

Integrated Market Risk and Regulatory Backtesting

An Empirical Study Based on Bitcoin Returns

Author: *Jiajun Tang*

Institution: Columbia University

Note: This study is developed based on the instructional framework and course materials of IEOR 4745 *Applied Financial Risk Management* at Columbia University. The analysis is conducted strictly for academic learning, research discussion, and educational demonstration purposes. Any form of commercial use is not permitted.

Summary

This project develops an integrated financial risk management framework for Bitcoin, combining volatility modeling, market risk measurement, tail risk assessment, and regulatory-style backtesting within a unified analytical pipeline. Daily Bitcoin log returns are constructed over a ten-year horizon, and conditional volatility is modeled using an Exponentially Weighted Moving Average (EWMA) approach to capture volatility clustering and persistence.

Based on EWMA volatility estimates, both parametric and historical simulation methods are employed to estimate Value at Risk (VaR) and Expected Shortfall (ES) at multiple confidence levels. The results show that parametric VaR under normality assumptions systematically understates tail risk, while historical simulation produces substantially larger risk estimates, reflecting the heavy-tailed nature of Bitcoin returns. Expected Shortfall consistently exceeds VaR and provides a more informative measure of extreme loss severity.

Regulatory-style backtesting using the Kupiec unconditional coverage test reveals that the EWMA-based parametric VaR model achieves acceptable coverage at moderate confidence levels but fails under extreme confidence levels, particularly for short positions. Overall, the findings highlight the limitations of classical parametric risk models when applied to cryptocurrencies and underscore the importance of tail-sensitive measures and empirical validation in managing Bitcoin market risk.

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 2 |
| 2 | Data and Return Construction | 3 |
| 2.1 | Data Source and Sample | 3 |
| 2.2 | Return Definition | 3 |
| 2.3 | Descriptive Statistics and Distributional Properties | 3 |
| 3 | Volatility Modeling | 4 |
| 3.1 | Bitcoin Returns and Volatility Clustering | 4 |
| 3.2 | EWMA Conditional Volatility | 5 |
| 3.3 | Autocorrelation Evidence | 6 |
| 3.4 | Shock–Volatility Relationship | 8 |
| 4 | Market Risk Measurement: Value at Risk | 8 |
| 4.1 | Parametric Value at Risk (EWMA–Normal) | 8 |
| 4.2 | Historical Simulation Value at Risk | 9 |
| 4.3 | Comparison of VaR Scenarios in Return Space | 9 |
| 5 | Tail Risk Measurement: Expected Shortfall | 10 |
| 5.1 | Parametric Expected Shortfall (EWMA–Normal) | 10 |
| 5.2 | Historical Simulation Expected Shortfall | 10 |
| 5.3 | Comparison of Tail Risk Across Methods | 11 |
| 6 | Regulatory Backtesting: Unconditional Coverage Test | 11 |
| 6.1 | Backtesting Setup | 11 |
| 6.2 | Kupiec Unconditional Coverage Results | 11 |
| 6.3 | Interpretation and Regulatory Implications | 12 |

1 Introduction

Over the past decade, Bitcoin has evolved from a niche digital experiment into a globally traded financial asset with substantial market capitalization and liquidity. Despite its growing adoption by institutional investors, Bitcoin remains characterized by extreme volatility, pronounced tail risk, and non-Gaussian return distributions. These features pose significant challenges for traditional risk management frameworks that were originally developed for more stable financial instruments such as equities, bonds, and foreign exchange.

Effective risk management requires not only the measurement of potential losses but also an understanding of how those losses behave under extreme market conditions and whether risk models remain reliable from a regulatory perspective. In conventional financial institutions, market risk is commonly quantified using Value at Risk (VaR), while tail risk is increasingly assessed using Expected Shortfall (ES). Regulatory frameworks further require that these risk measures be validated through formal backtesting procedures to ensure their adequacy in capturing realized losses. However, the applicability and robustness of such methodologies in the context of cryptocurrencies—particularly Bitcoin—remain open questions.

