GBM Performance on Market Data

Up to this point, we've being developing a framework to understand how the GBM can be useful to model a stock price $S(t)$ given that it expresses a deterministic trend in the possible paths of the underlying asset (associated with the expected return μ on the stock) as well a random component, considering the unpredictability of the market (associated with the volatility of the stock σ) [18], but how well it performs in practice?

In order to test the performance of the GBM model, we gather daily historical price data from the stock AAPL for 2 years (501 values from 2022 to 2024). Using the AAPL prices, we calculate the log-returns $\ln[S(t+1)/S(t)]$[5] for each day timestep considered, which, according to the GBM (7), should follow [18, 11]:

$$\ln\left[\frac{S(t+1)}{S(t)}\right] = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma[\Delta W^{\mathbb{P}}(t+1) - \Delta W^{\mathbb{P}}(t)], \tag{22}$$

or in other words, to follow a normal distribution of mean $(\mu - \sigma^2/2)\Delta t$ and variance $\sigma^2 \Delta t$ analogously to (8) [18, 11]:

$$\ln\left[\frac{S(t+1)}{S(t)}\right] \sim \mathcal{N}\left[\left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \sigma^2 \Delta t\right]. \tag{23}$$

For trading days timestamps, we have $\sigma^2_{\text{daily}} \cdot 1 = \sigma^2_{\text{daily}}$, i.e. the variance of the log returns is equal to the squared value of the daily volatility. Analogously, the annual variance can be calculated as $\sigma^2_{\text{annual}} \cdot \sqrt{1/252}$ in the 252-trading days framework. Through a proportion analysis one can figure out that [13]:

$$\sigma_{\text{annual}} = \sigma_{\text{daily}}\sqrt{252}, \tag{24}$$

meaning the annual volatility can be achieved through a $\sqrt{252}$ factor upon the daily volatility or σ_{daily} calculated by multiplying σ_{annual} by a factor $\sqrt{1/252}$.

From the log-return of the stock prices we can calculate the 2-year *historical volatility* [15, 13] of AAPL by computing the standard deviation of the log return data[6]:

$$\sigma_{\text{daily}} = \text{Std}(R_t) = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N-1}(R_t - \bar{R})^2}, \quad R_t = \ln\left(\frac{S_{t+1}}{S_t}\right), \tag{25}$$

where N is the sample size of 500 log-returns and $\bar{R}$ stands for the arithmetic mean of the log-return data:

$$\bar{R} = \frac{1}{N}\sum_{t=1}^{N} R_t. \tag{26}$$

The estimative can then be multiplied by $\sqrt{252}$ to yield the annual volatility according to (24):

$$\sigma_{\text{daily}} \approx 0.018294 \quad \Rightarrow \quad \sigma_{\text{annual}} = 0.018294\sqrt{252} \approx 0.290416, \tag{27}$$

or roughly a 29% volatility per year on the AAPL stock price.

[5]Notice that the list of log returns will be one unit shorter than the list of prices, since the price $S(t)$ in $\ln[S(t+1)/S(t)]$ won't exist for the first day (there will be 500 log-returns in the context of the experiment).

[6]Here I've used t as a discrete index, such that $t \in \{1, 2, \dots, N\}$, representing the daily time steps.

The historical volatility can be an easy way to estimate stock volatility: It only relies on historical price data using basic statistics, making no assumption about the data distribution and being thus purely empirical [15, 13]. Although one could try to use it to estimate the range in which the prices would be in the future, its important to state that it only reflects the past behavior of the prices, not necessarily what's to come. Financial markets are non-stationary, thus past patterns may not hold in the future [11].

Next we can plot a histogram together with a normal PDF, both using the log returns from the AAPL prices as input data, observing how well their area match each other (for the PDF the area under the curve). The described plot is shown in figure (1). For the purpose of showing this graphic comparison, the histogram was created in such a way for it to have the area under all bars adding to 1, as in a probability density function [15].

