[STOCHASTIC CALCULUS AND THE GBM]

PRESENTING EUROPEAN OPTIONS AND IMPLEMENTATION OF BS MODEL IN PYTHON AND C++

Felipe Gimenez Souza

Abstract: This text introduces stochastic calculus applied to financial modeling, focusing on the Geometric Brownian Motion (GBM) as a model for stock prices. After reviewing Wiener and Itô processes, we analyze the GBM's distribution, expected value, variance, PDF, and CDF, with graphical support. We explore the distinction between the real-world measure $\mathbb{P}$ and the risk-neutral measure $\mathbb{Q}$, emphasizing martingale properties and their role in arbitrage-free pricing. The theoretical framework culminates in a performance evaluation of the GBM using historical data from AAPL.

Introducing Wiener and Itô Stochastic Processes

Definition 1.1. *A stochastic process is a sequence of random variables $X(t)$ indexed by the time variable $t \in [t_0, t_n = T]$:*

$$\{X(t)\}_{t_0 \leq t \leq T} = \{X(t_1), X(t_2), X(t_3), \dots, X(T)\}.$$

In the context of stochastic processes, the filtration $\{\mathcal{F}(t)\}_{t\geq 0}$ represents the evolution of information over time: $\mathcal{F}(t)$ encodes all events observable up to and including time t. A random variable $X(t)$ is said to be $\mathcal{F}(s)$-measurable if its value is completely determined by the information available at time s. In particular, if $t < s$, then any $\mathcal{F}(s)$-measurable random variable, such as $X(t)$, must already be known at time s, since it refers to a value from the past. On the other hand, $X(s)$ is $\mathcal{F}(s)$-measurable because its value is revealed at time s, but it is not $\mathcal{F}(t)$-measurable for $t < s$, as it represents future uncertainty from the perspective of time t.

This framework naturally leads to the concept of conditional expectation. The conditional expectation $\mathbb{E}[X \mid \mathcal{F}(t)]$ represents the best estimate of a random variable X given the information available up to time t. It is itself an $\mathcal{F}(t)$-measurable random variable, meaning it depends only on what is known at that time. For example, if X is $\mathcal{F}(s)$-measurable with $s > t$, then $\mathbb{E}[X \mid \mathcal{F}(t)]$ captures our forecast of X at time s, using all available information up to time t.

A key concept when dealing with stochastic processes $X(t)$ is the *martingale property*:

Definition 1.2. *Consider a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, where Ω is the set of all possible outcomes, $\mathcal{F}(t)$ is the sigma-algebra and $\mathbb{Q}$ the probability measure. A stochastic process $X(t)$ for $t \in [t_0, t_n = T]$ is said to be a martingale with respect to the filtration $\mathcal{F}(t)$ under measure $\mathbb{Q}$, if for all $t < \infty$ if:*

- **(1)***: The process $X(t)$ is adapted, i.e. $X(t)$ is $\mathcal{F}(t)$-measurable.*
- **(2)***: The absolute expected value of the stochastic process is finite:*

$$\mathbb{E}[|X(t)|] < \infty \tag{1}$$

- **(3)**: *The conditional expectation under probability measure $\mathbb{Q}$ with filtration on time t' of $X(t)$ equals the process in time t':*

$$\mathbb{E}[X(t)|\mathcal{F}(s)] = X(s). \tag{2}$$

Property **(3)** from definition (1.2) can be stated as "the best prediction of the expectation of a martingale's future is its present value". The martingale process express a "fair game" condition to the stochastic process, meaning the expected future value equals the current value, conditional on current information.

The simplest of the stochastic process is the Wiener process (or Brownian Motion) — a process widely used as a building block for SDEs, defined by the following properties:

Definition 1.3. *A Wiener process $W(t)$ can be defined by the following properties:*

- **(1)**: *$W(t_0) = 0$ for t_0 being the process initial time.*
- **(2)**: *$W(t)$ is a almost surely continuous process, but nowhere differentiable* [1].
- **(3)**: *For any $0 \le t_0 < t_1 < t_2 < \cdots < t_n$ with $t_i - t_{i-1} = \Delta t$, the increments*

$$W(t_1) - W(t_0),\ W(t_2) - W(t_1),\ \ldots,\ W(t_n) - W(t_{n-1})$$

are random variables independent between themselves where

$$W(t + \Delta t) - W(t) \sim \mathcal{N}(0, \Delta t),$$

this is, the distribution of the difference of the increments depends only on Δt and not on t.

An useful way of writing a Wiener process[2] is by using a standard normal random variable $z(0,1)$ (mean $= 0$ and variance $= 1$):

$$\Delta W(t) = z(0,1)\sqrt{\Delta t}. \tag{3}$$

This must be valid from property **(3)** from definition (1.3). Suppose we had $\Delta W(t) = \alpha z(0,1)$, for $\alpha \in \mathbb{R}$ constant, then evidently the expected value would be 0 for any $\alpha \in \mathbb{R}$ anyway:

$$\mathbb{E}[\alpha z(0,1)] = \alpha \mathbb{E}[z(0,1)] = 0.$$

If property **(3)** from (1.3) holds, then we must have

$$\mathrm{Var}[\Delta W(t)] = \Delta t,$$

but for that to happen, $\alpha = \sqrt{\Delta t}$, since

$$\mathrm{Var}[\alpha z(0,1)] = \alpha^2 \mathrm{Var}[z(0,1)] = \alpha^2,$$

yielding (3). From this demonstration we conclude that

$$\begin{cases} \mathbb{E}[\Delta W(t)] = 0 \\ \mathrm{Var}[\Delta W(t)] = \Delta t \end{cases} \tag{4}$$

[1] The almost surely convergence is defined by a sequence of random variables X_n fo which $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$

[2] This alternative way of writing the Wiener process is vastly used in numerical approaches of the GBM such as Monte Carlo simulations by sampling pseudo-random standard normal variables $z(0,1)$.

Notice too that under property **(3)** from the Wiener process definition (1.3), choosing $t+\Delta t \to t$ and $t \to t_0$, we have that

$$W(t) - W(t_0) = W(t) - 0 \sim \mathcal{N}(0,t)$$

according to property **(1)**, thus:

$$W(t) \sim \mathcal{N}(0,t), \tag{5}$$

i.e. not only the increments $W(t+\Delta t) - W(t)$ have a normal distribution as in property **(3)** from (1.3), but also the Wiener process itself $W(t)$.

It's easy to demonstrate that a Wiener process $W(t)$ is a martingale. To show that this statement holds, lets prove each property from definition (1.2) for a Wiener process described by definition (1.3).

By construction, $W(t)$ is $\mathcal{F}(t)$-measurable, being thus adapted, concluding that property **(1)** is true. We can calculate the expected value of $|W(t)|$, knowing that it follows a normal distribution of mean 0 and variance t according to (5), using the symmetry of the product of functions inside the integral and a change of variables $x = w^2/2t$ (so that $t\mathrm{d}w = x\mathrm{d}x$):

$$\begin{aligned}
\mathbb{E}[|W(t)|] &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} |w| e^{-\frac{(w-0)^2}{2t}} \mathrm{d}w \\
&= \frac{2}{\sqrt{2\pi t}} \int_{0}^{\infty} w e^{-\frac{(w-0)^2}{2t}} \mathrm{d}w \\
&= \frac{2t}{(2\pi t)^{1/2}} \int_{0}^{\infty} e^{-x} \mathrm{d}x \\
&= \sqrt{\frac{2t}{\pi}} (0 + e^{-0}) \\
&= \sqrt{\frac{2t}{\pi}}.
\end{aligned}$$

thus the expected value on the absolute value of $W(t)$ is finite proving property **(2)** from (1.2):

$$\mathbb{E}[|W(t)|] = \sqrt{\frac{2t}{\pi}} < \infty. \tag{6}$$

For the third martingale property in (1.2), we use property **(3)** from the Wiener process definition: Since the increments $W(s) - W(t)$ are independent of the past, their future increments will be independent of all information available up to time s, which is exactly what $\mathcal{F}(s)$ encodes, hence:

$$\mathbb{E}[W(s) - W(t) | \mathcal{F}(s)] = \mathbb{E}[W(s) - W(t)] = 0 \tag{7}$$

where the last equality is given by the fact that $W(s) - W(t) \sim \mathcal{N}(0, s-t)$ thus having expected value 0.

