Beginner's Analysis of Financial Stochastic Process Models

David Garcia
Harvey Mudd College

Follow this and additional works at: https://scholarship.claremont.edu/hmc_theses

Part of the Education Economics Commons, Other Mathematics Commons, Partial Differential Equations Commons, and the Statistics and Probability Commons

Recommended Citation
https://scholarship.claremont.edu/hmc_theses/269

This Open Access Senior Thesis is brought to you for free and open access by the HMC Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in HMC Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.
Beginner’s Analysis of Financial Stochastic Process Models

David Garcia

Jina Kim, Advisor

Alfonso Castro, Reader

Department of Mathematics

May, 2023
Abstract

This thesis explores the use of geometric Brownian motion (GBM) as a financial model for predicting stock prices. The model is first introduced and its assumptions and limitations are discussed. Then, it is shown how to simulate GBM in order to predict stock price values. The performance of the GBM model is then evaluated in two different periods of time to determine whether its accuracy has changed before and after March 23, 2020.
# Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Abstract</strong></td>
<td>iii</td>
</tr>
<tr>
<td><strong>Acknowledgments</strong></td>
<td>xi</td>
</tr>
<tr>
<td><strong>1 Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>1.1 Statement of Objectives</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Motivations and Interests</td>
<td>3</td>
</tr>
<tr>
<td>1.3 The Story of Stochastic Processes and Finance</td>
<td>6</td>
</tr>
<tr>
<td><strong>2 Background</strong></td>
<td>9</td>
</tr>
<tr>
<td>2.1 The Stock Market (and relevant financial topics)</td>
<td>9</td>
</tr>
<tr>
<td>2.2 Basic Mathematical Concepts</td>
<td>12</td>
</tr>
<tr>
<td>2.3 Stochastic Processes</td>
<td>16</td>
</tr>
<tr>
<td>2.4 Stochastic Differential Equations (SDE)</td>
<td>19</td>
</tr>
<tr>
<td><strong>3 Introduction to Brownian Motion and Geometric Brownian Motion</strong></td>
<td>21</td>
</tr>
<tr>
<td>3.1 Brownian Motion</td>
<td>21</td>
</tr>
<tr>
<td>3.2 Geometric Brownian Motion (GBM)</td>
<td>26</td>
</tr>
<tr>
<td><strong>4 Analysis of Geometric Brownian Motion</strong></td>
<td>33</td>
</tr>
<tr>
<td>4.1 Obtaining Data</td>
<td>33</td>
</tr>
<tr>
<td>4.2 Assumptions of our Model</td>
<td>35</td>
</tr>
<tr>
<td>4.3 Model Description</td>
<td>37</td>
</tr>
<tr>
<td>4.4 Model Simulation</td>
<td>39</td>
</tr>
<tr>
<td>4.5 Model Prediction</td>
<td>42</td>
</tr>
<tr>
<td>4.6 Analysis of Model</td>
<td>45</td>
</tr>
<tr>
<td><strong>5 Future Work</strong></td>
<td>55</td>
</tr>
<tr>
<td>5.1 Improvements to Model</td>
<td>55</td>
</tr>
</tbody>
</table>
5.2 Application of Model ......................... 57
5.3 Other Models .............................. 58

6 Conclusion ................................... 59

Bibliography .................................. 61
# List of Figures

2.1 Graphical Representation of a Normal Distribution . . . . . 14
2.2 Mapping diagram showing the indexing of $A$ by set $I$ . . . . 15
2.3 Stochastic Process Discrete Example ($I \subseteq \mathbb{Z}$) . . . . 17
2.4 Stochastic Process Continuous Example ($I \subseteq \mathbb{R}_{\geq 0}$) . . . 18
2.5 Stochastic Process Discrete Example ($I \subseteq \mathbb{Z} \times \mathbb{Z}$) . . . 18

3.1 Random walk of particle (due to water molecule collisions). 23
3.2 Example Solution(s) to Geometric Brownian Motion . . . . . 27
3.3 Example PDFs for Lognormal Distribution . . . . . . . . . . 29
3.4 Example CDFs for Lognormal Distribution . . . . . . . . . . 30

4.1 Log(Adj. Close Value of ^GSPC/$S_0$) vs. Random Walk . . . 36
4.2 GBM vs Actual Stock (^GSPC) . . . . . . . . . . . . . . . . . . 39
4.3 GBM vs Actual Stock (^DJI) . . . . . . . . . . . . . . . . . . 39
4.4 GBM vs Actual Stock (^IXIC) . . . . . . . . . . . . . . . . . 40
4.5 GBM vs Actual Stock (^RUT) . . . . . . . . . . . . . . . . . . 40
4.6 95% Confidence Interval for the Normal Distribution ($\alpha = 0.05$) 42
4.7 ^GSPC Prediction Error Distributions for $I_{pre}$ and $I_{post}$ . . . 49
4.8 ^DJI Prediction Error Distributions for $I_{pre}$ and $I_{post}$ . . . . 49
4.9 ^IXIC Prediction Error Distributions for $I_{pre}$ and $I_{post}$ . . . 50
4.10 ^RUT Prediction Error Distributions for $I_{pre}$ and $I_{post}$ . . . . 50
List of Tables

1.1 Average return on S&P 500 across several years . . . . . . . . 4
4.1 Derived $\mu, \sigma$ values from (02/15/19) to (04/15/19) . . . . . 40
4.2 95% Confidence Intervals for Stocks Values on April 29th, 2019 43
4.3 Probability of Positive/Negative ROI on Stocks After 2 Weeks 44
4.4 Prediction Error for GBM on Index Funds (2018) . . . . . . . 47
4.5 Summary of Prediction Error for GBM on Index Funds (2018) 47
4.6 Prediction Error for GBM on Index Funds (2022) . . . . . . . 48
4.7 Summary of Prediction Error for GBM on Index Funds (2022) 48
4.8 Difference of Prediction Errors (2022 - 2018) . . . . . . . . . 52
4.9 Summary of Difference of Prediction Errors (2022 - 2018) . . 52
4.10 t-statistic for each stock . . . . . . . . . . . . . . . . . . . . . 53
Acknowledgments

I would like to thank the following people:

• Izabela Joy Quintas, for all the incredible things you have done for me – you have been such an inspiration and I will always be grateful for the joy ;) that you bring into the world!

• My family, for supporting me throughout my educational experience. I will always be more than happy to help debug any technical issue you may have and I promise to provide the absolute best IT support imaginable.

• The many Harvey Mudd College professors, who have helped and guided me during my time here. It was here and because of you all that I found this deep passion for Mathematics and the support to pursue it.

• Prof. Jina Kim, for all your amazing contributions and help in writing this thesis. Without your help, none of this would have been possible!

• All the people I have met along this beautiful journey, especially my friends who I’m sure will go on to do amazing things.

• Last but certainly not lease, I’d like to thank YOU for taking the time to read this thesis.
Chapter 1

Introduction

To begin, this paper is aimed at the layman mathematician who may generally know what the stock market is, but does not necessarily have any expertise in the area.

In this section, we outline the objectives of this article, the main motivations of this topic, why the project is of interest, and a comprehensive review of the literature (including properly cited references) used in the article.

1.1 Statement of Objectives

The goals of this paper is three-fold:

1. Examine certain Stochastic Differential Equation (SDE) financial models,

2. Learn how to apply these SDE models to real American stock market data, and

3. Analyze the accuracy of these models on both pre and post COVID-19 pandemic stock market data.

Without going into excruciating detail at this time, we will be examining the stochastic differential equation financial model geometric Brownian motion. While we do not have time to go into other stochastic differential equation financial models, a good follow-up to this paper would include an analysis of the famous options-pricing Black-Scholes model as well as
several others.

Once we have an understanding of each of these models, we will then apply these models to some well-known stocks, including top index funds like the S&P 500 (^GSPC) and the Dow Jones Industrial Average (^DJI).

We will focus predominately on comparing and contrasting the time periods of stock data taken before and after the pandemic (notably, before and after March 23, 2020, as this was the date in which, debatably, the stock market crashed the hardest).

Finally, we will analyze the accuracy of these models and the effect the pandemic had on the accuracy as well as work to understand why, if at all, our models changed in accuracy.
1.2 Motivations and Interests

Within my lifetime, there have been three major stock market crashes (admittedly, only two that I can truthfully remember). The stock market not only acts as an indicator of the country’s economy, but can also convey the relative health of companies within particular markets. Right now is an especially unique time in the stock market history as we have just recently experienced a major stock market correction, followed by a period of rapid growth (potentially due to inflation) for certain industries, followed again by the worries of an upcoming recession.

Although none of these stock market crashes have been as devastating for the United States as the crash of 1920, which happened over 100 years ago, the effects of each were and are still incredibly significant. Though crashes (such as the one caused by the COVID-19 pandemic) aren’t the focus of the paper, an understanding of how the accuracy of our mathematical models may change in these events will generally allow us to be better versed in the world of finance and hopefully build better intuitions about these models and their relation to the stock market.

One of the largest motivations of the paper is to better understand the economy, albeit through a better understanding of the stock market. The stock market, and the economy with it, can have drastic effects on our lives. These positive and negative affects are felt especially strongly by investors (which accounts for consistently more than 50% of Americans[1]) and can drastically influence anything from one’s own personal investment portfolios to one’s retirement accounts and plans.

Statistically speaking, there is significant value to investing. As of February 2023, investing in index funds (specifically, the S&P 500) would all yield a return rate of roughly 10%.[^1]

[^2]: Table 1.1 from https://tradethatswing.com/average-historical-stock-market-returns-for-sp-500-5-year-up-to-150-year-averages
# Introduction

<table>
<thead>
<tr>
<th>Years Averaged</th>
<th>Average Return / Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>10.382%</td>
</tr>
<tr>
<td>10</td>
<td>10.326%</td>
</tr>
<tr>
<td>20</td>
<td>10.326%</td>
</tr>
<tr>
<td>30</td>
<td>9.749%</td>
</tr>
<tr>
<td>50</td>
<td>10.432%</td>
</tr>
<tr>
<td>100</td>
<td>10.331%</td>
</tr>
<tr>
<td>150</td>
<td>9.079%</td>
</tr>
</tbody>
</table>

*Table 1.1*  Average return on S&P 500 across several years

Therefore, the stock market not only helps keep up with the rate of inflation (so that one’s money is not constantly losing value), but actually has high chances of growing one’s money given a sufficiently long enough period of time.