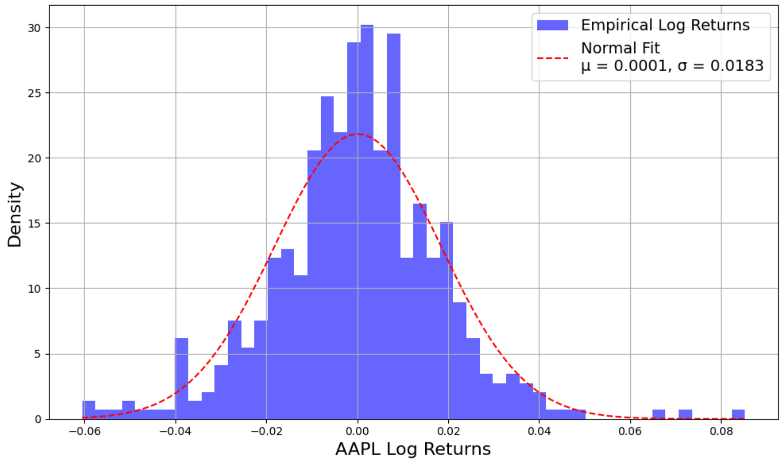

Figure 1: Plot of histogram and a fitted normal distribution both from AAPL historical price log-returns from 2022 to 2024 (501 empirical prices). The x-axis shows the lognormal return values $\ln[S(t+1)/S(t)]$ meanwhile the y-axis represents the density values of the histogram (height of the bars).

The height of the histogram bars in figure (1) represents their density. For a particular bar: being f the frequency the log return points in the x-axis fall into the bar interval, N the total number of log return points in the x-axis (since we have 501 prices, then $N = 500$ log return values) and w the bin width, we have that the density or height of the histogram bar ρ is given by

$$\rho = \frac{f/N}{w}. \tag{28}$$

The bin width w serves as a normalization of the relative frequency of the bin f/N, ensuring that the area below the bar represents probability [15].

Analyzing figure (1), the first benefit from the GBM model for us to take notice of is that the model correctly avoids the possibility of negative stock prices [18]. A key property of lognormally distributed processes is that it has probability 1 of being positive, this is $S(t) > 0, \ \forall\, t$ which wouldn't happen if we assumed $S(t)$ to have a normal distribution. When looking at figure (1) we can see that the histogram area adheres tightly to the area under the Gaussian PDF curve, showing us that stock log-returns $\ln S(t+1)/S(t)$ empirically match a normal distribution, at least in a short time horizon (such as the study's 2 years interval) [17].

An interesting way to estimate how well the normal PDF curve adheres to the empirical stock prices, would be to evaluate the standard deviation parameter from the best fitted Gaussian curve shown in figure (1) by the use of the optimization parameter technique Maximum Likelihood Estimation (MLE), obtaining the fitted version of the volatility $\hat{\sigma}_{\text{daily}}$. The resulting standard deviation found and volatility estimation are the following:

$$\hat{\sigma}_{\text{daily}} \approx 0.018276 \quad \Rightarrow \quad \hat{\sigma}_{\text{daily}} = 0.018276\sqrt{252} \approx 0.290126. \tag{29}$$

Hence the historical volatility, purely empirical, calculated in (27) reaches pretty close to the normal PDF fit of the log-return data in (29), having the latter an error of $\sigma_{\text{daily}} - \hat{\sigma}_{\text{daily}} \approx 0.00029$ when compared with the former.

Although we have a good fit of the AAPL empirical data in figure (1), the model has many downsides. Real markets exhibit sudden jumps or shocks in stock price values due to unforeseen events such as news, crises, opening up of the market etc. meanwhile the GBM assumes continuous paths[7]. The model presuppose that drift/expected return μ and volatility σ are constant through time which also doesn't reflect reality [13]. Volatility can better modeled as stochastic [12], having sometimes properties such as mean-reversion [8, 7] (tend to fluctuate around their long-term average or mean). The drift/expected on the other hand may change overtime due to macroeconomic factors.