This project aims to construct an integrated risk modeling framework for Bitcoin that combines market risk measurement, tail risk assessment, and regulatory-style validation into a coherent analytical pipeline. Rather than treating individual risk measures in isolation, the analysis emphasizes the logical progression from return construction and volatility modeling to loss quantification and model evaluation. In doing so, the project demonstrates how classical financial risk management tools can be adapted, tested, and critically assessed when applied to a highly volatile and non-traditional asset.

Specifically, this study addresses the following research question:

How can an integrated framework combining market risk measures, tail risk metrics, and regulatory backtesting be constructed and evaluated for Bitcoin?

To answer this question, the project proceeds in a structured manner. First, daily Bitcoin returns are constructed and explored to highlight key empirical features such as volatility clustering and fat-tailed behavior. Next, conditional volatility is modeled using an Exponentially Weighted Moving Average (EWMA) approach, providing the foundation for parametric risk estimation. Building on this volatility model, both parametric and historical simulation approaches are employed to estimate Value at Risk at multiple confidence levels. Expected Shortfall is then introduced as a complementary tail risk measure that captures the severity of losses beyond the VaR threshold. Finally, the reliability of the VaR models is assessed through an unconditional coverage backtesting framework, reflecting regulatory standards commonly applied in financial risk supervision.

The contributions of this project are threefold. First, it provides a clear and reproducible risk modeling pipeline tailored to Bitcoin, illustrating how traditional market risk tools perform in a cryptocurrency setting. Second, it offers a comparative analysis of parametric and non-parametric approaches to risk measurement, highlighting their respective strengths and limitations in the presence of heavy-tailed returns. Third, by incorporating regulatory-style backtesting, the study bridges the gap between theoretical risk modeling and practical risk governance considerations.

2 Data and Return Construction

2.1 Data Source and Sample

This study uses a daily time series of the U.S. dollar closing price of Bitcoin obtained from Bloomberg (ticker: XBTUSD BGN Curncy) and provided in the course data file `bitcoin.xlsx`. The full historical sample spans from 31 July 2015 to 31 July 2025.

The empirical analysis is conducted over two horizons:

- **Full sample (10 years):** prices from 31 July 2015 to 31 July 2025, generating 3421 daily log-return observations from 02 August 2015 to 31 July 2025.
- **Recent 2-year window:** prices from 31 July 2023 to 31 July 2025, generating 731 daily log-return observations from 01 August 2023 to 31 July 2025.

This two-window design is intentional. The full sample provides a long-run characterization of Bitcoin's return distribution, while the recent window is aligned with the risk forecasting and backtesting exercises conducted in later sections.

2.2 Return Definition

Let S_t denote the Bitcoin closing price on trading day t . The daily logarithmic price return is defined as

$$r_t = \ln \left(\frac{S_t}{S_{t-1}} \right).$$

Log returns are used because they aggregate additively over time and constitute the standard input for volatility modeling and parametric risk estimation. Throughout this paper, returns are computed in decimal form for calculation and are reported in percent where appropriate for interpretability (e.g., “3.59%” rather than “0.0359”).

2.3 Descriptive Statistics and Distributional Properties

To summarize Bitcoin's short-horizon risk characteristics, descriptive statistics are computed for the log-return series under consistent definitions. Specifically:

- Root mean square (RMS): $\sqrt{\frac{1}{T} \sum_{t=1}^T r_t^2}$.
- Bias-corrected standard deviation: $\sqrt{\frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})^2}$.
- q -quantile: the $\lceil qT \rceil$ -th order statistic using the ceiling rule.
- Skewness and kurtosis: moment-based coefficients computed using T observations, where kurtosis is reported as raw kurtosis (not excess kurtosis).

Table 1 reports descriptive statistics for daily Bitcoin log returns. All return statistics are expressed in percent.

