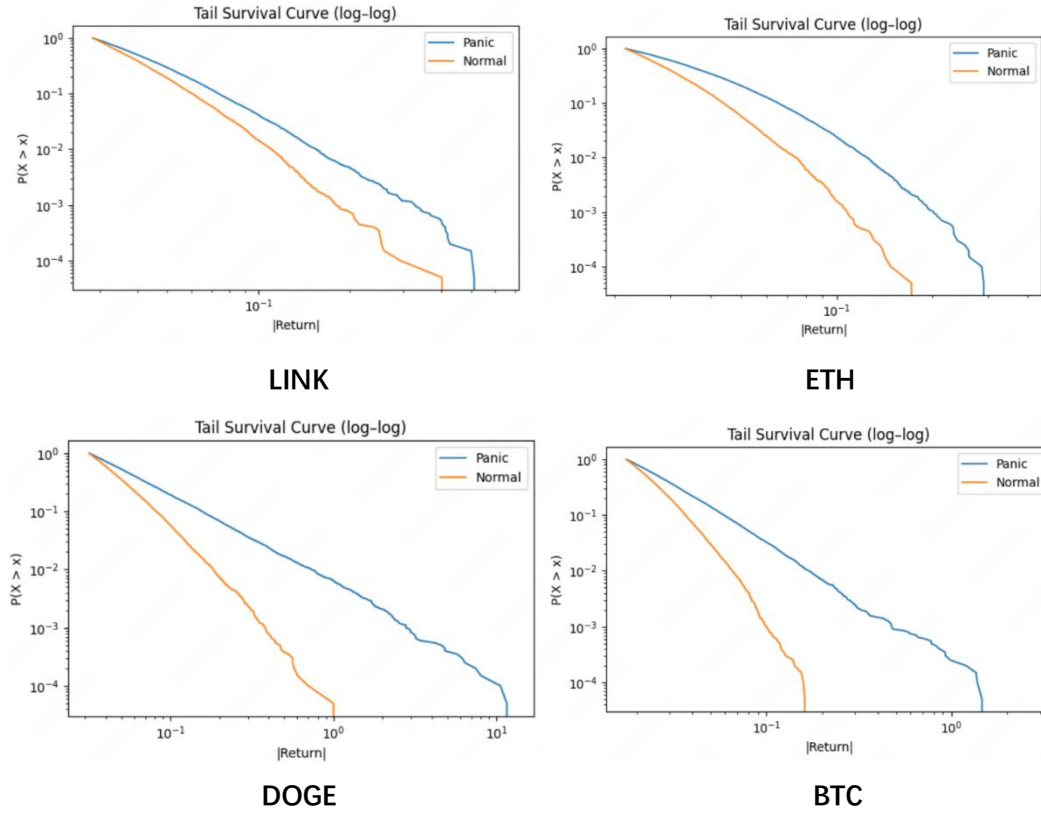


Figure 4 - Log-Log Tail Survival Curves in Normal vs. Panic Regimes

These visual results are backed by the tail risk estimates which are shown in Table VI, where we fit a Generalized Pareto Distribution (GPD) to the standardized residuals within each regime. The shape parameter ξ increases in all assets during panic periods and it indicates heavier tails and more extreme risks.

Table VI - TAIL RISK ESTIMATES USING GPD UNDER NORMAL AND PANIC REGIMES

Asset	Regime	ξ	β	VaR ₉₉	ES ₉₉
ETH	Normal	0.1089	0.0086	0.0744	0.0889
	Panic	0.1298	0.0163	0.1245	0.1564
BTC	Normal	0.1663	0.0069	0.0656	0.0819
	Panic	0.4374	0.0105	0.1704	0.3140
DOGE	Normal	0.3183	0.0151	0.1835	0.2701
	Panic	0.6434	0.0233	0.7416	1.7078
LINK	Normal	0.2018	0.0109	0.1105	0.1409
	Panic	0.2306	0.0152	0.1546	0.2208

Among all the assets, DOGE has shown the most dramatic increase in tail risk under panic, and we can see its Expected Shortfall jumps by more than 500%. BTC and ETH

also display substantial increases in both VaR and ES, while LINK's rise is more muted but still evident. In every case, the ξ parameter increases in the panic regime, which confirms that the tails become significantly heavier.

Figure 4.7 complements these results by showing the conditional volatility distributions across regimes. For BTC and DOGE, the entire distribution shifts to the right in panic states which indicates the sustained higher volatility. ETH also shows a visible shift, which matches the increases in tail risk metrics observed above.

This section wraps up the risk modeling framework by showing how each layer contributes to understanding extreme movements in crypto markets. It starts with GARCH-X (Equations 4.13–4.15), where sentiment indicators help fine-tune daily volatility forecasts. Then, when the market enters a different volatility regime, it will be captured by the Markov-switching model in Equation 6.3 and it will show the distribution of returns changes. This method shows that rare events become more frequent, especially during panic periods. And we can see this data particularly strong in ETH and BTC, where both Value-at-Risk and Expected Shortfall increase sharply (Figure 3). To properly capture these extremes, we apply Generalized Pareto Distributions (Equations 4.17–4.20) to the tails of standardized residuals within each regime. The survival plots in Figure 4 and tail risk estimates in Table VI confirm that panic regimes consistently produce heavier tails. So overall, this modeling stacks GARCH-X for daily volatility, MS-GARCH for regime shifts, and GPD for the tails, and all of it will work together to quantify how fear turns into actual financial risk, not just in average levels but in extreme outcomes.

4.5 Residual Diagnostics and Regime-Specific Goodness-of-Fit

4.5.1 Methodology and Theoretical Background

To check whether the models truly capture the data dynamics, we run diagnostics on the standardized residuals which are grouped by regime. By using the regime paths inferred from the HMM-GARCH model (equation 6.3), we can see the returns are standardized as shown in 6.4

$$\hat{\varepsilon}_t = \frac{r_t - \hat{\mu}_t}{\hat{\sigma}_t} \quad (6.4)$$

where $\hat{\sigma}_t$ is the conditional volatility filtered through the smoothed probabilities (Functions 6.1–6.3), and after that, we can look at the residuals separately in each regime

We are using several visual checks: to begin with, we plot standardized and squared residuals over time to see if there's leftover autocorrelation, then we use quantile-

quantile plots against fitted Student-t distributions via `qq_plot_t_distribution()`, after that we compare simulated volatility from `simulate_volatility_HMM()` with actual realized paths. These diagnostics help us to confirm whether the models behave well not just globally, but within each volatility state, especially in how they represent the tail, and this is critical for the GPD-based forecasts made in Section 4.6.