One of the major concerns regarding the GBM model is that it underestimates tail risk [19]: the lognormal distribution decays too rapidly in the tails. This implies that to extreme losses or gains (i.e., large absolute values of log-returns) are assigned very low probabilities, which contradicts empirical observations in financial markets [19]. In figure (1), we observe that extreme log-return values do occur with non-negligible frequency — particularly in the tails — challenging the assumption of light tails inherent to the GBM. Real-world return distributions are known to exhibit fatter tails and higher peaks than those predicted by the normal distribution assumed for GBM log-returns which would clearly be a better fit for the data shown in in figure (1).

[7]Being more precise, the GBM model is almost surely continuous, this is, a stochastic process where, for all possible outcomes (i.e., with probability 1), the function representing the process's evolution overtime is continuous. Meaning that while instances where the process isn't continuous occur, they will happen with probability 0 [14, 11].

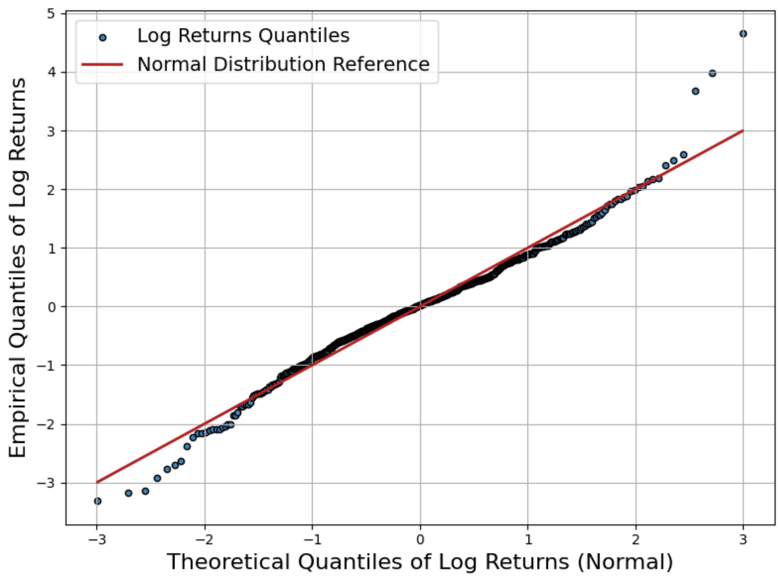

Figure 2: q-q plot from AAPL historical price log returns from 2022 to 2024. The x-axis shows the theoretical quantiles of the log returns (standard normal distribution values) meanwhile the y-axis represents empirical AAPL $\ln[S(t+1)/S(t)]$ quantile values (standardized). The 45° red line represents the standard normal distribution reference.

To further examine this underestimation of tail risk, figure (2) presents a Q–Q plot (quantile–quantile plot) comparing AAPL's empirical log-returns with the theoretical quantiles of a standard normal distribution. In this plot, quantiles represent the values below which a given proportion of the data falls — for instance, the 25th percentile is the value below which 25% of the log-returns lie. If the data were normally distributed, the points would fall along the red 45° reference line.

The Q–Q plot in figure (2) reinforces the earlier observations. The log-return data aligns well with the normal distribution in the central region, near the mean of zero, indicating that the GBM model captures the average behavior of returns reasonably well. However, deviations become clear in the extreme quantiles: points in the upper and lower tails stray from the reference line, revealing the presence of fatter tails in the empirical distribution. These tail deviations highlight the model's inability to account for large market moves [19]. Moreover, the slight S-shaped curve of the data suggests asymmetry — or skewness — that the symmetric normal distribution fails to capture.

In his book *The Black Swan* [19], professor Nassim Taleb strongly criticizes the reliance on the normal distribution in modeling financial phenomena — including assumptions such as those underlying the GBM. He argues that models based on Gaussian statistics dangerously underestimate the frequency and impact of rare, high-consequence events — the so-called "Black Swans." According to Taleb, this leads to a systemic underappreciation of real-world risk, particularly in the tails of the distribution where extreme events lie (sometimes called Black Swans) [19]. In the context of GBM, this critique is directly reflected in the Q–Q plot and histogram of AAPL log-returns, which visually expose the model's failure to capture the empirical distribution's heavy tails. Taleb advocates for adopting more robust, fat-tailed models that better reflect the reality of financial markets, where outliers are not just possible — they are inevitable.

