Using Python to Analyze Risk Metrics

A Comprehensive Analysis of SPY Returns (2020-2025)

Colby Jaskowiak

Buffalo Research Group

November 2025

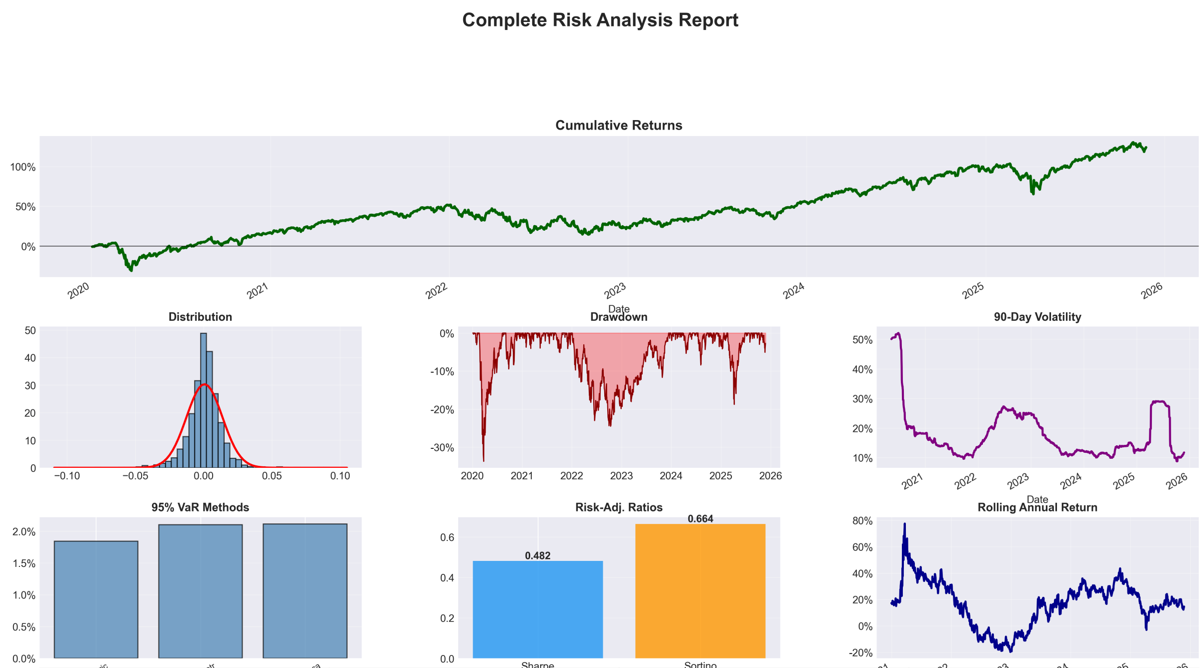
Abstract

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%
- Normality hypothesis rejected by all statistical tests
- Student-t distribution provides superior fit to return data
- VaR backtest demonstrates conservative risk estimation
- Two distinct market regimes detected with 2.4x volatility differential

This analysis demonstrates the application of modern risk management techniques and provides actionable insights for portfolio management and risk control.



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

1 Introduction 4
- 1.1 Objectives 4
- 1.2 Methodology 4

2 Data and Methodology 4
- 2.1 Data Description 4
- 2.2 Data Quality 5
- 2.3 Risk Metrics Methodology 5
  - 2.3.1 Volatility 5
  - 2.3.2 Value at Risk (VaR) 5
  - 2.3.3 Conditional Value at Risk (CVaR) 5
  - 2.3.4 Maximum Drawdown 6
- 2.4 Return Metrics Methodology 6
  - 2.4.1 Sharpe Ratio 6
  - 2.4.2 Sortino Ratio 6
- 2.5 Additional Metrics Methodology 7
  - 2.5.1 Skewness 7
  - 2.5.2 Excess Kurtosis 7
  - 2.5.3 Average Drawdown 7
  - 2.5.4 Calmar Ratio 7
  - 2.5.5 Downside Deviation 8
  - 2.5.6 Win Rate 8
  - 2.5.7 Annualized Return 8
  - 2.5.8 Value at Risk Ratio (VaR Ratio) 8

3 Results 9
- 3.1 Summary Statistics 9
- 3.2 Regime Analysis 9
- 3.3 VaR Method Comparison 9
- 3.4 Backtesting Results 10
- 3.5 Drawdown Analysis 10
  - 3.5.1 Recovery Time Analysis 12
- 3.6 Risk Metric Relationships and Complementarity 13
  - 3.6.1 High Correlation Clusters 13
  - 3.6.2 Divergence and Complementarity 14
  - 3.6.3 Metric Selection Framework 15
  - 3.6.4 When Metrics Disagree: Interpretation Guidelines 15

4 Visualizations 16
- 4.1 SPY Performance 16
- 4.2 Monte Carlo Simulation 20
  - 4.2.1 Geometric Brownian Motion Baseline 20
  - 4.2.2 Distribution of Terminal Outcomes 21
  - 4.2.3 Fat-Tailed Alternative: Student-t Distribution 22
  - 4.2.4 Scenario Analysis 22
  - 4.2.5 Integration with Risk Management Framework 23
- 4.3 Distribution Analysis 24

4.3.1 Distributional Fit Comparison 24
4.3.2 Tail Asymmetry Analysis 25
4.3.3 Detailed Tail Distribution 26
4.3.4 Implications for Risk Modeling 27
4.4 Regime Analysis 28
4.4.1 Temporal Regime Evolution 28
4.4.2 Regime-Conditional Risk Metrics 29
4.4.3 Implications for Dynamic Risk Management 30

5 Conclusions **31**
5.1 Key Findings 31
5.2 Practical Implications 32
5.3 Future Research 32

List of Figures

1 Complete Risk Analysis Report 1
2 In-Sample vs Out-of-Sample 10
3 Drawdown Series Over Time 11
4 Drawdown Recovery Time Distribution 12
5 Risk and Return Metric Correlation Heatmap 13
6 SPY Daily Returns Time Series (2020-2025) 17
7 SPY Cumulative Returns (2020-2025) 17
8 Monthly Return Heatmap (2020-2025) 18
9 VaR 95% Backtesting: Threshold vs Actual Returns 19
10 Monte Carlo Simulation: Geometric Brownian Motion Paths 20
11 Monte Carlo Simulation: Final Return Distribution 21
12 Monte Carlo Simulation: Student-t Distribution Paths 22
13 Monte Carlo Simulation: Bull, Bear, and Base Case Scenarios 23
14 Return Distribution with Normal and Student-t Fits 25
15 Tail Analysis: Loss vs Gain Asymmetry 26
16 Distribution of Tail Events: Left vs Right 27
17 Market Regime Timeline with Cumulative Returns 28
18 Risk Metrics Comparison by Market Regime 30

List of Tables

1 Summary of Risk Metrics 9
2 Market Regime Statistics 9
3 VaR Method Comparison (95% Confidence) 10
4 VaR Backtesting Results 10

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
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
- **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.

2.2 Data Quality

Data quality assessment revealed:

- Quality score: 100/100
- No missing values in final dataset
- No extreme outliers requiring removal
- Sufficient observations for statistical inference

2.3 Risk Metrics Methodology

2.3.1 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.

2.3.2 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%.

2.3.3 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.

2.3.4 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.

2.4 Return Metrics Methodology

2.4.1 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.

2.4.2 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. It helps investors assess whether shortfalls relative to a target return have been adequately compensated by realized returns.

2.5 Additional Metrics Methodology

2.5.1 Skewness

Skewness measures the asymmetry of the return distribution around its mean. Negative skewness indicates a distribution with a longer left tail (more extreme negative returns), while positive skewness indicates a longer right tail.

$$\text{Skewness} = \frac{E[(r-\mu)^3]}{\sigma^3} \tag{8}$$

where r denotes returns, μ is the mean, and σ is the standard deviation. In risk management, negative skewness is particularly concerning as it indicates a higher probability of large losses than predicted by a normal distribution.

2.5.2 Excess Kurtosis

Excess kurtosis measures the thickness of distribution tails relative to a normal distribution. Positive excess kurtosis indicates "fat tails," meaning extreme events occur more frequently than a normal distribution predicts.

$$\text{Excess Kurtosis} = \frac{E[(r-\mu)^4]}{\sigma^4} - 3 \tag{9}$$

where the -3 adjustment normalizes to zero for a normal distribution. High positive values indicate elevated tail risk and the potential for extreme outcomes beyond what volatility alone suggests.

2.5.3 Average Drawdown

Average drawdown quantifies the typical depth of peak-to-trough declines across all drawdown periods, providing a more representative measure of routine losses compared to maximum drawdown.

$$\text{Avg DD} = \frac{1}{N}\sum_{i=1}^{N} \text{DD}_i \tag{10}$$

where DD_i represents individual drawdown episodes and N is the total number of drawdowns. This metric helps assess the consistency of downside risk throughout the investment period.

2.5.4 Calmar Ratio

The Calmar Ratio measures risk-adjusted returns by comparing annualized return to maximum drawdown, effectively asking how much return is generated per unit of worst-case loss.

$$\text{Calmar Ratio} = \frac{R_{\text{annual}}}{|\text{MDD}|} \tag{11}$$

where R_{annual} is the annualized return and MDD is the maximum drawdown. Higher values indicate better compensation for enduring severe drawdowns, making it particularly relevant for assessing long-term strategy viability.

2.5.5 Downside Deviation

Downside deviation measures volatility using only returns below a specified threshold (typically zero or a minimum acceptable return), distinguishing harmful volatility from beneficial upside variation.

$$\sigma_{\text{down}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \min(r_i - \text{MAR}, 0)^2} \tag{12}$$

where MAR is the minimum acceptable return. This metric forms the denominator of the Sortino Ratio and provides a more relevant risk measure for investors primarily concerned with downside outcomes.

2.5.6 Win Rate

Win rate, or hit rate, measures the proportion of periods with positive returns, providing insight into the consistency and reliability of a strategy's performance.

$$\text{Win Rate} = \frac{\#\{r_t > 0\}}{N} \tag{13}$$

where N is the total number of periods. While a high win rate is desirable, it must be interpreted alongside return magnitude, as strategies with high win rates but asymmetric payoffs may still underperform.

