Outline: Geometric Brownian Motion as a Model for Stock Prices

Kai S. Wong (kaiwong@princeton.edu; kaiw@asagi.waseda.jp)

July 2025

1 Background

In 1872, the botanist Robert Brown was studying pollen grains suspended in water under a microscope when he observed small particles that exhibited seemingly random motion. Brown's observations of the 'random walk' laid the foundation for our modern understanding of atoms, random (stochastic) processes, and other fields of math and physics. Decades later, the mathematician Louis Bachelier demonstrated in his PhD dissertation *Théorie de la spéculation* that the price of a stock can be modeled as a stochastic process, based on the same random motion observed by Brown. Bachelier was one of the first to model a stochastic process, and his application of mathematics to finance forever altered the field of finance. He is now regarded as the forefather of mathematical finance.

Inspired by this story, I have recently become interested in the intersection of physics, mathematics, and finance. Like Bachelier, I want to use stochastic processes (Brownian motion) to simulate stock prices. From these simulations I will derive the datasets that I will use in this project.

2 Objective

1. Use a geometric Brownian motion (GBM) to simulate stock prices. It is obtained by solving the following stochastic differential equation (SDE):

   $$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \tag{1}$$

   where S_t is the asset price at time, μ and σ are respectively the expected return (drift) and volatility coefficients, and W_t denotes the Wiener process (Brownian motion).

   This model makes a few key assumptions, namely: log-normal returns, an efficient market, and constant parameters (drift and volatility). I will estimate the parameters in this model, such as μ and σ, from real stock data.

2. Apply the basic statistical tools we learned in class (hypothesis testing and linear regression) to test assumptions about GBM on the simulated data. The assumptions I will test are as follows:

   (i) Log returns are normally distributed.

(ii) Successive returns are independent (The efficient-market hypothesis (EMH))

(iii) Simulated volatility matches target volatility.

By doing so I will be able to empirically verify/refute the GBM assumptions, and highlight the discrepancies between a simplified theoretical model and complex financial reality.

3 Methodology

In order to simulate stock prices using a GBM based algorithm, we must solve our SDE. It is possible to approximate the solution by discretizing our SDE in time using numerical methods such as Euler-Maruyama, and then incrementing it by Δt. However, GBM is a well studied stochastic process, with an analytical solution that can be found using Itô's Lemma on the log transformation. This is an exact solution, meaning that even with relatively large time steps (Δt) the normal log-normal distribution is preserved. As such, we will use this approach. Beginning with out SDE for GBM, our proof is as follows:

Proof.

$$dS_t = \mu S_t \, dt + \sigma S_t \, dW_t \tag{1}$$

We have Itô's Lemma where $y(t, X_t)$ represents a time-dependent function of a stochastic process

$$dy(t, X_t) = \frac{\delta y}{\delta t}dt + \frac{\delta y}{\delta X_t}dX_t + \frac{\delta^2 y}{\delta X_t^2}dX_t^2 \tag{2}$$

Consider a log transformation on the stock process such that

$$g(S_t) = ln(S_t) \tag{3}$$

Apply Itô's Lemma to the log transformation function

$$dg(S_t) = \frac{\delta y}{\delta t}dt + \frac{\delta y}{\delta S_t}dS_t + \frac{1}{2}\frac{\delta^2 y}{\delta S_t^2}dS_t^2 \tag{4}$$

It follows from equations 3 and 1

$$\frac{\delta g}{\delta t} = 0$$

$$\frac{\delta g}{\delta S_t} = \frac{1}{S_t}$$

$$\frac{\delta^2 g}{\delta S_t^2} = \frac{1}{S_t^2}$$

$$dS_t^2 = 0 + 0 + \sigma^2 S_t^2 dt = \sigma^2 S_t^2 dt$$

By substitution of these into equation 4 we obtain

$$dg(S_t) = 0 + \frac{1}{S_t}(\mu S_t\, dt + \sigma S_t\, dW_t) + \frac{1}{2}(-\frac{1}{S_t^2})\sigma^2 S_t^2 dt$$

$$dg(S_t) = \mu\, dt + \sigma\, dW_t - \frac{1}{2}\sigma^2\, dt$$

$$dg(S_t) = (\mu\ - \frac{1}{2}\sigma^2)\, dt + \sigma\, dW \tag{5}$$

Integrate both sides from $[t_0, t]$

$$\int_{t_0}^{t} dg(S_t) = \int_{t_0}^{t} (\mu\ - \frac{1}{2}\sigma^2)\, dt + \int_{t_0}^{t} \sigma\, dW_t$$

$$g(S_t) - g(S_{t_0}) = (\mu\ - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W_t - W_{t_0})$$

