Using Python to Analyze Risk Metrics

A Consolidated Analysis of SPY Returns (2020-2025)

Colby Jaskowiak

Buffalo Research Group

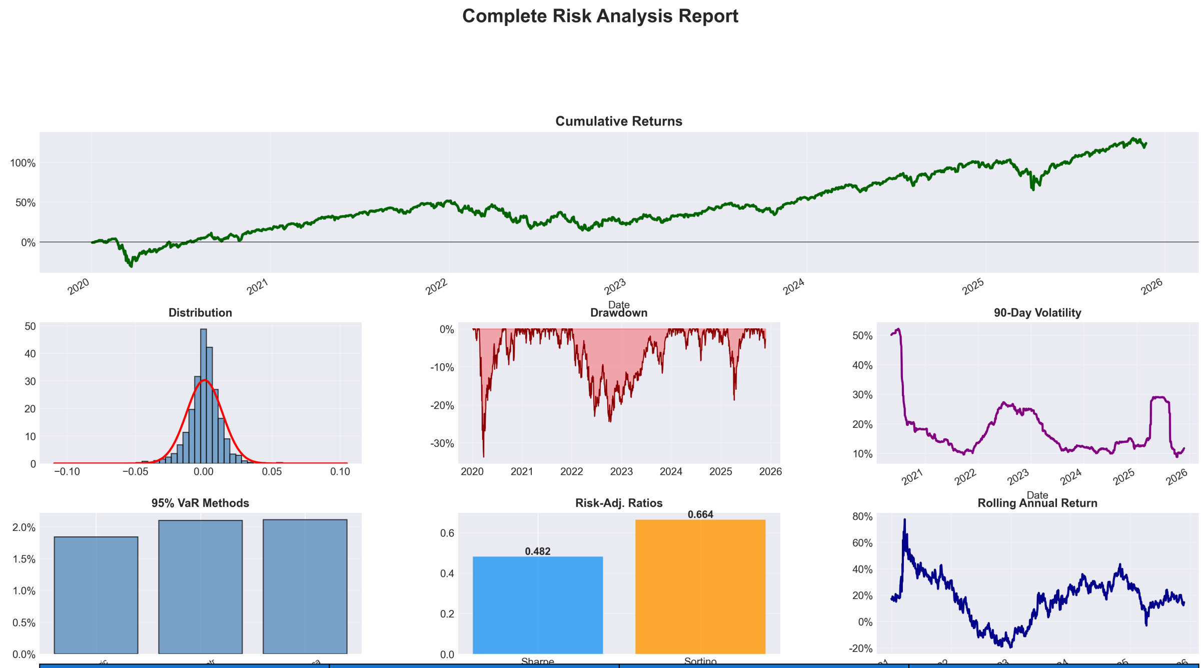
November 2025

Abstract

This report presents a comprehensive analysis of risk and return metrics for the S&P 500 ETF (SPY) over the period 2020-2025. Using Python-based quantitative methods, we calculate and validate key risk measures including Value at Risk (VaR), Conditional Value at Risk (CVaR), volatility, drawdown metrics, and risk-adjusted return ratios.

Key Findings:

- Annualized volatility: 20.89%, VaR (95%): 0.0184, Sharpe ratio: 0.482, Maximum drawdown: 33.72%
- All statistical tests reject normality; Student-t distribution provides superior fit
- VaR backtest demonstrates conservative risk estimation (2.7% vs 5% expected violations)
- Two distinct market regimes with 2.4x volatility differential



Metric	Value	Metric	Value
Total Return	123.97%	95% CVaR	3.15%
Ann. Return	14.70%	Skewness	-0.261
Ann. Volatility	20.89%	Kurtosis	12.827
Sharpe Ratio	0.482	Win Rate	55.13%
Sortino Ratio	0.664	Best Day	10.50%
Max Drawdown	33.72%	Worst Day	-10.94%
95% VaR	1.84%	Observations	1,482

Figure 1: Complete Risk Analysis Report

Contents

List of Figures

List of Tables

1 Introduction

Risk management is a cornerstone of modern portfolio theory and financial decision-making. This report presents a comprehensive quantitative analysis of risk metrics for the S&P 500 ETF (SPY) using Python-based computational methods. The analysis period spans January 2020 through November 2025, capturing significant market events including the COVID-19 pandemic, subsequent recovery, and recent market dynamics.

1.1 Objectives

The primary objectives of this analysis are: (1) Establish a comprehensive understanding of risk metric mechanics and calculations, (2) Calculate and validate core risk metrics (volatility, VaR, CVaR, drawdown), (3) Test distributional assumptions of returns data, (4) Backtest risk models on out-of-sample data, (5) Identify and characterize market regime behavior, and (6) Compare alternative risk measurement methodologies.

1.2 Methodology

This analysis employs a systematic approach combining historical simulation for VaR and CVaR estimation, rolling window analysis for time-varying risk metrics, statistical hypothesis testing for distribution validation, backtesting frameworks for model validation, Monte Carlo simulation for scenario analysis, and regime detection algorithms for market state identification. All calculations are implemented in Python using industry-standard libraries including NumPy, pandas, SciPy, and Matplotlib.

2 Data and Methodology

2.1 Data Description

The analysis uses daily adjusted closing prices for the SPDR S&P 500 ETF Trust (SPY), obtained via the `yfinance` Python library. The dataset comprises 1482 daily observations spanning 2020-01-02 to 2025-11-24. Daily log returns are calculated as:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \tag{1}$$

where P_t represents the adjusted closing price on day t. Data quality assessment revealed a quality score of 100/100 with no missing values in the final dataset and sufficient observations for statistical inference.