Additionally, investing has never been easier with the introduction of modern technology and software. It is becoming increasingly common for younger people to invest. As a result, more and more people are trying to “beat” the market and accurately predict the future direction of a stock. There is significant effort being made by multi-billion dollar corporations in order to maximize gains and minimize losses of trading stock each and every day. These money makers use highly complex mathematical models to properly value stock as well as extrapolate and predict trends that humans otherwise cannot. However, the general public knows surprisingly little about how these larger corporations analyze the stock market and make these more informed predictions. A majority of people know nothing about what models are commonly used, let alone how each model works and it’s general accuracy.

With this context in mind, this paper aims to introduce regular people (or mathematicians at least) to the basic building blocks of these more highly complex mathematical models. That being said, this paper is not
about learning how to beat the market or learning how to make advanced investments to make money. Simply, it would be nice for us all, myself as a young investor included, to have a better understanding of what these models are and be able to analyze the effectiveness of each model on predicting stock market data.
1.3 The Story of Stochastic Processes and Finance

The story of stochastic processes and stochastic differential equations (SDEs) in financial modeling is long, but incredibly fascinating. We briefly summarize a couple crucial moments in the history of stochastic processes and finance.

In the early twentieth century, a mathematician named Louis Bachelier was trying to find a way to model the random movements of stock prices. He realized that stock prices exhibit random behavior similar to a Brownian motion, with prices moving up and down seemingly at random due to factors such as market sentiment, news events, and changes in supply and demand. Because of this, he believed that the same mathematical techniques used to describe Brownian motion could be applied to model stock prices; he published these ideas in his doctoral thesis “Théorie de la Spéculation.”

This was genuinely a revolutionary idea at the time, since most economists and financial analysts believed that stock prices followed a predictable pattern and that it was possible to predict their movements with great accuracy. It was Bachelier’s work that helped to shift the focus of financial modeling away from these deterministic models and instead towards stochastic models that could better capture the randomness and uncertainty of financial markets. Although Bachelier’s thesis was far ahead of its time, it did lay the foundation for future research in the field.

About half a century later, in the 1950s, Harry Markowitz, an economist and mathematician, developed the field we now know as portfolio theory, which uses stochastic processes to model the behavior of assets in a portfolio. This theory considers the risk and return of a portfolio of assets and attempts to optimize the portfolio returns based on these factors.

A decade later in the 1960s, the mathematician Paul Samuelson and the economist Robert Merton both independently developed the Black-Scholes model, which is a stochastic differential equation used to value options contracts. The model considers the price of the underlying asset, the time until the option expires, and the volatility of the asset’s price. Having a confident estimate of the value of an option allows one to buy undervalued options and sell them once their true value is fully realized. To this day, the Black-Scholes model is the basis for many modern financial models.
(especially models involving options).

Since the development of the Black-Scholes model, researchers have continued to refine and expand the use of SDEs in finance. Recently, researchers have begun using machine learning algorithms in conjunction with stochastic processes to make more accurate predictions about stock prices. These algorithms can analyze vast amounts of historical data and identify patterns that are difficult for human analysts to detect. Note that we do not dive further into these machine learning models in this paper, so if you are interested in these types of financial models as well, I highly recommend doing some research yourself to see what exists out there!

In conclusion, the use of stochastic processes and SDEs in finance has a long and rich history that continues to evolve and expand. From Bachelier’s pioneering work to modern portfolio theory to the Black-Scholes model, stochastic processes and SDEs have played a crucial role in the development of financial modeling and stock market prediction. As financial markets continue to become more and more complex, these tools will undoubtedly continue to be essential for analysts and traders seeking to make more accurate predictions about the stock market.
Chapter 2

Background

In order to dive deeply into our paper, we must first provide sufficient context on stochastic processes, stochastic differential equations, and the stock market and define all relevant terminology so that we are all sure to be on the same page.

2.1 The Stock Market (and relevant financial topics)

The term stock market typically refers to the market in which shares of certain companies are traded. The buying and selling of these shares occur through regulated, formal exchanges. Traders may buy or sell from any of the stock exchanges that comprise the overall stock market, though the top stock exchanges in the United States are the NASDAQ and the New York Stock Exchange. The stock market is, in essence, the place where traders go to buy or sell shares of public companies from other traders.

Alternative trading systems exist and are regulated differently than the exchanges mentioned above; however, the stock market will be the primary focus of this paper as this is the way a majority of Americans interact with the shares of a public corporation.

The way people usually make a profit by trading is when they buy an asset at a lower price than the one they sell the asset for. The difference between the sell price versus the buy price is their net profit, often referred to as capital gain.
Up until this point, we have been referring abstractly to the things we trade as ‘shares’ or as ‘assets’, so let us clear that up a bit now. There are several types of assets that we are most interested in talking about. Specifically, stocks, which represent a small percentage of ownership of the company, and options, which represent a contract to buy the stock in the future.

Although the exact definitions of each of these assets are not as relevant for the purposes of this paper, the intuitive understanding of what each of these are and how the value of each changes is both relevant and important to the remainder of the paper.

The value of stocks comes from both the number of stocks and the assets and earnings of the company. Hence, when a company does well (i.e. its number of assets and earnings increase), the stock price rises and vice versa. We note that the total number of stocks a company has (remember that each stock represents a percentage of ownership) can in fact change. Although it happens infrequently, it occurs when the company performs a stock split (when each stock now represents half the percentage of the ownership that it previously represented, but the amount of stock each person owned has now doubled) or a reverse stock split (when each stock now represents double the percentage of the ownership that it previously represented, but the amount of stock each person owned has now half-ed). In essence, neither situation affects the overall value of the stocks in hand. Hence, when referring to the value of a stock, people primarily only consider the assets and earnings of the company (typically released each quarter).

The value of options is much more complicated since the underlying asset itself is much more complicated. As described before, an option gives the owner of the option the ability (but not the obligation) to buy or sell the stock at a previously agreed upon price, called the strike price.

For call options, those that allow the owner to buy the stock at a previously agreed price make the most return on investment if the value of the stock increases. This is because you could buy the now much more valuable stock for a lower price, then you could sell the stock you now own for a profit.

For put options, the ones allowing the owner to sell the stock at a previously agreed upon price, make the most return on investment if the stock value drops. This is because you could sell the now much less valuable stock for a
higher price than you otherwise would have been able to.

There is a lot of theory in being able to use options as an insurance, so that, however poorly a stock does, there is a floor to the amount of money you will lose. There is also a premium associated with the cost of a contract such that riskier contracts (ones that are less likely to make a lot of money) cost less than the ones that are more sure bets (the contracts are much more likely to make the owner money).

Additionally, there is a distinction between American and European options – American options allow one to exercise their contract early up until the option expiration date; however, European options can only be exercised on the expiration date and not earlier. Typically, this causes American options to be more expensive (specifically, to have higher premiums), as the flexibility to exercise the option earlier than a previously set date is more desirable.

Given this information, the value of an option is dependent on the ability to exercise the option early, the strike price, the time before the expiration of the option, how the stock is currently performing and how it is expected to perform in the future.
2.2 Basic Mathematical Concepts

2.2.1 Probability

In this section, we define key terms from the field of probability to use as a reference and refresher.

**Definition 2.2.1 (Random Variable).** Let \( \Omega \) be the set that represents the possible outcomes of an experiment. A random variable is a function \( f : \Omega \rightarrow \mathbb{R} \).

**Definition 2.2.2 (Bernoulli Random Variable).** A Bernoulli random variable is a random variable whose only possible values are 0 and 1.

\[ X : \Omega \rightarrow \{0, 1\} \]

For example, flipping a 2-sided coin can be described as a Bernoulli random variable, where the outcomes of a flip are \( \Omega = \{H, T\} \):

\[ X(\omega) = \begin{cases} 1, & \omega = H \\ 0, & \omega = T \end{cases} \]

**Definition 2.2.3 (Mean).** Given some finite collection of numbers \( X \), the mean (notationally referred to as \( \mu \)) of \( X \) is the average of \( X \).

\[ \mu = \left( \sum_{i=1}^{n} \frac{x_i}{n} \right) = \frac{x_1 + x_2 + \cdots + x_n}{n} \]

**Definition 2.2.4 (Expected Value).** Given some random variable \( X \), the expected value of \( X \) (notationally referred to as \( \mu \) or \( E[X] \)) is a weighted average where the value is the outcome and the weight is the probability of that outcome occurring.

\[ E[X] = \left( \sum_{i=1}^{n} x_i \cdot f(x_i) \right) \quad \text{for discrete } X \]

\[ E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \, dx \quad \text{for continuous } X \]

Note that mean and expected values are very similar and oftentimes used interchangeably.
**Definition 2.2.5** (Standard Deviation). Given some finite collection of numbers $X$ or some random variable $X$, the standard deviation (notationally referred to as $\sigma$) of $X$ is a measure of how far the values of $X$ are from the average value of $X$. Specifically,

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)}{n}} \quad \text{for discrete } X
\]

\[
\sigma = \sqrt{\int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \, dx} \quad \text{for continuous } X
\]

**Definition 2.2.6** (Variance). Given some finite collection of numbers $X$ or some random variable $X$, the variance of $X$ (notationally referred to as $\text{Var}[X]$) is simply the squared standard deviation of $X$.

\[
\text{Var}[X] = \sigma^2 = \left(\sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)\right) \quad \text{for discrete } X
\]

\[
\text{Var}[X] = \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x) \, dx \quad \text{for continuous } X
\]

Furthermore, we could use our expected value rules to find that

\[
\text{Var}[X] = \text{E}[X^2] - \text{E}[X]^2
\]

though we will not explicitly prove this here.

**Definition 2.2.7** (Normal Distribution). The normal distribution is a type of continuous probability distribution for some random variable with all real values. The mean of the distribution is $\mu$, standard deviation $\sigma$, and variance $\sigma^2$. 
We say that a random variable $X$ is ‘normally distributed’ if $X$ follows a normal distribution with mean $\mu$ and standard deviation $\sigma$ – or equivalently, $X \sim \mathcal{N}(\mu, \sigma^2)$.

**Theorem 2.2.1.** If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $E[e^X] = e^{\mu + \frac{\sigma^2}{2}}$.

**Definition 2.2.8** (Probability Density Function). The probability density function (PDF) (notationally expressed as $f_X(x)$) is strongly related to the probability that a random variable $X$ is equal to $x$. Note that for a continuous random variable, the probability of taking on any particular value (say $x$) is 0 – so more accurately, the PDF is more commonly used to specify the probability that the random variable falls within a particular range of values.