Table 1: Descriptive Statistics of Daily Bitcoin Log Returns

| Statistic | Entire sample (08/02/2015–07/31/2025) | Past 2 years (08/01/2023–07/31/2025) |
|--------------------------|---------------------------------------|--------------------------------------|
| Number of observations | 3421 | 731 |
| RMS of returns (%) | 3.5922 | 2.4653 |
| Std. dev. of returns (%) | 3.5885 | 2.4598 |
| Minimum return (%) | -31.7282 | -10.0112 |
| Maximum return (%) | 19.0903 | 9.6377 |
| 0.01 quantile (%) | -10.6467 | -6.0042 |
| 0.05 quantile (%) | -5.5390 | -4.0437 |
| 0.95 quantile (%) | 5.8358 | 4.5310 |
| 0.99 quantile (%) | 10.2669 | 6.9048 |
| Skewness | -0.3447 | 0.1189 |
| Kurtosis | 8.7475 | 4.5572 |

These results highlight several empirical features that directly motivate the modeling choices in later sections. First, Bitcoin exhibits extremely high daily volatility, and the RMS is nearly identical to the standard deviation, indicating that short-horizon risk is dominated by dispersion rather than mean drift. Second, the return distribution is strongly fat-tailed, with kurtosis far exceeding the Gaussian benchmark of 3, and tail thickness is more pronounced in the full sample than in the recent window. Third, full-sample returns display negative skewness, consistent with the presence of relatively more severe downside movements. Collectively, these distributional properties provide a quantitative foundation for the volatility modeling, VaR and ES estimation, and regulatory-style backtesting developed in subsequent sections.

3 Volatility Modeling

3.1 Bitcoin Returns and Volatility Clustering

Figure 1 presents daily Bitcoin log returns. Returns fluctuate around zero, indicating the absence of a persistent mean component at the daily horizon. The series exhibits extremely large return magnitudes, with frequent observations exceeding $\pm 5\%$ and several

extreme downside events below -30% , highlighting substantial market and tail risk. Large absolute returns are temporally clustered rather than randomly dispersed, providing clear visual evidence of volatility clustering and time-varying variance.