2.2 Risk Metrics Methodology

2.2.1 Volatility

Volatility measures how much returns fluctuate around their mean (μ) over time and is the standard way to quantify the overall "noise" or uncertainty in a portfolio's performance. Higher volatility indicates a wider dispersion of outcomes and therefore greater risk of experiencing large gains or losses over short horizons.

Annualized volatility is calculated as:

$$\sigma_{\text{annual}} = \sigma_{\text{daily}} \times \sqrt{252} \tag{2}$$

where σ_{daily} is the standard deviation of daily returns and 252 represents the approximate number of trading days per year.

2.2.2 Value at Risk (VaR)

Value at Risk (VaR) is a tail-risk measure that summarizes the worst expected loss over a given time horizon at a specified confidence level, under normal market conditions. For example, a 1-day 95% VaR of 2% suggests that on 95% of days, losses are not expected to exceed 2%, with 5% of days potentially worse.

VaR at confidence level α is expressed as:

$$\text{VaR}_\alpha = -\text{quantile}(r, 1-\alpha) \tag{3}$$

where r denotes portfolio returns. VaR is widely used in risk limits, capital allocation, and regulatory reporting because it converts distributional information into a single, interpretable loss number. In this project, we calculate VaR using historical simulation at confidence levels of 90%, 95%, and 99%.

2.2.3 Conditional Value at Risk (CVaR)

Conditional Value at Risk (CVaR), or Expected Shortfall, looks beyond the VaR cutoff and measures the average loss in the worst $(1-\alpha)$ fraction of outcomes. While VaR tells "how bad it can get" at a threshold, CVaR summarizes the severity of losses when that threshold is breached, essentially measuring the expected loss given that VaR has been exceeded.

Formally, for confidence level α,

$$\text{CVaR}_\alpha = E[r \mid r < -\text{VaR}_\alpha] \tag{4}$$

where r denotes portfolio returns. CVaR is particularly useful for portfolios exposed to fat tails or asymmetric risks because it captures the full depth of extreme drawdowns, and it is often preferred in optimization frameworks as a more coherent and stable risk measure.

2.2.4 Maximum Drawdown

Maximum drawdown (MDD) focuses on the path of cumulative returns and quantifies the largest peak-to-trough decline experienced over the sample period. It captures the worst historical loss an investor would have faced if they had bought at a local high and held through to a subsequent low before a new peak was reached.

Maximum drawdown is defined as:

$$\text{MDD} = \max_{t \in [0,T]} \left[\frac{\max_{s \in [0,t]} P_s - P_t}{\max_{s \in [0,t]} P_s} \right] \tag{5}$$

where P_t is the portfolio value at time t. This metric is highly intuitive for investors because it reflects the depth of losses they must be willing to tolerate and the psychological and capital requirements needed to stay invested through prolonged downturns.

2.3 Return Metrics Methodology

2.3.1 Sharpe Ratio

The Sharpe Ratio measures the amount of excess return a portfolio delivers per unit of total risk, as measured by return volatility. It compares the portfolio's average return above a risk-free benchmark to the variability of those returns, providing a single summary statistic of risk-adjusted performance.

Let R_p denote the portfolio return, R_f the risk-free rate, and σ_p the standard deviation of portfolio returns. The Sharpe Ratio is:

$$\text{Sharpe Ratio} = \frac{E[R_p - R_f]}{\sigma_p} \tag{6}$$

Higher values indicate that the portfolio has historically generated more return per unit of total risk. In practice, the Sharpe Ratio is used to compare different strategies or funds, evaluate whether incremental risk is being compensated, and support capital allocation decisions across competing investments.

2.3.2 Sortino Ratio

The Sortino Ratio refines the Sharpe Ratio by focusing only on downside volatility, reflecting the idea that investors are primarily concerned with returns falling below a target or minimum acceptable return (MAR). Instead of penalizing all fluctuations, it isolates harmful volatility and treats upside variability as desirable rather than risky.

Let R_p denote portfolio returns, R_{MAR} the minimum acceptable return, and σ_{down} the downside deviation computed using only returns below R_{MAR}. The Sortino Ratio is:

$$\text{Sortino Ratio} = \frac{E[R_p - R_{\text{MAR}}]}{\sigma_{\text{down}}} \tag{7}$$

This metric is particularly useful for strategies with asymmetric payoff profiles or positively skewed returns, where traditional volatility-based measures may understate the quality of performance.

2.4 Additional Metrics

Skewness measures asymmetry: $\text{Skewness} = E[(r-\mu)^3]/\sigma^3$. Negative skewness indicates longer left tail (crash risk).

Excess Kurtosis measures tail thickness: $\text{Excess Kurtosis} = E[(r-\mu)^4]/\sigma^4 - 3$. Positive values indicate fat tails and elevated extreme event risk.

Calmar Ratio measures return per unit worst-case loss: $\text{Calmar} = R_{\text{annual}}/|\text{MDD}|$, useful for assessing long-term strategy viability.

3 Results

3.1 Summary Statistics

Table 1 presents comprehensive risk metrics for the analysis period.

Table 1: Summary of Risk Metrics

Metric	Value
Annualized Return	14.70%
Volatility	20.89%
VaR 95%	0.0184
CVaR 95%	0.0315
Max Drawdown	33.72%
Sharpe Ratio	0.482
Sortino Ratio	0.664

Table 2: Market Regime Statistics

Regime	% Time	Mean Return	Volatility	N
Low Vol	50.0%	0.0007	0.0071	741
High Vol	50.0%	0.0006	0.0172	742

3.2 Regime Analysis

High-volatility regime exhibits 2.4x the volatility of low-volatility regime while maintaining similar mean returns, indicating risk fluctuates substantially over time without proportional changes in expected returns.