Python Implementation

In this section the python script used when creating the performance GBM example is presented. The code collects and analyzes historical stock data for Apple Inc. (AAPL) over a specified period using the `yfinance` API. After importing necessary libraries, the ticker and date range are defined. To avoid redundant API calls and ensure reproducibility, the script saves the fetched data into an Excel file inside the project directory. If the file already exists, the script loads it directly, improving efficiency and reducing external dependencies.

Once the data is prepared, the script calculates the daily log returns and estimates both daily and annualized historical volatility using the standard deviation. A normal distribution is then fitted to the log returns to check compatibility with the assumptions of the Geometric Brownian Motion (GBM) model. To visually compare the empirical distribution with the theoretical model, the script plots both the histogram with the fitted normal PDF and a Q–Q plot. These plots help verify whether the log returns exhibit normality and whether GBM is a suitable model for the stock's behavior.

The complete and modular version of this script — organized into reusable functions and classes — is available in the repository: Black-Scholes Derivations and Scripts (GitHub).

```python
# Import necessary packages
import datetime as dt
import os
import numpy as np
import pandas as pd
import scipy.stats as stats
import matplotlib.pyplot as plt
import yfinance as yf
from pathlib import Path

# Defining the equity ticker
ticker = "AAPL"
# Defining start and end dates
start_date_str = dt.date(2022, 1, 1).strftime("%Y-%m-%d")
end_date_str = dt.date(2024, 1, 2).strftime("%Y-%m-%d")

# Getting current folder (assuming script is inside the project folder)
computer_path = Path(__file__).parent.__str__()
# Defining folder and file to save historical prices
folder_to_save_name = "yfin_hist_prices"
prices_file_name = f"AAPL_historical_prices_{start_date_str}_{end_date_str
    }.xlsx"
path_prices = os.path.join(computer_path, folder_to_save_name,
    prices_file_name)

if not os.path.exists(path_prices):
    # Download historical price data if file doesn't exist
    df_eq_price_data = yf.download(
        tickers=ticker,
        interval="1d",
        start=start_date_str,
        end=end_date_str,
        progress=False
    ).dropna().droplevel(level='Ticker', axis=1).reset_index(drop=True)
    df_eq_price_data.to_csv(path_prices, index=False)
else:
    # Load from file, clean and reset index
```

```
    df_eq_price_data = pd.read_csv(path_prices)

# Get market close prices, forcing float65 dtype
prices = df_eq_price_data["Close"].astype('float64')
# Compute daily log returns
log_returns = np.log(prices / prices.shift(1)).dropna()

# Compute the historical volatility
sigma_hist_daily = log_returns.std()
sigma_hist_annual = sigma_hist_daily * np.sqrt(252)
print(f"Historical volatility: sigma (daily) = {sigma_hist_daily:.4f}, "
      f"sigma (annual: 252) = {sigma_hist_annual:.4f}.")

# Fit normal distribution to log returns data
mean, variance = stats.norm.fit(log_returns)
print(f"The estimated fitting parameters: mu_log (daily) = {mean:.4f}, "
      f"sigma (daily) = {variance:.4f}.")

# Histogram vs normal PDF plot configuration
x_vals = np.linspace(log_returns.min(), log_returns.max(), 100)
pdf_vals = stats.norm.pdf(x_vals, mean, variance)
plt.figure(figsize=(10, 6))
plt.hist(log_returns, bins=50, density=True, alpha=0.6,
         color='blue', label="Empirical Log Returns")
plt.plot(x_vals, pdf_vals, 'r--',
         label=f'Normal Fit\nmu = {mean:.4f}, sigma = {variance:.4f}')
plt.xlabel("AAPL Log Returns", fontsize=16)
plt.ylabel("Density", fontsize=16)
plt.legend(fontsize=14)
plt.grid(True)
plt.tight_layout()
plt.show()

# Q-Q Plot configuration
log_returns_std = (log_returns - log_returns.mean()) / log_returns.std()
qq_data = stats.probplot(log_returns_std, dist="norm")
theoretical_quantiles = qq_data[0][0]
empirical_quantiles = qq_data[0][1]
plt.figure(figsize=(8, 6))
plt.scatter(theoretical_quantiles, empirical_quantiles,
            color='steelblue', edgecolor='black', s=20,
            label='Log Returns Quantiles')
plt.plot(theoretical_quantiles, theoretical_quantiles,
         color='firebrick', linewidth=2,
         label='Normal Distribution Reference')
plt.xlabel("Theoretical Quantiles of Log Returns (Normal)", fontsize=16)
plt.ylabel("Empirical Quantiles of Log Returns", fontsize=16)
plt.grid(True)
plt.legend(fontsize=14)
plt.tight_layout()
plt.show()
```