We substitute equation 3 into 5 to obtain

$$ln(S_t) - ln(S_{t_0}) = (\mu\ - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W_t - W_{t_0})$$

$$ln(\frac{S_t}{S_{t_0}}) = (\mu\ - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W_t - W_{t_0})$$

$$e^{ln(\frac{S_t}{S_{t_0}})} = e^{(\mu\ - \frac{1}{2}\sigma^2)(t - t_0) + (\sigma(W_t - W_{t_0})}$$

$$\frac{S_t}{S_{t_0}} = e^{(\mu\ - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W_t - W_{t_0})}$$

$$S_t = S_{t_0}\, e^{(\mu\ - \frac{1}{2}\sigma^2)(t - t_0) + \sigma(W_t - W_{t_0})} \tag{6}$$

$\square$

The closed-form solution to our SDE in equation 1. We will employ this in our model. I will also employ the Euler-Maruyama method to compare the accuracy and time complexity of a numerical method to an exact solution.

We must estimate values for the constants μ and σ. We will obtain this from the sample mean and sample variance of real market data. In this report I will estimate these parameters from Apple's (AAPL) daily closing values from 4.1.2024 to 6.12.2025. Then, I will generate 2000 simulated paths using these parameters.

With these simulated paths, we will now employ statistical methods to check our assumptions about GBM. Below I have stated the null hypotheses and alternative hypotheses I will test for as well as the methods that I will use.

1. Log returns are normally distributed (Shapiro-Wilk, $\alpha = 0.05$).

   H_0: The log of our returns is distributed normally. H_a: The log of our returns is not distributed normally.

2. Successive returns are independent (Ljung-Box, $\alpha = 0.05$).

   H_0: Success returns are not correlated H_a: Successive returns are correlated.

3. Simulated volatility matches target volatility (One-sample t-test, $\alpha = 0.05$).

   H_0: $\mu = \mu_0$; H_a: $\mu \neq \mu_0$, where μ is the mean simulated volatility and μ_0 is the input volatility.

4 Data

The data in this project was obtained by running the GBM model in equation 6 to generate (N=2000) simulated paths of stock prices.

Real stock market data was obtained using Yahoo! Finance's API yfinance for AAPL from 4.1.2024 to 6.12.2025. This data is stored in a .xls file.

5 Results

Below are the results from executing my python code.

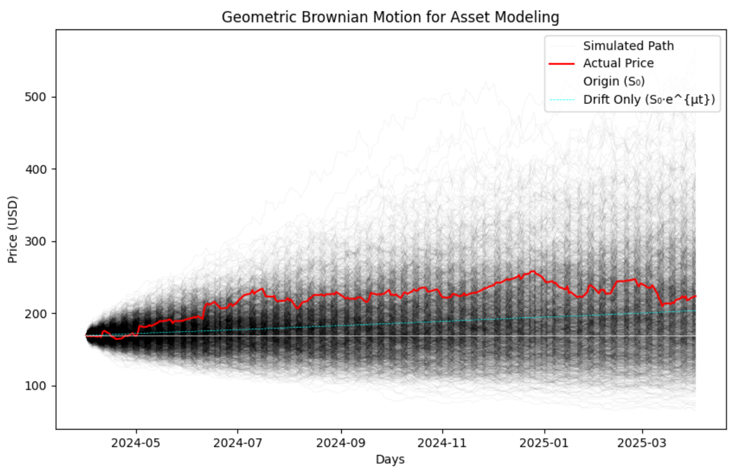

Figure 1: 2000 simulated paths compared to the actual path using the closed-form solution to GBM

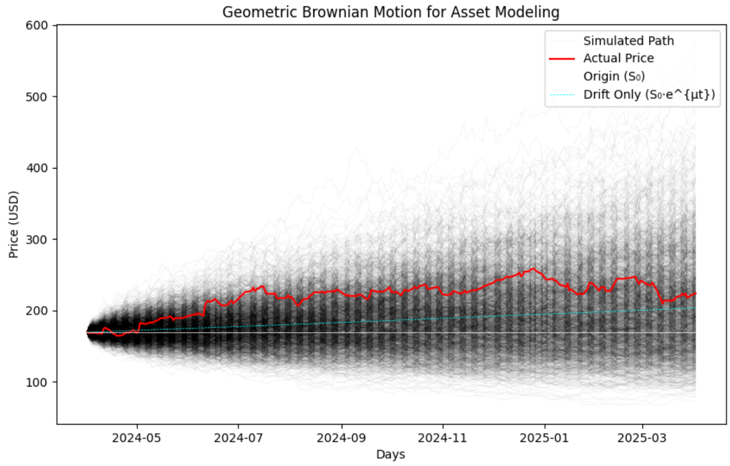

Figure 2: 2000 simulated paths compared to the actual path using the Euler-Maruyama method for GBM