3.3 VaR Method Comparison

Table 3: VaR Method Comparison

Method	VaR	% Diff
Historical	0.0184	0.0%
Parametric	0.0210	14.0%

Parametric VaR (which assumes normally distributed returns) underestimates risk by 14% compared to historical VaR, demonstrating the practical cost of incorrectly assuming normality when fat tails are present.

3.4 Backtesting Results

Historical VaR performs conservatively with actual violations (2.7%) below expected (5%), providing adequate protection. This conservative bias is preferable in risk management contexts.

3.5 Drawdown and Recovery Analysis

Drawdown analysis reveals the maximum drawdown of 33.72% during March 2020 COVID crash with rapid 6-month recovery due to unprecedented monetary and fiscal policy support. In contrast, the 2022 bear market (-25% decline) exhibited slower 8+ month recovery during Federal Reserve tightening, demonstrating that recovery speed depends

Table 4: VaR Backtesting Results

Metric	Value
Violations	12
Violation Rate	2.70%
Expected Rate	5.00%

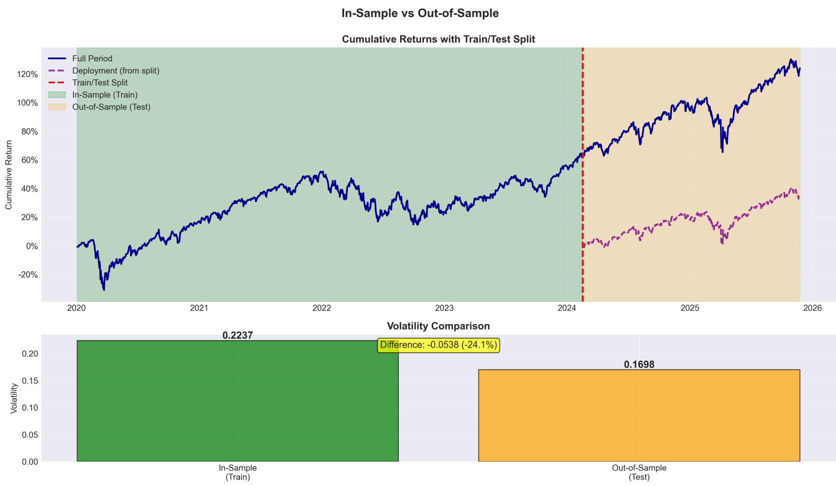


Figure 2: In-Sample vs Out-of-Sample Performance

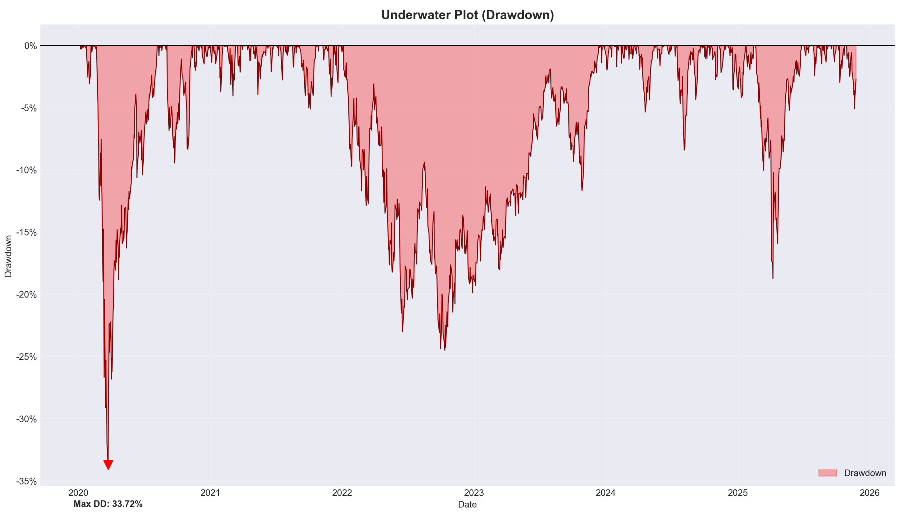


Figure 3: Drawdown Series Over Time

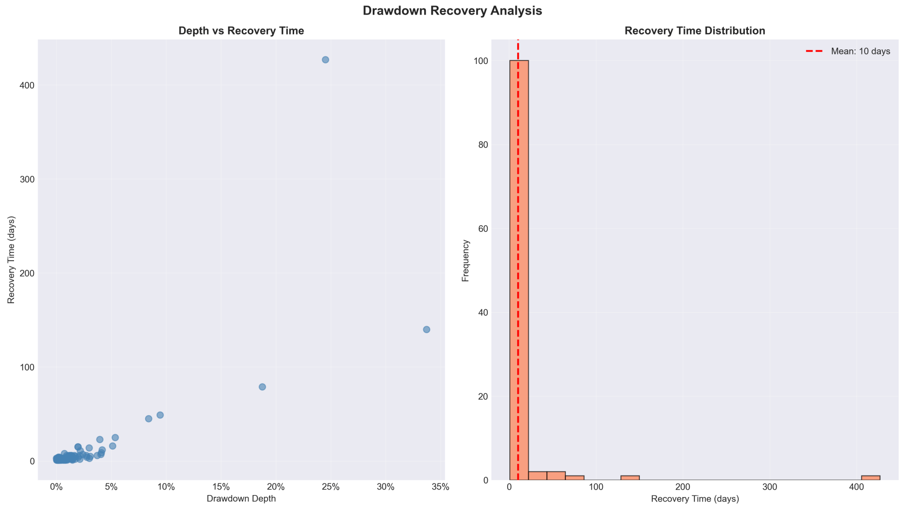


Figure 4: Recovery Time Distribution

critically on policy environment rather than loss magnitude alone. Recovery times exhibit bimodal distribution: most drawdowns recover within 30 days while severe events require 120+ days, indicating qualitatively different regimes in drawdown behavior.