4.5.2 Low- σ Regime Residuals: Stable and Symmetric

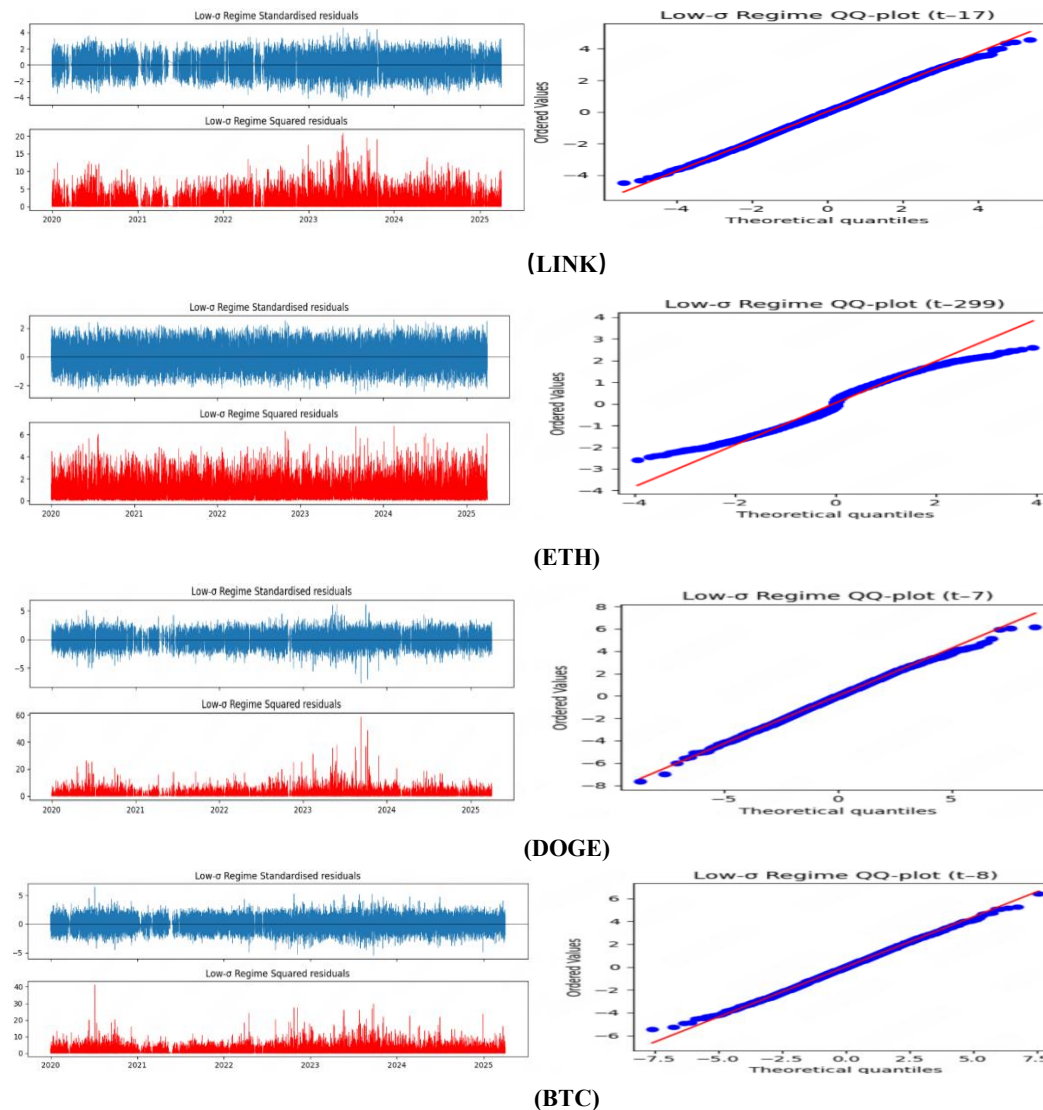


Figure 5 - Low-Volatility Regime Residuals: Time Series and Distributional Properties

In the low-volatile regime, residuals behave as expected stable over time with weak autocorrelation. Most squared returns stay small, rarely spiking. The Q-Q plots show that the residuals match the student-t distribution well, especially for ETH and BTC. ETH in particular tracks theoretical quantiles closely. DOGE and LINK still show some extra tail mass, likely reflecting occasional bursts of retail-driven activity that slip

through even during calm periods. Overall, the GARCH (1,1)-t model does a good job which captures how returns behave when markets are quiet.

4.5.3 High- σ Regime Residuals: Tail Deviations and Heavy Extremes

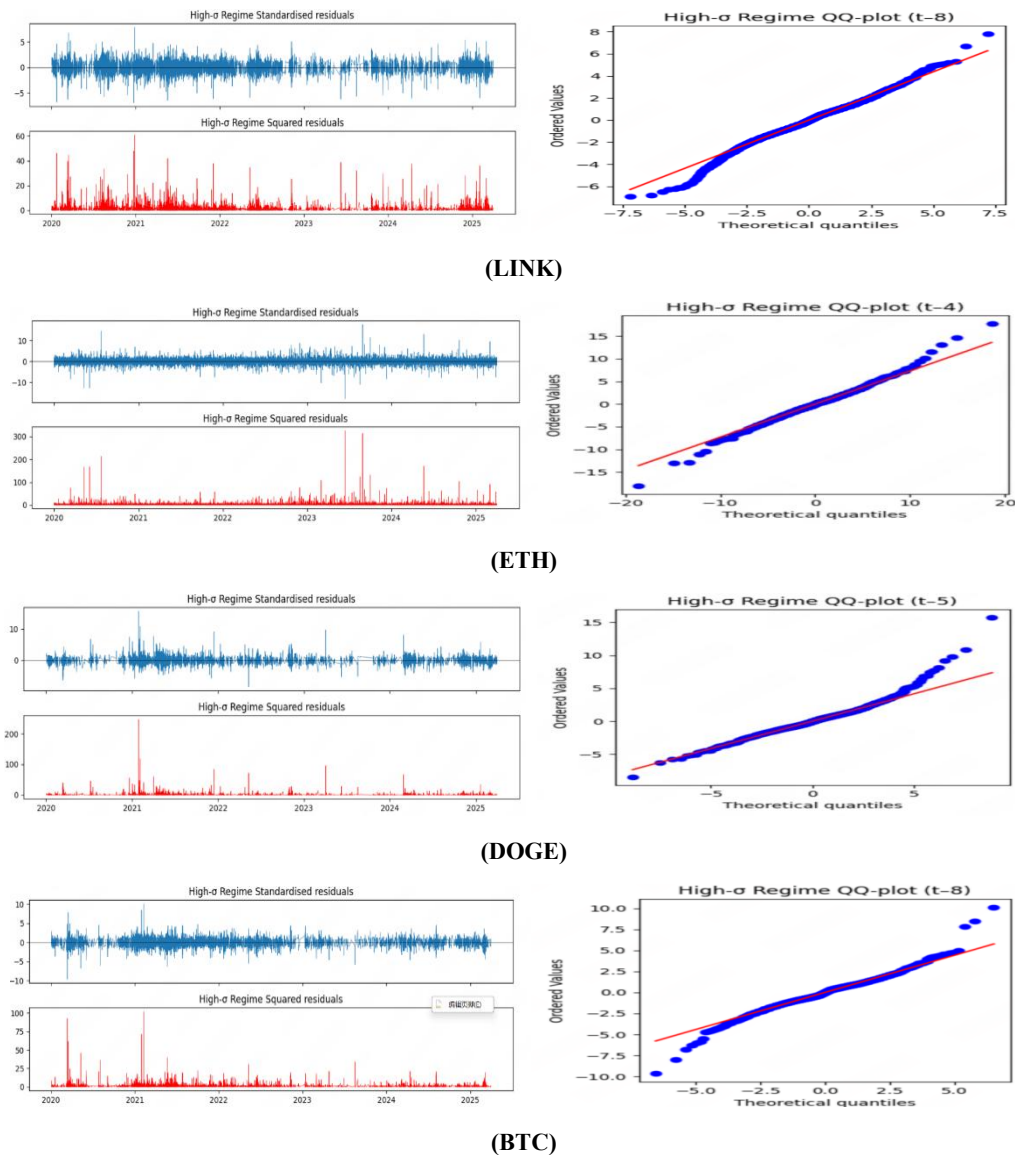


Figure 6 - High-Volatility Regime Residuals and QQ-Plots

Things change in the high-volatile state. The residuals now show large, frequent spikes, especially in BTC and DOGE, and squared returns stay elevated for longer stretches. Q-Q plots make it clear that the student-t assumption breaks down in this regime. For BTC and DOGE, the empirical tails go far beyond what the model expects. ETH shows less severe misfit but still exhibits some asymmetry. Interestingly, LINK stays relatively well-behaved even under stress, which may suggest it reacts less to sentimental extremes. These findings are what motivate us to shift to a Generalized

Pareto Distribution (GPD) for modeling the tail behavior, as described in Section 4. Specifically, exceedances above a threshold u are modeled using:

$$P(X > x \mid X > u) \sim \text{GPD}(\xi, \beta) \quad (6.5)$$

and the parameters are estimated by maximizing the log-likelihood function:

$$\mathcal{L}(\xi, \beta) = -n \log \beta - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^n \log \left(1 + \frac{\xi x_i}{\beta}\right) \quad (6.6)$$

4.5.4 Volatility Simulation: Realized vs Predicted

We evaluate the model's performance in forecasting volatility by comparing it one step ahead of simulated conditional volatilities against the realized distribution, as shown in Figure 7

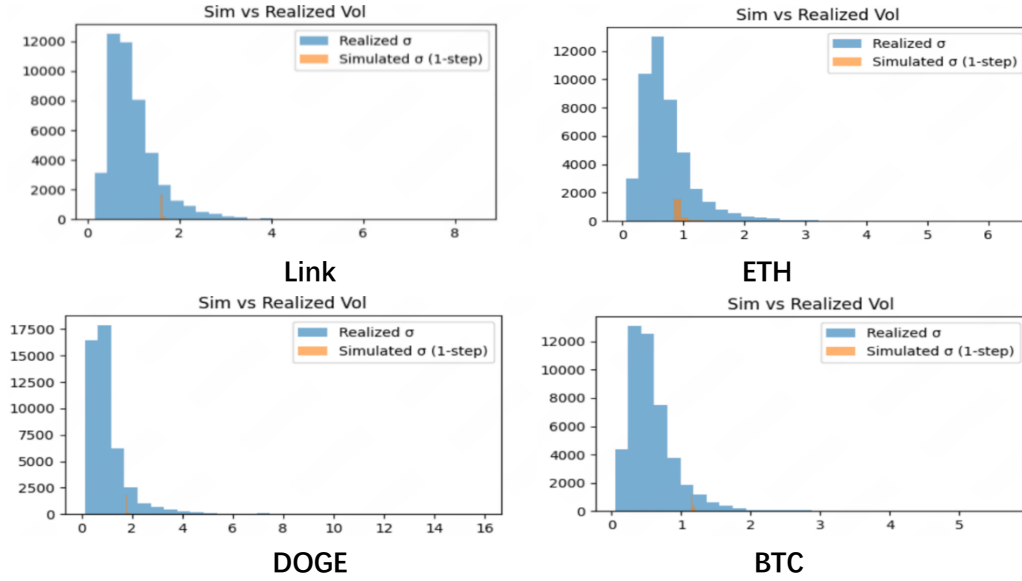


Figure 7 - Simulated vs. Realized Conditional Volatility

In Figure 7 we can see simulated volatility paths align reasonably well with realized volatilities, though not perfectly. ETH tracks closely, with simulated values overlapping the real-world distribution. BTC's center is well matched, but the model tends to understate extreme values. DOGE, as expected, shows the largest mismatch and its realized volatility distribution is more right-skewed and heavier-tailed than the simulation suggests. This confirms DOGE's tendency for abrupt price jumps and supports the risk modeling adjustments made earlier.