For $t \in [t_0, T]$ and $s \in [t_0, T]$, notice that $W(t) = W(s) + [W(t) - W(s)]$, using (7) we calculate the conditional expectation of $W(t)$:

$$\begin{aligned}
\mathbb{E}[W(t)|\mathcal{F}(s)] &= \mathbb{E}[W(s) + \{W(t) - W(s)\}|\mathcal{F}(s)] \\
&= \mathbb{E}[W(s)|\mathcal{F}(s)] + \mathbb{E}[W(t) - W(s)|\mathcal{F}(s)] \\
&\overset{(7)}{=} \mathbb{E}[W(s)|\mathcal{F}(s)] + 0 \\
&= W(s).
\end{aligned}$$

Thus we conclude that the Wiener process $W(t)$ is a martingale.

A more general stochastic process is known as *Itô's process* and it is defined as it follows.

Definition 1.4. *The solution $X(t)$ of the following SDE is known as Itô process:*

$$\mathrm{d}X(t) = a(X(t), t)\mathrm{d}t + b(X(t), t)\mathrm{d}W(t), \tag{8}$$

for $X(t_0) = X_0$ (constant), $\mathrm{d}W(t)$ a differential of the Wiener process:

$$\mathrm{d}W(t) = \lim_{\Delta t \to 0} [W(t + \Delta t) - W(t)] \tag{9}$$

and for $a = a(X(t), t)$ and $b = b(X(t), t)$ (usually called drift and diffusion term) being functions that don't increase to rapidly, meaning the functions $a = a(X(t), t)$ and $b = b(X(t), t)$ must satisfy the so called Lipschitz conditions:

$$\begin{cases} |a(x,t) - a(y,t)|^2 + |b(x,t) - b(y,t)|^2 \leq k_1|x-y|^2 \\ |a(x,t)|^2 + |b(x,t)|^2 \leq k_2(1 + |x|^2) \end{cases}$$

for $k_1,\ k_2 \in \mathbb{R}^+$ and $x, y \in \mathbb{R}$.

The drift and diffusion terms receive these names due to their influence in the time evolution of $\mathrm{d}X(t)$. Imagine that for an Itô's process (8) we had the diffusion term $b(X(t), t) = 0$ and a drift term exclusively dependent on time $a = a(t)$ (not as a function of $X(t)$). In this scenario the differential equation becomes purely deterministic:

$$\mathrm{d}X = a\mathrm{d}t.$$

Integrating both sides from t_0 to t ($t_0 \leq t \leq T$) yields a line

$$X(t) = a(t - t_0) + X(t_0)$$

such that a is its slope. This creates some intuition about the term since it relates it to the directional movement of the stochastic process (remembering that the usual process (8) will have the noisy influence of the stochastic term).

The expected value of an Itô's process (8) when $a = a(t)$ exclusively is given by:

$$\mathbb{E}(\mathrm{d}X) = a\mathrm{d}t. \tag{10}$$

which can be quickly derived as follows (here we adopt the Δ notation, which can be made into differential through the limit $\Delta t \to 0$):

$$\mathbb{E}(\Delta X) = a\mathbb{E}(\Delta t) + b\mathbb{E}(\Delta W) \overset{(4)}{=} a\Delta t + b \cdot 0 = a\Delta t.$$

From (10) we can see that the drift term can be regarded as the expected change of the Itô's process over an infinitesimal interval $\mathrm{d}t$. If we have a pure Brownian motion $\mathrm{d}X(t) = b\mathrm{d}W(t)$, the drift will be 0, meaning this process expected value won't change in time, or according to our intuition, there will be no directional movement of the stochastic process as a whole. Summarizing, the drift term is associated with the average behavior of the process over time.

The name drift can be associated to the nomenclature employee in physics for the average velocity of a particle perturbed by the influence of an applied field (namely electrons in a conductor): drift velocity. Among the perturbed motion (analogous to the Brownian motion although sometimes deterministic) there's also a net directional trend due to the field — this net trend is the drift.

The diffusion term $b(X(t),t)$ is evidently associated to the randomness of the process over time as it acts as the coefficient of the Wiener differential $\mathrm{d}W(t)$, or in other words as an amplitude or scale for the stochastic part of the process. If we have a large value of $b(X(t),t)$, then we say the process is *volatile*, meaning it has a higher degree of uncertainty associated with it and when $b(X(t),t) = 0$ it becomes deterministic, i.e. no uncertainty. This element is essential when modeling market prices in order to replicate it's innate randomness.

The variance of an Itô's process (8) when $b = b(t)$ exclusively is given by the following equation

$$\mathrm{Var}(\mathrm{d}X) = b^2\mathrm{d}t, \tag{11}$$

and it can be derived using the expected value (4):

$$\begin{aligned}
\mathrm{Var}(\Delta X) &= \mathrm{Var}(a\Delta t) + \mathrm{Var}(b\Delta W) + 2\mathrm{Cov}(a\Delta t, b\Delta W) \\
&= 0 + \mathrm{Var}(b\Delta W) + 0 \\
&\overset{(4)}{=} b^2\{\mathbb{E}[z^2\Delta t] - \mathbb{E}[z\sqrt{\Delta t}]^2\} \\
&= b^2\Delta t[\mathbb{E}(z^2) - \mathbb{E}(z)^2] \\
&= b^2\Delta t\mathrm{Var}(z) \\
&= b^2\Delta t,
\end{aligned}$$

where we've used the variance of the sum of two random variables (considered here $a\Delta t$ and $b\Delta W(t)$) and that the covariance term vanishes due to $a\Delta t$'s deterministic nature. Equation (11) shows that the squared diffusion term represents the variance of the Itô's process over an infinitesimal interval $\mathrm{d}t$, connecting it directly with the randomness and spread when compared to the average behavior (diffusion) of the stochastic process[3].

Notice that since the drift term is deterministic and $\mathrm{d}W(t) \propto z(0,1) \sim \mathcal{N}(0,1)$ the Itô's process must follow a normal distribution, with the expected value (10) and variance (11) (for non-stochastic drift and diffusion terms):

$$\Delta X \sim \mathcal{N}(a\Delta t, b^2\Delta t) \tag{12}$$

Given a function $f(X(t),t)$ of on an Itô's process (8) we can use a powerful tool called *Itô's lemma* in order to derive the function dynamics, i.e. $\mathrm{d}f(X(t),t)$. It works analogously to a Taylor expansion when dealing with randomic processes (when applying the $\Delta t \to 0$ limit and ignoring higher order terms, namely higher than first order). The demonstration of Itô's lemma will be displayed in the Appendice section.