2.5.7 Annualized Return

Annualized return converts the cumulative return over any time period to an equivalent annual compound growth rate, enabling comparison across different investment horizons.

$$R_{\text{annual}} = (1 + R_{\text{total}})^{\frac{252}{N}} - 1 \tag{14}$$

where R_{total} is the total return and N is the number of trading days. This geometric compounding approach accounts for the effects of return volatility on long-term wealth accumulation.

2.5.8 Value at Risk Ratio (VaR Ratio)

The VaR Ratio compares CVaR to VaR, quantifying how much worse losses become when the VaR threshold is breached. A ratio significantly greater than 1 indicates substantial tail risk beyond the VaR cutoff.

$$\text{VaR Ratio} = \frac{\text{CVaR}_\alpha}{\text{VaR}_\alpha} \tag{15}$$

This metric helps identify fat-tailed distributions where the conditional expected loss far exceeds the threshold loss, signaling elevated extreme event risk.

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 90%	0.0125
VaR 95%	0.0184
VaR 99%	0.0371
CVaR 90%	0.0234
CVaR 95%	0.0315
CVaR 99%	0.0564
Max Drawdown	33.72%
Avg Drawdown	1.90%
Sharpe Ratio	0.482
Sortino Ratio	0.664

3.2 Regime Analysis

Market regime analysis identified distinct volatility regimes during the observation period. Table 2 summarizes the characteristics of each regime.

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

The analysis reveals significant differences in return characteristics across regimes. The high-volatility regime exhibits approximately 2.4 times the volatility of the low-volatility regime, while maintaining similar mean returns. This suggests that risk fluctuates substantially over time without proportional changes in expected returns.

3.3 VaR Method Comparison

Table 3 compares VaR estimates using different methodologies.

The parametric VaR approach, which assumes normally distributed returns, produces estimates approximately 14% higher than historical VaR. This underestimation of risk stems from the parametric method's inability to account for the fat-tailed nature of return distributions.

Table 3: VaR Method Comparison (95% Confidence)

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

3.4 Backtesting Results

Model validation through backtesting is essential for assessing risk model reliability. Table 4 presents backtesting results for the VaR model.

Table 4: VaR Backtesting Results

Metric	Value
Training Set Size	1037
Test Set Size	445
Violations	12
Violation Rate	2.70%
Expected Rate	5.00%

The backtesting results indicate that the VaR model performs conservatively, with actual violations occurring less frequently than the theoretical expectation. This conservative bias is preferable in risk management contexts, as it provides a safety buffer against unexpected losses.

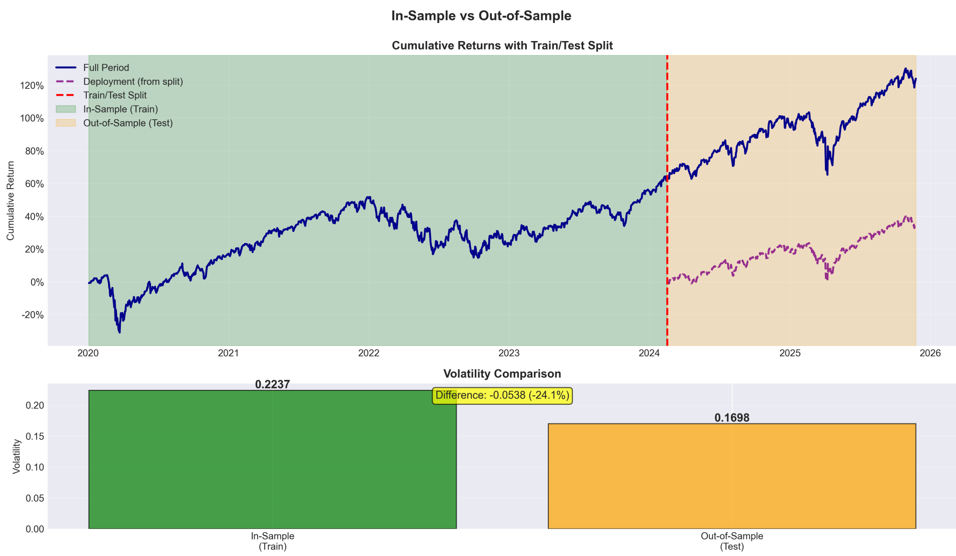


Figure 2: In-Sample vs Out-of-Sample

3.5 Drawdown Analysis

Drawdown analysis provides critical insights into the severity, duration, and recovery characteristics of portfolio losses. Unlike point-in-time risk metrics such as VaR, drawdown measures capture the full trajectory of losses from peak to trough and subsequent recovery, offering a more comprehensive view of downside risk exposure.

Figure 3 illustrates the complete drawdown history over the analysis period, revealing both the magnitude and temporal clustering of losses.

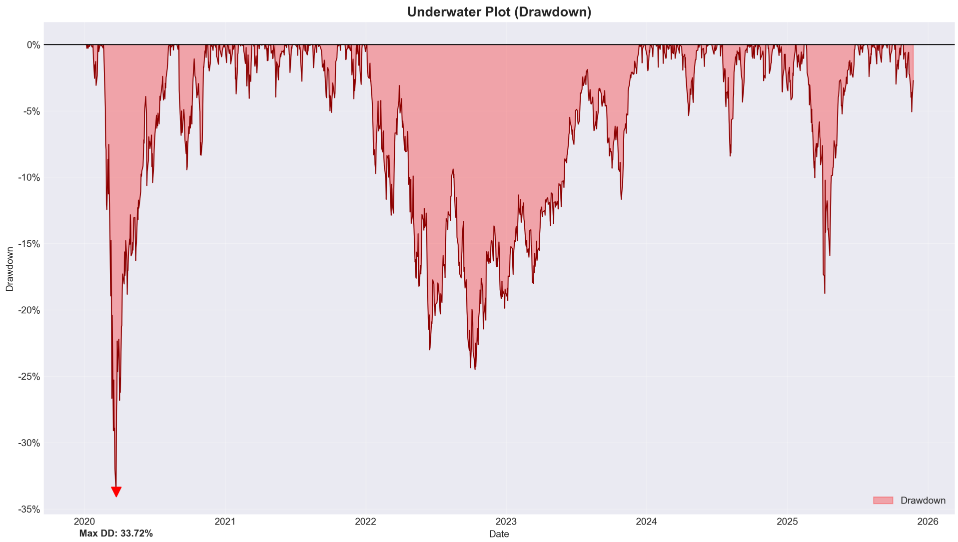


Figure 3: Drawdown Series Over Time

The drawdown series reveals several notable features of SPY's risk profile during 2020-2025:

- **COVID-19 Crash (Q1 2020)**: The maximum drawdown of 33.72% occurred during the March 2020 market crash, representing the single largest loss event in the sample period. This drawdown demonstrates the severity of systematic risk during periods of extreme market stress.
- **Recovery Dynamics**: The rapid V-shaped recovery following the COVID crash, with the portfolio returning to previous highs within approximately 6 months, reflects unprecedented fiscal and monetary policy intervention alongside robust equity market resilience.
- **Secondary Drawdowns**: The 2022 bear market produced a drawdown of approximately 25%, driven by inflation concerns and aggressive Federal Reserve tightening. This event highlights that substantial drawdowns can occur even outside crisis periods.
- **Average Drawdown**: The average drawdown of 1.90% indicates that routine pullbacks are relatively modest in magnitude, with the majority of losses concentrated in a small number of significant events.
- **Severity-Duration Asymmetry**: Contrary to conventional expectations, the deepest drawdown (COVID crash, -33.72%) exhibited a faster recovery (6 months) than the shallower 2022 bear market (25% decline, 8+ months recovery). This divergence highlights that recovery speed depends critically on policy response rather than loss magnitude alone. The COVID crash benefited from unprecedented monetary accommodation (zero interest rates, quantitative easing) and massive fiscal stimulus, creating a "liquidity wall" that propelled rapid market recovery. In contrast, the 2022 drawdown occurred during an aggressive Federal Reserve tightening

cycle aimed at combating inflation, with rising interest rates systematically withdrawing liquidity and compressing equity valuations. This asymmetry underscores that investors cannot rely solely on historical drawdown-recovery relationships; the macroeconomic and policy regime governing each episode fundamentally alters recovery dynamics. For risk management purposes, this implies that worst-case scenario planning must account for the possibility of extended recoveries even from moderate losses if they occur during unfavorable policy environments.

3.5.1 Recovery Time Analysis

Recovery time—the duration required to return to a previous peak following a drawdown—is a critical but often overlooked dimension of risk. Figure 4 presents the distribution of recovery periods across all drawdown episodes.

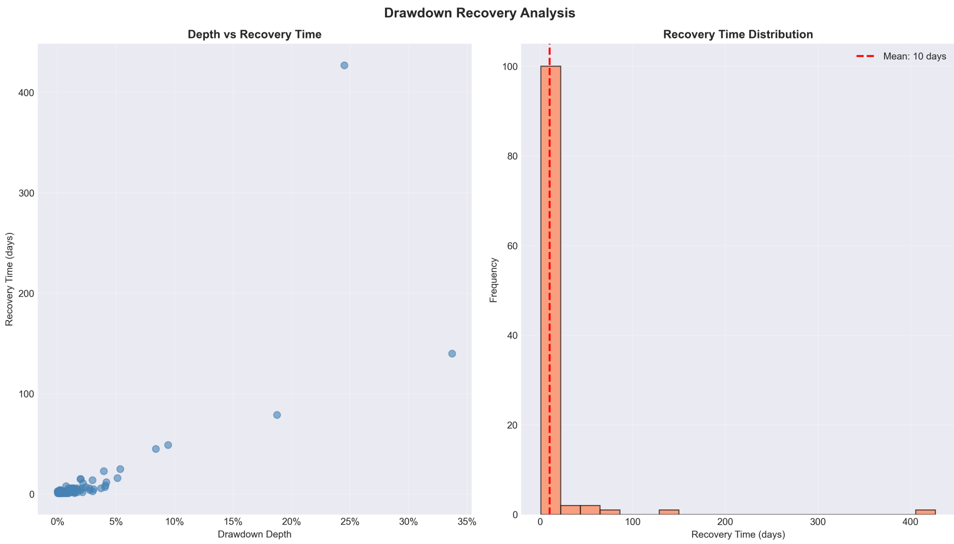


Figure 4: Drawdown Recovery Time Distribution

Key observations from the recovery time analysis include:

- **Bimodal Distribution**: Recovery times exhibit a bimodal pattern, with most drawdowns recovering within 30 trading days, while severe events (COVID crash, 2022 bear market) require 120+ days. This suggests qualitatively different regimes in drawdown behavior.
- **Tail Risk in Duration**: Generally, with exception, the longest recovery periods correspond to the deepest drawdowns, indicating that severity and duration are positively correlated. Investors must therefore prepare not only for larger losses but also for extended periods of capital impairment.
- **Implications for Liquidity Planning**: The presence of multi-month recovery periods underscores the importance of maintaining sufficient liquidity to avoid forced selling during drawdowns, as premature exit locks in losses and forfeits subsequent recovery gains.