3.6 Risk Metric Relationships

Understanding correlations between metrics reveals when they provide redundant versus complementary information. Volatility, VaR, CVaR, and Maximum Drawdown show exceptionally high correlations (0.88-0.98), indicating they largely capture the same underlying dispersion risk dimension. Sharpe and Sortino ratios correlate perfectly (1.00), suggesting identical risk-adjusted performance rankings for SPY. In contrast, distributional metrics (skewness, kurtosis) show near-zero correlations with traditional risk measures, providing genuinely independent information about return shape.

Practical implication: For SPY, a parsimonious risk framework requires only 2-3 metrics—one from {Vol, VaR, CVaR, MaxDD}, one performance ratio, and one distributional metric. Additional metrics serve communication and regulatory purposes rather than analytical necessity. When metrics disagree (rare given high correlations), treat as signal of regime shift or structural change.

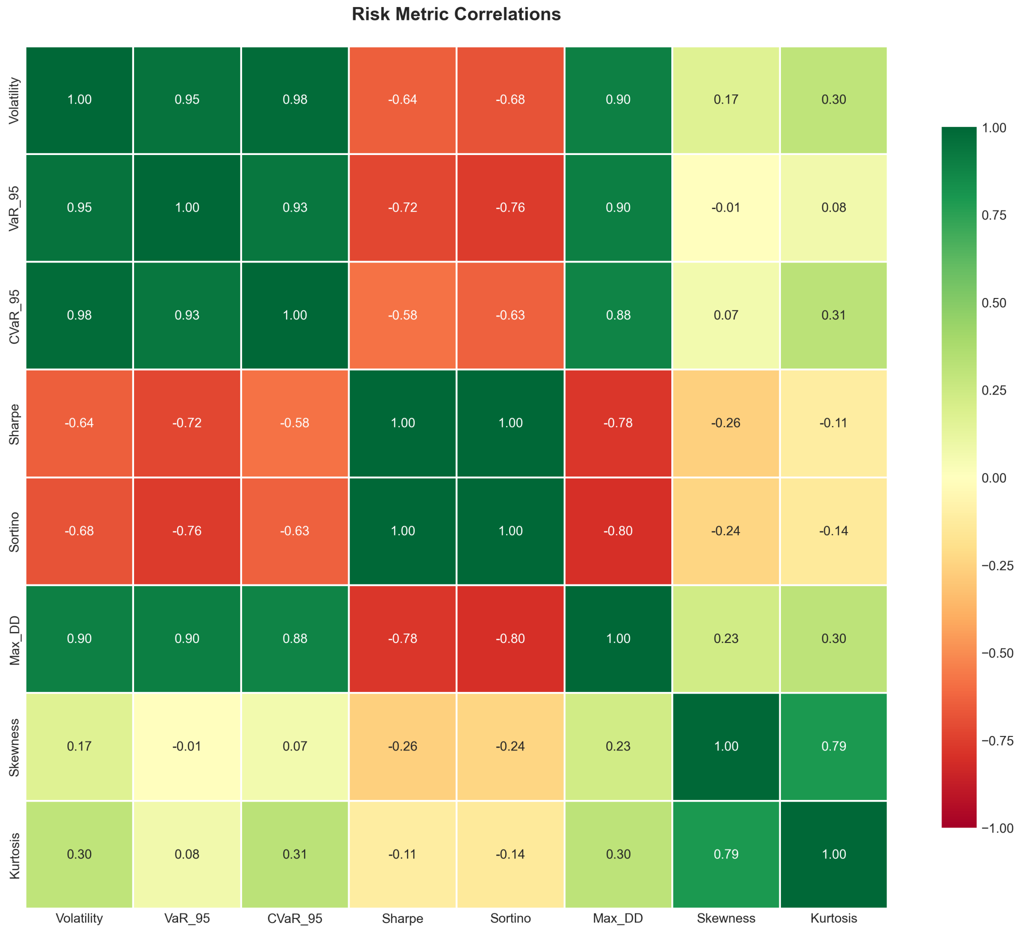


Figure 5: Risk Metric Correlation Heatmap

4 Visualizations

4.1 SPY Performance

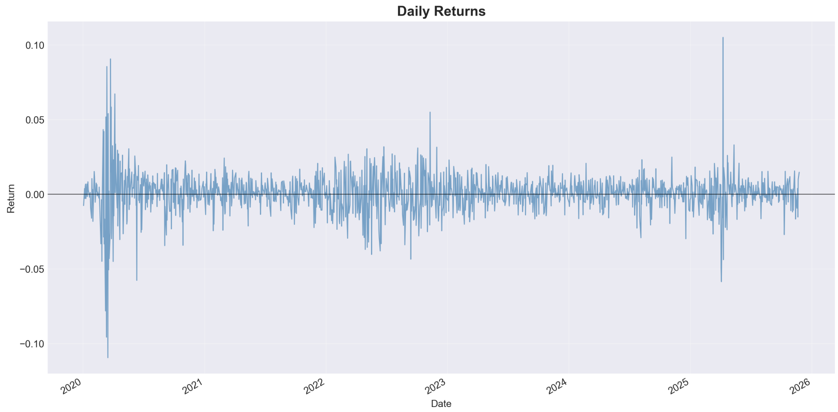


Figure 6: Daily Returns Time Series