Theorem 1.1. *Suppose a Itô process $X(t)$ (8), let $f = f(X(t),t)$ be a function of $X(t)$ (as independent variable) and t with continuous first and second partial derivatives in $X(t)$ and first derivative in t, then the stochastic process $f(X(t),t)$ follows a a dynamics:*

$$\mathrm{d}f(X(t),t) = \left[\frac{\partial f}{\partial t} + a(X,t)\frac{\partial f}{\partial X} + \frac{1}{2}b(X,t)\frac{\partial^2 f}{\partial X^2}\right]\mathrm{d}t + b(X,t)\frac{\partial f}{\partial X}\mathrm{d}W(t). \tag{13}$$

[3]Here we relate the variance with the uncertainty of the process given its definition: $\mathrm{Var}(X) = \mathbb{E}[(X-\mu)^2]$ for a random variable X with a distribution of mean μ. This expression captures the average squared deviation of X from its mean, quantifying how much the values of X typically fluctuate around μ

Notice that the lemma's equation is itself an Itô's process with

$$\text{drift} \rightarrow \frac{\partial f}{\partial t} + a(X,t)\frac{\partial f}{\partial X} + \frac{1}{2}b(X,t)\frac{\partial^2 f}{\partial X^2}, \quad \text{diffusion} \rightarrow b(X,t)\frac{\partial f}{\partial X}.$$

Introducing the GBM model

Definition 2.1. *A stochastic process* $\{S(t)\}_{t_0 \le t \le T}$ *is said to follow a Geometric Brownian Motion (GBM) if it is a solution for the SDE (that is a particular case from Itô's process (8)):*

$$\mathrm{d}S(t) = \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{P}}(t) \tag{14}$$

where $\mathrm{d}W(t)$ *is a differential Wiener process under* $\mathbb{P}$ *measure and we have* $\mu S(t)$ *and* $\sigma S(t)$ *as the drift and diffusion terms respectively, for* $\mu, \sigma \in \mathbb{R}$ *constants. This process is commonly used to model stock prices* $S(t)$ *from underlying assets in derivative contracts such as no-dividend paying stock options.*

Notice the GBM definition is such that drift and diffusion terms depends on the stochastic stock price $S(t)$, not allowing us to use the expected value (10) and variance (11) and consequently not the normal distribution of $\mathrm{d}S(t)$ as in (12). Considering equation (14) in the following format

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu\mathrm{d}t + \sigma\mathrm{d}W^{\mathbb{P}}(t) \tag{15}$$

we can see that $\mathrm{d}S(t)/S(t)$ must represent the percentage return from the stock, when the stock price changes from $S(t)$ to $S(t) + \mathrm{d}S(t)$ in a time interval t to $t + \mathrm{d}t$. Using equation (14) under the format (15), lets consider the dynamics of the percentage return on the stock $\mathrm{d}S(t)/S(t)$. Evidently the drift and diffusion terms become constant, enabling us to apply (12):

$$\frac{\mathrm{d}S(t)}{S(t)} \sim \mathcal{N}(\mu\mathrm{d}t, \sigma^2\mathrm{d}t). \tag{16}$$

Concluding that although we cannot say that $\mathrm{d}S(t)$ follows a normal distribution, the percentage return on the stock $\mathrm{d}S(t)/S(t)$ does follow one according to (16).

Suppose $\sigma = 0$ in (15), in this scenario we could write (14) as a deterministic equation:

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu\mathrm{d}t.$$

Integrating from t_0 to t ($t_0 \le t \le T$) yields:

$$S(t) = S(t_0)e^{\mu(t-t_0)}, \tag{17}$$

hence, when there's no uncertainty (diffusion term), the evolution of the stock price is going to be the future value under a continuously compounded rate μ. Here we can see that the coefficient μ is related to the log-return from the stock $S(t)$ [4]. From (16) equation's perspective, we have that the coefficient μ can be regarded as the average behavior of the return of the stock $\mathbb{E}[\mathrm{d}S(t)/S(t)] = \mu\mathrm{d}t$ over a small amount of time $\mathrm{d}t$ (from (10)) or,

[4] Equation (17) shows that $\mu = [\ln S(t) - \ln S(t_0)]/(t - t_0)$ or the average rate of change of the log-price per unit time - also known as the log-return on the stock. Which is essentially different from the percentage return on the stock price $\mathrm{d}S(t)/S(t)$.

as we expressed when describing the drift term, to the trend movement of this stochastic process through time. From these properties, the coefficient μ represents under the GBM model for stocks the *expected return* (annualized) earned by an investor on the stock in a short period o time $\mathrm{d}t$.

Given μ being the representation of the expected return from the stock of price $S(t)$, we can enumerate its dependencies. If an investment has higher risk (a large change of failure), the investor will usually require an evenly higher return from the stock, thus μ's value should be regulated by the risk of return from the stock. Another influence in μ is the interest rate (IR) from the economy where this investment is inserted in. A higher level of IR implies a higher expected return from any stock required by the investors. Luckily as we will see, the price of an vanilla option contract won't depend on μ.

In practice there will be uncertainty when dealing with stock prices, this statement aligns with the *efficient market hypothesis* which suggests that asset prices changes are essentially random. Under the GBM model this role is portrayed by the diffusion term σ and the Wiener process $\mathrm{d}W(t)$ with σ serving as an amplitude for the random fluctuations. When looking under equation (15) framework, the coefficient σ can be associated with the variance of the stock price return: $\mathrm{Var}[\mathrm{d}S(t)/S(t)] = \sigma^2 \mathrm{d}t$ over a small period of time $\mathrm{d}t$ (from (11)) implying that it has an influence over how abruptly the price of a particular stock deviates from the expected value in $\mathrm{d}t$. We conclude that parameter σ measures how volatile a stock price will be behave in a short period of time, naming it the *volatility* of the stock.

Subsequently, we apply the Itô's lemma to a function $f(S(t), t) = \ln S(t)$, taking the result into account:

$$\begin{aligned}
\mathrm{d}\left[\ln S(t)\right] &= \left[\frac{\partial}{\partial t}(\ln S) + \mu S \frac{\partial}{\partial S}(\ln S) + \frac{1}{2}\sigma^2 S \frac{\partial^2}{\partial S^2}(\ln S)\right]\mathrm{d}t + \sigma S \frac{\partial}{\partial S}(\ln S)\mathrm{d}W^{\mathbb{P}}(t) \\
&= \left[0 + \mu S\left(\frac{1}{S}\right) + \frac{1}{2}\sigma^2 S^2\left(-\frac{1}{S^2}\right)\right]\mathrm{d}t + \left(\frac{1}{S}\right)\sigma S \mathrm{d}W^{\mathbb{P}}(t)
\end{aligned}$$

$$\mathrm{d}\left[\ln S(t)\right] = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma \mathrm{d}W^{\mathbb{P}}(t). \tag{18}$$

Before we had the distribution of the percentage return (15) instead of $\mathrm{d}S(t)$, but here, since we have deterministic drift $(\mu - \sigma^2/2)$ and diffusion σ terms, we can apply (12) yielding

$$\mathrm{d}\left[\ln S(t)\right] \sim \mathcal{N}\left[\left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t, \sigma^2 \mathrm{d}t\right]. \tag{19}$$

When the logarithm of a random variable has a normal distribution $\ln X \sim \mathcal{N}(\cdot,\cdot)$, it's said that it follows a lognormal distribution. Hence we state that the GBM dynamics $\mathrm{d}S(t)$ is lognormally distributed according to (19), with mean $(\mu - \sigma^2/2)\mathrm{d}t$ and variance $\sigma^2\mathrm{d}t$.

In the continuation, we solve the SDE (19), finding the value of $S(t)$ according to the

GBM model (14). Integrating (19) from t_0 to t ($t_0 \leq t \leq T$) produces the following result[5]:

$$\int_{t_0}^{t} \mathrm{d}\,[\ln S(t')] = \int_{t_0}^{t} \left(\mu - \frac{\sigma^2}{2}\right) \mathrm{d}t + \sigma \int_{t_0}^{t} \mathrm{d}W^{\mathbb{P}}(t')$$
$$\ln S(t) - \ln S(t_0) = \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma(W^{\mathbb{P}}(t) - W^{\mathbb{P}}(t_0))$$

$$\ln S(t) = \ln S(t_0) + \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma(W^{\mathbb{P}}(t) - W^{\mathbb{P}}(t_0)), \tag{20}$$

which clearly yields, once more, a lognormal distribution for $S(t)$[6]:

$$\ln S(t) \sim \mathcal{N}\left[\ln S(t_0) + \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0),\ \sigma^2(t - t_0)\right]. \tag{21}$$