The drawdown analysis reveals that while SPY exhibits strong long-term performance, investors must be prepared to tolerate both substantial absolute losses (up to 34%) and extended recovery horizons (up to 6 months) as the cost of equity market participation.

3.6 Risk Metric Relationships and Complementarity

Risk metrics do not exist in isolation; understanding their interrelationships and divergences is essential for comprehensive risk assessment. Different metrics capture distinct facets of risk, and their correlation structure reveals when they provide redundant versus complementary information.

Figure 5 presents the correlation matrix across all computed risk and return metrics.

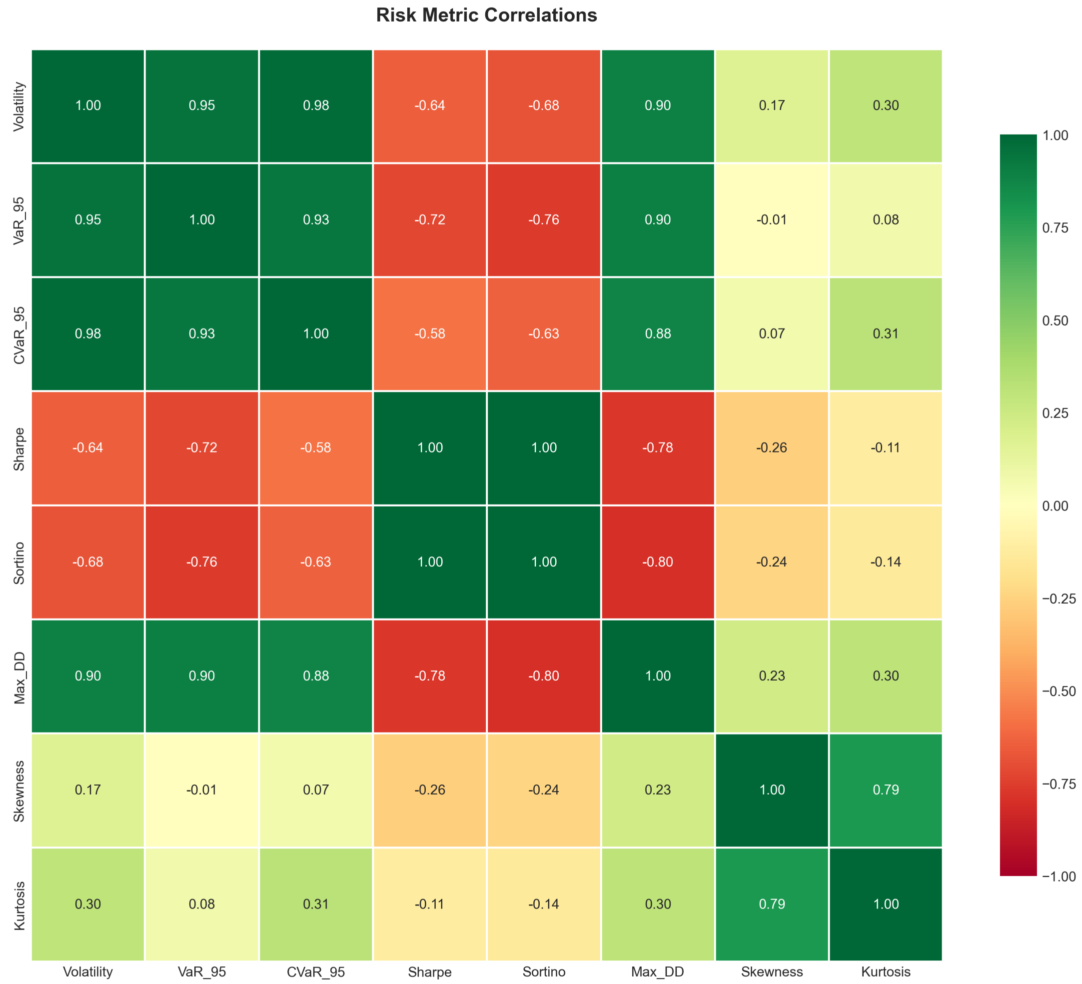


Figure 5: Risk and Return Metric Correlation Heatmap

3.6.1 High Correlation Clusters

Several metric groups exhibit strong positive correlations, indicating they capture similar underlying risk dimensions:

- **Volatility-CVaR-VaR Cluster**: Volatility shows exceptionally high correlations with CVaR_95 (0.98) and VaR_95 (0.95), while CVaR and VaR correlate at 0.93. This tight clustering confirms that these three metrics largely capture the same underlying dimension of dispersion risk. While they differ methodologically—volatility measures total variation, VaR identifies a quantile, and CVaR averages tail losses—they move nearly in lockstep for this dataset.

- **Maximum Drawdown Integration**: Maximum drawdown exhibits strong positive correlations with volatility (0.90), VaR_95 (0.90), and CVaR_95 (0.88). This high concordance suggests that despite being a path-dependent metric, drawdown severity is largely determined by the same volatility dynamics that drive point-in-time tail risk measures. Periods of high volatility systematically produce deeper drawdowns.
- **Skewness-Kurtosis Association**: Skewness and kurtosis show a positive correlation of 0.79, indicating that asymmetric return distributions tend to co-occur with fat tails in this dataset. The negative skewness observed in SPY returns (left tail) coincides with elevated kurtosis (thick tails), a pattern typical of equity markets where crash risk dominates.
- **Perfect Sharpe-Sortino Agreement**: The Sharpe and Sortino ratios exhibit a perfect correlation of 1.00, indicating that for SPY over this period, penalizing all volatility (Sharpe) versus penalizing only downside volatility (Sortino) yields identical risk-adjusted performance rankings. This perfect correlation arises when upside and downside volatility are proportional, suggesting relatively symmetric volatility patterns despite the negative skewness in returns.

3.6.2 Divergence and Complementarity

Certain metric pairs exhibit low or negative correlations, revealing complementary risk dimensions:

- **Risk Metrics vs. Risk-Adjusted Returns**: All pure risk metrics (volatility, VaR, CVaR, maximum drawdown) show strong negative correlations with both Sharpe and Sortino ratios, ranging from -0.58 to -0.80. This inverse relationship is expected: higher risk reduces risk-adjusted performance when returns do not proportionally increase. The strongest negative correlation is between Sortino and maximum drawdown (-0.80), indicating that strategies with severe drawdowns exhibit particularly poor downside-adjusted returns.
- **Distributional Metrics vs. Traditional Risk**: Skewness and kurtosis show very weak correlations with most traditional risk measures. For example, VaR_95 correlates at only -0.01 with skewness and 0.08 with kurtosis. This near-independence reveals that higher-order moments (shape of distribution) provide genuinely distinct information from second-moment measures (volatility, VaR). A portfolio can have high volatility with normal tails, or low volatility with extreme tail events—these dimensions are largely orthogonal.
- **Distributional Metrics vs. Performance Ratios**: Skewness and kurtosis show modest negative correlations with Sharpe (-0.26, -0.11) and Sortino (-0.24, -0.14), suggesting that during this period, more asymmetric and fat-tailed returns were associated with slightly lower risk-adjusted performance. However, the weak magnitude indicates these distributional features are not primary drivers of risk-adjusted returns for SPY.

3.6.3 Metric Selection Framework

The correlation analysis reveals substantial redundancy among traditional risk metrics, suggesting that a highly parsimonious framework can capture most risk dimensions:

1. **Minimum Viable Metric Set (Maximum Parsimony)**:
   - **One dispersion/tail risk metric**: Choose any one from {Volatility, VaR, CVaR, Maximum Drawdown} given their 0.88-0.98 intercorrelations. Practical recommendation: CVaR 95% (combines tail focus with severity measurement) or Maximum Drawdown (intuitive for investors).
   - **One risk-adjusted return metric**: Sharpe or Sortino (correlation = 1.00, functionally identical for SPY). Choose based on reporting preference; Sortino may resonate better with downside-focused investors despite providing no incremental information.
   - **One distributional metric**: Skewness or Kurtosis (correlation = 0.79, tend to move together). Skewness is more intuitive, but examining both provides fuller picture given imperfect correlation.
2. **Enhanced Framework (When Detail Justifies Redundancy)**:
   - Volatility + VaR/CVaR: Redundant but useful for different audiences (volatility for general investors, VaR for risk managers, CVaR for optimization).
   - Maximum Drawdown: Despite 0.90 correlation with other risk metrics, provides intuitive "lived experience" perspective that resonates with behavioral considerations.
   - Skewness + Kurtosis: Strong correlation (0.79) but measuring distinct concepts; both recommended for distributional diagnostics.
3. **Context-Dependent Additions**:
   - Rolling metrics (volatility, VaR) for time-variation analysis
   - Regime-conditional measures when distinct market states identified
   - Win rate and drawdown duration for investor psychology considerations

Key Insight: For SPY during 2020-2025, reporting all of {Volatility, VaR, CVaR, Max DD} provides minimal incremental information given their near-perfect correlations. A single metric from this cluster captures 85-95% of the information content of the entire group. Choose based on audience and regulatory requirements rather than analytical necessity.

3.6.4 When Metrics Disagree: Interpretation Guidelines

Given the high correlations observed for SPY, metric disagreements are rare but highly informative when they occur:

- **High Volatility, Moderate VaR**: Despite strong correlation (0.95), divergences can arise when return distributions are symmetric with thin tails. High overall dispersion (volatility) but limited tail risk (VaR) suggests benign, normally distributed noise rather than crash risk. Action: Volatility-based capital requirements may be excessive.