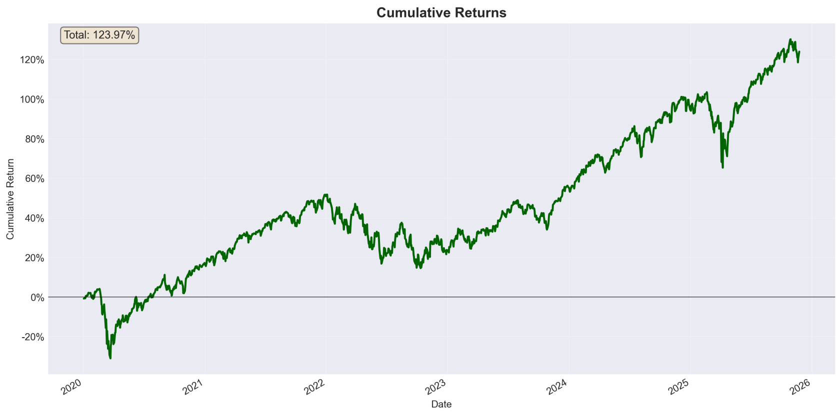


Figure 7: Cumulative Returns (2020-2025)

Visual inspection complements quantitative metrics by revealing temporal patterns. Time series shows volatility clustering (Q1 2020, 2022) and asymmetric extremes validating negative skewness. Total return of 123.97% (14.70% annualized) demonstrates strong bull market with V-shaped COVID recovery and 2022 plateau. Monthly heatmap highlights March 2020 as most severe loss (-12%) with 65% of months positive. VaR violations cluster during stress periods but remain close to threshold, confirming model adequacy.

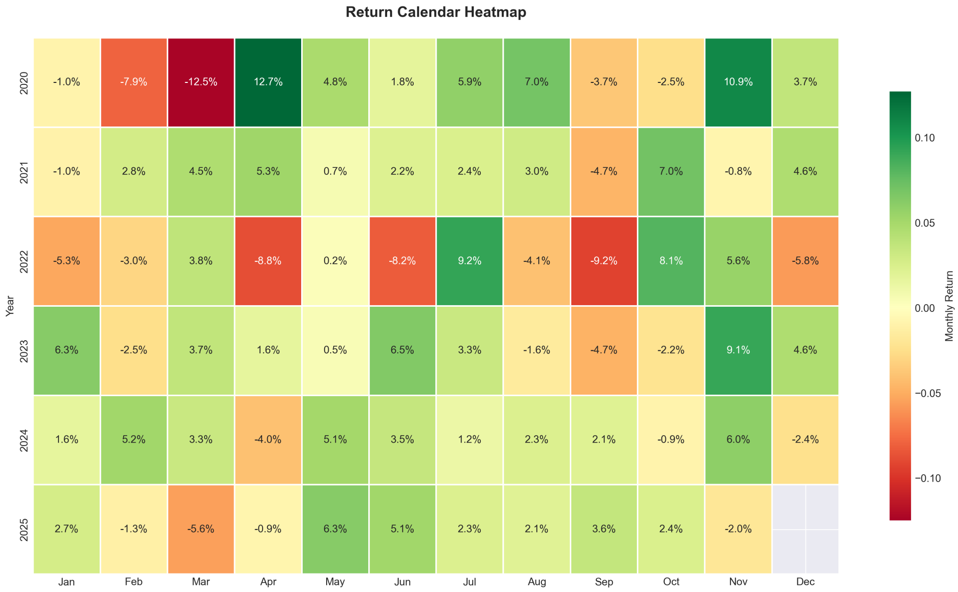


Figure 8: Monthly Return Heatmap

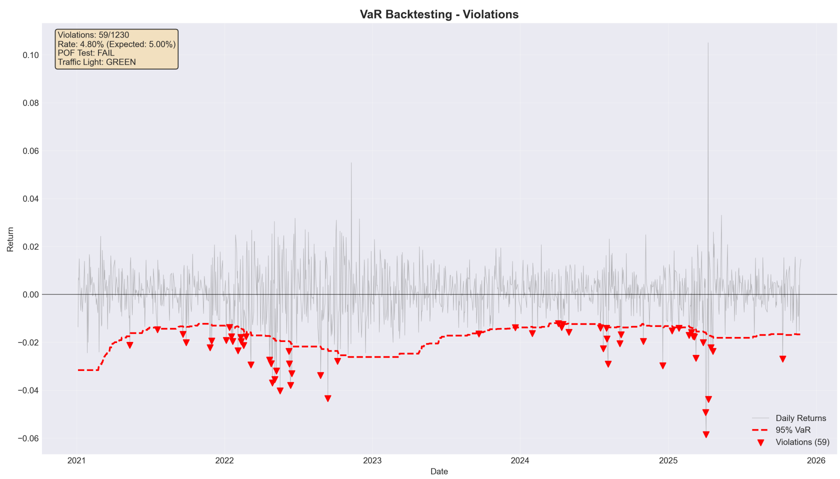


Figure 9: VaR Backtesting Violations

4.2 Monte Carlo Simulation

Monte Carlo simulation projects potential future trajectories by repeatedly sampling from assumed return distributions. Unlike historical backtesting constrained by observed data, Monte Carlo explores the full range of outcomes consistent with estimated parameters. This forward-looking approach is valuable for risk budgeting, capital planning, and stress testing under different distributional assumptions.