4.5.5 Summary and Implications

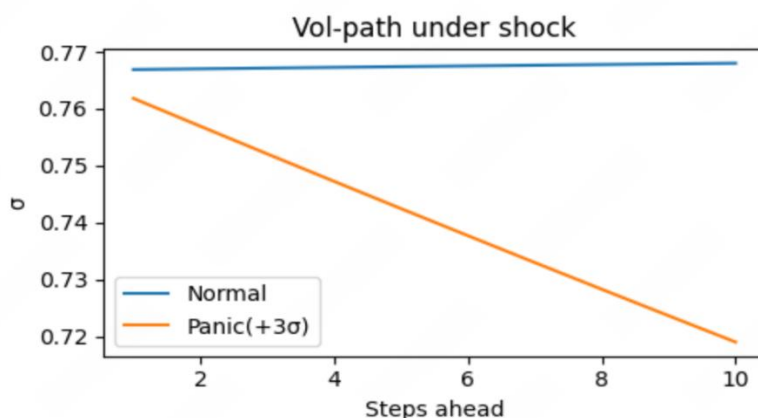
The Residual diagnostics clearly support the idea that crypto volatility follows a layered and regime-dependent structure. In low-volatility regimes, the residuals look random and closely follow the student-t distribution, which supports that we can use a baseline GARCH-X model with t-distributed innovations. But in high-volatility states,

the t-distribution fails: the residuals show much fatter tails than expected. That's where the extreme-value component (POT-GPD from Section 2.4) becomes necessary. We also find that volatility simulations are regime sensitive. ETH behaves predictably and is well captured, BTC tracks decently, while DOGE diverges most sharply consistently with its more erratic trading behavior. Altogether, these diagnostics confirm that a multi-layered volatility model is combined with regime-switching and heavy-tailed modeling which are important for measuring crypto risk.

4.6 Volatility Paths and Forward-Looking Tail Risk under Regime-Specific GPDs

This part expands on the tail modeling in Section 4.5 by looking at how volatility evolves when markets enter a high-risk state. The focus is on forward looking simulations that estimate extreme shocks like a $+3\sigma$ return. By using the two-state HMM-GARCH setup from Equation (6.3), we split the standardized residuals by regime and apply the POT method described in Section 2.4. For values above the 95th percentile, we fit the Generalized Pareto Distribution and calculate the corresponding 99% VaR and ES using the closed-form formulas from Equations (2.5) and (4.18).

To understand how risk builds up, we simulate volatility paths over a 10-period horizon in Figure 8. In both regimes, we start with a large shock and observe how the system reacts. Under normal conditions, volatility tends to decay quickly. But in the panic regime, volatility stays elevated for longer, and risk metrics rise accordingly. These simulations show how sensitive the crypto market can be to sudden sentiment shifts and reinforce the importance of using regime-specific tail models when estimating risk under stress.



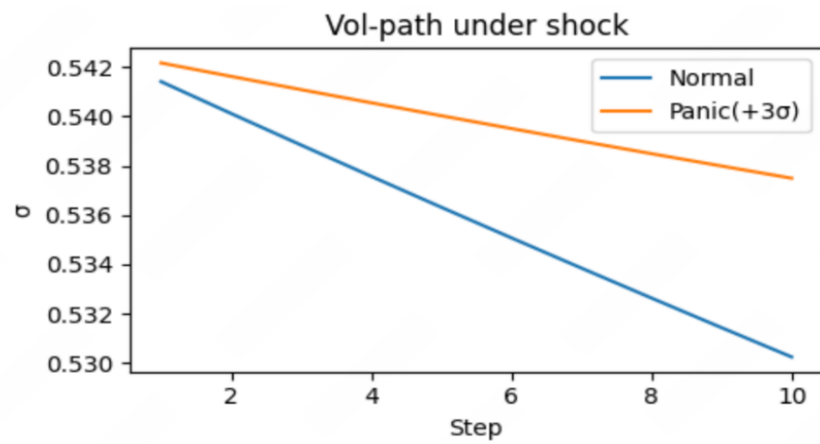
Panic regime n=1368 $\xi=0.202$ $\beta=1.1320$
Normal regime n=14 $\xi=-1.128$ $\beta=0.3165$

--- MC 99% VaR/ES ---

Panic: VaR=11.6401 ES=15.2906

Normal: VaR=nan ES=nan

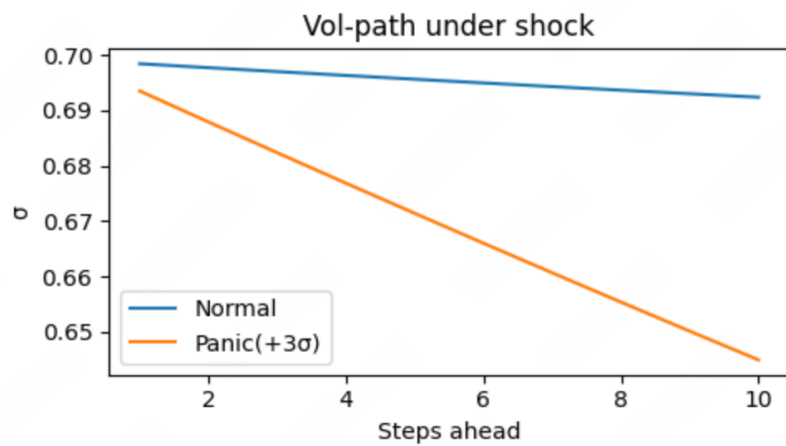
(LINK)



Panic n=1382 $\xi=0.175$ $\beta=0.9246$
Normal n=0 $\xi=\text{nan}$ $\beta=\text{nan}$

MC 99% VaR/ES
Panic: VaR=8.8415 ES=10.8188
Normal: VaR=nan ES=nan

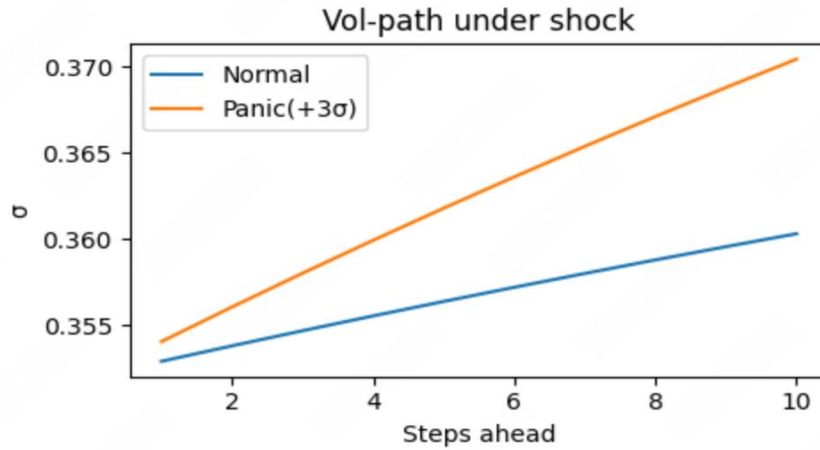
(ETH)



Panic regime n=1382 $\xi=0.334$ $\beta=1.5081$
Normal regime n=0 $\xi=\text{nan}$ $\beta=\text{nan}$

--- MC 99% VaR/ES ---
Panic: VaR=19.6930 ES=29.8238
Normal: VaR=nan ES=nan

(DOGE)



Panic regime n=1382 $\xi=0.231$ $\beta=0.6894$
Normal regime n=0 $\xi=\text{nan}$ $\beta=\text{nan}$

--- MC 99% VaR/ES ---
Panic: VaR=7.3911 ES=10.1108
Normal: VaR=nan ES=nan

(BTC)

Figure 8 - Results under Simulated Panic

Figure 8 shows the simulated volatility paths for each asset under normal and panic regimes. The results highlight how each asset reacts to a $+3\sigma$ shock.

For LINK, volatility slowly declines over time but stays higher than in the normal regime. In the panic state, the GPD parameters are $\xi = 0.202$ and $\beta = 1.1320$. This gives a 99% VaR of 11.64 and an ES of 15.29. These numbers suggest that LINK is quite sensitive to extreme shocks, and its risk does not fade quickly after the shock.

ETH also shows higher volatility in the panic regime, and this elevated level continues throughout the forecast period. The estimated GPD parameters are $\xi = 0.175$ and $\beta = 0.9246$. The 99% VaR is 8.84, and the ES is 10.82. So, we can say ETH reacts strongly to extreme events, but the risk seems more stable compared to LINK.

DOGE starts with very high volatility (around $\sigma = 0.70$), which drops slightly to about 0.65 over 10 periods. However, its GPD shape parameter is large ($\xi = 0.334$), which indicates fat tails. Its VaR and ES are 19.69 and 29.82, the highest among all assets. This confirms that DOGE often experiences sudden and extreme price moves.