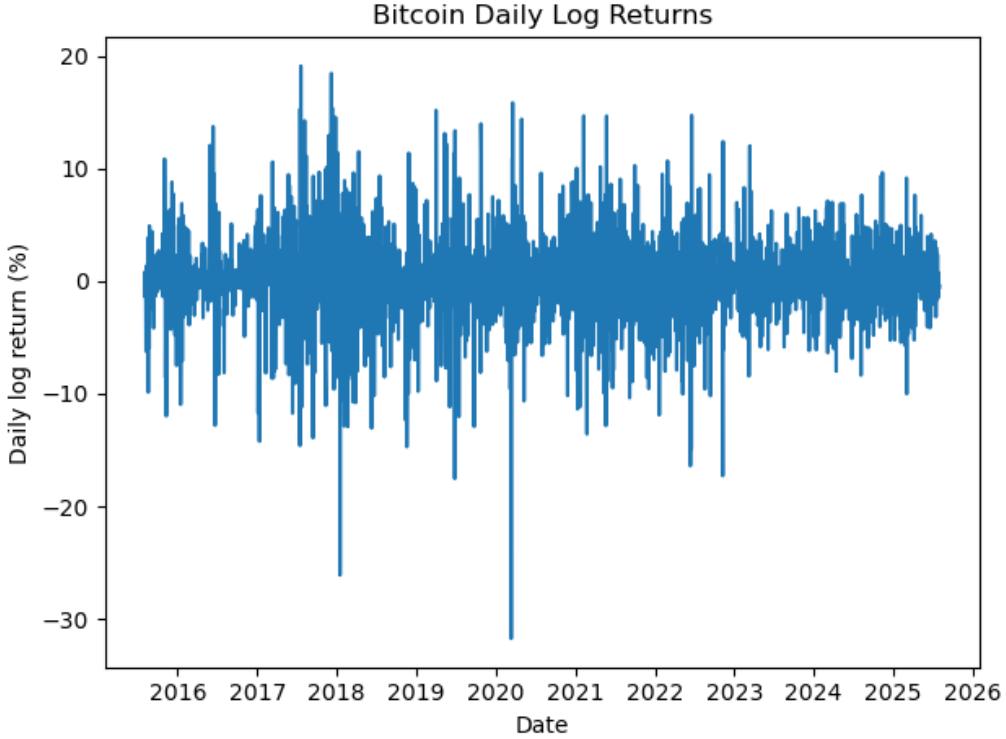


Figure 1: Bitcoin Daily Log Returns

3.2 EWMA Conditional Volatility

Figure 2 displays the Exponentially Weighted Moving Average (EWMA) conditional volatility with decay parameter $\lambda = 0.94$. Conditional volatility varies substantially over time and responds sharply to large return shocks, confirming the presence of time-varying risk in Bitcoin returns. Periods of elevated volatility persist over extended intervals, indicating strong volatility persistence consistent with the high decay parameter.

This behavior demonstrates that EWMA effectively captures both rapid shock absorption and gradual volatility decay, making it suitable as an input for parametric risk measures.

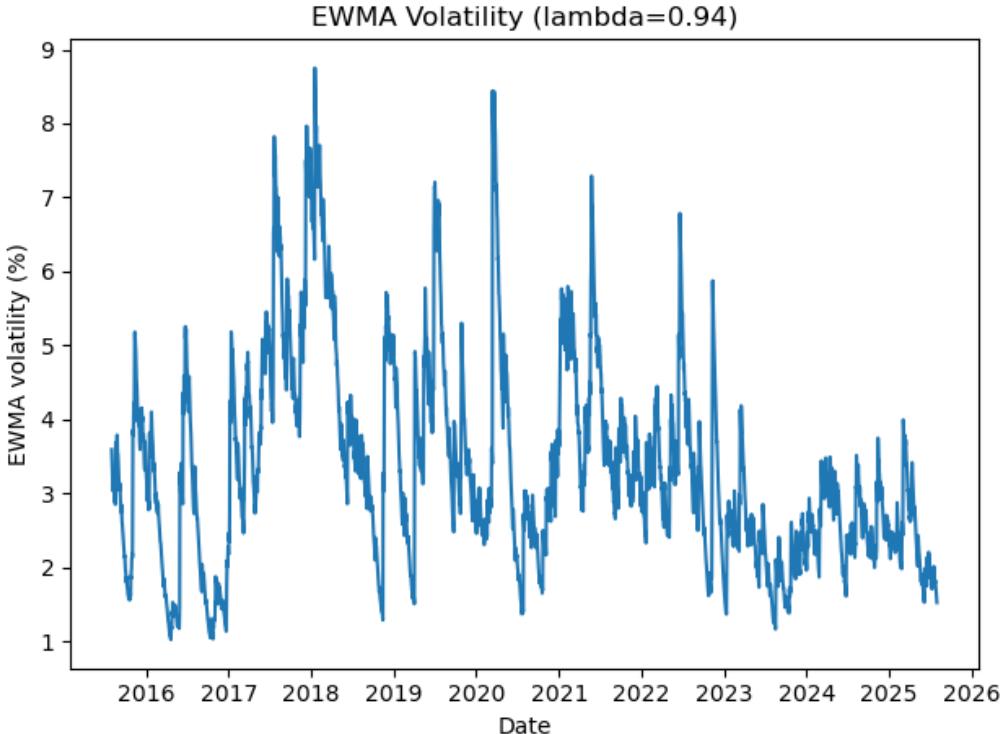


Figure 2: EWMA Conditional Volatility ($\lambda = 0.94$)

3.3 Autocorrelation Evidence

The autocorrelation of squared returns remains positive across the first ten lags, providing clear evidence of volatility clustering in Bitcoin returns despite weak linear dependence in raw returns. In contrast, the autocorrelation of EWMA variance is extremely high and decays slowly, remaining above 0.80 even at lag 10, indicating a highly persistent conditional volatility process.