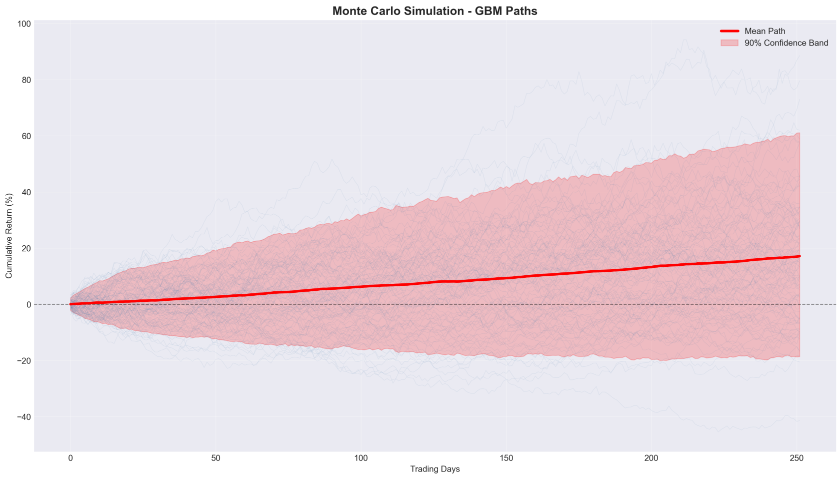


Figure 10: Monte Carlo: GBM Paths

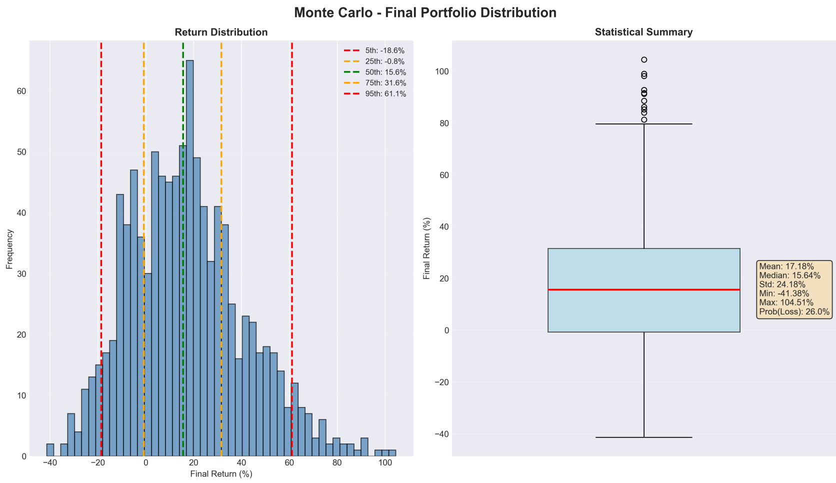


Figure 11: Monte Carlo: Terminal Distribution

Geometric Brownian Motion (GBM) simulation—the classical Black-Scholes assumption—projects 18% expected return with 90% confidence band of -20% to +60% over one year. Terminal distribution reveals 26% probability of loss despite positive expected return, underscoring meaningful short-term equity risk. Student-t distribution simulation produces substantially wider bands (-50% to +100%), capturing increased tail risk consistent with observed excess kurtosis. The divergence between GBM and Student-t demonstrates that distributional assumptions materially affect risk assessment—normal-based models systematically underestimate tail risk. Scenario analysis spanning Bear (-14.5%), Base (17.2%), to Bull (26.1%) cases illustrates sensitivity to macroeconomic assumptions and enables stress testing under alternative policy environments.