References

[1] Tomas Björk. *Arbitrage theory in continuous time*. Oxford university press, 2009.

[2] Fischer Black and Myron Scholes. "The pricing of options and corporate liabilities". In: *Journal of political economy* 81.3 (1973), pp. 637–654.

[3] Fischer Black and Myron Scholes. "The valuation of option contracts and a test of market efficiency". In: *The Journal of finance* 27.2 (1972), pp. 399–417.

[4] Merton Robert C. "Theory of Rational Option Pricing". In: *The Bell Journal of Economics and Management Science* 4.3 (1973), pp. 167–183.

[5] Freddy Delbaen and Walter Schachermayer. "A General Version of the Fundamental Theorem of Asset Pricing". In: *Mathematische Annalen* 300.1 (1994), pp. 463–520. DOI: 10.1007/BF01450498.

[6] Eugene F. Fama. "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *The Journal of Finance* 25.2 (1970), pp. 383–417. DOI: 10.2307/2325486.

[7] Jean-Pierre Fouque and Matthew Lorig. "A Fast Mean-Reverting Correction to Heston's Stochastic Volatility Model". In: *SIAM Journal on Financial Mathematics* 2.1 (2011), pp. 118–150. DOI: 10.1137/090761458.

[8] Jean-Pierre Fouque, George Papanicolaou, and K. Ronnie Sircar. "Mean-Reverting Stochastic Volatility". In: *International Journal of Theoretical and Applied Finance* 3.1 (2000), pp. 101–142.

[9] Igor Vladimirovich Girsanov. "On transforming a certain class of stochastic processes by absolutely continuous substitution of measures". In: *Theory of Probability & Its Applications* 5.3 (1960), pp. 285–301.

[10] Paul Glasserman. *Monte Carlo methods in financial engineering*. Vol. 53. Springer, 2004.

[11] Lech A. Grzelak and Cornelis W. Oosterlee. *Mathematical Modeling and Computation in Finance: With Exercises and Python and MATLAB Computer Codes*. World Scientific Publishing Europe Ltd, 2019. ISBN: 978-1786348050.

[12] Steven L. Heston. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". In: *Review of Financial Studies* 6.2 (1993), pp. 327–343.

[13] John C. Hull. *Options, Futures, and Other Derivatives*. 10th ed. Pearson, 2022. ISBN: 978-1-292-40856-2.

[14] Ioannis Karatzas and Steven E Shreve. *Brownian motion*. Springer, 2021, pp. 47–127.

[15] Peter E Kloeden. *Stochastic Differential Equations*. Springer, 2011, pp. 1520–1521.

[16] Bernt Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[17] Krishna Prasad et al. "Effectiveness of Geometric Brownian Motion Method in Predicting Stock Prices: Evidence from India". In: *Asian Journal of Accounting & Governance* 18 (2022), pp. 121–134. DOI: 10.17576/AJAG-2022-18-09.

[18] Steven E Shreve et al. *Stochastic calculus for finance II: Continuous-time models*. Vol. 11. Springer, 2004.

[19] Nassim Nicholas Taleb. *The Black Swan: The Impact of the Highly Improbable*. 2nd. Harlow, England: Penguin Books, 2008. ISBN: 9780141034591.

[20] Paul Wilmott, Jeff Dewynne, and Sam Howison. *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press, 1994. ISBN: 978-0952208204.