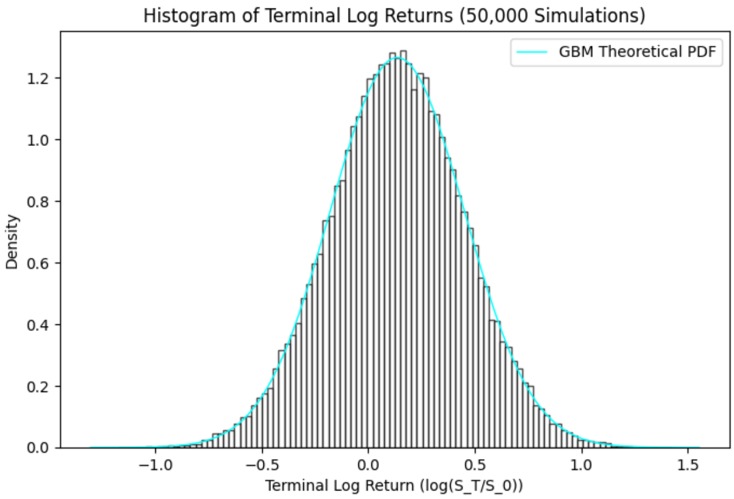

Figure 3: The terminal log returns plotted on a histogram along with the theoretical PDF

Method (2000 simulations)	**Time Complexity**	**Time (s)**
Closed-Form	O(n)	0.0577 seconds
Euler-Maruyama	O(n)	0.2630 seconds

Table 1: Comparison of Time Complexity and Execution Time for Different Methods

6 Analysis

We begin with Test 1. To check if the log returns are normally distributed, we first calculated the log returns for each day and each of the 2000 simulated paths, using the formula $r_t = \log \frac{S_t}{S_{t-1}}$. We found that the proportion of simulated paths with a Shapiro–Wilk p-value greater than $\alpha = 0.05$ is 0.953. Therefore, for the vast majority (95.3%) of simulations, we do not reject the null hypothesis (H_0: The log of our returns is distributed normally). This is demonstrated visually in Figure 3.

We conducted Test 2. To check if successive returns are independent, we first calculated the log returns for each day and each of the 2000 simulated paths, using the formula $r_t = \log \frac{S_t}{S_{t-1}}$. We found that the proportion of simulated paths with a Ljung-Box (lag = 10) p-value greater than $\alpha = 0.05$ is 0.942. Therefore, for the vast majority (94.2%) of simulations, we do not reject the null hypothesis (H_0: Success returns are not correlated).

We conducted Test 3. To check if simulated volatility matches target volatility, we first calculated the sample mean of the volatiles for the simulated paths. We found that with a one-sample t-test using the sample mean and the input volatility, the p-value is $0.8981 > \alpha = 0.05$. Therefore, we do not reject the null hypothesis (H_0: The mean simulated volatility equals the input volatility).

Additionally, it was observed that the MSE between the exact closed-form solution and the Euler-Maruyama method is 0.384457 which means that RMSE ≈ 0.62. The time complexity of both methods is the same, however the time it takes to execute the closed-form solution is faster. This is because it utilizes vector operations instead of for loops.

7 Conclusion

The results from the three statistical tests indicate that geometric Brownian motion (GBM) is a reasonable model for capturing key characteristics of simulated stock price behavior. In Test 1, the Shapiro–Wilk test showed that 95.3% of the simulated log return series did not significantly deviate from normality, supporting the GBM assumption of log-normal returns. In Test 2, the Ljung–Box test confirmed that 94.2% of simulated paths exhibited no significant autocorrelation, aligning with the efficient-market hypothesis (EMH) embedded in the GBM framework. Finally, in Test 3, a one-sample t-test showed no statistically significant difference between the mean simulated volatility and the target volatility (p = 0.8981), reinforcing the accuracy of the model's parameterization.

Moreover, when comparing the closed-form solution of GBM to the Euler–Maruyama numerical method, the root-mean-square error (RMSE) was approximately 0.62. This suggests that while the numerical method introduces some error, it remains a practical alternative when analytical solutions are unavailable. For visualization, if the terminal price of the stock is \$100, the Euler-Maruyama method has an error of \$0.62. This is less than 1%.

Overall, the GBM model—despite its simplicity—demonstrates strong internal consistency and replicates its assumptions well in simulation. These results highlight the usefulness of GBM as a theoretical benchmark in quantitative finance, while also laying the foundation for more complex models that account for market imperfections.

8 Comments

It should be noted that geometric Brownian motion has its limitations and is often not a good predictor for the future price of an asset. It is a fairly rudimentary model. One of GBM's biggest pitfalls is that it oversimplifies market dynamics. It assumes constant drift and volatility, and it assumes a lognormal distribution of returns, which is just simply not the case in the real market. Moreover, it does not consider external factors such as economic news, company-specific events, or global events that can have significant impacts on stock prices.