**Definition 2.2.9** (Cumulative Distribution Function). The cumulative distribution function (CDF) (notationally expressed as $F_X(x)$) of a probability distribution is the probability that a random variable $X$ is less than or equal to $x$.

$$F_X(x) = P[X \leq x] = \int_{-\infty}^{x} f_X(u) \, du$$

**Definition 2.2.10** (Quantile Function). The quantile function of a probability distribution (notationally expressed as $F_X^{-1}(p)$) specifies the value $x$ such that $P[X \leq x] = p$ for a given probability $p$.

### 2.2.2 Probability Space

A probability space or a probability triple $(\Omega, \mathcal{F}, P)$ is a mathematical construct that provides a formal model of a random process or “experiment”.

![Graphical Representation of a Normal Distribution](image-url)
Definition 2.2.11 (Probability Space). A probability space is a tuple consisting of the following three elements:

1. A sample space, $\Omega$, which is the set of all possible outcomes.
2. An event space, which is a set of events $2^\Omega$, an event being a set of outcomes in the sample space.
3. A probability measure $P$, which assigns each event in the event space a real number between 0 and 1.

2.2.3 Indexing Set

Definition 2.2.12 (Indexing Set). An indexing set $I$ is a set whose elements label (or index) the members of another set $A$ (adapted from Munkres, 2000).

Although not necessarily always the case, an indexing set can define the ordering of elements of the set $A$.

Figure 2.2 Mapping diagram showing the indexing of $A$ by set $I$

If there exists a surjective function from the elements of set $I$ onto the elements of set $A$ (i.e. if we are able to label the elements in set $A$ using the elements in set $I$), then we say $I$ is an indexing set. We write $\{A_i\}_{i \in I}$ to denote $A$ being indexed by $I$. We can similarly modify this notation slightly to consider only a subset of $A$, which is labeled by a subset of $I$ (for example, $\{A_j\}_{j \in J}$ where $J \subseteq I$).

2.2.4 Differential Equations

Definition 2.2.13 (Differential Equation). A differential equation is an equation that relates one or more unknown functions and their derivatives.
2.3 Stochastic Processes

We give formal definitions for a stochastic process below and then work to build up our intuition from there.

Definition 2.3.1. Given a probability space \((\Omega, 2^\Omega, P)\), a stochastic process is a collection of random variables defined on this probability space (Florescu, 2014).

More specifically, we consider a stochastic process to be the collection of random variables \(\{X(t)\}_{t \in I}\) (or alternatively written \(\{X(t) : t \in I\}\)), where \(I\) is an indexing set. Some common examples of stochastic processes are Markov chains, Brownian motion (which we define in the next section), Poisson processes, and many more!

Notationally speaking, we will alternate between the notation \(X_t\) and \(X(t)\) to denote the value of the stochastic process at time \(t\).

2.3.1 “Heads or Tails” as a Stochastic Process

Since a stochastic process is just a collection of random variables with an indexing set, a stochastic process can be as simple as flipping a fair two-sided coin repeatedly (ignoring and re-rolling in the unlikely case that the coin lands on it’s side rather than on a face).

Let’s work to understand what exactly this stochastic process would look like using this more concrete example:

As we would expect, flipping a fair coin once could have the outcomes heads or tails. Therefore, the sample space is

\[ \Omega = \{H, T\} \]

As we described earlier in Definition 2.2.2, we can represent flipping a 2-sided coin as a Bernoulli random variable:

\[ X(\omega) = \begin{cases} 1, & \text{if } \omega = H \\ 0, & \text{if } \omega = T \end{cases} \]
Each element of $\Omega$ represents a possible outcome. We have here the outcome of flipping a head or flipping a tail, represented by $H$ and $T$ respectively. Therefore, the event space is

$$2^\Omega = \{\emptyset, \{H\}, \{T\}, \{H, T\}\}$$

The event $\emptyset$ represents flipping neither a head nor a tail. The events $\{H\}$ and $\{T\}$ represent flipping only a head or only a tail, respectively. The events $\{H, T\}$ represent flipping either a head or a tail. The probability mass function $p: X(\Omega) \to [0, 1]$ is:

$$p(X(\omega)) = \begin{cases} 0.5, & \text{if } X(\omega) = 1, \text{ or equivalently, } \omega = H \\ 0.5, & \text{if } X(\omega) = 0, \text{ or equivalently, } \omega = T \end{cases}$$

Therefore, if we let $X_i$ represent the $i^{th}$ coin flip and $I_n = \{1, 2, \cdots, n\}$ (where $n$ is the total number of coin flips), then our stochastic process would be a collection of all of these coin flips $\{X_i : i \in I_n\}$.

### 2.3.2 Indexing Set $I$

The indexing set of a stochastic process, the parameter $I$, not only labels our collection of random variables, but in most cases also induces an ordering on our collection. For the purposes of this paper, the ordering we will be inducing using our indexing set will be an ordering of time. The shape and size of $I$ also determine the type of stochastic process we are dealing with.

![Stochastic Process Discrete Example ($I \subseteq \mathbb{Z}$)](image)

If $I$ is discrete (i.e. $I = \{0, 1, 2, \cdots\}$ or similar), then our stochastic process is a discrete-time stochastic process. In these cases, it is common to write our collection of random variables as $\{X_n\}_{n \in \mathbb{N}}$. 
If $I$ is continuous (i.e. $I = [0, \infty)$ or similar), then our stochastic process is a continuous-time stochastic process. In these cases, it is common to write our collection of random variables as $\{X_t\}_{t \geq 0}$.

There are many more sets that $I$ can be, but the examples above are likely the ones we will be dealing with for the remainder of the paper. For the sake of completeness, we give a couple more abstract examples now.

If $I = \mathbb{Z} \times \mathbb{Z}$ (or similarly discrete), then we may be describing a discrete random field (like the figure above).
2.4 Stochastic Differential Equations (SDE)

Stochastic Differential Equations are incredibly similar to Differential Equations – the key difference is that they have an additional stochastic term(s).

Consider the differential equation

\[
\frac{dX_t}{dt} = AX_t
\]

where \( X_t \in \mathbb{R}^n \) and \( A \in \mathbb{R}^{n \times n} \). For a realistic effect, we add a random behavior to this classical model such that

\[
\frac{dX_t}{dt} = A_t + \sigma \cdot \text{"noise"}
\]

where \( \sigma \in \mathbb{R}^{n \times n} \). For a Brownian motion \( B_t \in \mathbb{R}^n \), if we let

\[
\text{"noise"} = \frac{dB_t}{dt}
\]

then we have the following SDE in the differential form

\[
dX_t = AX_t dt + \sigma dB_t.
\]

Here, we call \( A \) the drift coefficient matrix and \( \sigma \) the diffusion coefficient.

In general, the deterministic term in a stochastic differential equation describes the ‘average’ dynamical behavior of the phenomenon under study whereas the stochastic term describes the ‘noise’ (i.e. the random perturbations that influence the phenomenon).

2.4.1 Itô’s Calculus

Now, how do we solve a differential equation that depends on some random variable? The principles that traditional (Newtonian) calculus is based upon do not always work when considering a stochastic differential equation.

As it turns out, we need new tools to consider these types of differential equations and this area of mathematics.

During World War II, a Japanese mathematician, Kiyosi Itô, devised a way of extending classical methods from calculus (for example, integration
and derivation) to stochastic processes. This field was named Itô’s Calculus (or Stochastic Calculus) in his honor.

A well-known stochastic process that can be solved with stochastic calculus is the Wiener process, which is used to model Brownian motion. As a quick side note, the terms Wiener Process and Brownian motion are often used interchangeably, so we will try sticking to just using ‘Brownian motion’ throughout the rest of this paper, for the sake of simplicity.

Since Itô was at the forefront of this new mathematical field, he delved deep into building and proving each piece of stochastic calculus in order to build a foundation for an equation that we will just be taking for granted. If you are interested in reading more about Itô’s Calculus or seeing the full proof for Itô’s Formula, check out Lawler, 2016! It provides a really in-depth understanding of stochastic calculus and rigorous proof of Theorem 2.4.1.

**Theorem 2.4.1 (Itô’s Formula).** Let \( f(t, x) \) be a function that is once differentiable with regards to \( t \) and twice differentiable with regards to \( x \), and let \( B_t \) be standard Brownian motion. Then,

\[
f(t, B_t) = f(0, B_0) + \int_0^t \partial_x f(s, B_s) \, dB_s + \int_0^t \left[ \partial_t f(s, B_s) + \frac{1}{2} \partial_{xx} f(s, B_s) \right] \, ds
\]

Note that this is the extended version of Itô’s Formula and it depends on both time \( t \) and position \( x \). We can describe it in its differentiable form as well:

\[
df(t, B_t) = \partial_x f(t, B_t) \, dB_t + \left[ \partial_t f(t, B_t) + \frac{1}{2} \partial_{xx} f(t, B_t) \right] \, dt
\]

The applications of Theorem 2.4.1 become useful in the next section.
Chapter 3

Introduction to Brownian Motion and Geometric Brownian Motion

3.1 Brownian Motion

First note that there are several constructions of the Brownian motion. The main mathematical constructions are 1) Paul Lévy’s construction and 2) a consequence of the Wiener process (Mörers and Peres, 2010).

In this paper, we follow Paul Lévy’s construction of Brownian motion (Mörers and Peres [2010]):

In general, Brownian motion is closely related to normal distribution.

Definition 3.1.1. We say that a stochastic process \( \{B(t) \mid t \in I\} \) (where \( I \) is the indexing set) is a Brownian motion if the following properties are true:

- \( B(t) \in \mathbb{R} \) for all values \( t \in I \)
- Let \( B_{\Delta t_i} = \left( B(t_i) - B(t_{i-1}) \right) \). For all times \( t_0 \leq t_1 \leq \cdots \leq t_n \), the increments \( B_{\Delta t_i}, \cdots, B_{\Delta t_n} \) are independent random variables.
- For all \( t \geq 0 \) and \( h > 0 \), the increment \( B(t + h) - B(t) \) is normally distributed with \( \mu = 0 \) and the variance \( h \). Alternatively said, the increment follows \( \mathcal{N}(0, h) \).
• Almost surely\(^1\) the function \(t \to B(t)\) is continuous.

Note that the last property is by far the most technically complex since it requires a bit more theory in and a greater understanding of the field of probability than we choose to go into within this body of work. As a result, we gloss over a bit of detail and do our best to stick to an intuitive understanding of this property.

Additionally, we say that the Brownian motion \(\{B(t) \mid t \in I\}\) is standard Brownian motion if \(B(0) = 0\).

### 3.1.1 History of Brownian Motion

The biologist Robert Brown, when looking at pollen grains in water through a microscope, noticed that the gains moved around within the water despite the container of water being perfectly still.