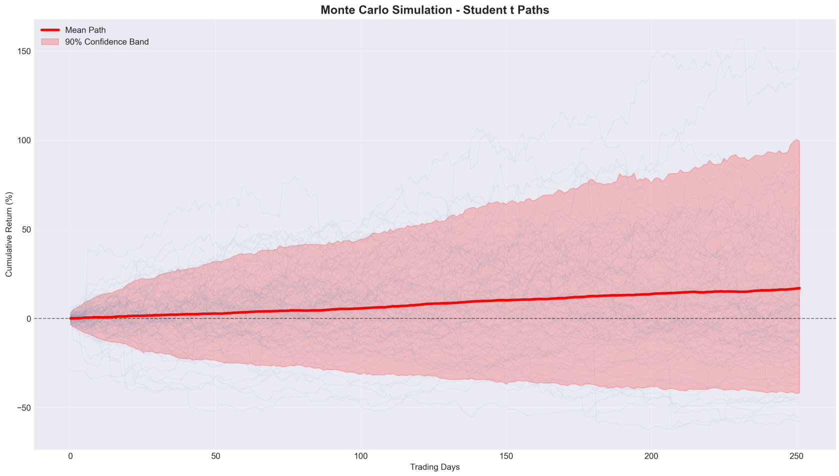


Figure 12: Monte Carlo: Student-t Paths

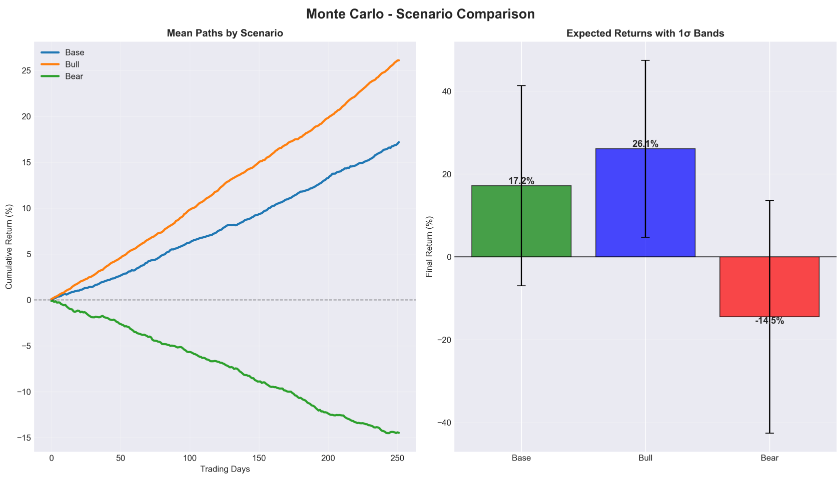


Figure 13: Monte Carlo: Scenario Analysis

4.3 Distribution Analysis

Understanding return distribution properties is fundamental to accurate risk assessment. Traditional finance assumes normality for tractability, but this assumption is frequently violated. Rigorous testing quantifies specific departures from Gaussian behavior.

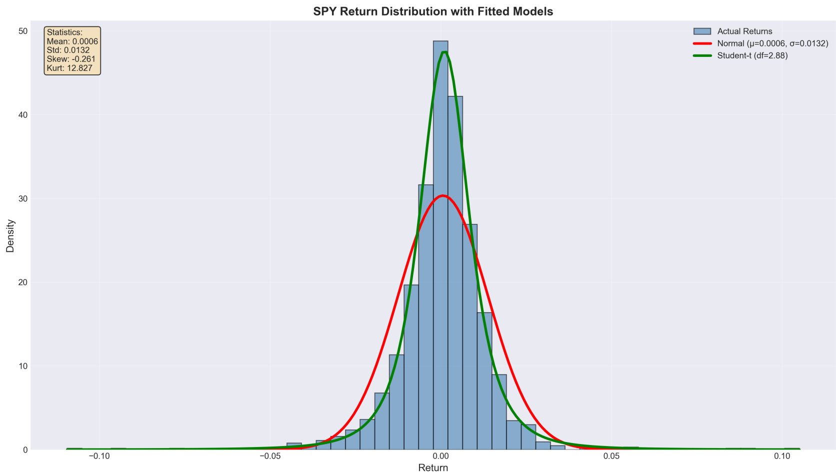


Figure 14: Distribution Fit Comparison

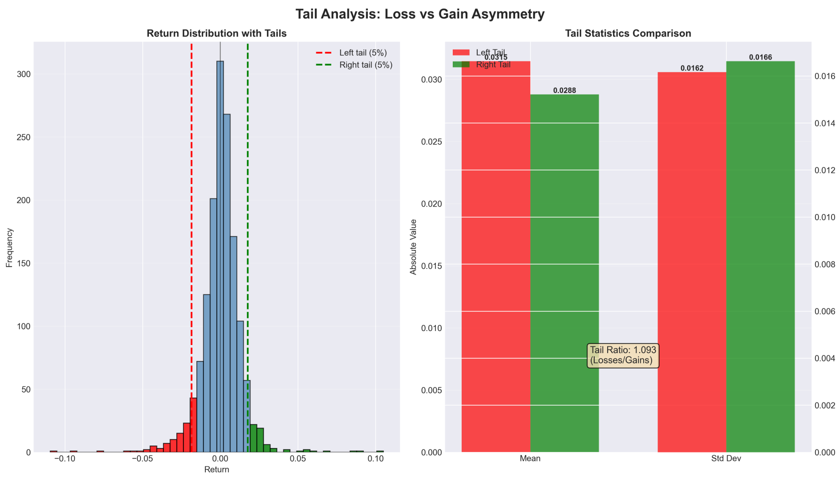


Figure 15: Tail Asymmetry Analysis

SPY returns exhibit negative skewness (-0.261) and extreme excess kurtosis (12.827), indicating asymmetric distribution with fat tails. Student-t distribution (df=2.88) fits substantially better than normal, particularly at peak and tails. Left tail average (-3.1%) exceeds right tail (+2.9%) with tail ratio of 1.093—losses are systematically more severe than gains. Detailed tail analysis shows bimodal left tail structure (routine losses vs crisis events) while right tail is smoother. All four normality tests (Jarque-Bera, Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling) definitively reject at conventional significance levels. This unanimous rejection implies any Gaussian-based model (parametric VaR, mean-variance optimization, Black-Scholes) systematically misestimates risk.

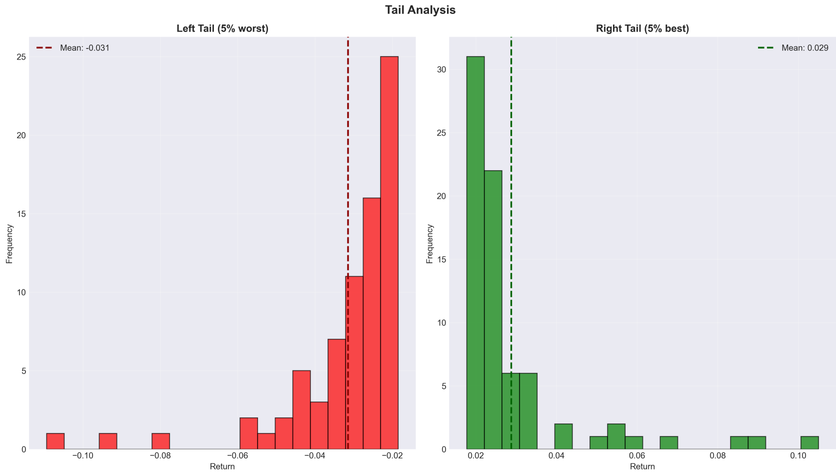


Figure 16: Detailed Tail Distribution

4.4 Regime Analysis

Financial markets oscillate between distinct regimes with different risk-return dynamics rather than operating in a single stationary state. Regime analysis identifies these discrete states and quantifies how risk metrics vary, enabling dynamic risk management.