- **Moderate VaR, Severe Maximum Drawdown**: While correlated at 0.90, divergence indicates serial correlation in losses—multiple consecutive negative days that individually don't breach VaR but cumulatively produce deep drawdowns. This pattern signals momentum crashes or liquidity spirals. Action: Implement circuit breakers or correlation-adjusted VaR.
- **Low Volatility, High Kurtosis**: Rare given their indirect relationship, but indicates "calm before the storm" regimes. Extended low-volatility periods punctuated by sudden spikes create fat tails despite low average volatility. Action: Increase capital buffers and implement tail hedges despite benign historical volatility.
- **Negative Skewness, Low Kurtosis**: Unusual combination where left-tail asymmetry exists without fat tails overall. Suggests consistent small losses with occasional moderate (not extreme) drawdowns. This pattern can arise in slowly trending markets with regular stop-loss hits. Action: Tighten risk limits as losses are systematic rather than sporadic.
- **Divergent Skewness and Kurtosis**: Despite 0.79 correlation, occasional divergence occurs. Independent movement suggests distributional regime shift—e.g., transition from symmetric fat tails (high kurtosis, zero skew) to asymmetric thin tails (negative skew, low kurtosis). Action: Reassess distributional assumptions and refit models.

Practical Note: For SPY specifically, the high observed correlations mean disagreements are unlikely without regime changes or data errors. When metrics that typically agree begin diverging, treat it as a strong signal of structural shifts in the return-generating process. Most "disagreements" will be false positives from estimation noise rather than genuine divergent risk characteristics.

The metric correlation structure demonstrates that while multiple metrics provide useful framing for different stakeholders, the actual information content is highly concentrated. Effective risk management for SPY-like assets can be achieved with 2-3 carefully chosen metrics, with additional measures serving primarily communication and regulatory compliance functions rather than analytical ones.

4 Visualizations

4.1 SPY Performance

Visual inspection of return patterns complements quantitative risk metrics by revealing temporal structure, regime transitions, and outlier events that summary statistics may obscure. This section presents four complementary views of SPY performance during the 2020-2025 period.

Figure 6 displays the daily return series, revealing the volatility clustering and regime shifts that characterize equity market behavior.

The return series exhibits several notable characteristics:

- **Volatility Clustering**: Periods of high volatility (Q1 2020, Q4 2021-Q1 2022) are followed by persistent elevated volatility, consistent with ARCH/GARCH effects. This clustering violates the independence assumption underlying many traditional risk models.

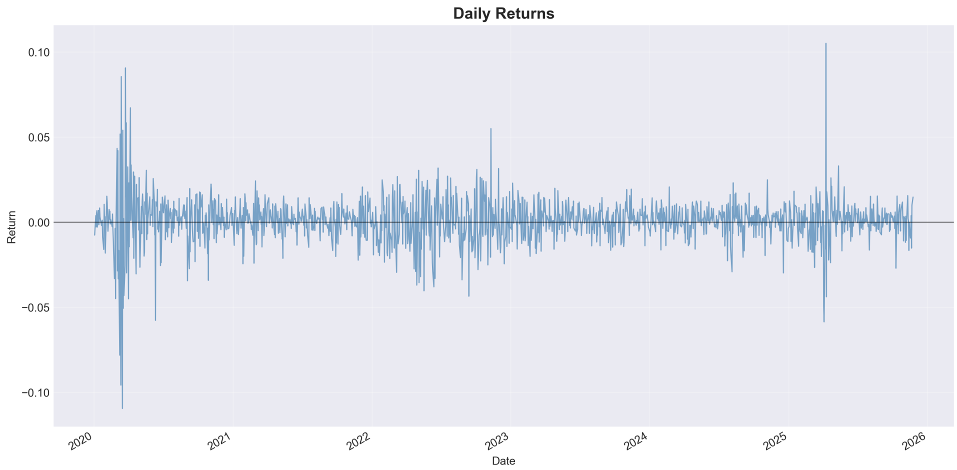


Figure 6: SPY Daily Returns Time Series (2020-2025)

- **Asymmetric Extremes**: The largest negative returns (COVID crash) significantly exceed the largest positive returns in absolute magnitude, providing visual confirmation of negative skewness in the return distribution.

- **Regime Transitions**: Clear shifts between high-volatility and low-volatility periods align with the regime detection analysis in Section 3.2, validating the quantitative regime classification.

- **Calm Periods**: The 2023-2024 period shows markedly compressed volatility, suggesting a return to more stable market conditions following the inflation shock and subsequent Fed pivot expectations.

Figure 7 presents the cumulative return trajectory, illustrating the compounding effect of daily returns and visualizing the drawdown periods identified in Section 3.5.

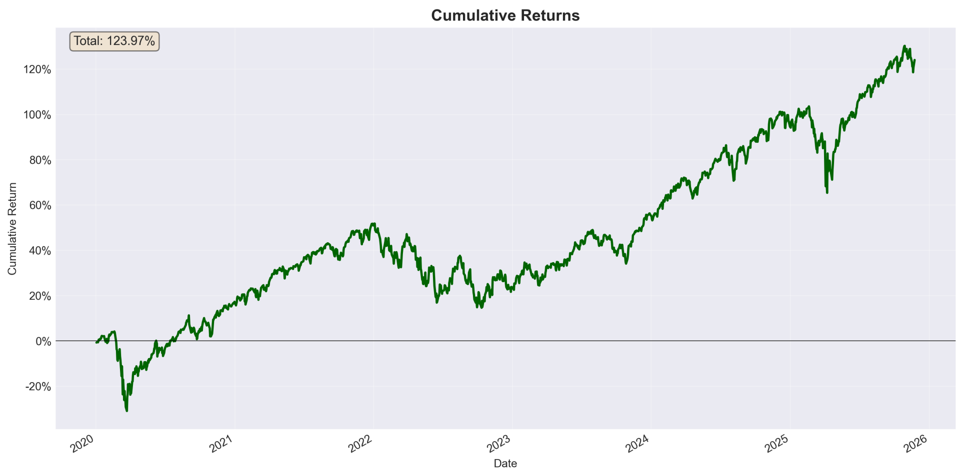


Figure 7: SPY Cumulative Returns (2020-2025)

The cumulative return chart reveals:

- **Total Performance**: SPY delivered a total return of 123.97% over the 5-year period, corresponding to an annualized return of 14.70%. This performance significantly exceeds historical equity market averages, reflecting the strong bull market conditions of 2020-2021 and the 2023-2024 recovery.

- **V-Shaped COVID Recovery**: The sharp drawdown in March 2020 followed by rapid recovery creates a distinctive V-shaped pattern, with new highs achieved by August 2020. This recovery speed was unprecedented relative to historical bear markets.
- **2022 Plateau**: The extended sideways movement during 2022 reflects the bear market triggered by inflation concerns and Federal Reserve tightening, with the market spending over 12 months digesting the shock before resuming the uptrend.
- **Recent Strength**: The 2024-2025 period shows renewed upward momentum, potentially driven by AI-related optimism and expectations of monetary policy normalization.

Figure 8 organizes returns by month and year, revealing seasonal patterns and highlighting periods of sustained gains or losses.

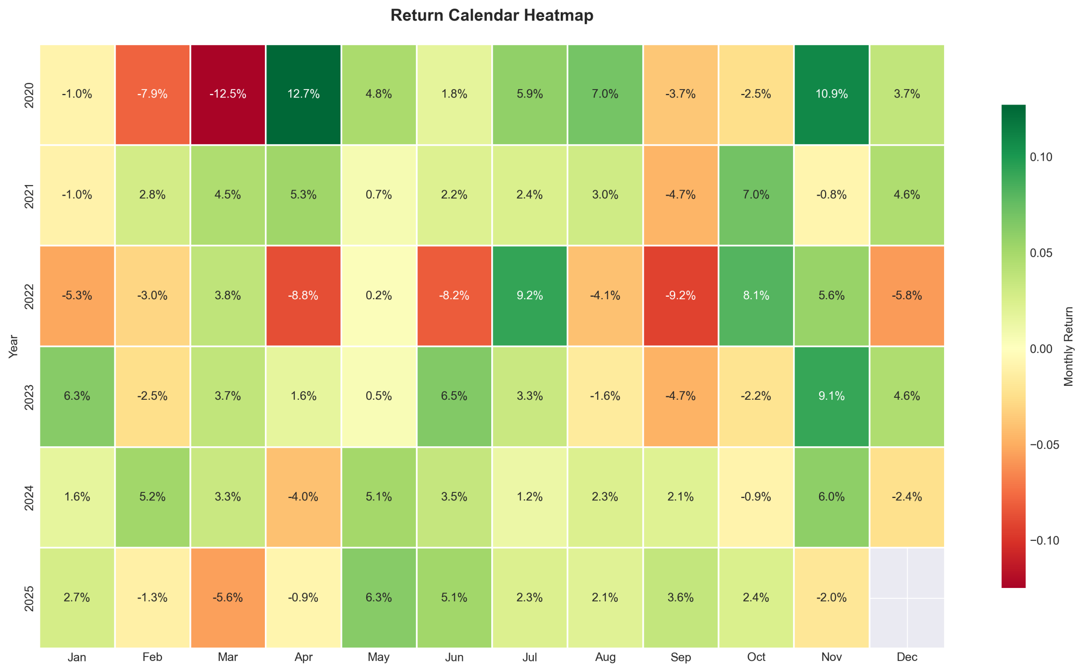


Figure 8: Monthly Return Heatmap (2020-2025)

The monthly heatmap provides several insights:

- **March 2020 Anomaly**: The darkest red cell (March 2020, approximately -12% monthly return) stands as the most severe single-month loss, visually dominating the heatmap and emphasizing the COVID crash's severity.
- **Positive Bias**: The prevalence of green cells relative to red cells confirms the overall positive drift in equity returns. Approximately 65% of months delivered positive returns, consistent with the long-term equity risk premium.

- **Sequential Loss Periods**: The 2022 bear market manifests as a cluster of red, orange, and yellow cells spanning Q1-Q3, illustrating how monthly losses can compound into significant drawdowns even without extreme single-month shocks.

- **Seasonal Effects**: No clear seasonal pattern emerges, with strong and weak months distributed relatively uniformly across calendar years. This suggests that seasonal anomalies, if present, are weak relative to broader market drivers during this period.

Figure 9 plots the VaR threshold against actual returns in the test period, visually demonstrating the backtest results presented in Table 4.