Brown’s original hypothesis was that the movement he noticed was caused somehow by organic matter; however, after repeating the experiment again with inorganic matter, the movement persisted. Despite not being able to determine the root cause of the movement, Brown’s discovery led to talk about atoms far earlier than they were discovered.

What we know now is that the movement that Brown noticed was caused by water molecules, which are constantly moving and colliding with the grain of pollen. Although a pollen grain is significantly larger than water molecules (roughly 10,000 times larger), the collective effect creates a force large enough to actually move the grain. When the pollen of grain is hit on one side or area more than its others, then the direction of this resulting force ends up being in the opposite direction, and thus the pollen of grain moves away from these collisions (a more intuitive description of these collisions may be that the pollen grain is bouncing off of the water molecules and ends up bouncing in one direction more than the others at any given time). Although probability tells us that the pollen is equally likely to be hit on all of its sides, this isn’t what ends up happening due to the random nature of motion and instead there tends to be some ever changing directionality in the collisions.

\(^1\)an event is said to happen almost surely if it happens with probability 1.
The random nature of this phenomenon (in this case, the randomness of motion or entropy) is the great discovery behind stochastic processes. In honor of Robert Brown, this type of phenomenon was named Brownian Motion.

3.1.2 Examples of Brownian Motion

We consider a toy example in order to more concretely understand each of the properties of Brownian motion. Unfortunately, our coin toss example is unable to be made into Brownian motion since it is a discrete stochastic process (which fails the almost surely continuous property of Brownian motion), so we must consider a different example.

Taking inspiration from the section History of Brownian Motion, let us together imagine a single grain of pollen suspended in still water. Similarly to Robert Brown, when we watch this grain of pollen closely, we also notice that it moves in the water in a seemingly random fashion. For simplicity, we choose to focus only on the location of the pollen grain along the x axis as time passes. Now, pretend (though impossible) that we have a machine that knows the exact location of this grain so that it measures and captures the location perfectly and continuously throughout time. We call the path that the particle takes throughout time a “random walk”.

![Random walk of particle (due to water molecule collisions)](image)

**Figure 3.1** Random walk of particle (due to water molecule collisions)

Say we record the gain of pollen’s location from times $t_0$ to $t_f$ such that
for all \( t_i \in [t_0, t_f] \), the location of the particle (say \( B(t_i) \)) at time \( t_i \) is given to us by this machine.

Hence, our stochastic process would include all the locations of the particle throughout the time interval described above. Mathematically written, our stochastic process would be

\[
\{ B(t) \mid t \in [t_0, t_f] \}
\]

where \([t_0, t_f]\) is our indexing set.

Now that we have a stochastic process in mind, let us go over each of the properties to ensure that they are true!

We explain how this example follows each of the properties of Brownian motion by focusing more on the intuition behind these properties rather than being extremely technical with proving each of these properties are true for our toy example.

1. First, all of our values \( B(t) \) are in fact real as they represent a location along the x-axis for all values of \( t \) in our indexing set \([t_0, t_f]\).

2. Next, we assume that the movement of these water molecules is random. Let \( B_{\Delta t_i} = (B(t_i) - B(t_{i-1})) \) for times \( t_0 \leq t_1 \leq \cdots \leq t_n \). First, take for granted that \( B_{\Delta t_i} \sim N(0, \Delta t_i) \) as we will show this explicitly in the next step. In this case, since \( B_{\Delta t_i} \) follows a normal distribution, its’ outcome in no way, shape, or form depends on the other increments. Therefore, \( B_{\Delta t_i} \) is independent of \( B_{\Delta t_j} \) (where \( i \neq j \)). Thus, we have independent increments, as desired!

3. Then, for all \( t \geq 0 \) and \( h > 0 \), let us consider the increment \( (B(t + h) - B(t)) \).

We expect the displacement of the grain of pollen between times \( t \) and \( t + h \) to be 0 as the probability the grain moves in any direction over another is zero, so it is expected to be in the same location at each of these times.

Additionally, the more time that passes, the greater the chance there is of this displacement being nonzero. Hence, we have that the variability is dependant on time, so although we do not prove it
technically, intuitively a variability of $h$ sounds reasonable. Hence, the increment follows $\mathcal{N}(0, h)$.

4. Finally, we consider that the property that we stated was the most technically complex to prove. Rather than proving that the event $t \to B(t)$ being continuous has probability 1, we settle with showing that $t \to B(t)$ is continuous. Since a grain of pollen cannot teleport, in order to move from one location to another, it must move continuously. Hence, as time progresses and as $t$ increases, we have that $B(t)$ is continuous (so $t \to B(t)$ is continuous).

Therefore, our toy example does in fact fulfill all the properties of Brownian motion and can be useful in considering the question “What is Brownian motion?”
3.2 Geometric Brownian Motion (GBM)

A stochastic process $S_t$ is said to be a geometric Brownian motion with drift $\mu$ and volatility $\sigma$ if it satisfies the following stochastic differential equation:

$$dS_t = \mu S_t \, dt + \sigma S_t \, dB_t$$

where $B_t$ is Brownian motion.

Let us consider Theorem 2.4.1 again when $f(t, x) = S_0 \cdot \exp(at + bx)$, which has partial derivatives:

$$\partial_t f(t, x) = a \cdot f(t, x), \quad \partial_x f(t, x) = b \cdot f(t, x), \quad \text{and} \quad \partial_{xx} f(t, x) = b^2 \cdot f(t, x).$$

Consider what happens when allowing the position to be described by Brownian motion, when $x = B_t$. We can set $S_t = f(t, B_t)$ to represent this and simplify the differential form of Itô's Formula as follows

$$df(t, B_t) = \partial_x f(t, B_t) \, dB_t + \left[ \partial_t f(t, B_t) + \frac{1}{2} \partial_{xx} f(t, B_t) \right] \, dt$$

$$dS_t = b f(t, B_t) \, dB_t + \left[ a f(t, B_t) + \frac{1}{2} b^2 f(t, B_t) \right] \, dt$$

$$dS_t = b S_t \, dB_t + \left[ a S_t + \frac{1}{2} b^2 S_t \right] \, dt$$

$$dS_t = b S_t \, dB_t + \left[ a + \frac{b^2}{2} \right] S_t \, dt$$

Although it took choosing a very clever selection for $f(t, x)$, this is exactly our definition for GBM where $\mu = a + \frac{b^2}{2}$ and $\sigma = b$. Thus, we can now easily solve our original equation:

$$S_t = f(t, B_t) = S_0 \cdot \exp(at + bB_t)$$

Plugging in $\mu$ and $\sigma$, we have our final solution for geometric Brownian motion as

$$S_t = S_0 \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right)$$

**Example 3.2.1.** Setting $S_0 = 1$, $\mu = 0.01$, and $\sigma \in \{0.001, 0.01, 0.05\}$, we get the following:
Figure 3.2  Example Solution(s) to Geometric Brownian Motion

We use the code found in (citation) in order to simulate geometric Brownian motion. In total, we generated 5 different samples (for each sigma value) and plot those samples together in the same graph.

We see here that at this scale, that all five simulations for ($\sigma = 0.001$) look really similar (as it’s hard to even distinguish that there were 5 samples plotted in that graph).

For the simulations where ($\sigma = 0.01$), we begin to see a bit more volatility between the samples and are able to see the differences between each of the samples. However, the sigma value does not seem to drastically affect the general trend of the simulation; the samples seem to be increasing at a similar rate.

Finally, the simulations for ($\sigma = 0.05$) are where we really begin to see the effects $\sigma$ has on geometric Brownian motion and in the last graph, we can really see how volatility affects the shape of the line.
3.2.1 Properties of GBM

Now that we have defined geometric Brownian motion and have some intuition behind what exactly it is, let’s consider what the expected value and variation of geometric Brownian motion is at each time \( t \).

Next, we get the expected value of GBM, or more specifically, \( S_t \) using 2.2.4:

\[
E[S_t] = E \left[ S_0 \cdot \exp \left( \left( \mu - \frac{\sigma^2}{2} \right) t + \sigma B_t \right) \right] \\
= S_0 e^{(\mu - \frac{\sigma^2}{2})t} \cdot E[e^{\sigma B_t}] \\
= S_0 e^{(\mu - \frac{\sigma^2}{2})t} \cdot e^{\frac{\sigma^2}{2}} \\
= S_0 e^{(\mu - \frac{\sigma^2}{2})t + \frac{\sigma^2}{2}} \\
= S_0 e^{\mu t}
\]

Next, we get the variance of \( S_t \)

\[
\text{Var}[S_t] = E[S_t^2] - (E[S_t])^2 \\
= E[S_t^2] - (S_0 e^{\mu t})^2 \\
= E\left[ S_0^2 \cdot e^{(\mu - \frac{\sigma^2}{2}) t + 2\sigma B_t} \right] - (S_0 e^{\mu t})^2 \\
= S_0^2 e^{2(\mu - \frac{\sigma^2}{2}) t} \cdot E[e^{2\sigma B_t}] - (S_0 e^{\mu t})^2 \\
= S_0^2 e^{2(\mu - \frac{\sigma^2}{2}) t} \cdot e^{\sigma^2 t} - (S_0 e^{\mu t})^2 \\
= S_0^2 e^{2(\mu - \frac{\sigma^2}{2}) t} \cdot e^{\sigma^2 t - 1} \\
= (S_0 e^{\mu t})^2 \cdot (e^{\sigma^2 t} - 1)
\]

The following theorem gives us the PDF for \( S_t \), given time \( t \).

**Definition 3.2.1.** We say \( X \) follows a lognormal distribution\(^2\) with mean \( \mu \)

\(^2\)We say \( X \) is lognormally distributed if \( Y = \ln(X) \) is normally distributed.
and standard deviation $\sigma$. Then, the probability density function $f$ of $X$ is

$$f_X(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \left( -\frac{[\ln(x) - \mu]^2}{2\sigma^2} \right), \quad x \in (0, \infty).$$

We can visualize this by plugging in values for $\mu$ and $\sigma$:

![Figure 3.3 Example PDFs for Lognormal Distribution](image)

Setting $X = S_t$, we see that $S_t$ is lognormally distributed with mean $\mu = \ln(S_0) + \left( \mu' - \frac{\sigma'^2}{2} \right)t$ and standard deviation $\sigma = \sigma' \sqrt{t}$ (where $\mu'$ is the drift parameter and $\sigma'$ is the volatility parameter of GBM). To show this explicitly, let $Y = \ln(S_t)$:

$$Y = \ln(S_t)$$

$$= \ln(S_0) + \left( \mu' - \frac{\sigma'^2}{2} \right)t + \sigma' B_t$$

$$= \ln(S_0) + \left( \mu' - \frac{\sigma'^2}{2} \right)t + X$$

where $X \sim N(0, \sigma'^2t)$ and $\left( \ln(S_0) + \left( \mu' - \frac{\sigma'^2}{2} \right)t \right)$ are constant for fixed $t$.