Table 2: Autocorrelation of Squared Returns and EWMA Variance

| Lag | ACF of r_t^2 | ACF of σ_t^2 |
|-----|----------------|---------------------|
| 1 | 0.1493 | 0.9830 |
| 2 | 0.0742 | 0.9642 |
| 3 | 0.0851 | 0.9464 |
| 4 | 0.1404 | 0.9290 |
| 5 | 0.1309 | 0.9099 |
| 6 | 0.0648 | 0.8894 |
| 7 | 0.1182 | 0.8699 |
| 8 | 0.0552 | 0.8493 |
| 9 | 0.0734 | 0.8298 |
| 10 | 0.0606 | 0.8109 |

Figures 3 and 4 provide a visual representation of these autocorrelation patterns.

Together, they confirm that Bitcoin returns exhibit clustered volatility and that the EWMA model effectively captures the persistence of risk over time, supporting its use in subsequent parametric risk measurement.

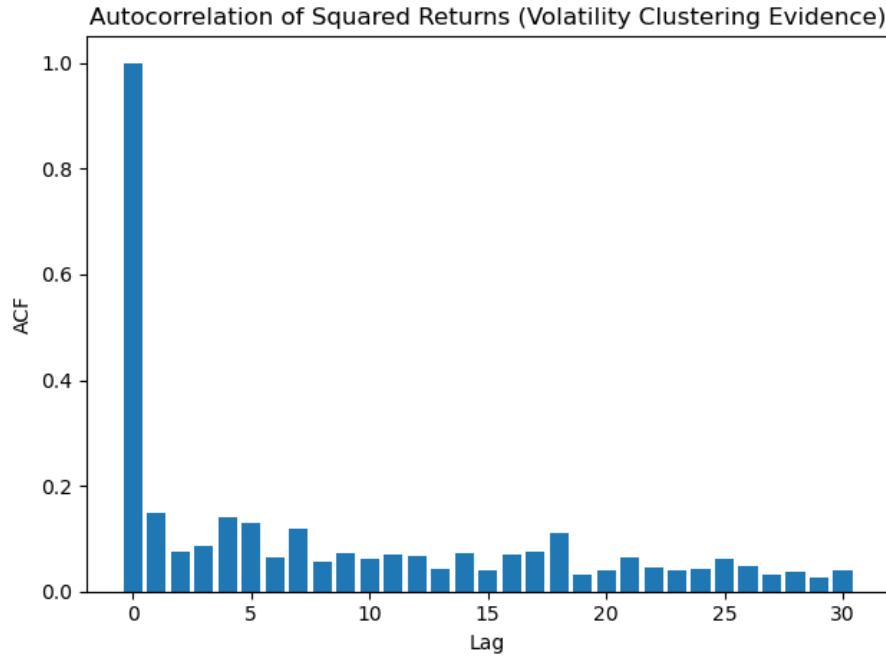


Figure 3: Autocorrelation of Squared Returns

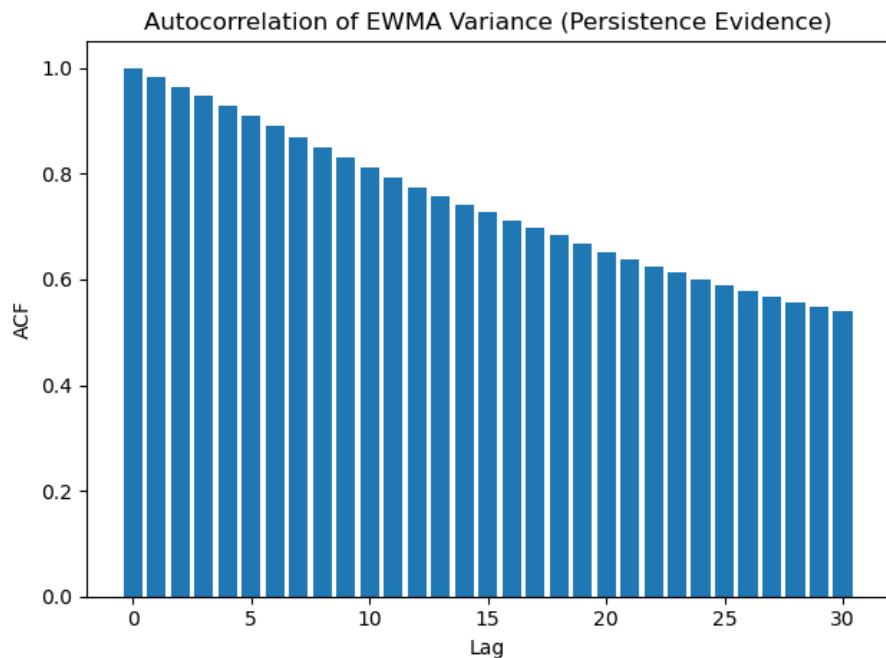


Figure 4: Autocorrelation of EWMA Variance

3.4 Shock–Volatility Relationship

Table 3 summarizes the relationship between return shocks and EWMA volatility. The correlation between lagged squared returns and contemporaneous EWMA variance equals 0.4044, indicating that large return shocks translate directly into higher subsequent volatility estimates. Moreover, average conditional volatility increases from 4.703% one day before extreme shocks to 6.161% within five days after such events, indicating a rapid and economically significant volatility response.

Table 3: Shock–Volatility Relationship and EWMA Output Summary

| Panel A: Shock–Volatility Statistics | |
|--|---------------------------------|
| Corr(r_{t-1}^2, σ_t^2) | 0.4044 |
| Avg. EWMA volatility (%) one day before extreme shock | 4.703 |
| Avg. EWMA volatility (%) five days after extreme shock | 6.161 |
| Panel B: EWMA Output (First 5 Observations) | |
| Date | ($r_t, \sigma_t^2, \sigma_t$) |
| 2015-08-02 | (-0.010949, 0.001290, 0.035922) |
| 2015-08-03 | (0.008135, 0.001220, 0.034931) |
| 2015-08-04 | (0.004989, 0.001151, 0.033926) |
| 2015-08-05 | (-0.008730, 0.001083, 0.032915) |
| 2015-08-06 | (-0.014315, 0.001023, 0.031984) |

These results demonstrate that the EWMA model is highly sensitive to extreme return realizations and effectively captures the dynamic transmission of market shocks into future risk levels, providing a robust volatility input for the VaR and ES analysis developed in subsequent sections.