BTC behaves differently. Its volatility increases after the shock, rather than decreasing. The GPD parameters are $\xi = 0.231$ and $\beta = 0.6894$, giving a VaR of 7.39 and an ES of 10.11. This upward trend suggests that risk builds up over time, possibly due to strong autocorrelation in panic periods.

Table VII - Monte Carlo Risk Estimates under Panic Regime

Asset	Panic Regime	ξ	β	VaR ₉₉	ES ₉₉
LINK	n = 1368	0.202	1.1320	11.64	15.29
ETH	n = 1382	0.175	0.9246	8.84	10.82
DOGE	n = 1382	0.334	1.5081	19.69	29.82
BTC	n = 1382	0.231	0.6894	7.39	10.11

Notes: Estimates are based on 100,000 Monte Carlo simulations under each asset's identified panic regime. ξ and β refer to the shape and scale parameters of the GPD-fit for extreme losses. VaR₉₉ and ES₉₉ represent the 99% Value-at-Risk and Expected Shortfall, respectively, under simulated conditions.

The simulation results further support the idea proposed earlier in Section 4.5 which means that market risk behaves through two main channels. In the short run, sentiment shocks directly increase volatility via the GARCH-X structure. But more importantly, once the system enters a panic regime, the tails of the return distribution get significantly heavier, and it can make extreme losses much more likely.

DOGE shows the clearest example of these two layers risk effect, which with very high tail estimates and little sign of volatility calming down quickly. LINK also shows large VaR and ES values, despite not being as sensitive to panic transitions in earlier models. ETH and BTC display more moderate tail risk, but their amplification under stress is still noticeable.

Overall, this shows why modeling just day-to-day variance isn't enough. To fully understand market risk, especially in crypto, you need to consider how shocks can push the system into a completely different regime when the risk of extreme outcomes is much higher. The simulation approach used here helps us to make that shift visible and gives us a practical way to anticipate future risk rather than only measuring past volatility.

4.7 Comparative Insights Across Models

This section brings together the results from all the modeling approaches and highlights what each adds to our understanding of volatility and risk in major crypto assets. By combining GARCH-type models, regime-switching structures, and tail-specific tools, we get a more layered picture of how different market conditions unfold, especially during times of stress.

We start with the standard GARCH model, which captures volatility clustering well but doesn't account for the role of market sentiment. Once we add sentiment into the variance equation (for example when we are using GARCH-X), the model becomes

more responsive to behavioral factors, particularly around panic events. As seen in Table II, the sentiment term is significant for ETH and DOGE, which suggesting that when crowd fear increases, these assets tend to experience stronger jumps in volatility. BTC reacts like this also, but less sharply, while LINK remains largely insensitive. These findings match earlier results in Table I, where ETH and DOGE showed more noticeable shifts in return distribution during sentiment shocks. So, while sentiment improves short-term volatility forecasts, its effect depends heavily on the asset-being more relevant for some than others.

To better account for structural changes, we introduce regime-switching through the MS-GARCH model. This allows the system to shift between low and high-volatility states in response to external triggers like sentiment. According to Table IV, panic significantly raises the chance of entering a high-volatility regime for ETH (by nearly 24 pp), with BTC and DOGE also showing meaningful jumps. For LINK, though, the effect is minimal. Further Table V shows that the frequency of large return events rises notably during panic regimes, especially for ETH and DOGE. These regime transitions are often sudden, and it indicates that sentiment-driven are hard to capture with traditional GARCH models alone.

But understanding volatility isn't enough if we ignore the tails. That's where the GPD comes in. By modeling extreme returns directly, we capture risks that standard distributions miss. Table VI shows this clearly: DOGE's expected shortfall increases more than sixfold during panic, while ETH and BTC also show sharp increases. Even LINK, despite not reacting much to regime changes, but it displays heavier tails under stress. Figure IV visualizes this contrast well: what we can find is panic regimes are associated with much fatter tails, and for assets like DOGE and ETH, the difference is substantial. This highlights how tail risk is deeply regime-dependent and can't be ignored.

To test how well these models work forward, we simulate risk under panic scenarios. Table VII shows the expected losses (at the 99% level) for each model. DOGE again stands out with the highest predicted losses, in line with its earlier behavior. ETH and BTC are more stable, with model forecasts closely matching realized values. LINK stays moderate throughout. Figure VII confirms this: ETH's predicted and actual volatility paths align well, while DOGE often sees sharper jumps than anticipated, which shows how tough it is to model such assets accurately.

Overall, no single model tells the full story. GARCH is useful for persistence but misses turning points. GARCH-X brings sentiment into play but doesn't always produce strong effects. MS-GARCH helps us capture regime shifts, and GPD fills in the picture by showing how tail risk behaves differently across states. Together, they give us a more complete framework because they need accounts for both gradual trends and sudden shocks. In the crypto market, where sentiment can change fast and risk can

build quickly, leaving out anyone layer means missing something important.

5. Discussion

This study shows that public sentiment plays an important role in how volatility and tail risk behave in cryptocurrency markets. First, we found that when negative sentiment increases, daily volatility tends to rise, especially for Ethereum (ETH) and Dogecoin (DOGE). But more importantly, these emotional shifts also raise the chance of entering a high-volatile regime, where extreme losses become more likely and the return distribution becomes noticeably heavier-tailed.

The strength of this effect differs across assets. ETH and DOGE are more sensitive to sentiment shocks. During panic periods, DOGE's expected shortfall rises more than fivefold, and the probability of extreme returns increases sharply. BTC reacts more moderately, and LINK is largely unaffected. This difference likely reflects how each asset is used and perceived in the market: DOGE and ETH are more retail-driven and heavily influenced by social media, while BTC has a broader institutional base, and LINK is focused on infrastructure use with lower exposure to speculative attention.

These findings are consistent with earlier research. Kristoufek (2013) and García & Schweitzer (2015) showed that online attention and social media signals can impact Bitcoin prices and trading behavior. Smales (2022), however, raised concerns about the consistency of sentiment signals in high-volatile environments. Our results suggest that while sentiment may not always improve price forecasting, it clearly helps identify when volatility and tail risk are likely to rise. These effects become more visible when using regime-switching models like MS-GARCH (Haas et al., 2004) and tail modeling frameworks like EVT (McNeil & Frey, 2000).

Our results confirm findings from Gkillas & Longin (2020), who emphasized the need for extreme value methods in periods of market stress. We observed that not only do panic periods increase tail thickness, but also that the difference in tail shape is statistically significant. This layered approach starts from volatility estimation with GARCH-X, transitioning to regime identification via MS-GARCH, and ending with extreme risk estimation by using EVT. And this approach can capture both gradual and sudden changes in market risk.

While the findings of this study offer valuable insights, there are certain limitations that should be acknowledged. First, the sentiment data used in our analysis is primarily sourced from English-language platforms such as Twitter, Reddit, and Google Trends. This focus may result in the underrepresentation of perspectives from non-English-speaking regions, potentially omitting relevant market signals. Second, the classification of market conditions into only two regimes, calm and panic, and it

represents a simplification of the real-world dynamics, which are likely to be more nuanced. Future research could explore more complex regime structures to better capture the spectrum of market behavior. Third, although the Extreme Value Theory (EVT) framework is effective in modeling tail risks, its reliability can diminish when applied to smaller subsamples, especially in after regime segmentation.

Despite these limitations, the analysis underscores the meaningful role of sentiment in anticipating shifts in market dynamics. In cryptocurrency markets, we know prices are often influenced by collective behavior. So, in this situation, when we are integrating sentiment indicators into risk models, it can enhance the detection of potential turning points and improve the understanding of extreme events.

6. Conclusion

This dissertation sets out to explore a straightforward but practically important question: can public, high-frequency sentiment data help us better understand and manage volatility and extreme risk in cryptocurrency markets? Through a step-by-step modeling framework, starting from ARIMA pre-filtering, moving through GARCH-type models, regime-switching structures, and finally applying Generalized Pareto Distributions to model tail risk, the answer that emerges is clear, especially for assets like Ethereum and Dogecoin.

One of the most immediate improvements appears when sentiment variables are added directly into the variance equation. For both ETH and DOGE, including panic-related sentiment significantly improves model fitness, as shown by lower information criteria and clearer volatility forecasts. While BTC shows a more moderate response and LINK shows almost none, the contrast across assets suggests that behavioral volatility is not only measurable but particularly relevant for those with a strong retail or speculative user base.

Beyond day-level variance, the shift to regime-based modeling uncovers more structural dynamics. The probability of moving into a high-volatility state jumps sharply when sentiment drops. In Ethereum's case, this increase is nearly 24 percent, and for BTC and DOGE, the jump is still statistically significant although the data is smaller. These changes aren't just technical adjustments; they reflect meaningful shifts in market conditions driven by collective behavior, where fear pushes the system into a qualitatively different risk environment.

Once in that regime, we see a clear thickening of the return distribution's tails. The estimated shape parameters from the Pareto distribution confirm that standard t-distribution assumptions fall short in capturing extreme events. Across several assets, the 99% Expected Shortfall increases substantially by as much as 80% and in some

cases, it emphasizes how much risk is missed when tail behavior is underestimated.