Therefore, $Y$ is normally distributed with mean $\left( \ln(S_0) + \left( \mu' - \frac{\sigma'^2}{2} \right)t \right)$ and standard deviation $\sigma' \sqrt{t}$. Thus, the probability density function of $S_t$ is

$$f_{S_t}(x) = \frac{1}{\sqrt{2\pi t} \sigma' x} \exp \left( -\frac{\left[ \ln\left( \frac{x}{S_0} \right) - \left( \mu' - \frac{\sigma'^2}{2} \right)t \right]^2}{2\sigma'^2t} \right), \quad x \in (0, \infty)$$
Next, we introduce the following theorem to find the CDF for $S_t$, given time $t$.

**Definition 3.2.2.** Let $X$ follow a lognormal distribution with mean $\mu$ and standard deviation $\sigma$. Then, the CDF $F$ of $X$ is

$$F_X(x) = \Phi \left( \frac{\ln(x) - \mu}{\sigma} \right)$$

where $\Phi$ is the CDF of the standard normal distribution.

We can visualize this by plugging in values for $\mu$ and $\sigma$.

![Figure 3.4 Example CDFs for Lognormal Distribution](image)

Therefore, the CDF of $S_t$ is

$$F_{S_t}(x) = \Phi \left( \frac{\ln \left( \frac{x}{S_0} \right) - \left( \mu' - \frac{\sigma'^2}{2} \right) t}{\sigma' \sqrt{t}} \right)$$

Finally, we introduce the following theorem to find the quantile function for GBM.

**Definition 3.2.3.** Let $X$ follow a lognormal distribution with mean $\mu$ and standard deviation $\sigma$. Then, the quantile function $F^{-1}(p)$ of $X$ is

$$F_X^{-1}(p) = \exp \left( \mu + \sigma \cdot \Phi^{-1}(p) \right)$$
Thus, the quantile function of $S_t$ is

$$F_{S_t}^{-1}(p) = \exp \left( \ln(S_0) + \left( \mu' - \frac{\sigma'^2}{2} \right) t + \sigma' \sqrt{t} \cdot \Phi^{-1}(p) \right)$$

We get into the importance of the stochastic differential equation for geometric Brownian motion (specifically, the solution associated with it) later in the paper.
We first analyze the accuracy of perhaps the most simplistic financial model, the geometric Brownian motion, to develop a baseline and a consistent analytical structure for our other models. We’ll spend a bit more time in this section justifying the structure and analytical tools we use to test the validity of our financial model. Note that we do not expect this model to be all that accurate, but that’s okay since that is not what we are measuring!

4.1 Obtaining Data

First, since we cannot reasonably test the model on every single stock, we must choose which stocks to prioritize. Since we really want to test the accuracy of the model on “the stock market”, we will be using some of the most popular U.S. index funds (which are constructed with the purpose of matching or tracking the state of a given financial market, in our case the U.S. stock market). Specifically, we consider

• ^GSPC (S&P 500),
• ^DJI (Dow Jones),
• ^IXIC (Nasdaq Composite), and
• ^RUT (Russell 2000).

Although our data should be the same if we were to use any other historical stock market database, we get all of our historical data from Yahoo
Finance due to my familiarity and comfort with the site, as well as its ability to download and query vast amounts of free historical data programmatically for specific stocks.\footnote{We’re using Python, so to programmatically get the historical data from Yahoo Finance, we use the package \texttt{yfinance}.}

We will also be looking only at the adjusted closing value for a given stock when considering the asset’s price for a given day. The adjusted close value of a stock is similar to its closing value, but adjusts the stock’s value to account for any corporate actions that may have affected the price (such as stock splits, merges, or dividends). For this reason, the adjusted closing price is a more accurate measure of the stock’s true value (at least when compared to the stock’s normal closing price) as it accounts for any changes in the number of outstanding shares or cash distributions to shareholders.

Generally speaking, using adjusted closing prices can help avoid biases in our model that may arise if we use other types of data, such as the raw closing price or intra-day prices.\footnote{For example, if we use raw closing prices, we may not account for changes in the number of outstanding shares due to stock splits, which could make it appear as though the stock has increased in value when in reality it has not. Similarly, using intra-day prices can introduce biases if the stock is highly volatile or experiences sharp price movements during the day.}

By using adjusted closing prices, we can create a more accurate representation of the stock’s true value over time. This can help our models make more accurate predictions about future price movements and avoid biases that may be introduced by other types of data.
4.2 Assumptions of our Model

To understand the usability of our model as a predictor of the stock market, we must first examine the assumptions we are making about the model.

First and foremost, we assume that the stock market follows geometric Brownian motion. Although there are a wide array of factors that influence any given stock and thus the stock market, most financial models make the assumption that the randomness of the stock market can be modeled using Brownian motion, so (at the very least) it isn’t unreasonable for me to be making this assumption as well. Generally speaking, we tend view the stock market as essentially a continuous random walk (Brownian motion). Though it is likely a flawed representation of reality, it is nevertheless a useful assumption.

Next, for geometric Brownian motion, we assume that the log of the asset price follows a Brownian motion. Again, this assumption ignores reality, as there are likely outside influences that we can consider to help us make a more informed prediction (i.e., knowing the price of an index fund may be more likely to decrease rather than increase when the country is in a recession), but it is still a surprisingly reasonable assumption.

Below, we plot the log-adjusted close values of some stock. We choose ^GSPC as it was the first stock on our list from 4.1. Granted, this selection is a bit arbitrary, but we are doing our best not to input any bias into the selection. Additionally, we plot from February 15, 2019 to April 15, 2019. The range of these dates is to allow for roughly 40 distinct data points (days the stock market is open), which helps give us a sufficiently long time-span to visualize the data without being overly complicated. Additionally, we chose this time of the year in order to eliminate even more selection bias. Finally, we chose to start in 2019 because this was the year prior to COVID, which is a year of particular interest to us. We also chose this date range to avoid certain seasonal patterns (craze around the winter holidays and summer months). Though the graph below in no way proves the validity of geometric Brownian motion, it does allow us to build the intuition behind why we tend to use GBM to model stocks in the first place.
Analysis of Geometric Brownian Motion

Figure 4.1  Log(Adj. Close Value of ^GSPC/\(S_0\)) vs. Random Walk

If \(S\) represents the adjusted close values of the stock \(^\text{GSPC}\) and \(t = 0\) represents February 15th 2019 and \(t_f\) represents April 15th, 2019, then the darker line in Figure 4.1 shows \(y(t) = \ln\left(\frac{S_t}{S_0}\right)\) for \(t \in [t_0, t_f]\).

Additionally, the lighter lines show 10 random walks, where each step is normally distributed, centered at 0 and has a standard deviation of the average step sizes of \(S\), \(\sigma = \frac{\sum_{i=t_1}^{t_f} (S_{t_i} - S_{t_{i-1}})}{t_f - t_1}\).

We see here that it seems entirely plausible that the above graph is just a random walk (given Brownian motion has accurate parameters \(\mu\) and \(\sigma\), which we really loosely approximated above).

Another assumption made by GBM is that drift \(\mu\) and volatility \(\sigma\) are constant throughout a given time frame. This is unfortunately, not a realistic assumption as drift (the appreciation or depreciation an asset has over time) empirically changes given a long enough time interval. Note that this follows our intuitive understanding about the potential price of any asset, since if the price of something increases forever at the same rate, there will be some time in the future when the stock’s value exceeds the finite resources (and thus monetary equivalent) on Earth. To make this assumption a little more realistic, we will focus on a sufficiently short time interval such that the values of \(\mu\) and \(\sigma\) do not change drastically. Though there isn’t an exact time interval for this, we will be limiting our time frame to 10 weeks. This interval seems short enough for the \(\mu\) and \(\sigma\) values to not change too drastically for our purposes, yet long enough to get more consistent and reliable data.
4.3 Model Description

Now that we understand what data we will be using and what assumptions our model is making, let’s dive more deeply into how to actually apply our financial model (in this case GBM) to the data.

Recall our solution from 3.2 of the geometric Brownian motion.

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

Here,

- \( S_T \) represents the stock’s price at time \( t = T \),
- \( S_0 \) represents the initial value of the stock,
- \( \mu \) represents the expected rate of return that a stock will earn over a short period of time,
- \( \sigma \) represents the expected volatility of the stock
- \( B_T \) represents the random volatility of the stock, which increases as \( t \) increases.

We use the following method to calculate the values for \( \mu \) and \( \sigma \) on the window \( t_{start} \) to \( t_{end} \):

1. We gather our historical data for a given stock (described in 4.1).
2. We simulate geometric Brownian motion for selected initial values of \( \mu \) and \( \sigma \).

Note from the previous equation that when \( S_T \approx S_0 \), then

\[
\exp \left( \left( \mu - \frac{\sigma^2}{2} \right) T + \sigma B_T \right) \approx 1
\]

meaning, for arbitrarily large \( T \),

\[
\left( \mu - \frac{\sigma^2}{2} \right) T + \sigma B_T \approx 0
\]
Thus, for time frames \((t_{\text{start}}, t_{\text{end}})\) where our stock’s value does not change that drastically, we should expect our values for \(\sigma\) to be near 0, and thus our value for \(\mu\) to also be near 0 as well. Though this certainly is not exact, it gives us a rough, but good, baseline to use when thinking about the values for these two constants.

3. We run the simulation over and over again until we achieve values of \(\mu\) and \(\sigma\) that optimizes (minimizes) the error between our GBM model’s predicted stock market values and the actual values over the time frame.

Recall that we make the assumption that \(\mu\) and \(\sigma\) are constants throughout a given time period. Since, in reality, these values likely change over time, it is only reasonable to assume \(\mu\) and \(\sigma\) are constant over a sufficiently short time interval. Hence, if we assume that our constants are only fixed over a 10-week time frame, and we use an 8-week time frame to calculate these values, then we may only use GBM with these parameters to model (and thus predict) the stock’s value for a 2-week time frame.

Additionally, once we obtain values for \(\mu\) and \(\sigma\), we should begin this time frame directly after our previous one (so our predicted time frame begins at \(t = t_{\text{end}}\) and ends at \(t = t_{\text{end+2-weeks}}\)).

\[\sum_{i=1}^{n} |S_i - X_i| \]

\(\) where \(S_i, X_i\) is the real and predicted stock value, respectively, at time \(t = i\).
4.4 Model Simulation

Much like how we simulated GBM before to find approximations for $\mu$ and $\sigma$, we run the simulation again with our newly found parameters. Specifically, we run 2 different simulations for each of our index stocks, which differ only in their start date.