Equivalently to (20) can be written as:

$$S(t) = S(t_0)\exp\left[\left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) + \sigma(W^{\mathbb{P}}(t) - W^{\mathbb{P}}(t_0))\right], \tag{22}$$

where, for $\sigma = 0$, we retrieve the deterministic result (17). The expected value of $S(t)$ (22) is given by

$$\mathbb{E}[S(t)] = S(t_0)e^{\mu(t-t_0)}. \tag{23}$$

Here we notice that the average behavior of the stock price over time is an analogous result for the deterministic version of the $S(t)$ process proposed in equation (17), with $\sigma = 0$. Concluding that $\mathbb{E}[S(t)]$ can be regarded as the FV of $S(t_0)$ with rate μ from t_0 to t, not being dependent on the volatility of the process. The variance of $S(t)$ is given by the following equation:

$$\mathrm{Var}[S(t)] = S(t_0)^2 e^{2\mu(t-t_0)}(e^{\sigma^2(t-t_0)} - 1), \tag{24}$$

where for $\sigma = 0$ or no uncertainty, we evidently get a null result as it would be expected. In summary, $S(t)$ follows a lognormal distribution of the form[7]:

$$S(t) \sim \mathrm{LN}\left[S(t_0)e^{\mu(t-t_0)},\ S(t_0)^2 e^{2\mu(t-t_0)}(e^{\sigma^2(t-t_0)} - 1)\right]. \tag{25}$$

PDF and CDF for the GBM model

Through our progress of the understanding of the GBM model for the stock price $S(t)$, we figured out that $\ln S(t)$ follows a normal distribution according to (21). Lets then use this distribution to determine the CDF (cumulative density function) and PDF

[5]Notice that the following equation is valid for an Itô's stochastic process such as $\{X(t)\}_{t_0 \leq t \leq T}$ according to Itô calculus:

$$\int_{t_0}^{t} \mathrm{d}X(t') = X(t) - X(t_0)$$

[6]Notice that this could be written differently using $W^{\mathbb{P}}(t_0) = 0$ according to Wiener's process definition (1.3) property **(1)**.

[7]Here the notation used is such that for a lognormally distributed random variable X of mean $\bar{\mu}$ and variance $\bar{\sigma}^2$, we can say that $X \sim \mathrm{LN}(\bar{\mu}, \bar{\sigma}^2)$.

(probability density function) for $\ln S(t)$. For $X(t) \sim \mathcal{N}(\bar{\mu}, \bar{\sigma})$ a random process under physical probability measure $\mathbb{P}$, the CDF is given by

$$F_{X(t)}(x) = \mathbb{P}[X(t) \le x] = \frac{1}{\sqrt{2\pi\bar{\sigma}}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{x' - \bar{\mu}}{\bar{\sigma}}\right)^2\right] \mathrm{d}x'. \tag{26}$$

with PDF

$$f_{X(t)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{X(t)}(x) = \frac{1}{\sqrt{2\pi\bar{\sigma}}} \exp\left[-\frac{1}{2}\left(\frac{x - \bar{\mu}}{\bar{\sigma}}\right)^2\right]. \tag{27}$$

for $\ln S(t)$'s normal distribution (21), we'll have

$$X(t) = \ln S(t),\ \bar{\mu} = \ln S(t_0) + \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0) \text{ and } \bar{\sigma} = \sigma(t - t_0)$$

yielding the CDF:

$$F_{\ln S(t)}(x) = \mathbb{P}[\ln S(t) \le x] = \frac{1}{\sigma\sqrt{2\pi(t - t_0)}} \int_{-\infty}^{x} \exp\left\{-\frac{\left[x' - \ln S(t_0) - \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0)\right]^2}{2\sigma^2(t - t_0)}\right\} \mathrm{d}x' \tag{28}$$

with PDF

$$f_{\ln S(t)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{\ln S(t)}(x) = \frac{1}{\sigma\sqrt{2\pi(t - t_0)}} \exp\left\{-\frac{\left[x - \ln S(t_0) - \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0)\right]^2}{2\sigma^2(t - t_0)}\right\}. \tag{29}$$

If instead of $\ln S(t)$ we want to express the CDF and PDF of $S(t)$ itself, first we consider $S(t) = e^{\ln S(t)} = e^{X(t)}$ keeping in mind that

$$\begin{aligned} S(t) &\le x \\ e^{X(t)} &\le x \\ X(t) &\le \ln x \end{aligned}$$

or in other words

$$F_{S(t)}(x) = \mathbb{P}[S(t) \le x] = \mathbb{P}[X(t) \le \ln x] = F_{X(t)}(\ln x).$$

Which is exactly (28) for $x \to \ln x$:

$$F_{S(t)}(x) = \frac{1}{\sigma\sqrt{2\pi(t - t_0)}} \int_{-\infty}^{\ln x} \frac{1}{x'} \exp\left\{-\frac{\left[\ln\left(\frac{x'}{S(t_0)}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0)\right]^2}{2\sigma^2(t - t_0)}\right\} \mathrm{d}x' \tag{30}$$

noticing that $\mathrm{d}(\ln x') = \mathrm{d}x'/x'$. The PDF could then be derived analogously to (29):

$$f_{S(t)}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{S(t)}(x) = \frac{1}{\sigma x\sqrt{2\pi(t - t_0)}} \exp\left\{-\frac{\left[\ln\left(\frac{x}{S(t_0)}\right) - \left(\mu - \frac{\sigma^2}{2}\right)(t - t_0)\right]^2}{2\sigma^2(t - t_0)}\right\}. \tag{31}$$

In figure (1) we have in the xy-plane several GBM paths following accordingly to (20). For different fixed dates (from $t = 0$ to $t = 1$ year with timestep 0.25 or a quarter), the PDF from $\ln S(t)$ (29) is shown.

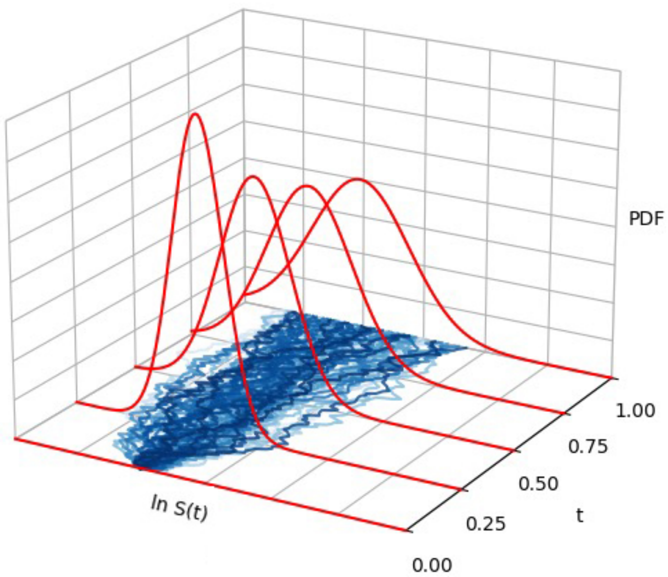

Figure 1: GBM paths from $\ln S(t)$ (xy-plane) and their probability density on the z-axis for fixed dates. The figure utilizes $S(t_0) = e^4$, $\mu = 0.05$, $\sigma = 0.4$, $t_0 = 0$, $T = 1$, number of paths = 100 and $\mathrm{d}t = 1/100$.

As we can see in figure (1), as uncertainty raises as time passes by, which can be noticed from the tendency of divergence of the paths in later times, the variance (width) of the PDF increases. This widening happens due to the linear time dependence of the variance of the distribution (21): $\sigma(t - t_0)$. This conclusion implies is that in later horizons, there's a broader set of possible values of $\ln S(t)$ and consequently $S(t)$, becoming less possible to address the value of $S(t)$ as time progresses.

The PDF (29) gets flatter as time passes by in figure (1), or in other words, the probability of x being a specific value $\ln S(t)$ in (29) becomes similar for a wider range of values of x in the real line. Given that $\ln S(t)$'s variance also depends on the volatility σ, this parameters will also have a heavy influence in how fast the spread of the paths will happen.