4 Market Risk Measurement: Value at Risk

4.1 Parametric Value at Risk (EWMA–Normal)

Using the EWMA conditional volatility estimate as of 31 July 2025 ($\sigma = 1.48199\%$), parametric Value at Risk is computed under the normality assumption for a portfolio value of \$1,000,000. VaR increases monotonically with the confidence level and is asymmetric between long and short positions due to the lognormal return transformation.

At the 99.5% confidence level, the exact parametric VaR reaches approximately \$37,454 for a long position and \$38,911 for a short position. The approximation-based VaR closely tracks the exact lognormal VaR but slightly overestimates losses at higher confidence levels. This difference becomes more pronounced as tail risk increases, reflecting the nonlinear effect of the exponential return transformation.

Table 4: Parametric Value at Risk (as of 31 July 2025, Portfolio Value = \$1,000,000)

| Confidence level | Long VaR (exact, \$) | Short VaR (exact, \$) | VaR (approx, \$) | Log-return scenario (%) |
|------------------|----------------------|-----------------------|------------------|-------------------------|
| 95.0% | 24,081.82 | 24,676.06 | 24,376.53 | 2.4377 |
| 99.0% | 33,888.66 | 35,077.38 | 34,476.19 | 3.4476 |
| 99.5% | 37,454.05 | 38,911.44 | 38,173.47 | 3.8173 |

4.2 Historical Simulation Value at Risk

Historical simulation VaR is computed using the most recent two-year window comprising 731 daily returns. Empirical quantile ranks are determined using the ceiling rule, ensuring consistency with regulatory-style historical VaR implementation.

Compared with the parametric approach, historical simulation produces substantially larger VaR estimates at all confidence levels. At the 99.5% confidence level, historical VaR reaches approximately \$66,403 for a long position and \$79,483 for a short position, significantly exceeding the corresponding parametric estimates. This discrepancy reflects the presence of heavy tails and extreme observations in Bitcoin returns, which are not fully captured by the normal distribution assumption.