For example, the first pair of graphs below both plot 100 simulations of geometric Brownian motion (in gray) with the derived values for $\mu$, $\sigma$ along with the true asset price for the given stock (in purple). The first graph begins on February 15th, 2019 ($t_0$) and lasts until April 15th, 2019 ($t_f$), which we plot in order to confirm that our chosen $\mu$, $\sigma$ values are reasonable. The second graph begins on April 15, 2019 ($t_f$) and lasts for a total of 14 days, which we plot to show how our predictions compare with the true asset price.

![Figure 4.2 GBM vs Actual Stock (^GSPC)](image1)

![Figure 4.3 GBM vs Actual Stock (^DJI)](image2)
Analysis of Geometric Brownian Motion

Figure 4.4  GBM vs Actual Stock (^IXIC)

Figure 4.5  GBM vs Actual Stock (^RUT)

For your convenience, we have the derived $\mu$ and $\sigma$ values in the below table.

<table>
<thead>
<tr>
<th>Stock</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>^GSPC</td>
<td>8.919e-4</td>
<td>2.938e-3</td>
</tr>
<tr>
<td>^DJI</td>
<td>1.344e-4</td>
<td>3.251e-3</td>
</tr>
<tr>
<td>^IXIC</td>
<td>1.452e-3</td>
<td>2.996e-3</td>
</tr>
<tr>
<td>^RUT</td>
<td>-3.13e-4</td>
<td>5.542e-3</td>
</tr>
</tbody>
</table>

Table 4.1  Derived $\mu, \sigma$ values from (02/15/19) to (04/15/19)

Note that any similarities in the graphs and $\mu, \sigma$ values are almost certainly due to the fact that all of these index funds are trying to accurately
reflect the state of the stock market. If each of the index funds were able to perfectly predict the state of the stock market, then we would expect them all to have the exact same $\mu$, $\sigma$ values. However, since they measure the state of the stock market through different means, index funds (at least in this time interval) are distinct enough to warrant differing values of $\mu$ and $\sigma$. 
4.5 Model Prediction

Recall from 3.2.1 that $E[S_t] = S_0 e^{\mu t}$ and $\text{Var}[S_t] = (S_0 e^{\mu t})^2 \left(e^{\sigma^2 t} - 1\right)$.

Now, we can use this information to find a 95% confidence interval for the expected stock price after 2 weeks. As a more simplistic example, if we were to do this for the normal distribution (rather than a lognormal distribution like GBM), it would look like the following.

![Figure 4.6 95% Confidence Interval for the Normal Distribution ($\alpha = 0.05$)](image)

For all distributions (including the lognormal distribution GBM), we get the lower and upper bound value of our confidence interval using the quantile function $Q(0.025) = \text{Lower}$ and $Q(0.975) = \text{Upper}$.

By doing this, we can find the 95% confidence intervals for each of our stocks to predict the price of our stock for April 29th 2019 (the 14th day after $t_f$).

In our scenario, we’re predicting the stock price starting at the end of our previous interval April 15th, 2019 and ending 14 days later at on April 29th, 2019 ($t_f' = t_f$ and $t_f' = t_f + 14$). Hence, our initial value for this interval will
be the value of the stock on April 15th, 2019 ($S'_0 = X_{t_f}$). We will be using the same derived $\mu$, $\sigma$ values and assume that they remain constant throughout this new window.

![Table 4.2](image)

Now, this means that we can be 95% confident that the stock price on April 29th, 2019 will fall between the lower and upper bounds of our confidence interval. Our highest lower bound is 99.83%, which is actually still pretty close to 100% (the full price of the initial asset) – meaning that buying ^IXIC would likely be our best bet here, though it’s still not a guaranteed profit for investors.

Using our CDF, we could calculate the probability of the stock dropping below it’s initial value ($X_t < S'_0$) and the probability of the stock staying at or rising above it’s initial value ($X_t \geq S'_0$). This yields the following:
Analysis of Geometric Brownian Motion

Now this means that we can be 99% confident of seeing a non-negative return on investment for the stock ^IXIC, which is consistent with what we saw before. Additionally, this means that we can be even more confident in buying this stock. If we were to prioritize buying stocks when there is a high probability of a positive return on investment, we’ll maximize our overall return on investment.

<table>
<thead>
<tr>
<th>Stock</th>
<th>$P(X_t &lt; S'_0)$</th>
<th>$P(X_t \geq S'_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>^GSPC</td>
<td>12.92%</td>
<td>87.08%</td>
</tr>
<tr>
<td>^DJI</td>
<td>44.09%</td>
<td>55.91%</td>
</tr>
<tr>
<td>^IXIC</td>
<td>3.53%</td>
<td>96.47%</td>
</tr>
<tr>
<td>^RUT</td>
<td>58.77%</td>
<td>41.23%</td>
</tr>
</tbody>
</table>

Table 4.3  Probability of Positive/Negative ROI on Stocks After 2 Weeks
4.6 Analysis of Model

Now, in order to get the accuracy of our model, we will:

1. Consider some fixed time interval \([t_s, t_f]\)

2. Randomly choose \(n\) dates from this time interval, and for each chosen date \((t \in [t_s, t_f])\)
   
   (a) Predict the value of the asset 14 days after \(t\)
   
   (b) Calculate the prediction error \(p_t\) for a date \(t\)

   \[ p_t = \frac{S_{t+14} - X_{t+14}}{X_{t+14}} \]

   where \(S_{t_i}\) is the predicted asset price at \(t_i\) and \(X_{t_i}\) is the actual asset price at \(t_i\).

3. Using a statistical test, we compare the prediction errors from pre and post pandemic.

4.6.1 Fixed Time Interval

Since one of our goals was to compare the accuracy of the model before and after the COVID-19 pandemic, we will focus on two fixed time intervals:

\[ I_{\text{pre}} = [1/1/2018, 12/31/2018] \quad I_{\text{post}} = [1/1/2022, 12/31/2022] \]

These time intervals are both exactly one year long and take place roughly a year before and after the peak of the pandemic’s impact on the stock market (which, for the purposes of this paper, we consider to be March 23rd, 2020).

\[^{5}\text{We use prediction error as a means of representing the model’s accuracy.}\]
4.6.2 Prediction Error

Given this date interval, we will consider all the dates the stock market is open (usually, but not always, Monday through Friday) and randomly select \( n \) unique dates. Note that there are a total of 251 dates for which the stock market is open in each interval. We choose \( n = 50 \) many random integers in the range \([0, 250]\) (where \( i \in [0, 250] \) represents the \( i^{th} \) day that the stock market is open for in the given year) to correspond to 50 dates throughout the given year. We choose 50 in particular, as it is a nice round integer which gives us solid representation of the year, without having to actually calculate the prediction error for every single date the NYSE is open for in each interval.

From there, we will run the simulation of the model and obtain the prediction errors (%). First, we get the prediction error for the pre-COVID interval \( I_{pre} \), then also for the post-COVID interval \( I_{post} \).

---

\(^6\)We’re using Python, so to create repeatable results, we use the \texttt{random} package with an arbitrarily chosen seed (42).
4.6.3 Prediction Error \( I_{\text{pre}} = [1/1/2018, 12/31/2018] \)

<table>
<thead>
<tr>
<th>(i^{th})-day</th>
<th>Date (2018)</th>
<th>(^GSPC)</th>
<th>(^DJI)</th>
<th>(^IXIC)</th>
<th>(^RUT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Jan 05</td>
<td>-0.19</td>
<td>0.31</td>
<td>-1.94</td>
<td>-1.43</td>
</tr>
<tr>
<td>9</td>
<td>Jan 16</td>
<td>0.62</td>
<td>1.37</td>
<td>0.77</td>
<td>0.45</td>
</tr>
<tr>
<td>10</td>
<td>Jan 17</td>
<td>1.57</td>
<td>2.44</td>
<td>1.71</td>
<td>1.86</td>
</tr>
<tr>
<td>12</td>
<td>Jan 19</td>
<td>4.7</td>
<td>5.51</td>
<td>5.09</td>
<td>5.41</td>
</tr>
<tr>
<td>15</td>
<td>Jan 24</td>
<td>8.81</td>
<td>9.17</td>
<td>8.75</td>
<td>8.99</td>
</tr>
<tr>
<td>21</td>
<td>Feb 01</td>
<td>6.36</td>
<td>7.3</td>
<td>5.06</td>
<td>4.45</td>
</tr>
<tr>
<td>27</td>
<td>Feb 09</td>
<td>-2.41</td>
<td>-2.01</td>
<td>-3.31</td>
<td>-3.86</td>
</tr>
<tr>
<td>29</td>
<td>Feb 13</td>
<td>-1.44</td>
<td>-1.12</td>
<td>-2.21</td>
<td>-2.74</td>
</tr>
<tr>
<td>42</td>
<td>Mar 05</td>
<td>-2.28</td>
<td>-1.96</td>
<td>-1.49</td>
<td>-4.82</td>
</tr>
<tr>
<td>44</td>
<td>Mar 07</td>
<td>-2.56</td>
<td>-2.94</td>
<td>-1.13</td>
<td>-3.73</td>
</tr>
<tr>
<td>45</td>
<td>Mar 08</td>
<td>0.52</td>
<td>0.44</td>
<td>1.93</td>
<td>-1.45</td>
</tr>
<tr>
<td>51</td>
<td>Mar 16</td>
<td>1.69</td>
<td>0.27</td>
<td>5.23</td>
<td>2.22</td>
</tr>
<tr>
<td>61</td>
<td>Apr 02</td>
<td>-4.64</td>
<td>-5.87</td>
<td>-3.59</td>
<td>-4.04</td>
</tr>
<tr>
<td>66</td>
<td>Apr 09</td>
<td>-4.3</td>
<td>-4.8</td>
<td>-4.14</td>
<td>-3.36</td>
</tr>
<tr>
<td>72</td>
<td>Apr 17</td>
<td>0.05</td>
<td>0.97</td>
<td>-0.17</td>
<td>1.67</td>
</tr>
<tr>
<td>78</td>
<td>Apr 25</td>
<td>-5.47</td>
<td>-4.35</td>
<td>-9.06</td>
<td>-5.23</td>
</tr>
<tr>
<td>79</td>
<td>Apr 26</td>
<td>-5.29</td>
<td>-4.14</td>
<td>-8.3</td>
<td>-5.15</td>
</tr>
<tr>
<td>84</td>
<td>May 03</td>
<td>-4.81</td>
<td>-4.14</td>
<td>-6.29</td>
<td>-5.73</td>
</tr>
<tr>
<td>95</td>
<td>May 18</td>
<td>1.0</td>
<td>1.55</td>
<td>0.39</td>
<td>1.67</td>
</tr>
<tr>
<td>98</td>
<td>May 23</td>
<td>0.72</td>
<td>0.62</td>
<td>-0.24</td>
<td>0.55</td>
</tr>
<tr>
<td>99</td>
<td>May 24</td>
<td>0.59</td>
<td>-0.05</td>
<td>0.43</td>
<td>1.07</td>
</tr>
<tr>
<td>104</td>
<td>Jun 01</td>
<td>-1.68</td>
<td>-2.36</td>
<td>-1.89</td>
<td>-0.89</td>
</tr>
<tr>
<td>105</td>
<td>Jun 04</td>
<td>0.12</td>
<td>-0.22</td>
<td>0.53</td>
<td>0.05</td>
</tr>
<tr>
<td>106</td>
<td>Jun 05</td>
<td>0.69</td>
<td>1.0</td>
<td>1.63</td>
<td>0.95</td>
</tr>
<tr>
<td>122</td>
<td>Jun 27</td>
<td>-2.08</td>
<td>-2.61</td>
<td>-1.41</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4 Prediction Error for GBM on Index Funds (2018)