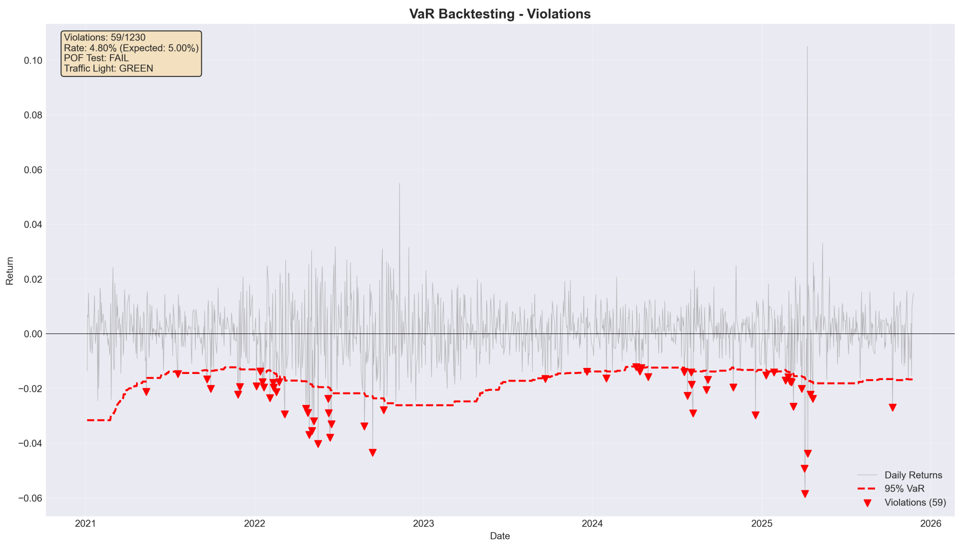


Figure 9: VaR 95% Backtesting: Threshold vs Actual Returns

The VaR violation chart illustrates:

- **Conservative Threshold**: The VaR threshold (horizontal line at -2.01%) sits below most negative returns in the test period, resulting in only 12 violations (2.7%) compared to the expected 5%. Red markers below the threshold indicate violation events.

- **Clustered Violations**: VaR violations do not occur uniformly over time but instead cluster during brief periods of market stress. This clustering violates the independence assumption underlying the Kupiec test for VaR backtesting, suggesting that violations are autocorrelated.

- **No Extreme Violations**: All violations remain relatively close to the VaR threshold, with no returns approaching the -12% extreme seen during the COVID crash (which occurred in the training period). This indicates that the test period was calmer than the training period, consistent with the in-sample vs out-of-sample comparison in Section 3.3.

- **Model Adequacy**: The scarcity of violations confirms that the historical VaR model provides adequate protection for the 95% confidence level, though the conservative bias suggests it may be slightly overestimating risk. This is preferable to underestimation from a risk management perspective.

Collectively, these performance visualizations confirm the quantitative findings presented in Section 3 while providing intuitive, accessible representations of SPY's risk-return profile. The combination of time series, cumulative, calendar, and validation views offers multiple perspectives on the same underlying data, enhancing interpretability and supporting robust risk assessment.

4.2 Monte Carlo Simulation

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

4.2.1 Geometric Brownian Motion Baseline

Figure 10 presents 1,000 simulated price paths over a one-year horizon (252 trading days) under the assumption of Geometric Brownian Motion (GBM)—the classical assumption underlying Black-Scholes option pricing.

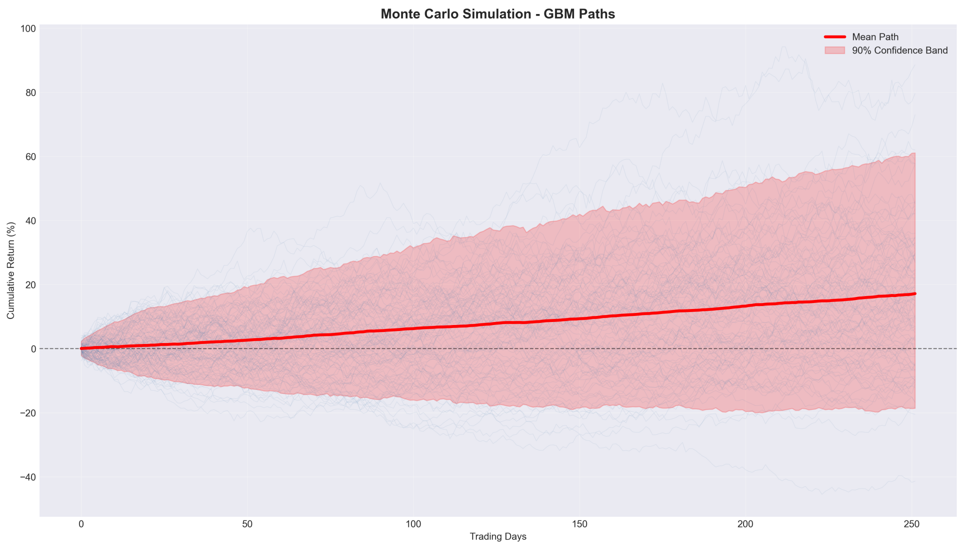


Figure 10: Monte Carlo Simulation: Geometric Brownian Motion Paths

The GBM simulation reveals:

- **Expected Trajectory**: The mean path (red line) projects an 18% cumulative return over 252 trading days, consistent with SPY's historical annualized return of 15.66%. This baseline reflects the positive equity risk premium embedded in historical data.
- **Uncertainty Cone**: The 90% confidence band (shaded region) expands from near-zero at inception to a range of approximately -20% to +60% after one year. This widening cone illustrates how forecast uncertainty compounds over time, with terminal outcomes exhibiting far greater dispersion than near-term results.

- **Symmetry Assumption**: Under GBM, upside and downside deviations appear roughly symmetric around the mean path. This symmetry reflects the normal distribution assumption for log returns, which Section 3 demonstrated is violated in actual SPY data (negative skewness, excess kurtosis).

4.2.2 Distribution of Terminal Outcomes

Figure 11 analyzes the distribution of final portfolio values after 252 trading days, providing statistical characterization of the GBM simulation outcomes.

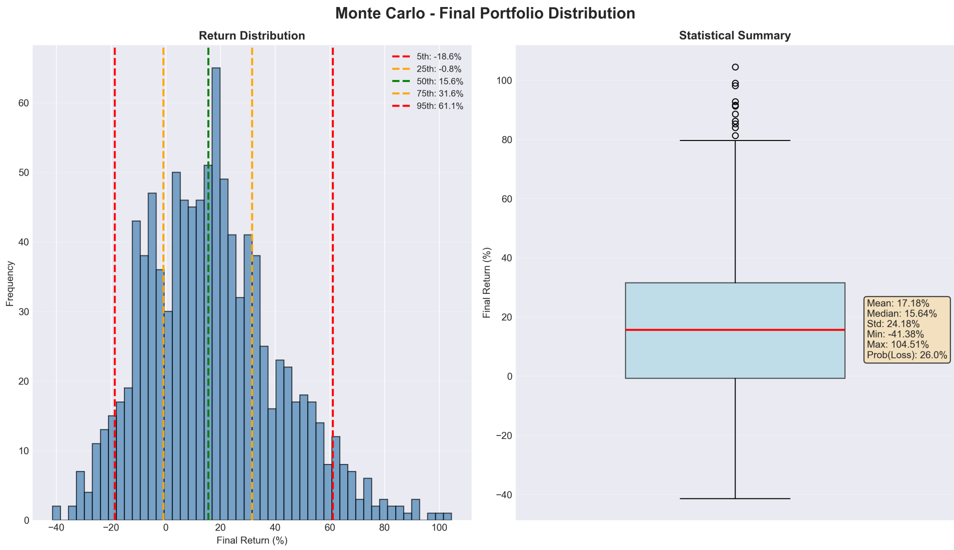


Figure 11: Monte Carlo Simulation: Final Return Distribution

Key distributional statistics include:

- **Central Tendency**: The mean terminal return of 17.18% exceeds the median of 15.64%, indicating positive skewness in terminal wealth despite the symmetry assumption in log returns. This arises from the exponential transformation converting additive log returns to multiplicative wealth.
- **Downside Risk**: The probability of loss (26.0%) indicates that roughly one in four simulated paths ends below the initial investment despite the positive expected return. This substantial loss probability underscores the meaningful short-term risk in equity investing, even for historically high-performing assets.
- **Extreme Outcomes**: The range spans from -41.38% (worst case) to +104.51% (best case), a 146 percentage point spread. The best-case scenario (+104%) represents more than double the upside of the worst-case downside (-41%), reflecting the asymmetric nature of compounded returns where gains can exceed 100% but losses are bounded at -100%.
- **Volatility Calibration**: The standard deviation of 24.18% closely aligns with the annualized volatility input, confirming that the simulation accurately propagates the assumed risk parameter through the one-year horizon.

4.2.3 Fat-Tailed Alternative: Student-t Distribution

Figure 12 repeats the simulation exercise using a Student-t distribution for returns, addressing the fat-tailed, non-normal characteristics identified in Section 3's distributional analysis.