Table 5: Historical Simulation Value at Risk (Last 2 Years, Portfolio Value = \$1,000,000)

| Confidence level | Rank (long) | Long VaR (\$) | Rank (short) | Short VaR (\$) |
|------------------|-------------|---------------|--------------|----------------|
| 95.0% | 37 | 39,630.72 | 695 | 46,351.86 |
| 99.0% | 8 | 58,275.38 | 724 | 71,488.12 |
| 99.5% | 4 | 66,403.16 | 728 | 79,482.71 |

4.3 Comparison of VaR Scenarios in Return Space

Comparing VaR scenarios in log-return space highlights systematic differences between parametric and historical approaches. Historical quantiles are consistently more extreme than parametric scenarios at all confidence levels, particularly in the upper tail.

For example, at the 99.5% confidence level, the historical log-return quantile exceeds 7.6% for short positions, while the parametric scenario remains below 3.9%. These results indicate that parametric VaR materially understates tail risk for Bitcoin, whereas historical simulation more fully reflects empirical extreme movements.

Table 6: VaR Scenarios in Log-Return Terms

| Confidence level | Historical (long, %) | Parametric (%) | Historical (short, %) |
|------------------|----------------------|----------------|-----------------------|
| 95.0% | -4.0437 | 2.4377 | 4.5310 |
| 99.0% | -6.0042 | 3.4476 | 6.9048 |
| 99.5% | -6.8711 | 3.8173 | 7.6482 |

5 Tail Risk Measurement: Expected Shortfall

5.1 Parametric Expected Shortfall (EWMA–Normal)

Based on the EWMA conditional volatility as of 31 July 2025 ($\sigma = 1.48199\%$), parametric Expected Shortfall (ES) is computed under the normality assumption with an exact lognormal profit-and-loss transformation. At all confidence levels, ES exceeds the corresponding Value at Risk, reflecting the average severity of losses beyond the VaR threshold.

The ES-to-VaR ratio decreases as the confidence level increases, falling from approximately 1.25 at the 95% level to about 1.12 at the 99.5% level for long positions. This pattern indicates that although tail losses increase with higher confidence levels, their incremental severity relative to VaR diminishes. Similar behavior is observed for short positions, with slightly higher ES/VaR ratios reflecting asymmetric tail exposure under lognormal returns.

Table 7: Parametric Expected Shortfall (as of 31 July 2025, Portfolio Value = \$1,000,000)

| Confidence level | Long VaR (\$) | Long ES (\$) | ES / VaR | Short VaR (\$) | Short ES (\$) | ES / VaR |
|------------------|---------------|--------------|----------|----------------|---------------|----------|
| 95.0% | 24,081.82 | 30,091.97 | 1.2496 | 24,676.06 | 31,056.86 | 1.2586 |
| 99.0% | 33,888.66 | 38,718.06 | 1.1425 | 35,077.38 | 40,299.66 | 1.1489 |
| 99.5% | 37,454.05 | 41,943.88 | 1.1199 | 38,911.44 | 43,799.86 | 1.1256 |

5.2 Historical Simulation Expected Shortfall

Historical simulation ES is computed using the most recent two-year window comprising 731 daily returns. At all confidence levels, historical ES substantially exceeds parametric ES, indicating much heavier empirical tails in Bitcoin returns than implied by the normal distribution.

At the 99.5% confidence level, historical ES reaches approximately \$79,676 for long positions and \$93,700 for short positions. The corresponding ES/VaR ratios remain close to 1.20, suggesting that losses beyond VaR remain economically significant even at very high confidence levels. Compared with the parametric approach, historical simulation produces both higher absolute tail losses and higher ES/VaR ratios, underscoring the importance of non-parametric methods for assets with extreme return behavior.