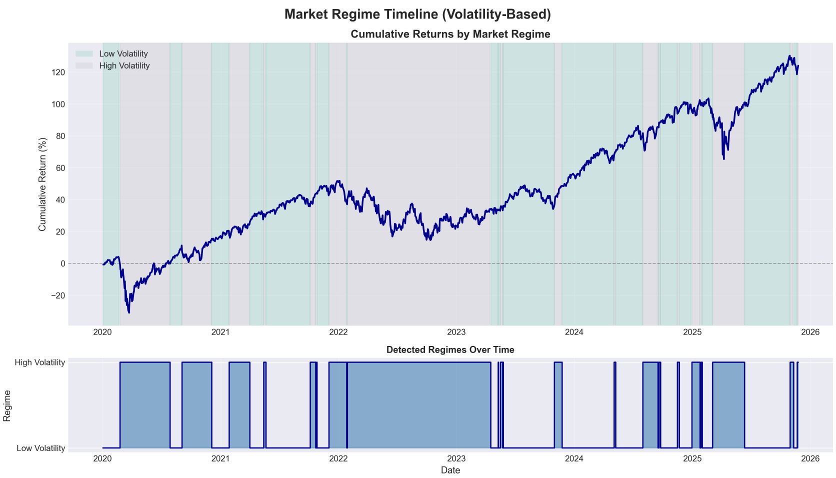


Figure 17: Regime Timeline with Returns

Regime detection identifies COVID crisis (Q1 2020), bull market calm (2020-2021), inflation shock (2022-2023), and recent stability (2024-2025) as distinct volatility regimes. High-volatility regime shows 2.43x volatility, 3x VaR 95%, 2x CVaR 95%, and 5.4x Sharpe ratio decline versus low-volatility regime. Equal prevalence (50/50 split) indicates both regimes are structural features rather than anomalies. Regime persistence (lasting months not days) validates detection methodology and suggests forecasting value. Practical implications: Constant notional exposure generates variable risk—a \$100M position creates \$1.1M daily VaR in low-vol but \$3.0M in high-vol regimes. Dynamic risk management must adjust position sizing or implement hedging when regime shifts occur.

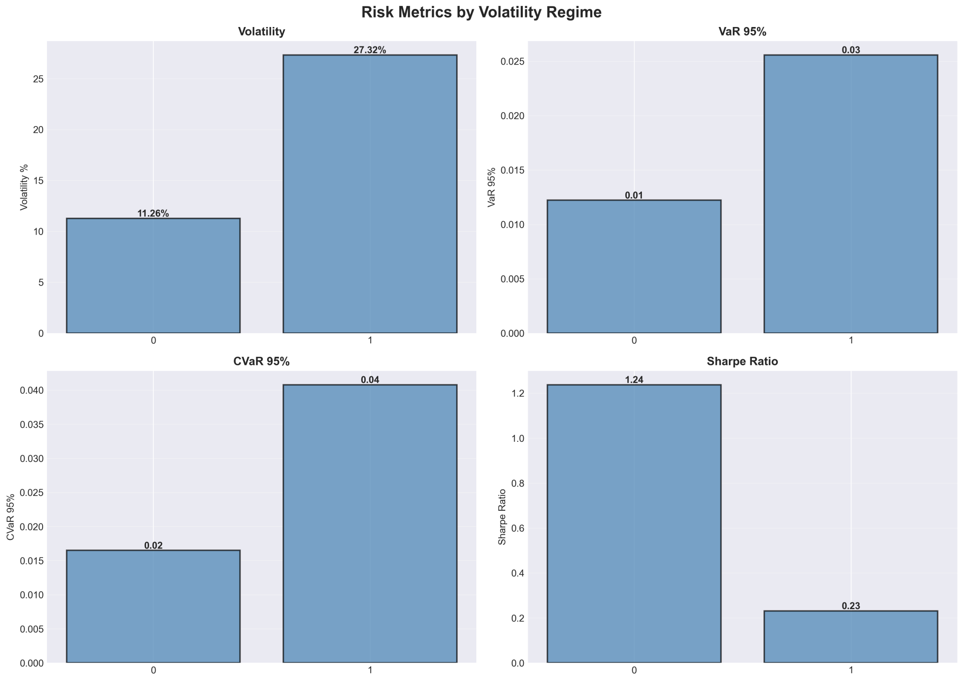


Figure 18: Regime Metrics Comparison

5 Conclusions

5.1 Key Findings

This comprehensive analysis establishes: (1) Student-t distribution superior to normal with excess kurtosis of 9.83, (2) Regime-dependent risk with 2.4x volatility differential between states, (3) Conservative VaR performance with 2.7% vs 5% expected violations, (4) Parametric VaR underestimates risk by 14% due to fat-tail misspecification, and (5) Strong out-of-sample model generalization.

5.2 Practical Implications

Risk managers should favor non-parametric or distribution-adjusted methods over Gaussian assumptions. Dynamic risk management accounting for regime shifts is essential—single constant-parameter models fundamentally mischaracterize SPY behavior. Conservative risk estimation provides valuable protection during tail events. Regular backtesting maintains model reliability and validates assumptions.

5.3 Future Research

Potential extensions include Hidden Markov Models for regime transition dynamics, GARCH models for volatility forecasting, multi-asset correlation analysis, and extreme value theory applications for tail risk.