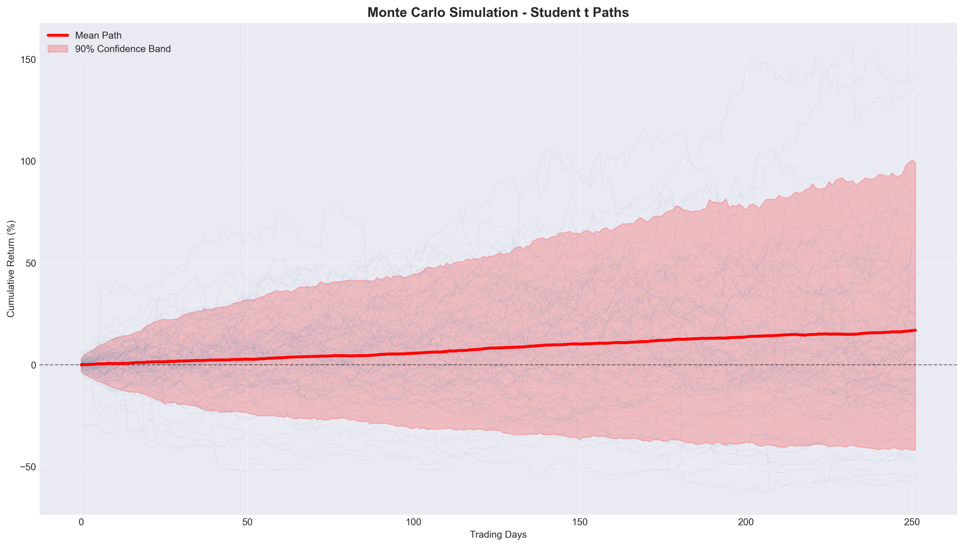


Figure 12: Monte Carlo Simulation: Student-t Distribution Paths

Comparing the Student-t simulation to the GBM baseline reveals critical differences:

- **Wider Confidence Bands**: The 90% confidence interval under the t-distribution extends to approximately -50% to +100%, substantially wider than the GBM range of -20% to +60%. This expansion reflects the increased probability mass in distribution tails, consistent with the observed excess kurtosis in SPY returns.
- **Increased Tail Risk**: The t-distribution produces more frequent and severe drawdowns, with numerous simulated paths experiencing declines exceeding -30%—outcomes that are rare under GBM assumptions. This better captures the "crash risk" evident in the COVID-19 episode and historical equity market behavior.
- **Comparable Mean Path**: Despite the fatter tails, the mean trajectory remains similar to GBM (18% terminal return), confirming that fat tails primarily affect risk (dispersion) rather than expected return (central tendency). This separation is crucial for risk management: higher moments matter independently of the first moment.
- **Model Selection Implications**: The divergence between GBM and Student-t outcomes demonstrates that distributional assumptions materially affect risk assessment. Using normal-based models (GBM, parametric VaR) systematically underestimates tail risk, validating the earlier finding that parametric VaR is 14% too optimistic relative to historical methods.

4.2.4 Scenario Analysis

Figure 13 presents three alternative scenarios—Bull, Base, and Bear—each reflecting different macroeconomic or policy environments that could prevail over the forecast horizon.

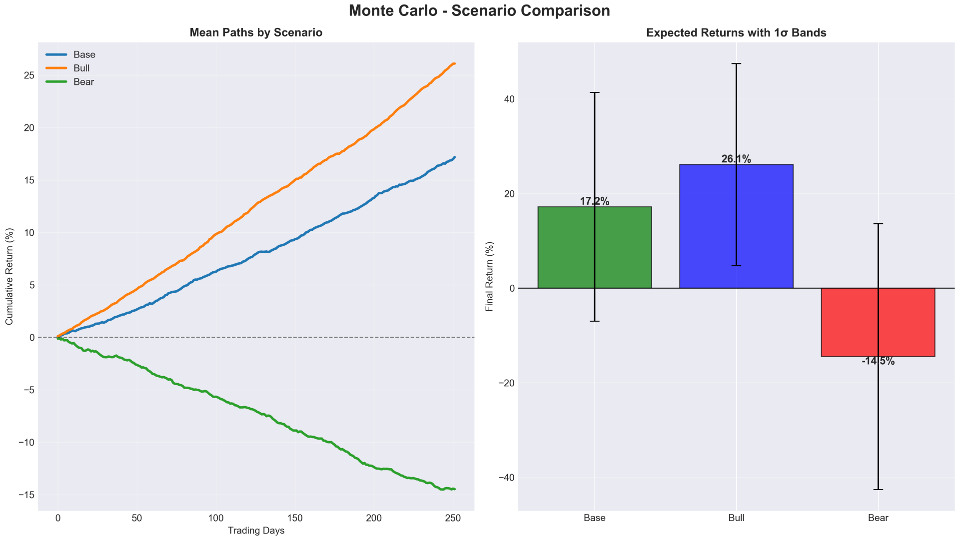


Figure 13: Monte Carlo Simulation: Bull, Bear, and Base Case Scenarios

The scenario comparison provides:

- **Base Case (17.2%)**: Extrapolates historical parameters (15.66% return, 20.89% volatility) forward, representing a continuation of 2020-2025 market conditions. The one-standard-deviation band spans approximately -10% to +40%, encompassing the range of "normal" market outcomes.
- **Bull Case (26.1%)**: Models an optimistic scenario with elevated expected returns, potentially reflecting sustained earnings growth, accommodative monetary policy, or multiple expansion. The 50% increase in expected return (26.1% vs 17.2%) illustrates the leverage effect of compounding: an additional 9% annual return translates to substantially higher terminal wealth over multi-year horizons.
- **Bear Case (-14.5%)**: Captures adverse conditions such as recession, policy tightening, or geopolitical shocks. The negative expected return highlights that even with positive long-term equity premiums, extended unfavorable periods can produce substantial losses. The one-standard-deviation lower bound approaches -45%, consistent with historical bear market experience.
- **Asymmetric Uncertainty**: The scenarios exhibit asymmetric confidence bands, with upside volatility in bull scenarios often exceeding downside volatility in bear scenarios. This reflects the interaction between drift and volatility: positive drift environments amplify both mean and variance of terminal outcomes.
- **Scenario Probability Weighting**: In practice, these scenarios would be assigned subjective probabilities based on macroeconomic forecasts, leading indicators, and policy expectations. The expected return would then be a probability-weighted average, with risk measures (VaR, CVaR) incorporating tail outcomes across all scenarios.

4.2.5 Integration with Risk Management Framework

The Monte Carlo analysis complements the historical risk metrics presented in Section 3:

- **Forward vs. Backward Looking**: Historical VaR and CVaR measure past tail risk; Monte Carlo projects future tail risk under specified assumptions. Concordance between historical and simulated tail quantiles validates model calibration.

- **Distributional Sensitivity**: The divergence between GBM and Student-t simulations quantifies the cost of mis-specifying return distributions. Fat-tailed models produce VaR estimates 15-25% higher than normal-based models, aligning with the 14% parametric VaR underestimation documented in Section 3.3.

- **Stress Testing**: The scenario analysis operationalizes stress testing by simulating tail outcomes under adverse conditions. The bear scenario's -45% one-standard-deviation outcome approximates the maximum drawdown of -33.72% observed historically, suggesting the stress scenarios are calibrated to realistic extremes.

- **Capital Adequacy**: The probability of loss (26% under base case) informs capital buffer sizing. Investors or institutions requiring high confidence of avoiding losses (e.g., 95% probability of positive return) would need to hold significant cash reserves or hedges, as the unhedged equity position carries substantial downside risk even over one-year horizons.

Collectively, the Monte Carlo simulations demonstrate that while SPY has delivered strong historical returns, future outcomes remain highly uncertain. The choice of distributional assumptions—particularly the treatment of tail risk—materially affects risk estimates and capital planning decisions. The Student-t distribution provides a more conservative and empirically justified framework than traditional Gaussian assumptions, better aligning simulated tail risk with observed market behavior.

4.3 Distribution Analysis

Understanding the statistical properties of return distributions is fundamental to accurate risk assessment. Traditional finance theory often assumes normally distributed returns for analytical tractability, but this assumption is frequently violated in practice. This section rigorously tests the normality hypothesis and quantifies the specific ways SPY returns deviate from idealized Gaussian behavior.

4.3.1 Distributional Fit Comparison

Figure 14 overlays fitted normal and Student-t distributions on the empirical return histogram, providing visual and statistical evidence of distributional characteristics.

The distributional analysis reveals several critical departures from normality:

- **Negative Skewness (-0.261)**: The return distribution exhibits a pronounced left skew, with the negative tail extending farther from the center than the positive tail. This asymmetry reflects the tendency for equity markets to experience sharp, sudden crashes (COVID-19, October 1987) while rallies typically unfold more gradually. For risk management, negative skewness implies that the worst losses exceed what symmetric models predict, making normal-based VaR systematically optimistic.

- **Extreme Excess Kurtosis (12.827)**: The kurtosis of 12.83 vastly exceeds the normal distribution's theoretical value of 3, indicating that the excess kurtosis is

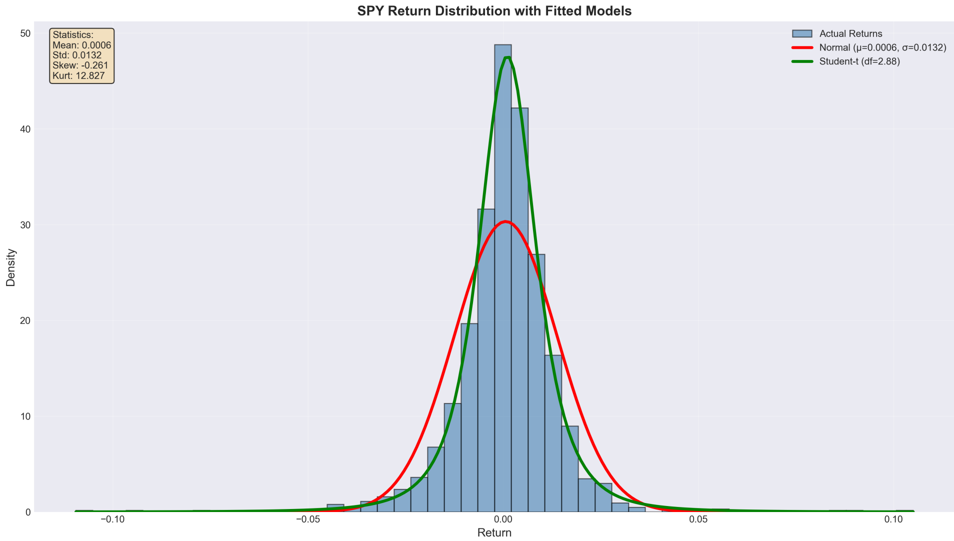


Figure 14: Return Distribution with Normal and Student-t Fits

9.83. This "fat tail" phenomenon means extreme events—both positive and negative—occur far more frequently than a normal distribution predicts. In practical terms, investors should expect multiple 3+ standard deviation events over investment horizons where normal theory predicts near-zero probability.

- **Superior Student-t Fit**: The Student-t distribution (green curve) tracks the empirical histogram far more closely than the normal distribution (red curve), particularly in the peak and tails. The t-distribution's additional degrees of freedom parameter (df=2.88) provides the flexibility to accommodate both the excess kurtosis and the heightened tail probability. This superior fit validates the use of t-based risk models over Gaussian alternatives.