Although hard to notice from figure (1) the normal distributions (29) mean tends to shift towards smaller values in the $\ln S(t)$-axis as time progresses (not reaching negative values since $\ln S(t) > 0$ always). Given that $\ln S(t)$ mean, showed in (21), is linear dependent on time with an argument $(\mu - \sigma^2/2)$, in the scenario from figure (1), the mean of $\ln S(t)$ follows a decrescent line $\ln S(t_0) + (\mu - \sigma^2/2)(t - t_0)$ since $(\mu - \sigma^2/2) < 0$. This happens since we chose $\mu = 0.05$ and $\sigma = 0.4$ yielding a negative drift of -0.03 and a mean of $\ln\left(e^4\right) - 0.03t = 4 - 0.03t \approx 3.97$.

Keeping $\mu = 0.05$ and setting a smaller volatility such as $\sigma = 0.2$, would give us a positive drift value of $+0.03$ and a crescent line $\ln S(t)$'s mean: $4 + 0.03t$. Since the drift $-0.03t$ is a line with a small angular parameter, the change in the average direction of the paths is almost unnoticed in figure (1) as pointed out before.

Lets now drawn conclusions regarding $S(t)$'s behavior in a similar manner as we did for figure (1), recalling that it follows a lognormal distribution (25) with PDF (31), thus having a skewed shape when compared to the normal distribution of $\ln S(t)$ as we can see from figure (2).

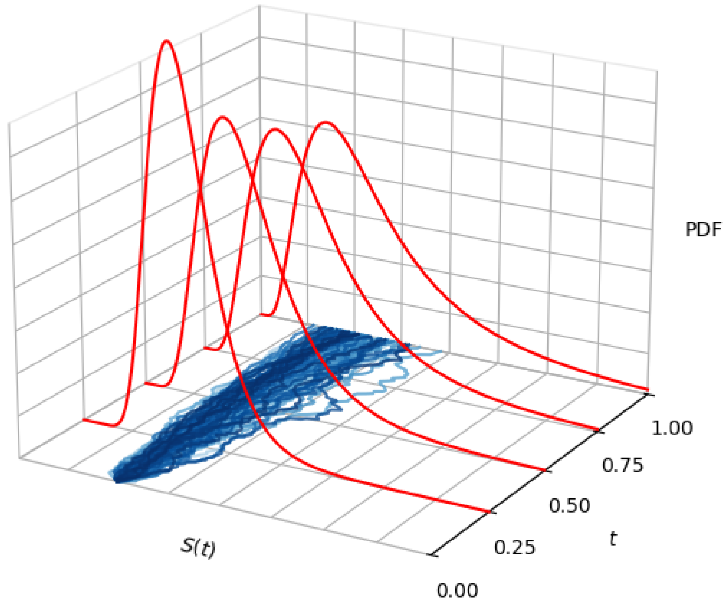

Figure 2: GBM paths from $S(t)$ (xy-plane) and their probability density on the z-axis for fixed dates. The figure utilizes $S(t_0) = 100$, $\mu = 0.08$, $\sigma = 0.25$, $t_0 = 0$, $T = 1$, number of paths $= 100$ and $\mathrm{d}t = 1/100$.

Observing figure (2), we can see that the red lognormal distributions (31) allow a wider range of values of $S(t)$ than $\ln S(t)$'s normal distributions in figure (1), shifting some of the paths to be under its long tail. This is caused by a positive skewness of the PDFs, implying a higher probability of movements towards larger values of $S(t)$ than for the smaller ones. This can be illustrated by some of the paths right below the tail of the curve in the figure.

The range of values expressed by the distributions width increases with time t, drawing the same conclusion we had for $\ln S(t)$'s distributions: the uncertainty or the set of possible values of $S(t)$ raises as time progresses due to a time dependence in the variance of $S(t)$. Although this time, variance of $S(t)$ doesn't change in a linear way, instead varying exponentially according to (24) as we can see from figure (3). Since (24) depends on the square of the volatility's value in the exponential argument, it will also have a high influence in the spreading of the paths.

As we can see from figure (3), for usually small values of μ, the expected value of $S(t)$ (23) tends to become closer to a constant line, even though it is an exponential. Meaning μ won't have that much influence in the behavior of the paths and their PDFs represented in figure (3). The consequence of a larger value of μ would be to create a small tendency in the paths by on average shifting them towards larger values of $S(t)$ and attenuating the peaks from the lognormal PDFs, making the probabilities of different possible values of $S(t)$ more and more homogeneous with time.

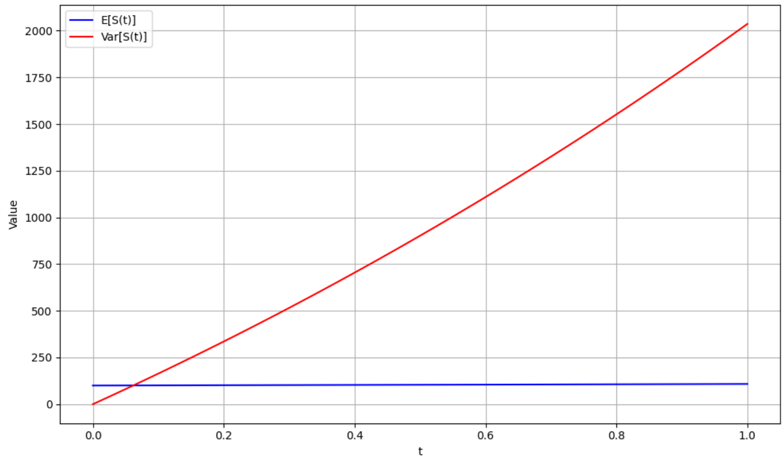

Figure 3: Graphic of the mean and variance of $S(t)$ (lognormally distributed) according to (23) and (24). The figure utilizes $S(t_0) = 100$, $\mu = 0.08$, $\sigma = 0.40$, $t_0 = 0$, $T = 1$, number of paths = 100.

Martingales and the Measure $\mathbb{Q}$

We may ask ourselves if the stochastic process $S(t)$ is a martingale under definition (1.2). The short answer is no, stock prices usually are not martingales, meaning there are at some level a degree of predictability of the future states of $S(t)$. Up to this point we've being using exclusively the Wiener process $\mathrm{d}W^{\mathbb{P}}(t)$ under probability measure $\mathbb{P}$ when describing $S(t)$, which is usually called *physical probability measure* or *real-world probability measure*. It reflects actual probabilities of events based on observed frequencies, considering investor's risk preferences in the expected return parameter μ.

In order to see that under the physical probability measure $\mathbb{P}$, the GBM process of $S(t)$ (22) isn't a martingale, we utilize the definition (1.2). Property **(1)** from the martingale definition (1.2) must hold since it implies that the expected value of the process must exist, which is something crucial for the usage of expected values in derivatives payoffs such as in option contracts. Recalling equation (23), we figure out that the conditional expectation of $S(t)$ under sigma-algebra $\mathcal{F}(t_0)$ is given by

$$\mathbb{E}[S(t+\Delta t)|\mathcal{F}(t)] = S(t_0)e^{\mu\Delta t} \tag{32}$$

in the same way as (23). This is because $S(t_0)$ is already $\mathcal{F}(t_0)$-measurable and $W^{\mathbb{P}}(t) - W^{\mathbb{P}}(t_0)$ is independent (by the same arguments explained when proving that $W(t)$ is a martingale), making the conditional expectation into a ordinary expected value. Equation (32) shows that

$$\begin{cases} \mathbb{E}[S(t+\Delta t)|\mathcal{F}(t)] > S(t), & \text{if } \mu > 0 \\ \mathbb{E}[S(t+\Delta t)|\mathcal{F}(t)] < S(t), & \text{if } \mu < 0, \end{cases} \tag{33}$$

implying that property **(3)** doesn't hold for $S(t)$ under physical probability measure $\mathbb{P}$ and thus it is not a martingale. In (33) we have a super-martingale for positive μ and a sub-martingale for negative μ reflecting the investor preferences innate to the expected return and the risk premium associated to it.