Table 8: Historical Simulation Expected Shortfall (Last 2 Years, Portfolio Value = \$1,000,000)

| Confidence level | Rank (long) | Long VaR (\$) | Long ES (\$) | ES / VaR | Rank (short) | Short VaR (\$) | Short ES (\$) | ES / VaR |
|------------------|-------------|---------------|--------------|----------|--------------|----------------|---------------|----------|
| 95.0% | 37 | 39,630.72 | 51,445.89 | 1.2981 | 695 | 46,351.86 | 60,781.72 | 1.3113 |
| 99.0% | 8 | 58,275.38 | 69,724.84 | 1.1965 | 724 | 71,488.12 | 84,025.41 | 1.1754 |
| 99.5% | 4 | 66,403.16 | 79,676.10 | 1.1999 | 728 | 79,482.71 | 93,700.48 | 1.1789 |

5.3 Comparison of Tail Risk Across Methods

Comparing parametric and historical Expected Shortfall highlights systematic differences in tail risk assessment. Historical ES is consistently larger than parametric ES across all confidence levels and for both long and short positions. This divergence reflects the presence of extreme observations and fat tails in Bitcoin returns, which are not fully captured by parametric models relying on normality assumptions.

These findings confirm that Expected Shortfall provides a more informative measure of tail risk than Value at Risk, particularly for assets characterized by high volatility and non-Gaussian return distributions.

Expected Shortfall reveals substantially heavier tail risk for Bitcoin than suggested by parametric VaR, especially under historical simulation, motivating formal backtesting and regulatory evaluation in the next section.

6 Regulatory Backtesting: Unconditional Coverage Test

6.1 Backtesting Setup

Value at Risk backtesting is conducted using the most recent two-year window from 01 August 2023 to 31 July 2025. One-day-ahead parametric VaR forecasts are generated using EWMA volatility estimates based on information available at time $t - 1$. After aligning forecast volatility with realized returns, the effective backtesting sample consists of 730 observations.

Exceedances are identified separately for long and short positions, and model adequacy is evaluated using the Kupiec unconditional coverage test at the 5% significance level.

6.2 Kupiec Unconditional Coverage Results

For the 95% confidence level, the number of long-position exceedances (34) lies within the 5% acceptance interval [25, 48], and the Kupiec test fails to reject the model. In contrast, the number of short-position exceedances (50) exceeds the upper acceptance bound, leading to rejection of the VaR model for short positions.

At the 99% confidence level, the long-position exceedance count (13) remains within the acceptance interval [3, 13], with a p -value slightly above 5%, indicating borderline but acceptable coverage. However, the short-position exceedance count (20) substantially exceeds the acceptable range, resulting in strong rejection.

At the 99.5% confidence level, the VaR model fails for both long and short positions. The exceedance counts exceed the acceptance region, and the Kupiec test rejects

unconditional coverage with high statistical significance.

Overall, the backtesting results reveal pronounced asymmetry between long and short positions. While the parametric EWMA-based VaR provides adequate coverage for downside risk at lower confidence levels, it systematically underestimates upside tail risk, particularly at extreme confidence levels.

Table 9: Kupiec Unconditional Coverage Backtest (5% Test)

| Confidence level | T | Acceptance interval | Long exceedances | p -value (long) | Reject (long) | Short exceedances | p -value (short) | Reject (short) |
|------------------|-----|---------------------|------------------|-------------------|---------------|-------------------|--------------------|----------------|
| 95.0% | 730 | [25, 48] | 34 | 0.6677 | No | 50 | 0.0295 | Yes |
| 99.0% | 730 | [3, 13] | 13 | 0.0561 | No | 20 | 0.0001 | Yes |
| 99.5% | 730 | [0, 8] | 9 | 0.0181 | Yes | 16 | 0.0000 | Yes |

6.3 Interpretation and Regulatory Implications

The unconditional coverage tests indicate that the EWMA–normal VaR model performs reasonably well for downside risk at moderate confidence levels but fails to capture extreme tail risk, especially for short positions. The systematic rejection on the short side reflects the presence of heavy right tails and large positive return shocks in Bitcoin markets.

These findings underscore the limitations of parametric VaR under normality assumptions and motivate the use of tail-sensitive risk measures and supplementary validation frameworks in regulatory risk management.

Backtesting results show that while EWMA-based parametric VaR can satisfy regulatory coverage requirements at moderate confidence levels, it fails to adequately capture extreme and asymmetric tail risk in Bitcoin returns.