Monte Carlo Analysis of SPY Returns and Black–Scholes Option Pricing

Model Validation, Monte Carlo Convergence, and Structural Risk

Milan Fusco

Fall 2025

Research Objective

Primary Question

Can Geometric Brownian Motion (GBM), simulated via Monte Carlo, adequately describe empirical SPY return dynamics?

Secondary Questions

- Does Monte Carlo pricing converge with small statistical error?
- Are deviations from historical returns due to sampling noise or structural misspecification?
- How does the change of measure ($\mathbb{P}$ vs. $\mathbb{Q}$) affect valuation?

Data and Parameter Estimation

Dataset

- SPY daily prices (Yahoo Finance)
- 248 trading days (Dec 2024–Dec 2025)

Log Returns

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

Estimated Parameters (Real-World Measure)

Spot Price S_0	\$681.38
Annual Drift μ	0.1384
Annual Volatility σ	0.1969
Daily Mean	0.000549
Daily Std. Dev.	0.012401

Geometric Brownian Motion

$$\frac{dS_t}{S_t} = \mu\, dt + \sigma\, dW_t$$

Closed-Form Solution

$$S_T = S_0 \exp\left[\left(\mu - \frac{1}{2}\sigma^2\right) T + \sigma W_T\right]$$

Distributional Implication

$$\ln\left(\frac{S_T}{S_0}\right) \sim \mathcal{N}\left(\left(\mu - \tfrac{1}{2}\sigma^2\right) T,\ \sigma^2 T\right)$$

Limitations

- No volatility clustering
- No jumps
- Zero skewness and excess kurtosis

Monte Carlo Framework

Simulation Engine

- `numpy.random.SeedSequence`
- Independent Philox streams
- Deterministic parallel splitting

Convergence Rate

$$\mathsf{SE} = O(N^{-1/2})$$

Increasing simulations reduces variance, but does not correct structural model error.

Outputs include confidence intervals and diagnostic statistics.

Distributional Comparison

Historical SPY Returns

- Mean: 0.000549
- Std. Dev.: 0.012376
- Skewness: 1.014
- Excess Kurtosis: **19.182**

GBM Simulated Returns

- Mean and variance matched by construction
- Skewness ≈ 0
- Excess kurtosis ≈ 0

Key Observation

Higher-order moments strongly reject Gaussian assumptions. Excess kurtosis of 19.182 indicates extreme heavy tails— far beyond what GBM can accommodate.

Hypothesis Testing

Mean Equality (Two-Sample z-Test)

$$z = 0.198, \quad p = 0.843$$

Fail to reject equality of means.

Variance Equality (Levene's Test)

$$W = 23.826, \quad p \approx 0$$

Reject equality of variances — consistent with heavy tails in historical returns.

Normality (Shapiro–Wilk)

$$W = 0.8014, \quad p = 4.45 \times 10^{-17}$$

Strong rejection of Gaussian log-return assumption.

Monte Carlo Forecast Distribution

One-Year Forecast (Real-World Drift)

- Mean: \$782.16
- Median: \$767.26
- 5th percentile (VaR): \$555.33
- 95th percentile: \$1,058.55

$$\mathbb{P}(S_T > S_0) \approx 0.73$$

Lognormal asymmetry implies Mean $>$ Median.

Risk-Neutral Pricing Framework

Under the risk-neutral measure $\mathbb{Q}$:

$$\frac{dS_t}{S_t} = r\,dt + \sigma\,dW_t^{\mathbb{Q}}$$

Drift μ is replaced by risk-free rate r.

This change of measure is justified by the absence of arbitrage (Girsanov's theorem).

Black–Scholes Formula

$$C = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

ATM Option Valuation Results

European Call

$$\hat{C} = \$70.94$$

SE $= 0.3119$ 95% CI: [$70.33, $71.55]

European Put

$$\hat{P} = \$36.24$$

SE $= 0.1813$ 95% CI: [$35.88, $36.60]

Put–call parity holds within Monte Carlo tolerance (residual $= 0.151$), confirming internal pricing consistency.

Option Greeks (Finite Differences)

Sensitivities estimated via central finite differences under the risk-neutral measure. Each Greek requires two simulations per perturbation.

Greek	Definition	ATM Estimate
Delta	$\partial V/\partial S$	0.6396
Gamma	$\partial^2 V/\partial S^2$	0.002763
Vega	$\partial V/\partial \sigma$ (per 1% $\Delta\sigma$)	2.5153
Theta	Daily time decay	-0.1204
Rho	$\partial V/\partial r$ (per 1% Δr)	3.6498

Finite-difference estimates introduce additional Monte Carlo noise, but confidence intervals remain stable relative to price magnitude.

Structural vs. Sampling Error

Monte Carlo Error

$$\text{SE} \sim O(N^{-1/2})$$

- Decreases with simulation count
- Quantifiable via confidence intervals

Structural Model Error

- Cannot be reduced by more simulations
- Arises from incorrect distributional assumptions
- Observed via skewness $(= 1.014)$ and excess kurtosis $(= 19.182)$

Conclusions

- GBM reproduces first and second moments by calibration.
- Historical SPY returns exhibit significant skewness and extreme heavy tails (excess kurtosis $= 19.182$).
- Variance is statistically inconsistent with GBM (Levene's test, $p \approx 0$).
- Monte Carlo pricing converges with small statistical error.
- Structural model misspecification dominates sampling error.

Final Takeaway

Black–Scholes is internally consistent and arbitrage-free, but externally misspecified relative to observed return dynamics.

Future Research Directions

- Heston stochastic volatility
- GARCH dynamics
- Jump-diffusion models (Merton)
- Calibration to implied volatility surface
- Out-of-sample validation

References

- Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities.
- Fusco, M. (2026). mcFramework. `https://github.com/milanfusco/mcFramework`
- NumPy Developers.
- SciPy Developers.
- Matplotlib Developers.
- yfinance.