Consider r being the risk-free interest rate. This would be the return an investor would get in a completely riskless asset such as government bond (like U.S. Treasuries) maturing in short time with no default risk. In contrast with the expected return μ, which takes investors risk preferences into account, in summary:

$$\mu = r + \rho, \tag{34}$$

where $\rho > 0$ stands for the risk premium or the compensation for a riskier asset (higher return), such as stocks. Notice that for the investment to be a rational decision for the investor we must have $\mu > r$ (which can be stated from (34) since $\rho > 0$).

With the goal of establishing a probability measure framework capable of describing the process $S(t)$ as a martingale, we define the *risk-neutral* measure $\mathbb{Q}$, under which we have the same Geometric Brownian Motion SDE (14) with the expected return μ replaced by the risk-free rate r:

$$\mathrm{d}S(t) = rS(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{Q}}(t), \tag{35}$$

from which all our previous results can be derived changing μ to r. To show that (14) becomes (35) by changing measure $\mathbb{P}$ to $\mathbb{Q}$ we introduce Girsanov theorem.

Theorem 4.1. *Roughly speaking*[8]*, the Girsanov theorem formulates how stochastic processes $X(t)$ change under changes in its measure. In particular, for a Wiener process $W(t)$, the change in measure reflects in the addition of a drift term of coefficient θ:*

$$\tilde{W}(t) = W(t) + \theta \mathrm{d}t, \tag{36}$$

where $\tilde{W}(t)$ represents the process under the new probability measure.

Choosing the Girsanov theorem (4.1) parameter θ to be such that

$$\mathrm{d}W^{\mathbb{Q}}(t) = \mathrm{d}W^{\mathbb{P}}(t) + \theta\mathrm{d}t = \mathrm{d}W^{\mathbb{P}}(t) + \left(\frac{\mu - r}{\sigma}\right)\mathrm{d}t \tag{37}$$

allow us to see how the change from real-world measure $\mathbb{P}$ to risk-neutral measure $\mathbb{Q}$ implies the Geometric Brownian Motion SDE (14) to become (35):

$$\begin{aligned}
\mathrm{d}S(t) &= \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{P}}(t) \\
&= \mu S(t)\mathrm{d}t + \sigma S(t)[\mathrm{d}W^{\mathbb{Q}}(t) - \theta\mathrm{d}t] \\
&\overset{(37)}{=} \mu S(t)\mathrm{d}t + \sigma S(t)\left[\mathrm{d}W^{\mathbb{Q}}(t) - \frac{\mu - r}{\sigma}\mathrm{d}t\right] \\
&= \mu S(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{Q}}(t) - S(t)(\mu - r)\mathrm{d}t \\
&= rS(t)\mathrm{d}t + \sigma S(t)\mathrm{d}W^{\mathbb{Q}}(t).
\end{aligned}$$

Using the new probability measure $\mathbb{Q}$ we may attempt to check if $S(t)$ can now be stated as a martingale process. Since practically the change from measure $\mathbb{P}$ to $\mathbb{Q}$ only changes the drift term making $\mu \to r$, the third property from martingale definition (1.2) will still not be satisfied:

$$\mathbb{E}[S(t + \Delta t)|\mathcal{F}(t)] = S(t_0)e^{r\Delta t}, \tag{38}$$

and thus

$$\begin{cases} \mathbb{E}[S(t + \Delta t)|\mathcal{F}(t)] > S(t), & \text{if } r > 0 \\ \mathbb{E}[S(t + \Delta t)|\mathcal{F}(t)] < S(t), & \text{if } r < 0, \end{cases} \tag{39}$$

[8]The formal Girsanov theorem can be seen in reference (!!!).

analogously to (33).

Observe though that if we instead use a discounted by the risk-free rate version of the stock price $e^{-r\Delta t}S(t)$, we satisfy property **(3)** from (1.2), achieving a martingale (properties **(1)** and **(2)** where already satisfied under probability measure $\mathbb{P}$ as we've seen):

$$\mathbb{E}[e^{-r\Delta t}S(t+\Delta t)|\mathcal{F}(t)] = e^{-r\Delta t}[S(t_0)e^{r\Delta t}] = S(t_0). \tag{40}$$

Notice that the same wouldn't hold for the process under physical probability measure $\mathbb{P}$, since the exponential factors wouldn't cancel out.

When pricing derivatives, one of the main goals is to avoid arbitrage opportunities, i.e. guarantee risk-free profits without taking on any risk and without committing any capital, these opportunities undermine the market. In an efficient market, prices fully reflect all available information, meanwhile arbitrage opportunities suggest that assets are mispriced relative to each other.

For a martingale such as the discounted stock price $\tilde{S}(t) = e^{-r\Delta t}S(t)$, arbitrage is prevented. If $S(t)$ were expected to rise, one could buy the stock and expect a profit without risk, if the opposite was expected to happen, one could short sell the asset and expect a profit, either case there would be an arbitrage opportunity, since the profit is riskless and with no initial investment. By enforcing (40) we are saying "Given what I known now, I can't expect the discounted price to go up or down in the future", which removes the possibility of arbitrage via buying, selling or holding the discounted stock price $\tilde{S}(t)$ without taking risk.

The connection between martingales and no-arbitrage opportunities such as the GBM modeled discounted stock price under risk-neutral measure $\mathbb{Q}$ can be extended through the Fundamental Theorem of Asset Pricing (FTAP):

Theorem 4.2. *The Fundamental Theorem of Asset Pricing (FTAP) states that a financial market is arbitrage-free if and only if there exists an equivalent martingale measure $\mathbb{Q}$ under which the discounted price processes $e^{-rt}S(t)$ (adjusted by the risk-free rate r) of all tradable assets are martingales. Where equivalency stands for measures $\mathbb{P}$ and $\mathbb{Q}$ agreeing on what events are possible (both assign 0 probability to the the same sets).*

The FTAP theorem (4.2) then concludes that if discounted prices are martingales under some measure $\mathbb{Q}$, then there's no way to guarantee a profitable operation without involving risk and therefore the market is arbitrage-free.

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 as a random component, considering the unpredictability of the market (associated with the volatility of the stock σ), 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 for each day timestep considered $\ln[S(t+1)/S(t)]$[9],

[9]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).

which, according to the GBM (20), should follow:

$$\ln\left[\frac{S(t+1)}{S(t)}\right] = \left(\mu - \frac{\sigma^2}{2}\right)\Delta t + \sigma[\Delta W^{\mathbb{P}}(t+1) - \Delta W^{\mathbb{P}}(t)], \tag{41}$$

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 (21):

$$\ln\left[\frac{S(t+1)}{S(t)}\right] \sim \mathcal{N}\left[\left(\mu - \frac{\sigma^2}{2}\right)\Delta t, \sigma^2\Delta t\right]. \tag{42}$$

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:

$$\sigma_{\text{annual}} = \sigma_{\text{daily}}\sqrt{252}, \tag{43}$$

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* of AAPL by computing the standard deviation of the log return data[10]:

$$\sigma_{\text{daily}} = \text{Std}(R_t) = \sqrt{\frac{1}{N-1}\sum_{t=1}^{N-1}(R_t - \bar{R})^2}, \quad R_t = \ln\left(\frac{S_{t+1}}{S_t}\right), \tag{44}$$

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

$$\bar{R} = \frac{1}{N}\sum_{t=1}^{N} R_t. \tag{45}$$

The estimative can then be multiplied by $\sqrt{252}$ to yield the annual volatility according to (43):

$$\sigma_{\text{daily}} \approx 0.018294 \quad \Rightarrow \quad \sigma_{\text{annual}} = 0.018294\sqrt{252} \approx 0.290416, \tag{46}$$

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

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. 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.

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 (4). 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.