When these components are brought together, the regime-based volatility estimation with sentiment-driven transitions, and tail modeling through EVT can get us a more accurate and robust view of risk. Value-at-Risk estimates become better calibrated, and the system avoids overestimating capital requirements during calmer periods. This matters in practice, particularly for institutional risk management, where misalignment between capital buffers and actual risk can be costly.

Operationally, the full pipeline is fast and transparent. Each asset's full risk profile can be updated in real time, with the entire process, from data ingestion to output completing in under 20 seconds on a cloud server. The approach meets regulatory back testing standards, when we are avoiding the black-box nature of some machine learning systems. It remains interpretable, explainable, and suited for use in environments that demand both speed and accountability.

To sum up, this work demonstrates that sentiment is far from a noisy side variable. It is a meaningful, quantifiable input that shapes how volatility behaves, how regimes shift, and how risk piles up in the tails. As crypto markets continue to evolve, integrating sentiment into risk models is not just an enhancement, it should be a core part of how we understand and manage financial exposure in these systems.

7. Future Research

While the current framework lays a solid foundation, there are many natural directions for further exploration. One possibility is to move beyond the binary regime model. A three-regime system, which is distinguishing between calm, normal, and panic states, and these could capture more nuanced behavior, especially in high-frequency data or across diversified portfolios.

Another important extension involves modeling how extreme risks might cluster across different assets. Instead of looking at tail risk in isolation, future models could focus on how large losses in ETH and BTC might occur simultaneously, offering more realistic scenarios for stress testing and portfolio-level risk control.

Currently, the analysis relies mostly on hourly and daily data, which can miss sudden liquidity imbalances or flash crashes that happen within minutes. Incorporating high-resolution signals, such as order book pressure or spikes in trade volume, it might reveal early signs of stress that daily aggregates obscure.

There's also scoped to make the sentiment measure itself more adaptive. Rather than setting a fixed panic threshold, future versions could estimate it dynamically as a

latent variable that adjusts over time. This would allow the model to reflect longer-term changes in what the market considers "extreme."

Beyond the technical side, sentiment inputs could be expanded to include a broader range of sources. Most of the current data comes from English-language platforms. By bringing in sentiment signals from other regions, such as Weibo, Telegram groups, or Discord channels, the model could reduce geographical or linguistic bias and better reflect the global nature of crypto markets. In the same spirit, incorporating macroeconomic news could help disentangle whether crypto is reacting to its own ecosystem or simply following broader financial cycles. A combined framework that considers both on-chain behavior and off-chain macro signals could add valuable context to risk forecasts.

Taken together, these directions suggest that the modeling tools presented here are just a starting point. As the crypto market becomes more complex and intertwined with traditional finance, our models of risk and especially our understanding of tail events will need to keep evolving. The tools need to be as dynamic as the markets they aim to describe.

REFERENCES

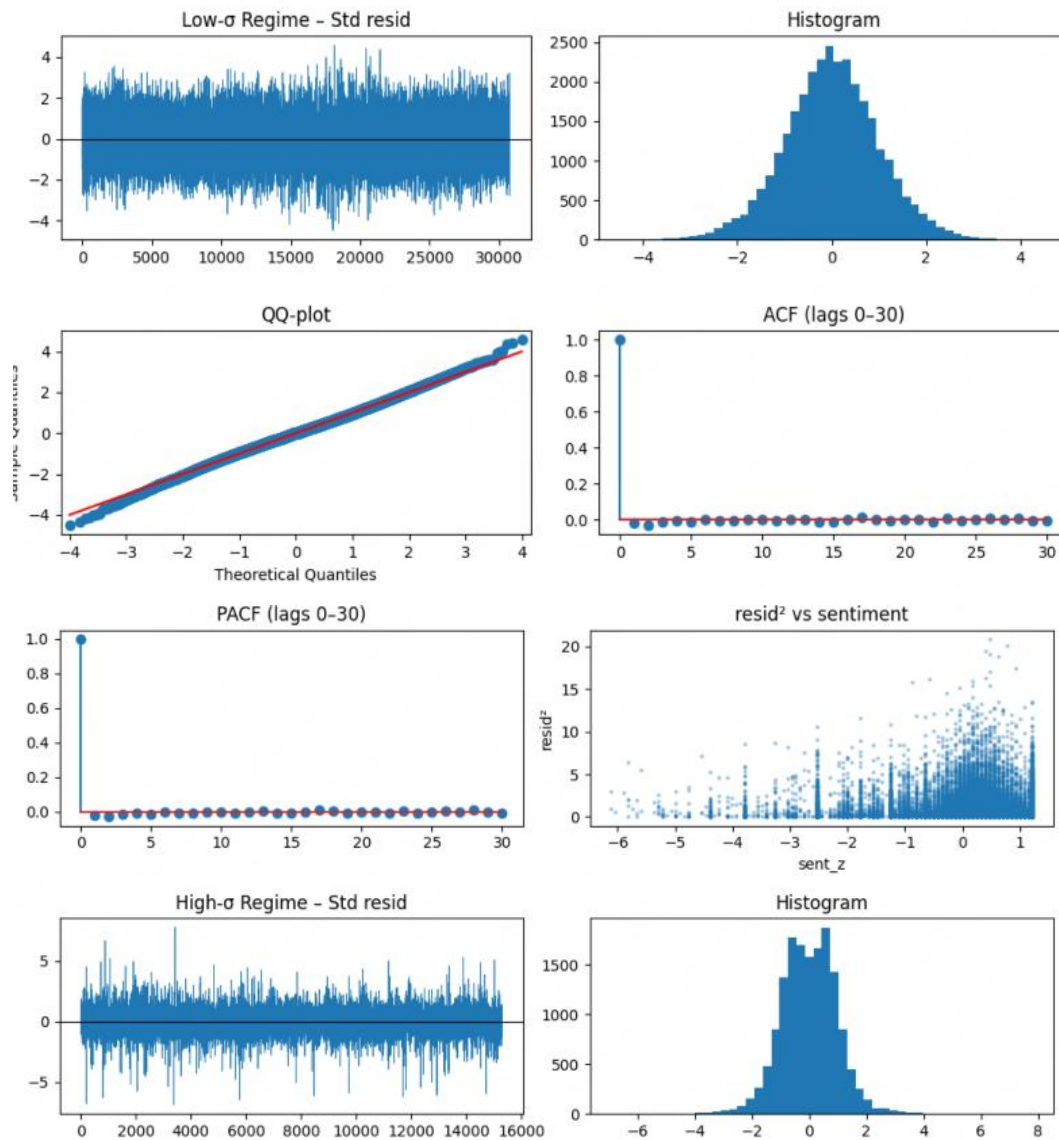
- Almgren, R. & Chriss, N. (2001). Optimal execution of portfolio transactions. *Journal of Risk*, 3(2), pp.5–39. <https://doi.org/10.21314/JOR.2001.041>.
- Alexander, C. (2008). *Market Risk Analysis, Volume IV: Value at Risk Models*. Chichester: John Wiley & Sons.
- Aparicio, J. Romao, M & Costa (2022), Predicting Bitcoin prices : The effect of interest rate, search on the internet, and energy prices, *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, Madrid, Spain, pp. 1-5, <https://doi.org/10.23919/CISTI54924.2022.9820085>
- Baur, D.G. & Dimpfl, T. (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters*, 173, pp.148–151. <https://doi.org/10.1016/j.econlet.2018.10.002>
- Baker, M. & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), pp.1645–1680. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>
- Basel Committee on Banking Supervision (2019). *Minimum capital requirements for market risk*. Basel: Bank for International Settlements. <https://www.bis.org/bcbs/publ/d457.pdf>.
- Baruník, J. & Křehlík, T. (2018). Measuring the frequency dynamics of financial connectedness and systemic risk. *Journal of Financial Econometrics*, 16(2), pp.271–296. <https://doi.org/10.1093/jjfinec/nby001>.
- Black, F. & Scholes, M. (1973). Pricing of options and corporate liabilities. *Journal of*