This is summarized in the following table:

<table>
<thead>
<tr>
<th></th>
<th>(^GSPC)</th>
<th>(^DJI)</th>
<th>(^IXIC)</th>
<th>(^RUT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (%)</td>
<td>-0.65</td>
<td>-0.48</td>
<td>-0.9</td>
<td>-1.09</td>
</tr>
<tr>
<td>Standard Deviation (%)</td>
<td>3.52</td>
<td>3.61</td>
<td>4.19</td>
<td>4.34</td>
</tr>
</tbody>
</table>

Table 4.5 Summary of Prediction Error for GBM on Index Funds (2018)
4.6.4 Prediction Error  ($I_{\text{post}} = [1/1/2022, 12/31/2022]$)  

This is summarized in the following table:

<table>
<thead>
<tr>
<th>$i^\text{th}$-day</th>
<th>Date (2022)</th>
<th>$^\text{GSPC}$</th>
<th>$^\text{DJI}$</th>
<th>$^\text{IXIC}$</th>
<th>$^\text{RUT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Jan 06</td>
<td>5.17</td>
<td>5.08</td>
<td>4.88</td>
<td>5.15</td>
</tr>
<tr>
<td>9</td>
<td>Jan 14</td>
<td>7.55</td>
<td>6.85</td>
<td>7.24</td>
<td>10.2</td>
</tr>
<tr>
<td>10</td>
<td>Jan 18</td>
<td>2.73</td>
<td>1.95</td>
<td>1.56</td>
<td>2.29</td>
</tr>
<tr>
<td>12</td>
<td>Jan 20</td>
<td>1.71</td>
<td>0.52</td>
<td>1.84</td>
<td>1.18</td>
</tr>
<tr>
<td>15</td>
<td>Jan 25</td>
<td>-4.15</td>
<td>-3.12</td>
<td>-7.09</td>
<td>-3.02</td>
</tr>
<tr>
<td>21</td>
<td>Feb 02</td>
<td>2.28</td>
<td>2.82</td>
<td>-0.7</td>
<td>-4.21</td>
</tr>
<tr>
<td>27</td>
<td>Feb 10</td>
<td>0.86</td>
<td>3.87</td>
<td>-2.48</td>
<td>-4.04</td>
</tr>
<tr>
<td>29</td>
<td>Feb 14</td>
<td>-3.62</td>
<td>-0.58</td>
<td>-7.15</td>
<td>-8.29</td>
</tr>
<tr>
<td>42</td>
<td>Mar 04</td>
<td>-5.68</td>
<td>-5.33</td>
<td>-7.66</td>
<td>-6.48</td>
</tr>
<tr>
<td>44</td>
<td>Mar 08</td>
<td>-8.06</td>
<td>-6.74</td>
<td>-10.23</td>
<td>-6.52</td>
</tr>
<tr>
<td>45</td>
<td>Mar 09</td>
<td>-4.72</td>
<td>-3.82</td>
<td>-6.02</td>
<td>-2.4</td>
</tr>
<tr>
<td>51</td>
<td>Mar 17</td>
<td>-5.7</td>
<td>-3.5</td>
<td>-8.47</td>
<td>-1.03</td>
</tr>
<tr>
<td>61</td>
<td>Mar 31</td>
<td>3.31</td>
<td>-0.17</td>
<td>6.05</td>
<td>4.07</td>
</tr>
<tr>
<td>66</td>
<td>Apr 07</td>
<td>5.02</td>
<td>1.07</td>
<td>8.36</td>
<td>3.55</td>
</tr>
<tr>
<td>72</td>
<td>Apr 15</td>
<td>8.37</td>
<td>6.28</td>
<td>9.60</td>
<td>7.26</td>
</tr>
<tr>
<td>78</td>
<td>Apr 26</td>
<td>9.47</td>
<td>7.23</td>
<td>13.13</td>
<td>10.61</td>
</tr>
<tr>
<td>79</td>
<td>Apr 27</td>
<td>11.04</td>
<td>8.2</td>
<td>16.13</td>
<td>12.58</td>
</tr>
<tr>
<td>84</td>
<td>May 04</td>
<td>7.55</td>
<td>7.14</td>
<td>9.82</td>
<td>6.08</td>
</tr>
<tr>
<td>98</td>
<td>May 24</td>
<td>-10.9</td>
<td>-7.56</td>
<td>-14.99</td>
<td>-13.44</td>
</tr>
<tr>
<td>99</td>
<td>May 25</td>
<td>-9.04</td>
<td>-6.21</td>
<td>-13.09</td>
<td>-10.4</td>
</tr>
<tr>
<td>104</td>
<td>Jun 02</td>
<td>7.48</td>
<td>6.53</td>
<td>6.27</td>
<td>7.96</td>
</tr>
<tr>
<td>105</td>
<td>Jun 03</td>
<td>4.32</td>
<td>4.49</td>
<td>0.61</td>
<td>4.63</td>
</tr>
<tr>
<td>106</td>
<td>Jun 06</td>
<td>7.66</td>
<td>6.73</td>
<td>4.89</td>
<td>8.24</td>
</tr>
<tr>
<td>122</td>
<td>Jun 29</td>
<td>-1.71</td>
<td>-1.1</td>
<td>-2.31</td>
<td>-0.93</td>
</tr>
<tr>
<td>...</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
<td>:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$i^\text{th}$-day</th>
<th>Date (2022)</th>
<th>$^\text{GSPC}$</th>
<th>$^\text{DJI}$</th>
<th>$^\text{IXIC}$</th>
<th>$^\text{RUT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>124</td>
<td>Jul 01</td>
<td>-4.73</td>
<td>-3.61</td>
<td>-6.62</td>
<td>-4.08</td>
</tr>
<tr>
<td>125</td>
<td>Jul 05</td>
<td>-4.92</td>
<td>-4.36</td>
<td>-4.91</td>
<td>-4.99</td>
</tr>
<tr>
<td>131</td>
<td>Jul 13</td>
<td>-10.84</td>
<td>-8.89</td>
<td>-11.37</td>
<td>-11.58</td>
</tr>
<tr>
<td>161</td>
<td>Aug 24</td>
<td>9.73</td>
<td>8.82</td>
<td>13.05</td>
<td>13.92</td>
</tr>
<tr>
<td>164</td>
<td>Aug 29</td>
<td>3.62</td>
<td>3.33</td>
<td>4.92</td>
<td>6.64</td>
</tr>
<tr>
<td>167</td>
<td>Sep 01</td>
<td>7.17</td>
<td>6.32</td>
<td>8.7</td>
<td>7.07</td>
</tr>
<tr>
<td>170</td>
<td>Sep 07</td>
<td>7.65</td>
<td>6.18</td>
<td>8.64</td>
<td>7.69</td>
</tr>
<tr>
<td>175</td>
<td>Sep 14</td>
<td>5.22</td>
<td>3.8</td>
<td>4.86</td>
<td>7.51</td>
</tr>
<tr>
<td>185</td>
<td>Sep 28</td>
<td>-3.3</td>
<td>-4.9</td>
<td>-3.48</td>
<td>-6.72</td>
</tr>
<tr>
<td>196</td>
<td>Oct 13</td>
<td>-8.66</td>
<td>-10.78</td>
<td>-7.27</td>
<td>-9.72</td>
</tr>
<tr>
<td>200</td>
<td>Oct 19</td>
<td>-5.6</td>
<td>-8.28</td>
<td>-3.05</td>
<td>-6.2</td>
</tr>
<tr>
<td>201</td>
<td>Oct 20</td>
<td>-5.24</td>
<td>-7.96</td>
<td>-1.89</td>
<td>-6.79</td>
</tr>
<tr>
<td>202</td>
<td>Oct 21</td>
<td>-4.22</td>
<td>-6.72</td>
<td>-0.84</td>
<td>-5.76</td>
</tr>
<tr>
<td>210</td>
<td>Nov 02</td>
<td>-7.81</td>
<td>-4.6</td>
<td>-10.14</td>
<td>-5.65</td>
</tr>
<tr>
<td>216</td>
<td>Nov 10</td>
<td>0.18</td>
<td>3.83</td>
<td>-2.12</td>
<td>5.08</td>
</tr>
<tr>
<td>217</td>
<td>Nov 11</td>
<td>1.54</td>
<td>4.17</td>
<td>0.22</td>
<td>5.59</td>
</tr>
<tr>
<td>220</td>
<td>Nov 16</td>
<td>-0.34</td>
<td>3.31</td>
<td>-2.4</td>
<td>2.07</td>
</tr>
<tr>
<td>221</td>
<td>Nov 17</td>
<td>-0.49</td>
<td>3.84</td>
<td>-2.73</td>
<td>1.55</td>
</tr>
<tr>
<td>224</td>
<td>Nov 22</td>
<td>6.78</td>
<td>10.42</td>
<td>4.08</td>
<td>8.33</td>
</tr>
<tr>
<td>225</td>
<td>Nov 23</td>
<td>7.61</td>
<td>10.62</td>
<td>5.68</td>
<td>8.77</td>
</tr>
<tr>
<td>244</td>
<td>Dec 21</td>
<td>3.48</td>
<td>2.22</td>
<td>5.75</td>
<td>0.62</td>
</tr>
<tr>
<td>249</td>
<td>Dec 29</td>
<td>-4.14</td>
<td>-2.91</td>
<td>-6.95</td>
<td>-8.35</td>
</tr>
<tr>
<td>250</td>
<td>Dec 30</td>
<td>-5.49</td>
<td>-3.63</td>
<td>-8.89</td>
<td>-10.27</td>
</tr>
</tbody>
</table>

**Table 4.6** Prediction Error for GBM on Index Funds (2022)

This is summarized in the following table:

<table>
<thead>
<tr>
<th></th>
<th>$^\text{GSPC}$</th>
<th>$^\text{DJI}$</th>
<th>$^\text{IXIC}$</th>
<th>$^\text{RUT}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (%)</td>
<td>-0.25</td>
<td>0.05</td>
<td>-0.8</td>
<td>-0.3</td>
</tr>
<tr>
<td>Standard Deviation (%)</td>
<td>6.5</td>
<td>6.06</td>
<td>7.8</td>
<td>7.48</td>
</tr>
</tbody>
</table>

**Table 4.7** Summary of Prediction Error for GBM on Index Funds (2022)
4.6.5 Distribution of Stock’s Prediction Errors (from $I_{\text{pre}}$ and $I_{\text{post}}$)

We see that, on average, the prediction errors have a negative expected value. We recall the equation for prediction error $p_t$:

$$p_t = \frac{S_{t+14} - X_{t+14}}{X_{t+14}}.$$

If this value is positive, then we have overestimated the asset’s value. If the value is negative, then we have underestimated the asset’s value. Finally, if the value is zero, then we have perfectly predicted the asset’s value.