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

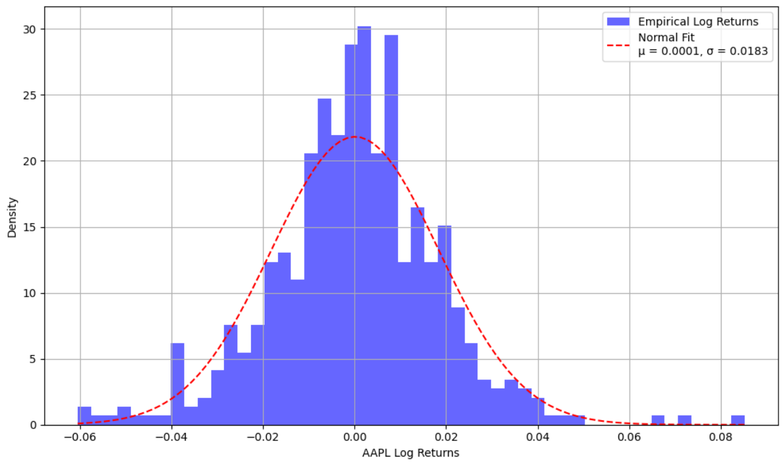

Figure 4: Plot of histogram and a fitted normal distribution both from AAPL historical price log-returns from 2022 to 2024 (501 empirical prices). The x-axis shows the lognormal return values $\ln[S(t+1)/S(t)]$ meanwhile the y-axis represents the density values of the histogram (height of the bars).

The height of the histogram bars in figure (4) 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

$$\rho = \frac{f/N}{w}. \tag{47}$$

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.

Analyzing figure (4), 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. 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 (4) 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).

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 (4) 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 founded and volatility estimation are the following:

$$\hat{\sigma}_{\text{daily}} \approx 0.018276 \quad \Rightarrow \quad \hat{\sigma}_{\text{daily}} = 0.018276\sqrt{252} \approx 0.290126. \tag{48}$$

Hence the historical volatility, purely empirical, calculated in (46) reaches pretty close to the normal PDF fit of the log-return data in (48), 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 (4), 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[11]. The model presuppose that drift/expected return μ and volatility σ are constant through time which also doesn't reflect reality. Volatility can better modeled as stochastic, having sometimes properties such as mean-reversion (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: 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. In figure (4), 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 (4).

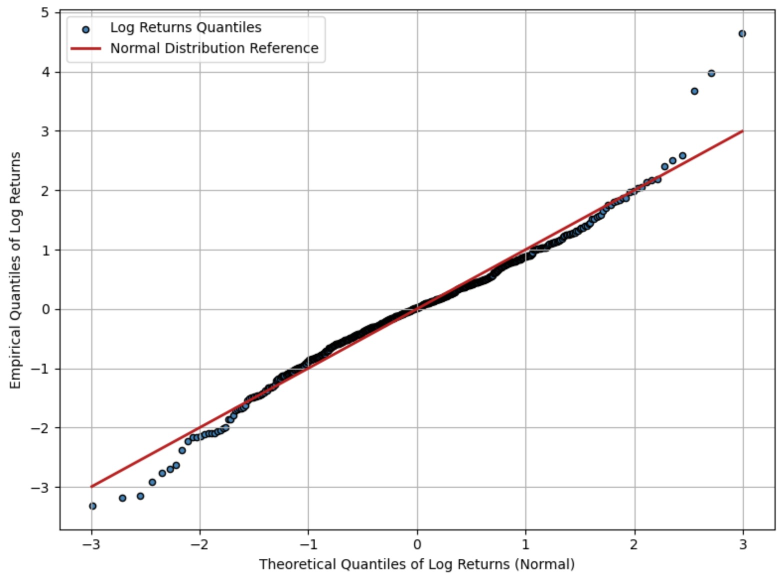

Figure 5: q-q plot from AAPL historical price log returns from 2022 to 2024. The x-axis shows the theoretical quantiles of the log returns (standard normal distribution values) meanwhile the y-axis represents empirical AAPL $\ln[S(t+1)/S(t)]$ quantile values (standardized). The 45° red line represents the standard normal distribution reference.

To further examine this underestimation of tail risk, figure (5) 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.

[11] 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.

The Q–Q plot in figure (5) 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. 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*, 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). 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.

Implementation in Python

Appendices

Derivation of Itô's Lemma

The Taylor series expansion for a function $f = f(X(t), t)$ (of two variables) around (X_0, t_0) (with $X(t_0) = X_0$) with $\Delta X = X - X_0$ and $\Delta t = t - t_0$ is the given by

$$\begin{aligned} f(X,t) = f(X_0,t_0) + \left.\frac{\partial f}{\partial t}\right|_{t=t_0} \Delta t + \left.\frac{\partial f}{\partial X}\right|_{X=X_0} \Delta X + \\ + \frac{1}{2}\left.\frac{\partial^2 f}{\partial t^2}\right|_{t=t_0} (\Delta t)^2 + \frac{1}{2}\left.\frac{\partial^2 f}{\partial X^2}\right|_{X=X_0} (\Delta X)^2 \\ + \left.\frac{\partial^2 f}{\partial X \partial t}\right|_{(X,t)=(X_0,t_0)} \Delta X \Delta t + \cdots , \end{aligned}$$

where we take the limits $\lim_{t\to t_0} \Delta t = \mathrm{d}t$ and $\lim_{X\to X_0} \Delta X = \mathrm{d}X$. Neglecting higher order terms of $\mathrm{d}t$ (this is, up to $\mathrm{d}t$ and higher order terms), we arrive at

$$\mathrm{d}f(X(t),t) = \frac{\partial f}{\partial t}\mathrm{d}t + \frac{\partial f}{\partial X}\mathrm{d}X + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}(\mathrm{d}X)^2. \tag{49}$$

Notice that a differential Wiener process can be written as in (), thus $\mathrm{d}X(t)\mathrm{d}t$ according to (8) depends on $\mathrm{d}t^2$ (for the drift term) and $\mathrm{d}t^{3/2}$ (for the diffusion term), both higher than dt's order. The term $(\mathrm{d}X)^2$ must be developed carefully:

$$\begin{aligned}(\mathrm{d}X)^2 &\stackrel{(8)}{=} (a\mathrm{d}t + b\mathrm{d}W)(a\mathrm{d}t + b\mathrm{d}W)\\ &= a^2(\mathrm{d}t)^2 + 2ab\mathrm{d}t\mathrm{d}W + b^2(\mathrm{d}W)^2\\ &\stackrel{(3)}{=} a^2(\mathrm{d}t)^2 + 2abz(\mathrm{d}t)^{3/2} + b^2z^2\mathrm{d}t\\ &\approx b^2z^2\mathrm{d}t\end{aligned}$$

where we neglected the second and 3/2-order dt terms. We now study the nature of the $z^2\mathrm{d}t$ term by calculating its expected value through

$$\mathbb{E}(z^n) = \begin{cases}\frac{(n)!}{2^{n/2}(n/2)!}, & \text{if n is even,}\\ 0, & \text{if n is odd}\end{cases} \tag{50}$$

for $n \in \mathbb{Z}_{\geq 0}$ and $z \sim \mathcal{N}(0,1)$ a standard normal random variable [12]:

- **Expected Value**:

$$\mathbb{E}[(\mathrm{d}W)^2] = \mathbb{E}[z^2\mathrm{d}t] = \mathrm{d}t\ \mathbb{E}[z^2] \stackrel{(50)}{=} \frac{2!}{2^1 \cdot 1!}\mathrm{d}t = \mathrm{d}t$$

- **Variance**:

$$\begin{aligned}\mathrm{Var}[(\mathrm{d}W)^2] &= \mathbb{E}[(\mathrm{d}W)^4] - \mathbb{E}[(\mathrm{d}W)^2]^2\\ &= (\mathrm{d}t)^2\mathbb{E}(z^4) - (\mathrm{d}t)^2\mathbb{E}(z^2)^2\\ &\stackrel{(50)}{=} (\mathrm{d}t)^2\left[\frac{4!}{2^2 2!} - 1^2\right]\\ &= (\mathrm{d}t)^2(3-1)\\ &= 2(\mathrm{d}t)^2\\ &\approx 0.\end{aligned}$$