- Political Economy*, 81(3), pp.637–654. <https://doi.org/10.1086/260062>.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3), pp.307–327. [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1).
- Box, G.E.P. & Jenkins, G.M. (1976). *Time-Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), pp.223–236. <https://doi.org/10.1080/713665670>.
- Conrad, C. & Loch, K. (2015). Anticipating long-term stock market volatility. *Journal of Applied Econometrics*, 30(7), pp.1090–1114. <https://doi.org/10.1002/jae.2404>.
- Corbet, S., Lucey, B., Urquhart, A. & Yarovaya, L. (2020). Cryptocurrencies as a financial asset: A systematic analysis. *International Review of Financial Analysis*, 71, 101554. <https://doi.org/10.1016/j.irfa.2020.101554>.
- Costa, C. J. (2024). DeFi: Concepts and Ecosystem. *arXiv preprint arXiv:2412.01357*. <https://doi.org/10.48550/arXiv.2412.01357>
- Christoffersen, P.F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), pp.841–862. <https://doi.org/10.2307/2527341>.
- Duffie, D., Pan, J. & Singleton, K.J. (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6), pp.1343–1376. <https://doi.org/10.1111/1468-0262.00164>.
- Dyhrberg, A.H. (2016). Bitcoin, gold and the dollar – A GARCH volatility analysis. *Finance Research Letters*, 16, pp.85–92. <https://doi.org/10.1016/j.frl.2015.10.008>.
- Daglish, T., Hull, J. & Suo, W. (2007). Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance*, 7(5), pp.507–524. <https://doi.org/10.1080/14697680601087883>.
- Derman, E. & Kani, I. (1994). Riding on a smile. *Risk*, 7(2), pp.32–39. https://www.researchgate.net/publication/228924943_Riding_on_a_Smile.
- Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1), pp.18–20. https://www.researchgate.net/publication/228792324_Pricing_with_a_smile.
- Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), pp.987–1007. <https://doi.org/10.2307/1912773>.
- Engle, R.F. & Ng, V.K. (1993). Measuring and testing the impact of news on volatility. *Journal of Finance*, 48(5), pp.1749–1778. <https://doi.org/10.1111/j.1540-6261.1993.tb05127.x>.
- García, D. & Schweitzer, F. (2015). Social signals and algorithmic trading of Bitcoin. *Royal Society Open Science*, 2(9), 150288. <https://doi.org/10.1098/rsos.150288>.
- Giacomini, R. & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, 74(6), pp.1545–1578. <https://doi.org/10.1111/j.1468-0262.2006.00718.x>.

- 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), pp.1779–1801. <https://doi.org/10.1111/j.1540-6261.1993.tb05128.x> .
- Gkillas, K. & Longin, F. (2020). Extreme value theory in finance: A review. *Journal of Economic Surveys*, 34(3), pp.470–517. <https://doi.org/10.1111/joes.12363>
- Haas, M., Mittnik, S. & Paolella, M.S. (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics*, 2(4), pp.493–530. <https://doi.org/10.1093/jjfinec/nbh020>
- Hamilton, J.D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), pp.357–384. <https://doi.org/10.2307/1912559>
- Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatility. *The Journal of Finance*, 42(2), pp.281–300. <https://doi.org/10.1111/j.1540-6261.1987.tb02568.x>].
- Heston, S.L. (1993). A closed-form solution for options with stochastic volatility. *The Review of Financial Studies*, 6(2), pp.327–343. <https://doi.org/10.1093/rfs/6.2.327>
- Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3, 3415. <https://doi.org/10.1038/srep03415>.
- Kupiec, P.H. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2), pp.73–84. <https://doi.org/10.3905/jod.1995.407942>.
- Kim, J., Lee, K. & Park, J. (2020). Volatility parity in cryptocurrency portfolios. *Finance Research Letters*, 35, 101310. <https://doi.org/10.1016/j.frl.2019.101310>
- Mai, F., Shan, Z., Bai, Q., Wang, X. & Chiang, R.H.L. (2018). How does social media impact Bitcoin value? A test of the silent majority hypothesis. *Journal of Management Information Systems*, 35(1), pp.19–52. <https://doi.org/10.1080/07421222.2018.1440774>
- 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), pp.271–300. [https://doi.org/10.1016/S0927-5398\(00\)00012-8](https://doi.org/10.1016/S0927-5398(00)00012-8)
- Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2), pp.347–370. Available from: <https://doi.org/10.2307/2938260>
- Smales, L. (2022). Investor attention in cryptocurrency markets. *International Review of Financial Analysis*, 79, Article 101972. <https://doi.org/10.1016/j.irfa.2021.101972>
- Taleb, N.N. (2020). *Statistical consequences of fat tails: Real world presumptions, epistemology, and applications*. New York: STEM Academic Press.

APPENDICES

Appendix A: Data and Results



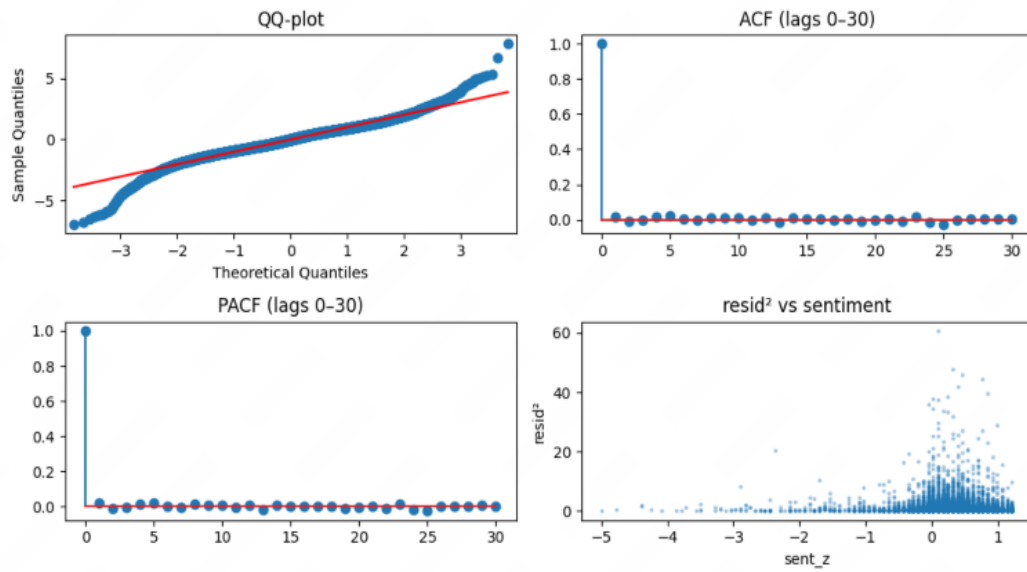
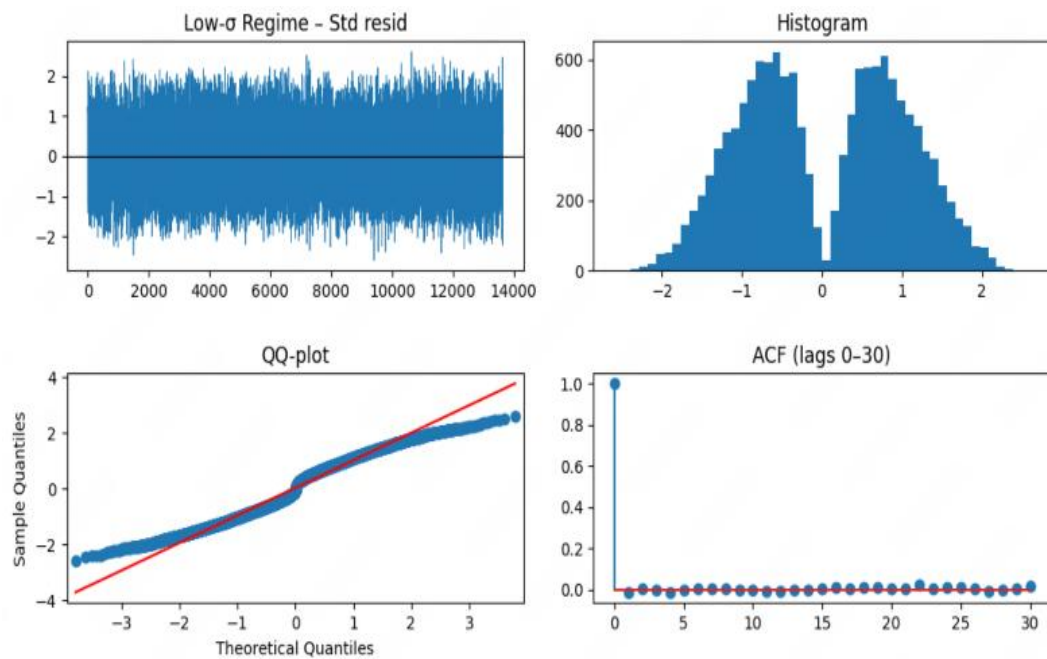