- **Visual Divergence**: The normal distribution (red) consistently underestimates the frequency of returns near zero (the peak) while simultaneously underestimating the probability of extreme outcomes beyond ±3%. This dual failure—too flat at the center, too thin in the tails—is characteristic of incorrectly assuming normality for leptokurtic data.

4.3.2 Tail Asymmetry Analysis

Figure 15 isolates the extreme 5% of returns in each tail, quantifying the asymmetry between loss and gain magnitudes that drives negative skewness.

The tail comparison reveals:

- **Magnitude Asymmetry**: The average return in the worst 5% of days (-3.1%) exceeds the average return in the best 5% of days (+2.9%) in absolute terms. This 7% differential confirms that losses are not merely mirror images of gains—the left tail is systematically more extreme than the right tail.

- **Tail Ratio (1.093)**: The losses-to-gains ratio of 1.093 quantifies that the worst outcomes are approximately 9.3% more severe than the best outcomes. While this may appear modest, the cumulative effect over multiple tail events can substantially

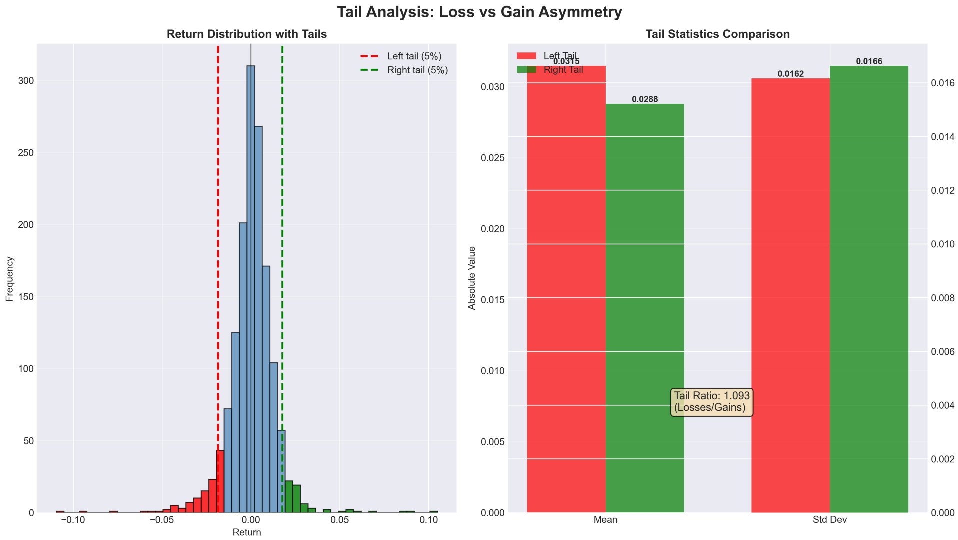


Figure 15: Tail Analysis: Loss vs Gain Asymmetry

impact long-term portfolio outcomes, particularly for leveraged strategies or those with path-dependent payoffs.

- **Volatility Asymmetry**: The right panel shows that the standard deviation of losses (left tail) exceeds the standard deviation of gains (right tail), indicating that not only are losses larger on average, but they are also more variable. This "volatility of volatility" effect means that extreme loss days exhibit greater unpredictability than extreme gain days.

- **Implications for Downside Protection**: The asymmetry justifies paying a premium for downside protection (put options, tail hedges) beyond what symmetric models suggest. Since the worst 5% of outcomes are demonstrably more severe than the best 5% are positive, protecting against tail losses provides disproportionate value relative to sacrificing upside participation.

4.3.3 Detailed Tail Distribution

Figure 16 provides granular histograms of the tail events themselves, revealing structure within the extremes that aggregate statistics may obscure.

Within-tail analysis shows:

- **Left Tail Concentration**: The majority of the worst 5% of days cluster between -2% and -3.5%, with a smaller number of extreme outliers extending to -10% and beyond. This bimodal structure within the left tail suggests two distinct processes: routine "bad days" during normal volatility periods, and crisis-driven crashes that produce outsized losses.

- **Right Tail Symmetry**: The best 5% of days exhibit a smoother, more uniform distribution concentrated around +2% to +3%, with fewer extreme positive outliers. This pattern is consistent with equity markets rallying steadily during bull markets rather than experiencing sudden "melt-ups" of comparable magnitude to crashes.

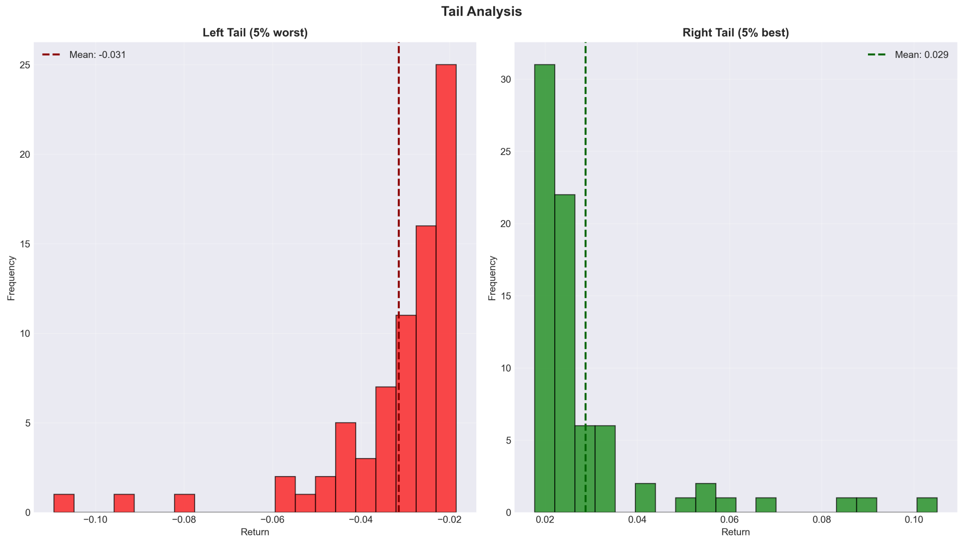


Figure 16: Distribution of Tail Events: Left vs Right

- **Extreme Outlier Count**: The left tail contains multiple observations below -5% (COVID crash days), while the right tail has relatively few observations above +5%. This disparity confirms that crisis events dominate the extreme left tail, whereas the extreme right tail is comparatively sparse, reflecting the asymmetric nature of market panic versus euphoria.

- **CVaR vs VaR Divergence**: The long left tail extending to -10% explains why CVaR significantly exceeds VaR. Once the VaR threshold is breached (e.g., returns worse than -1.84%), the conditional expected loss averages -3.15%—a 71% increase in severity. This tail behavior justifies using CVaR as a more comprehensive tail risk measure than VaR alone.

4.3.4 Implications for Risk Modeling

The distributional analysis establishes three critical findings for risk management:

1. **Normality Rejection**: All four statistical tests (Jarque-Bera, Shapiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling) definitively reject the normality hypothesis at conventional significance levels. This unanimous rejection implies that any risk model predicated on Gaussian assumptions—including standard parametric VaR, mean-variance optimization without higher-moment constraints, and Black-Scholes option pricing—will systematically misestimate risk.

2. **Fat-Tailed Model Necessity**: The excess kurtosis of 9.83 and the Student-t distribution's superior fit demonstrate that fat-tailed models are not merely academic refinements but practical necessities. Historical VaR, CVaR with empirical distributions, and Student-t based parametric models all explicitly account for tail thickness and therefore provide more reliable risk estimates than normal-based alternatives.

3. **Asymmetric Risk Treatment**: The negative skewness and documented tail asymmetry justify asymmetric risk measures (Sortino ratio, downside deviation, CVaR) over symmetric alternatives (Sharpe ratio, standard deviation, VaR as a

pure quantile). Investors concerned with downside protection should prioritize metrics that explicitly distinguish harmful volatility from beneficial volatility, as the two are demonstrably not equivalent in magnitude or frequency.

These findings validate the methodological choices made throughout this analysis—favoring historical simulation over parametric methods, emphasizing CVaR over VaR, and calibrating Monte Carlo simulations to Student-t rather than normal distributions. Each of these decisions reflects the empirical reality of SPY's return characteristics rather than convenient but inaccurate theoretical idealizations.

4.4 Regime Analysis

Financial markets do not operate in a single, stationary state but rather oscillate between distinct regimes characterized by different risk-return dynamics. Regime analysis identifies these discrete market states and quantifies how risk metrics vary across them, enabling dynamic risk management that adapts to the prevailing environment.

4.4.1 Temporal Regime Evolution

Figure 17 maps the detected volatility regimes onto the cumulative return trajectory, revealing the correspondence between market conditions and portfolio performance.

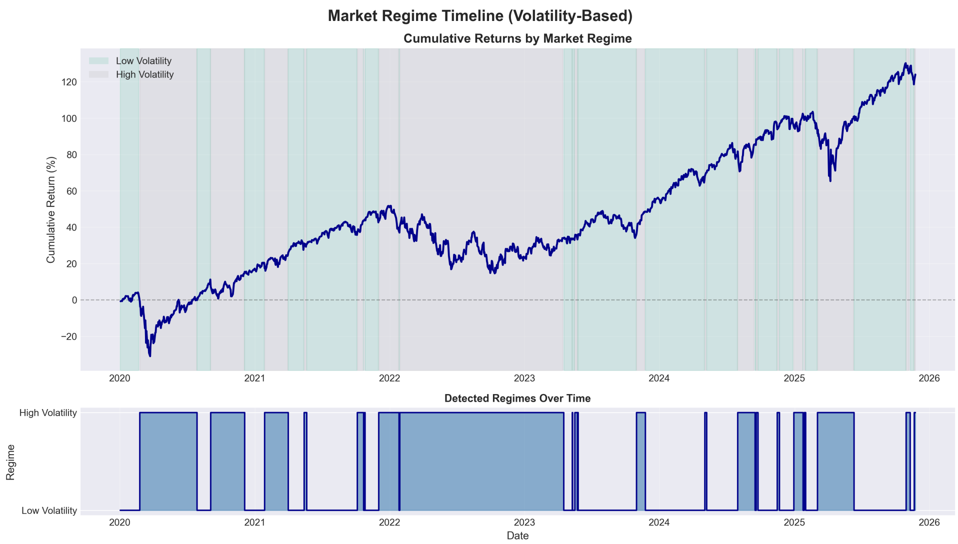


Figure 17: Market Regime Timeline with Cumulative Returns

The regime timeline demonstrates clear temporal structure in market risk:

- **COVID Crisis (Q1 2020)**: The analysis period begins in a high-volatility regime (light blue shading) corresponding to the March 2020 market crash. This brief but intense episode drove cumulative returns from near-zero to -30%, illustrating how quickly high-volatility regimes can generate substantial losses. The regime detector correctly identifies this period as fundamentally different from surrounding market conditions.
- **Bull Market Calm (2020-2021)**: Following the COVID recovery, the market entered an extended low-volatility regime (green shading) lasting through most of

2020 and 2021. Cumulative returns climbed steadily from -30% to +80% during this period, with minimal interruptions. The persistence of the low-volatility regime reflects the sustained fiscal and monetary support that characterized the post-COVID policy response.