This implies that the variance of $(\mathrm{d}W)^2$ converges to 0 quadratically while the expected value converges to 0 linearly in dt when d$t \to 0$, vanishing the variance according to our first order approximation in dt. Since the variance is 0 for d$t \to 0$, then $(\mathrm{d}W)^2$ is deterministic and equal to its expected value:

$$(\mathrm{d}W)^2 \approx \mathrm{d}t, \text{ when } \mathrm{d}t \to 0. \tag{51}$$

This statement can be intuitively derived graphically. For the purpose of the visualization and practicability, lets consider a normal distributed variable $X \sim \mathcal{N}(\mu, \sigma^2)$:

In the limit where $\sigma \to 0$, we have that the curve becomes more centered sharped, or equivalently there's a higher change of sampling a value close to μ. As we take the limit $\sigma \to 0$ further, the distribution becomes a spike around μ called Dirac delta function

[12] This derivation can be made by using the moment generating function of the standard normal distribution $M_X(\tau) = e^{\tau^2/2}$ and knowing that $\mathbb{E}(X^n) = \frac{\mathrm{d}^n M_X}{\mathrm{d}\tau^n}(\tau)\Big|_{\tau=0}$, noticing that $\frac{\mathrm{d}^n M_X}{\mathrm{d}\tau^n}(\tau) = H_n(\tau)M_X(\tau)$ where $H_n(\tau)$ is the nth Hermite polynomial.

[13], sampling exclusively the mean value μ. At this point our random variable becomes deterministic, with value $X = \mu$.

We summarize the derived results regarding the products of $\mathrm{d}t$ with $\mathrm{d}W(t)$ with the so called Itô's multiplication table:

	$\mathrm{d}t$	$\mathrm{d}W(t)$
$\mathrm{d}t$	0	0
$\mathrm{d}W(t)$	0	$\mathrm{d}t$

Table 1: Itô's multiplication table for a Wiener process $\mathrm{d}W(t)$.

With the derivation of $(\mathrm{d}W)^2 \approx \mathrm{d}t$ for $\mathrm{d}t \to 0$, we figure out that

$$(\mathrm{d}X)^2 \approx b^2 \mathrm{d}t \tag{52}$$

from (11). With this result in mind together and the definition of Itô's process (8), we continue from (49) by replacing $\mathrm{d}X$ and $(\mathrm{d}X)^2$:

$$\mathrm{d}f(X,t) = \frac{\partial f}{\partial t}\mathrm{d}t + \frac{\partial f}{\partial X}\,(a\mathrm{d}t + b\mathrm{d}W) + \frac{1}{2}\frac{\partial^2 f}{\partial X^2}\,(b^2\mathrm{d}t)$$
$$\mathrm{d}f(X,t) = \left(\frac{\partial f}{\partial t}\mathrm{d}t + a\frac{\partial f}{\partial X} + \frac{b^2}{2}\frac{\partial^2 f}{\partial X^2}\right)\mathrm{d}t + b\frac{\partial f}{\partial X}\mathrm{d}W,$$

matching (13) and concluding our demonstration.

Derivation expected value and variance of $S(t)$

First we derive the expected value for a lognormal process $Y(t) = e^{X(t)}$ with $X(t)$ being an integrated from t_0 to t Itô's process (8) with parameters a and b non-stochastic:

$$X(t) = a(t - t_0) + b[W(t) - W(t_0)]. \tag{53}$$

Noticing that $X(t)$ follows a normal distribution $X(t) \sim \mathcal{N}[a(t-t_0), b^2\sqrt{(t-t_0)}]$ for a and b non-stochastic (from property **(3)** in Wiener's process definition). We can use the moment generating function for a Gaussian distribution $Z(t) \sim \mathcal{N}(\bar{\mu}, \bar{\sigma}^2)$:

$$M_Z(\tau) = \mathbb{E}\left(e^{\tau Z}\right) = e^{\bar{\mu}\tau + \frac{\bar{\sigma}^2}{2}\tau^2}, \tag{54}$$

setting $\tau = 1$ for our Itô's process $X(t)$ with $\bar{\mu} = a(t - t_0)$ and $\bar{\sigma} = b\sqrt{(t - t_0)}$ yielding:

$$\mathbb{E}(Y) = \mathbb{E}\left(e^{X}\right) = e^{a(t-t_0) + \frac{b^2}{2}(t-t_0)}. \tag{55}$$

[13] Formally the Dirac delta $\delta(x - x_0)$ is not a function but a distribution, defined by

$$\delta(x - x_0) = \begin{cases} \infty, & \text{if } x = x_0, \\ 0, & \text{otherwise.} \end{cases}$$

We can then apply (55) for the particular case of the stock price $S(t)$ in (22), starting by taking the expected value on both sides:

$$\begin{aligned}
\mathbb{E}[S(t)] &= S(t_0)e^{\left(\mu-\frac{\sigma^2}{2}\right)(t-t_0)}\mathbb{E}\left[e^{\sigma(W^{\mathbb{P}}(t)-W^{\mathbb{P}}(t_0))}\right] \\
&\overset{(55)}{=} S(t_0)e^{\left(\mu-\frac{\sigma^2}{2}\right)(t-t_0)}e^{0+\frac{\sigma^2}{2}(t-t_0)} \\
&= S(t_0)e^{\mu(t-t_0)},
\end{aligned}$$

where $\sigma[W^{\mathbb{P}}(t)-W^{\mathbb{P}}(t_0)] \sim \mathcal{N}[0,\sigma^2(t-t_0)]$ according to Wiener's process definition property **(3)**, matching the presented result (23).

Next we derive the variance of the lognormal process $Y(t) = e^{X(t)}$ using (55) and noticing that $\mathbb{E}(2X) = 2\mathbb{E}(X)$ and $\mathrm{Var}(2X) = 4\mathrm{Var}(X)$:

$$\begin{aligned}
\mathrm{Var}(e^X) &= \mathbb{E}\left(e^{2X}\right) - \mathbb{E}\left(e^X\right)^2 \\
&\overset{(55)}{=} e^{2a(t-t_0)+\frac{4b^2}{2}(t-t_0)} - e^{2a(t-t_0)+2\frac{b^2}{2}(t-t_0)} \\
&= e^{2a(t-t_0)}\left[e^{2b^2(t-t_0)} - e^{b^2(t-t_0)}\right]
\end{aligned}$$

or

$$\mathrm{Var}(Y) = \mathrm{Var}\left(e^X\right) = e^{(2a+b^2)(t-t_0)}\left[e^{b^2(t-t_0)} - 1\right]. \tag{56}$$

This result can then be used to achieve the variance in the particular case $S(t)$, starting by taking the variance on both sides and keeping only the stochastic terms inside the $\mathrm{Var}(\cdot)$:

$$\begin{aligned}
\mathrm{Var}[S(t)] &= S(t_0)^2e^{2\left(\mu-\frac{\sigma^2}{2}\right)(t-t_0)}\mathrm{Var}\left[e^{\sigma(W^{\mathbb{P}}(t)-W^{\mathbb{P}}(t_0))}\right] \\
&= S(t_0)^2e^{(2\mu-\sigma^2)(t-t_0)}\left\{e^{(2\cdot 0+\sigma^2)(t-t_0)}\left[e^{\sigma^2(t-t_0)}-1\right]\right\} \\
&= S(t_0)^2e^{(2\mu-\sigma^2)(t-t_0)}\left\{e^{\sigma^2(t-t_0)}\left[e^{\sigma^2(t-t_0)}-1\right]\right\} \\
&= S(t_0)^2e^{2\mu(t-t_0)}\left[e^{\sigma^2(t-t_0)}-1\right],
\end{aligned}$$

again we used $\sigma[W^{\mathbb{P}}(t) - W^{\mathbb{P}}(t_0)] \sim \mathcal{N}[0,\sigma^2(t-t_0)]$ matching (24).