Figure 9 - Residual Diagnostics for LINK: Low- and High-Volatility Regimes



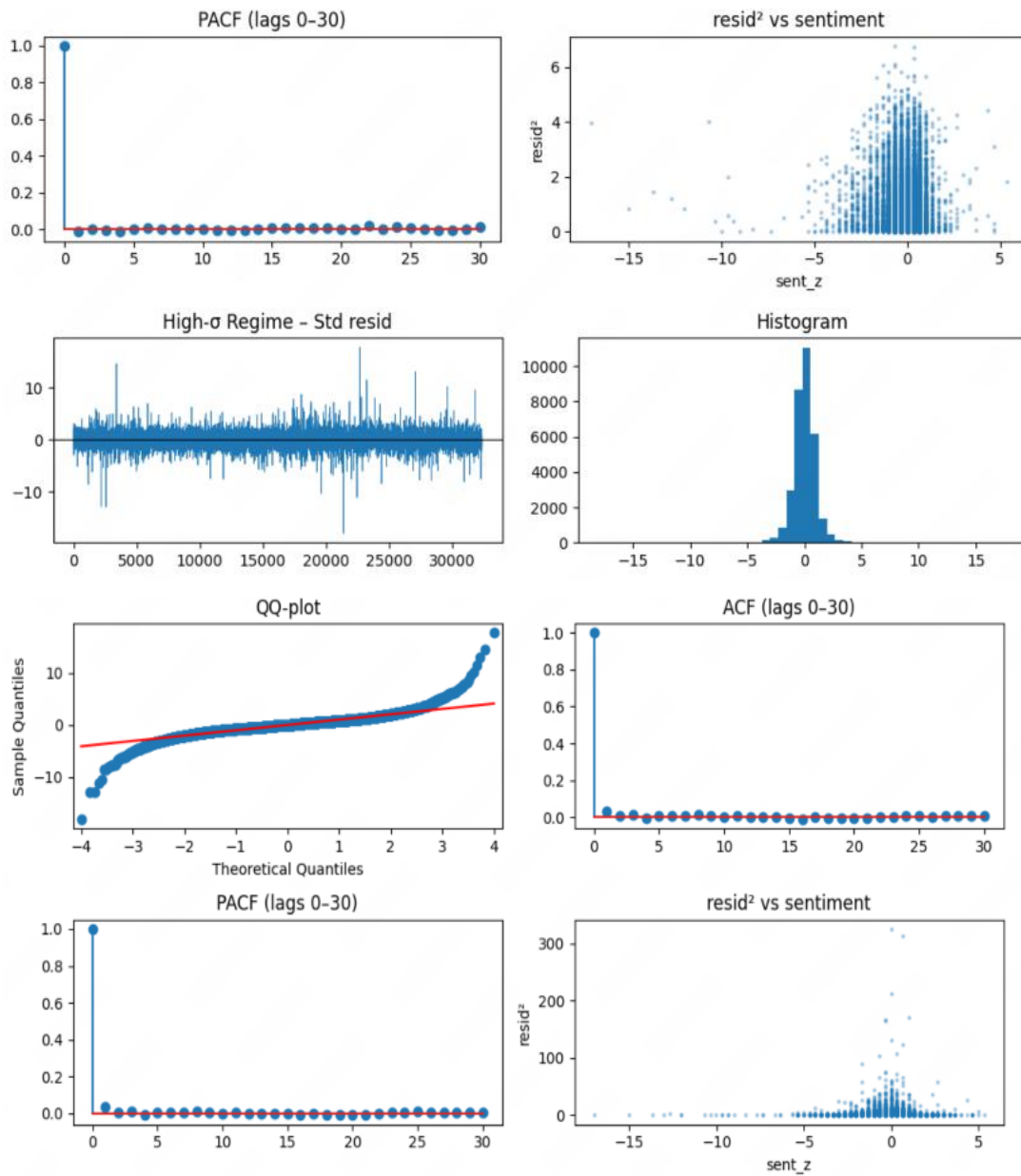
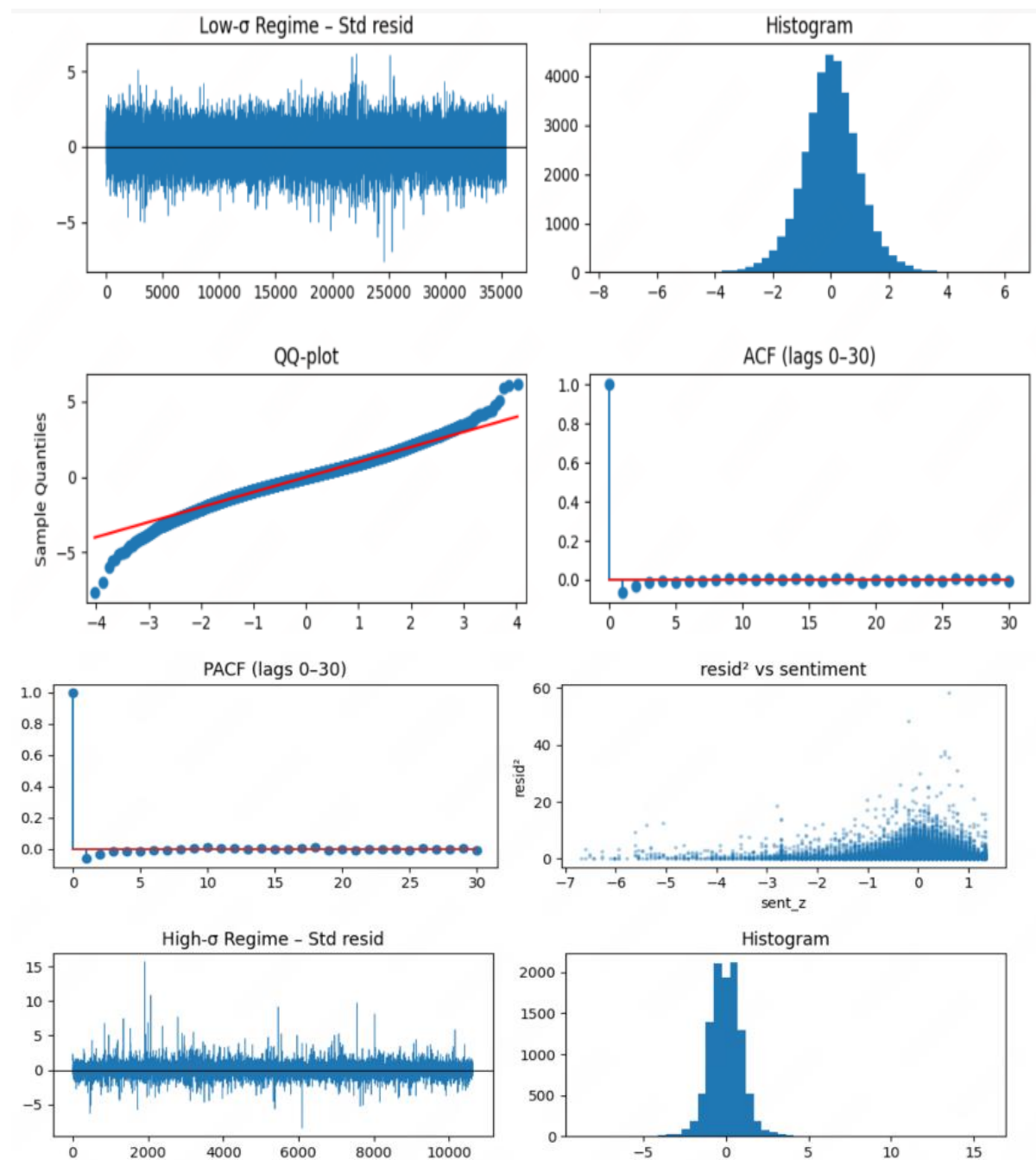