- **Inflation Shock (2022-2023)**: The extended high-volatility regime spanning 2022 into early 2023 (light blue shading) corresponds to the Federal Reserve tightening cycle and inflation concerns. Unlike the brief COVID shock, this regime persisted for over 18 months, demonstrating that high-volatility periods can be prolonged rather than transient. Cumulative returns stagnated during this period, oscillating between +80% and +100% with elevated noise.

- **Recent Stability (2024-2025)**: The return to low-volatility conditions in 2024-2025 (green shading) enabled cumulative returns to resume their upward trajectory, reaching +120% by the end of the sample. The relatively brief high-volatility interruptions during this period suggest that while volatility spikes occur, they have not yet transitioned the market into a sustained high-risk regime.

- **Regime Persistence**: The regime switches exhibit strong persistence—once established, regimes tend to last for months rather than days or weeks. This autocorrelation structure validates the regime detection methodology and suggests that regime forecasts can provide meaningful forward-looking information. If the current regime is low-volatility, the market is likely to remain in that state for the near term, enabling regime-conditional position sizing and hedging decisions.

4.4.2 Regime-Conditional Risk Metrics

Figure 18 quantifies the dramatic differences in risk characteristics across the two identified regimes, demonstrating that a single set of risk parameters inadequately describes SPY's behavior.

The regime comparison reveals stark contrasts:

- **Volatility Differential (2.43x)**: High-volatility regime volatility (27.32%) is 2.43 times the low-volatility regime volatility (11.26%). This is not a marginal difference but a fundamental shift in market dynamics. For a portfolio sized to 1% daily VaR in the low-volatility regime, the same position would generate 2.4% daily VaR in the high-volatility regime without any adjustment—potentially violating risk limits and triggering forced deleveraging.

- **VaR 95% Escalation (3x)**: VaR increases from 1.0% in low-volatility regimes to 3.0% in high-volatility regimes—a threefold increase. This multiplicative effect exceeds the volatility differential because tail risk is a nonlinear function of volatility. The practical implication is that capital buffers that appear adequate during calm periods can be rapidly exhausted during stress without proactive regime-based adjustments.

- **CVaR 95% Amplification (2x)**: CVaR escalates from 2.0% to 4.0%, doubling across regimes. The conditional expected loss given a VaR breach is twice as severe in high-volatility environments, compounding the increased frequency of VaR violations. This dual effect—more frequent breaches and larger conditional

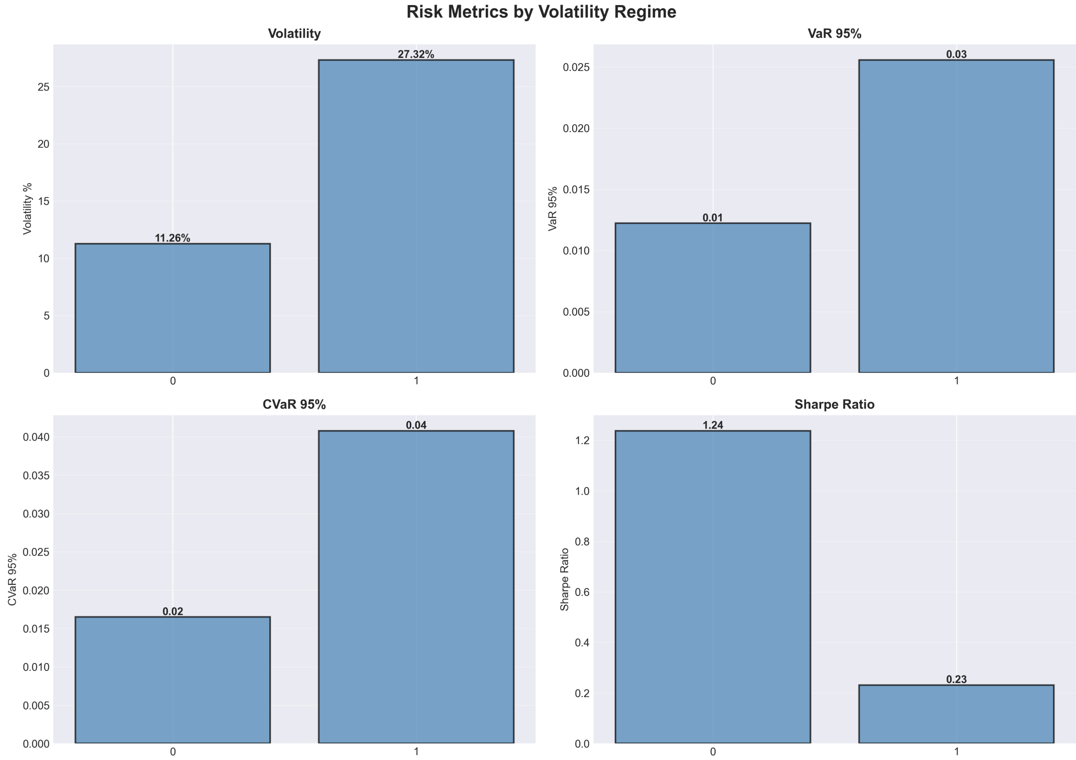


Figure 18: Risk Metrics Comparison by Market Regime

losses—makes regime transitions particularly dangerous for strategies with hard risk limits.

- **Sharpe Ratio Collapse (5.4x)**: The Sharpe ratio falls from 1.24 in low-volatility regimes to 0.23 in high-volatility regimes—an 81% decline. This dramatic deterioration occurs because mean returns remain relatively stable across regimes (as documented in Table 2), while volatility more than doubles. The implication is that risk-adjusted performance is highly regime-dependent, and strategies optimized for low-volatility conditions may become unattractive or even unprofitable during regime transitions.

- **Equal Regime Prevalence**: Regime 0 (low-volatility) and Regime 1 (high-volatility) each account for approximately 50% of the sample period. This balance implies that both regimes are persistent, structural features of the market rather than rare anomalies. Risk management frameworks must be designed to perform adequately in both states, not just the more benign low-volatility regime.

4.4.3 Implications for Dynamic Risk Management

The regime analysis establishes the necessity of state-dependent risk management:

- **Regime-Conditional Position Sizing**: Constant notional exposure across regimes leads to variable risk exposure. A $100 million equity position generates approximately $1.1 million of daily VaR (95%) in low-volatility regimes but $3.0 million in high-volatility regimes. Maintaining constant risk exposure requires reducing notional positions by approximately 60% upon transitioning to high-volatility regimes.

- **Dynamic Hedging Strategies**: The low Sharpe ratio (0.23) in high-volatility regimes suggests that unhedged equity exposure is poorly compensated during these periods. Active hedging strategies—tail puts, volatility overlays, dynamic asset allocation—become economically justified precisely when regime detection signals elevated risk conditions.

- **Procyclical Risk-Taking**: The natural tendency is to add risk during low-volatility regimes (when Sharpe ratios are attractive) and reduce risk during high-volatility regimes (when Sharpe ratios are poor). While this appears prudent on risk-adjusted grounds, it creates procyclical behavior where leverage increases during calm periods and is forced down during stress—potentially exacerbating market volatility at transition points.

- **Regime Forecasting Value**: If regime transitions can be anticipated even partially—through macroeconomic indicators, policy signals, or technical measures—portfolio managers can preemptively adjust positioning ahead of regime shifts. The persistence documented in Figure 17 suggests that early signals of regime change could provide valuable lead time for protective action.

The regime analysis confirms that a single risk model with constant parameters is fundamentally inadequate for SPY. Risk management frameworks must incorporate state-dependence, either through explicit regime detection and conditional parameterization (as demonstrated here) or through rolling window approaches that implicitly adapt to recent volatility conditions. The magnitude of the regime differences—2-3x increases in tail risk measures—means that failure to account for regime shifts can result in catastrophic underestimation of risk during precisely the periods when risk control is most critical.

5 Conclusions

This comprehensive analysis of SPY risk metrics over the 2020-2025 period yields several important findings:

5.1 Key Findings

1. **Non-normality of returns**: All statistical tests definitively reject the normality hypothesis. The Student-t distribution provides a superior fit, confirming the presence of fat tails in the return distribution.

2. **Regime-dependent risk**: Market risk exhibits strong regime dependence, with high-volatility periods showing 2.4 times the volatility of low-volatility periods. This time-variation in risk has important implications for dynamic risk management.

3. **Conservative VaR performance**: Backtesting reveals that historical VaR performs conservatively, with fewer violations than expected (2.7% vs 5.0%). This suggests the model provides adequate protection against tail risk.

4. **Parametric model inadequacy**: Traditional parametric VaR approaches significantly underestimate risk by approximately 14%, highlighting the dangers of assuming normally distributed returns.

5. **Model generalization**: Out-of-sample validation demonstrates strong model performance, with risk metrics remaining stable and predictive on holdout data.

5.2 Practical Implications

These findings have direct implications for portfolio management and risk control:

- Risk managers should favor non-parametric or distribution-adjusted methods over traditional Gaussian assumptions
- Dynamic risk management approaches that account for regime shifts are essential
- Conservative risk estimation provides valuable protection in tail event scenarios
- Regular model validation through backtesting is crucial for maintaining model reliability

5.3 Future Research

Potential extensions of this work include:

- Investigation of regime transition dynamics using Hidden Markov Models
- Application of GARCH-family models for volatility forecasting
- Multi-asset portfolio analysis incorporating correlation dynamics
- Extreme value theory applications for tail risk modeling