Since we, on average, underestimate the gains, this means that if we were to buy and sell these stocks during this time period with this model, we would likely have made positive gains.

Here, we show the distribution of our prediction errors for each of our index funds for both intervals $I_{\text{pre}}$ and $I_{\text{post}}$:

**Figure 4.7** GSPC Prediction Error Distributions for $I_{\text{pre}}$ and $I_{\text{post}}

**Figure 4.8** DJI Prediction Error Distributions for $I_{\text{pre}}$ and $I_{\text{post}}
In each of our figures, the top graph is a histogram of the prediction errors for $I_{\text{pre}}$ and the bottom graph is a histogram of the prediction errors for $I_{\text{post}}$ for each respective stock.

We remind ourselves here that the prediction errors of a stock act as a representation of the accuracy of our models, where a prediction error of 0% signifies the model is 100% accurate for a given stock and date. Thus, if the model’s accuracy has changed from $I_{\text{pre}}$ to $I_{\text{post}}$, then the distributions of the prediction errors from $I_{\text{pre}}$ to $I_{\text{post}}$ would also change.

### 4.6.6 Statistical Test on Prediction Error Distributions

Now, we are curious whether there is a significant difference in the difference between these two samples (the prediction error $p_t$ for $t \in I_{\text{pre}}$ and the prediction error $p_{t'}$ for $t' \in I_{\text{post}}$, where $t$ and $t'$ are the $i^{th}$ day that the stock market is open in given interval).

A paired t-test is a statistical test used to determine whether there is a significant difference between the means of two related samples. It is often used when the same group of individuals is tested under two different conditions, such as, in our case, before and after the pandemic. The paired t-test calculates the difference between each pair of values in the two samples and then determines whether the mean difference is significantly different from zero using a t-distribution. The test is called “paired” because the data points in each sample are paired with corresponding data points in the other sample.
Hence, we can pair a values from each sample (i.e., dates from 2018 and 2022) based off of the days the stock market is open for in each given year. We denote the difference of our pairs 

\[ d_i = (p_{t_i}') - (p_{t_i}) \]

for all 50 pairs of values \( i \) (where the \( i^{th} \) date of each interval can be found in Tables 4.4 and 4.6).

Since we are ultimately concerned with the difference between two measures in one sample, the paired t-test reduces to the one sample t-test where each value \( d_i \) in the one sample t-test is the difference of the 50 overall paired values.

For each stock, let us assume the following null and alternate hypothesis:

- \( H_0 \): The average difference in predicted error, for each stock, is 0 from \( I_{pre} \) to \( I_{post} \) (\( \mu_{d_{stock}} = 0 \)).

- \( H_A \): The average difference in the predicted error, for each stock, is \( \neq \) 0 from \( I_{pre} \) to \( I_{post} \) (\( \mu_{d_{stock}} \neq 0 \)).

The criterion for the decision is defined through the level of significance, which we define to be \( \alpha = .05 \) (a commonly chosen value for this type of statistical test). Note that our critical value of t is based on this (where \( df = n - 1 \)):

\[ t_{(df=49, \alpha=0.05)} = \pm 2.009575 \]

We get the following prediction error difference:
Table 4.8  Difference of Prediction Errors (2022 - 2018)

Which is summarized below:

Table 4.9  Summary of Difference of Prediction Errors (2022 - 2018)
Let $\bar{d}_{stock}$ be the mean difference and standard error $\frac{s_{d_{stock}}}{\sqrt{n}}$ (where $s_{d_{stock}}$ is the standard deviation of the differences) for the given stock. We can calculate the t-statistic below (note that if the absolute value of the calculated t statistic is greater than the critical value of t, we reject the null hypothesis).

$$t_{stock} = \frac{\bar{d}_{stock} - \mu_{d_{stock}}}{(s_{d_{stock}} / \sqrt{n})}$$

This yields

<table>
<thead>
<tr>
<th></th>
<th>^GSPC</th>
<th>^DJI</th>
<th>^IXIC</th>
<th>^RUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-score</td>
<td>0.36</td>
<td>0.49</td>
<td>0.08</td>
<td>0.63</td>
</tr>
<tr>
<td>Reject $H_0$?</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 4.10  \text{t-statistic for each stock}

Hence, for all of our stocks, we fail to reject the null hypothesis. Since we were using these index funds as a way of representing the stock market, we find that there is not a significant difference between the accuracy of the model at predicting stock prices pre- and post- pandemic! It is notable that the distributions for the interval $I_{post}$ appear to be less normally distributed than the distributions for the interval $I_{pre}$ – specifically, the histograms for $I_{post}$ appear to be flatter and have a subtle bimodal shape, though we need more data to say anything stronger than the bolded statement.
Chapter 5

Future Work

5.1 Improvements to Model

In order to improve our geometric Brownian motion (GBM) model, there are several avenues that we could explore. An area that could be investigated is the calibration of the model. Currently, we have assumed that the parameters of the model (i.e., the drift and volatility) are fixed and can be found for a sufficiently small time period. However, in reality, these parameters may fluctuate or be different than what we computed. By coming up with different, more accurate ways of finding these parameters, we could obtain more accurate estimates of the model, and thus improve the model’s performance.

Another area that could be explored is the use of more advanced stochastic processes. Although GBM is a widely used model in finance, it has some limitations. For example, it assumes that the asset price follows a lognormal distribution, which may not be appropriate for all types of financial assets. Other models, such as the Black-Scholes model, incorporate different types of stochastic processes and may be more appropriate for certain types of assets or financial instruments (such as options). By exploring different models, we could improve our understanding of how different factors affect financial markets and develop more accurate pricing models.

Furthermore, we could investigate the use of more sophisticated techniques to predict financial returns. Machine learning algorithms, such as neural networks and random forests, have been shown to be much more accurate at predicting asset prices and returns than these simpler stochastic
models. By incorporating these techniques into our analysis, we could potentially improve the accuracy of our predictions and better capture the complex relationships between different factors and financial returns.

Finally, we could explore the use of alternative data sources in our analysis. In recent years, there has been a growing trend towards the use of alternative data sources, such as social media data and even satellite imagery, in financial analysis. By incorporating these data sources into our analysis, we could potentially uncover new insights into financial markets and improve the accuracy of our predictions.

In general, there are many avenues that could be explored to improve our GBM model and our understanding of financial markets in a more general way. While our current model provides a good starting point for analyzing financial returns, there is always room for improvement and refinement.
5.2 Application of Model

Aside from simply determining whether the accuracy of the model changed before and after the pandemic, there are several other potential applications of our model.

One natural follow-up would be to determine the actual accuracy of the model. This would yield valuable insight as to how reliable actually using the model is.

Additionally, we could always use the model to develop a trading strategy. Using our model to make predictions about future stock prices, we could develop a trading strategy that seeks to exploit market inefficiencies and generate profits. Though we didn’t explore this in depth, it is an interesting avenue that would certainly fit into the scope of this paper.
5.3 Other Models

In this thesis, we used geometric Brownian motion (GBM) to model the stock prices. However, there are many other models that can be used to simulate stock prices. One of the most popular models is the Black-Scholes model, which is a mathematical model used to estimate the price of European-style options.

If we had more time to write this thesis, we could explore other models such as the stochastic volatility model, the jump diffusion model, and the Heston model. The stochastic volatility model takes into account the fact that volatility is not constant over time and varies stochastically. The jump diffusion model accounts for sudden jumps in stock prices that are not captured by the continuous-time models such as GBM. The Heston model is a stochastic volatility model that takes into account the fact that volatility is mean-reverting. Finally, we could look into the Black-Scholes model, as it is widely used and the building point for many other financial models.

Using different models would enable us to compare their accuracy in predicting stock prices and evaluate their suitability for different types of options. We could also explore the impact of different parameters on the accuracy of the models and investigate the implications of using different models for risk management strategies.

Overall, there is a vast amount of research that can be done in the area of financial modeling, and there are always new models and techniques being developed. Although we have focused on GBM in this thesis, there is still much more to explore in this field, and future research could build on the findings of this thesis and expand on them.
Chapter 6

Conclusion

In conclusion, this thesis explored the use of geometric Brownian motion (GBM) as a model for financial assets. We began by introducing the basics of stochastic calculus and presented the key features of GBM. We then demonstrated how GBM can be used to simulate asset prices and how to calibrate the model using real data.

Next, we applied the GBM model to the financial instrument ‘stock closing prices’ and presented how to use the model to price this instrument. We also discussed the limitations of the model and the possibility for using more advanced models, such as the Black-Scholes model.

We also presented an analysis of the performance of the GBM model using real data, showing that the model can capture the main characteristics of asset prices, such as volatility and mean reversion, but that it can also suffer from significant errors.

Finally, we discussed potential improvements to the GBM model and explored other applications of the model that we did not cover in this thesis, such as using the model in the development of a trading strategy.

Overall, this thesis provides a comprehensive introduction to GBM and its application to financial asset modeling. While GBM has limitations, it is a powerful and widely used model in finance, and this thesis provides a strong foundation for further research in the field.
Bibliography