Figure 10 - Residual Diagnostics for ETH: Low- and High-Volatility Regimes



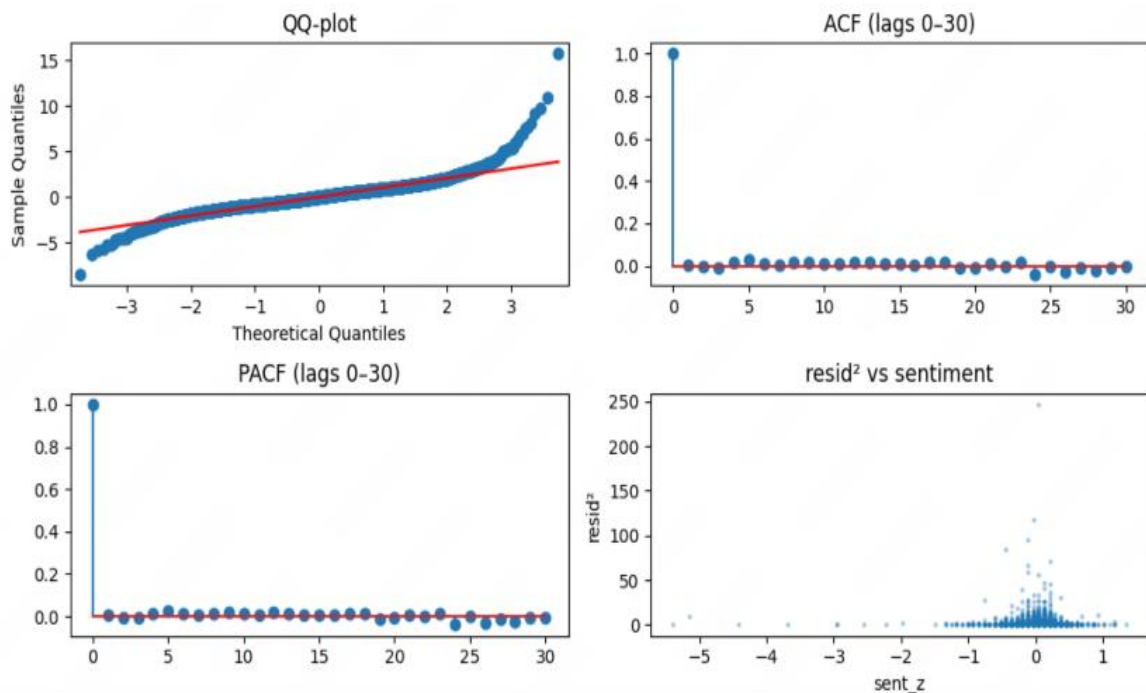
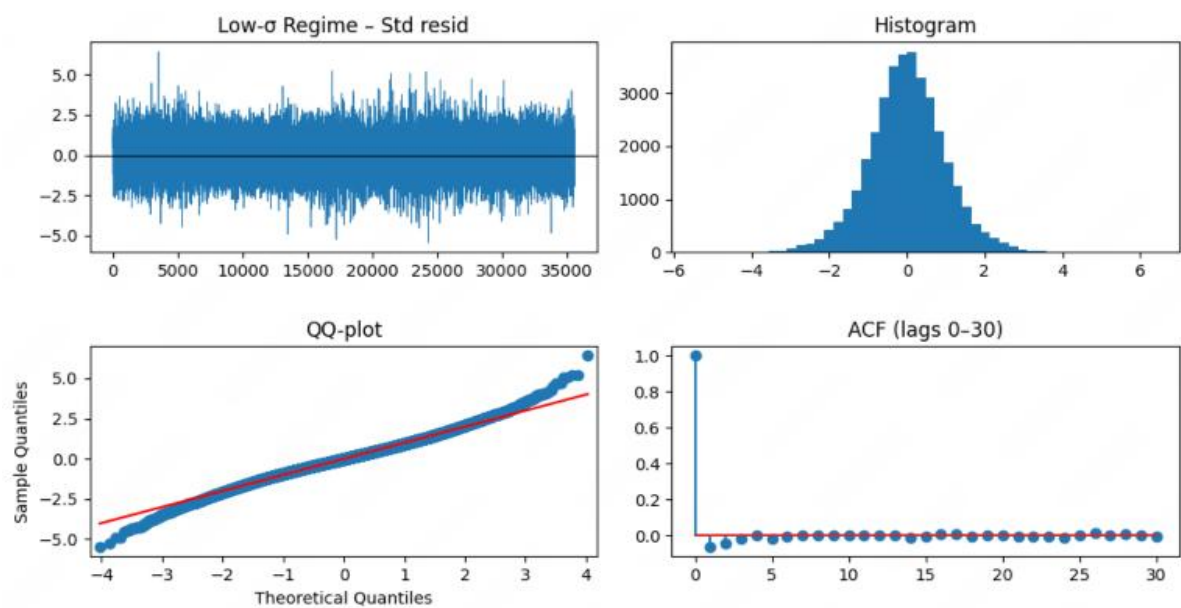


Figure 11 - Residual Diagnostics for DOGE: Low- and High-Volatility Regimes



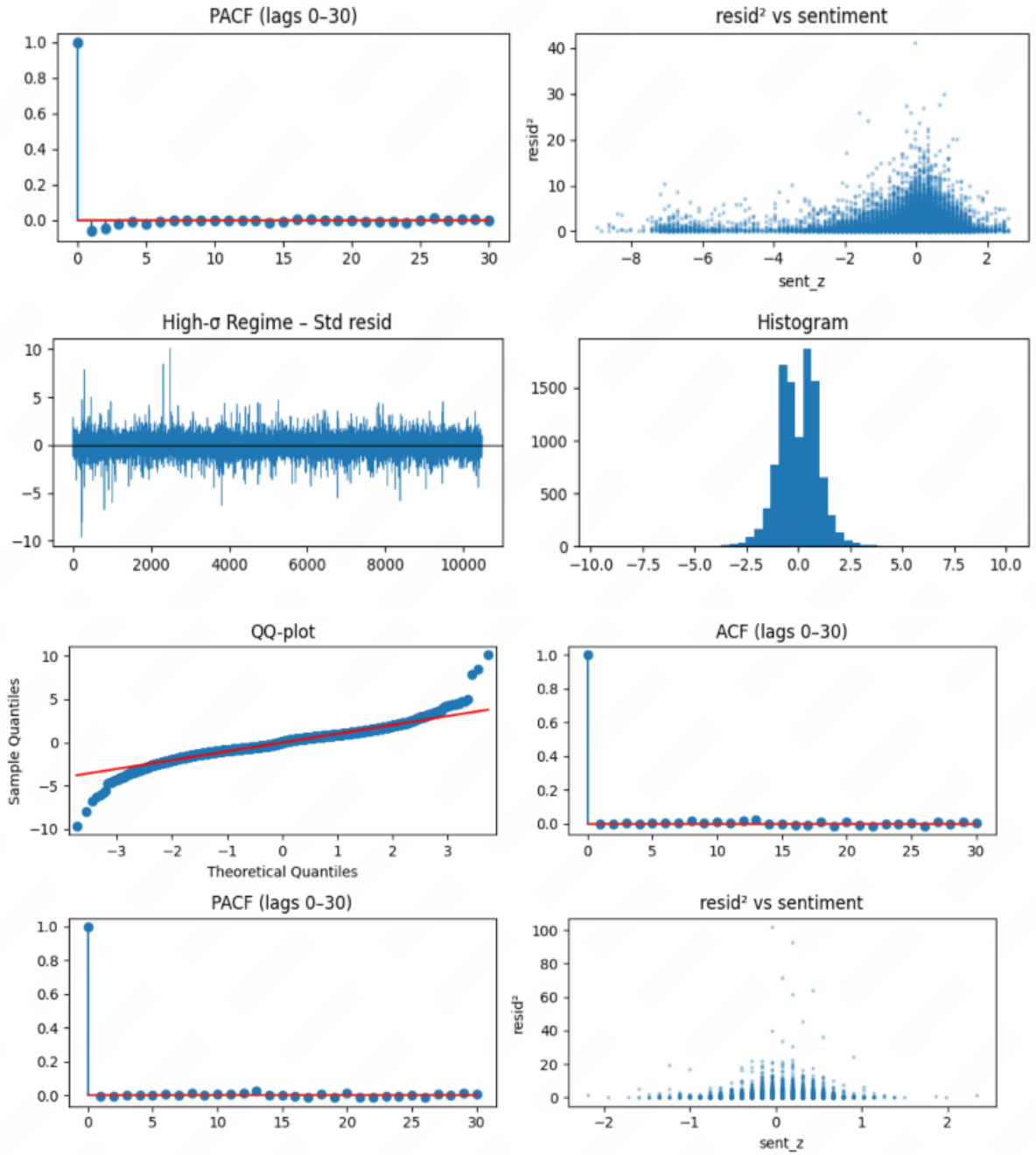


Figure 12 - Residual Diagnostics for BTC: Low- and High-Volatility Regimes

Appendix B: Programming

This appendix outlines how the cryptocurrency analysis was implemented using Python. All programming work was done in Jupyter Notebook, and the core dataset—sourced from LunarCrush which contains high-frequency sentiment and price data for five cryptocurrencies: Bitcoin (BTC), Ethereum (ETH), Chainlink (LINK), Dogecoin (DOGE), and Solana (SOL). The time series spans from 1 January 2020 to 2 April 2025, with most variables recorded on an hourly basis. The dataset includes key features such as Galaxy Score, AltRank, closing price, market cap, and custom sentiment scores,

which were used to build and evaluate volatility models across different coins.

The data was first imported from a structured Excel file, which included all five coins laid out side by side. Because the formatting required manual parsing, I split each coin into its own DataFrame and aligned columns like timestamp, price, Galaxy Score, and sentiment into a consistent long format time series. After checking for missing values and duplicated timestamps, I resampled the data to a daily frequency for modelling purposes which was used by daily means for continuous variables like price and Galaxy Score. Lagged variables were also created to reflect the natural delay between changes in sentiment and resulting market movements.

Model construction was done in two stages. The first stage used standard GARCH-X models, with sentimental features included in the variance equation. I used the arch package to fit these models on the log returns of each asset. The second stage extended this by incorporating a Markov regime-switching structure into the GARCH-X framework, allowing for dynamic shifts between high and low volatility states. For this, I built a custom implementation using hmmlearn to estimate state transitions and combined it with GARCH estimation to update volatility in each regime. Both types of models were evaluated using a rolling one step ahead to forecast procedure, and performance was measured using RMSE and RMSSE. These were computed across different forecast windows within the 2020–2025 range. Residual analysis and diagnostic tests were also conducted to check for autocorrelation and distributional assumptions.

In the final stage, I explored how outliers, especially during the early COVID-19 period and the 2021 bull run affected model performance. Two anomaly windows were defined: March 2020 to June 2021, and March 2020 to December 2021. Six different outlier adjustment methods were applied, ranging from simple mean-replacement to local smoothing, and the GARCH-X model was re-estimated on the cleaned data. Forecast accuracy was compared before and after each adjustment using RMSE. The results helped identify which cleaning strategy worked best for each coin and whether adjusting for sentiment driven outliers improved the stability of forecasts.

Overall, the entire workflow was modular and repeatable. All scripts were built from scratch and structured to allow easy switching between coins. The analysis relied on several core Python libraries including pandas, numpy, arch, matplotlib, statsmodels, and hmmlearn. The combination of sentiment driven features from LunarCrush and flexible volatility models formed the basis for the empirical results shown in Chapter 6.