A RANDOM WALK INTO THE BLACK SCHOLES MODEL

ASHER JIANG

Abstract. A discussion of the mathematical prerequisites and economic intuition necessary to understand the Black-Scholes options pricing model. We begin with a foray into the construction of the Wiener Process, laying the groundwork for a mathematical model of stock price behavior derived from the Efficient Market Hypothesis. We then construct the famous Black-Scholes Model and assess its empirical accuracy by analyzing historical options trading data.

Contents

1. Introduction 1
2. Stochastic Calculus 2
2.1. Wiener Process 2
2.2. Ito's Lemma 3
3. The Stochastic Nature of Stock Prices 3
4. An Introduction to Options 4
5. Arbitrage Pricing Theory and Portfolio Replication 5
6. The Black-Scholes Model 6
7. Empirical Testing of the Black-Scholes Formula 8
7.1. A Survey of Large-Cap Stock Options 9
7.2. High vs. Low Volatility 10
7.3. The Super Micro Computer Inc. Anomaly 12
Acknowledgments 14
References 14

1. Introduction

In the 1960's, economists such as Eugene Fama and Paul Samuelson theorized that the seemingly random movements of stock prices could be well-modeled by tools from stochastic calculus. This culminated in the *Efficient Market Hypothesis*– a collection of postulates which state, to varying degrees of extremity, that all relevant information pertaining to the "fair" value of a stock is incorporated nearly instantly by the large number of traders of that stock. The result is that stock prices seem to wander randomly around their "fair" price. Then, in 1973, Fischer Black and Myron Scholes derived a formula for pricing derivatives by building off of this stochastic model of stock prices. The result was the *Black-Scholes Model*, a partial differential equation governing the relationship between the price of an option and its underlying asset. This model now forms the backbone of modern financial markets. Starting from a discussion of stochastic calculus, we will examine the mathematical and economic concepts required to understand the Black-Scholes Model, before providing an intuitive derivation of the famous Black-Scholes PDE. Finally, we will quantify the accuracy of the model in predicting real-life options prices in various markets by analyzing historical data from the NASDAQ and New York Stock Exchange.

2. Stochastic Calculus

Definition 2.1. (Markov Process). A sequence consisting of random variables $X_1, X_2, ...X_n$ is a Markov process if

$$\mathrm{P}[X_{n+1} = x \,|\, X_1 = x_1, X_2 = x_2, ..., X_n = x_n] = \mathrm{P}[X_{n+1} = x \,|\, X_n = x_n]. \tag{2.2}$$

An example of a discrete-time Markov process is a simple random walk on $\mathbb{Z}$. Let $\varepsilon_1, \varepsilon_2, ...\varepsilon_n$ be a sequence of independent random variables with probability distribution $\mathrm{P}[\varepsilon_i = -1] = \mathrm{P}[\varepsilon_i = 1] = \frac{1}{2}$. Define $X_0 = x_0$, and let $X_n = \sum_{i=0}^{n} \varepsilon_i$, so that ε_i represents the ith step of the walk and X_i is the location after i steps. The key observation is that, due to the independent nature of the ε_i, the value of X_{n+1} depends only on the value of X_n and ε_{n+1}. Thus X_n satisfies the properties of a Markov process.

2.1. **Wiener Process.**

The Wiener process is a special type of Markov process whose characteristics are more suited to modeling real-life phenomena such as stock prices.

In the following discussion, we write $X \sim \mathcal{N}(\mu, \sigma^2)$ to mean that the random variable X has density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}.$$

Definition 2.3. (Wiener Process). We say that Z_t follows a Wiener process if it satisfies the following criteria:

(1) Z_t is a Markov process.
(2) $Z_0 = 0$ almost surely.
(3) The function $t \to Z_t$ is almost surely continuous with probability 1.
(4) $(Z_{t+u} - Z_t) \sim \mathcal{N}(0, u)$.

Starting from a simple discrete-time model, we will develop the intuition for the continuous Wiener Process. Consider a random variable Z which satisfies the conditions for a Wiener process. Consider two discrete times, t_0 and $t_1 = t_0 + 1$. By property (4) above, we know that $(Z_{t_0+1} - Z_{t_0}) \sim \mathcal{N}(0, 1)$. Dividing the elapsed time in half, we get that the sum of the increments $(Z_{t_0+\frac{1}{2}} - Z_{t_0}) + (Z_{t_0+1} - Z_{t_0+\frac{1}{2}})$ must also satisfy $\mathcal{N}(0, 1)$. Assuming symmetry between time periods, we can apply the linearity of variance for independent random variables to find that $(Z_{t_0+\frac{1}{2}} - Z_{t_0})$ and $(Z_{t_0+1} - Z_{t_0+\frac{1}{2}})$ satisfy $\mathcal{N}(0, \frac{1}{2})$. The standard deviation of Z_t on each subinterval is $\sqrt{\frac{1}{2}}$, and thus, we can rewrite $\mathcal{N}(0, \frac{1}{2})$ as $\sqrt{\frac{1}{2}} * \mathcal{N}(0, 1)$. We can continue this process of subdividing each time interval into multiple smaller intervals. The key observation is that during a small time interval Δt, the change in Z, ΔZ, is equal to $\epsilon * \sqrt{\Delta t}$, where $\epsilon \sim \mathcal{N}(0, 1)$. Taking the limit of this process, we arrive at a model for continuous-time Wiener processes.

We now introduce a bit of informal notation, mirroring the standard calculus of real variables. We write dZ to mean the infinitesimal change in Z, i.e. ΔZ as

$\Delta t \to 0$. With this tool, we may now define a generalized version of the continuous-time Wiener process, the Ito process.

Definition 2.4. X is an Ito process if it satisfies

$$dX = a(X,t)dt + b(X,t)dZ \tag{2.5}$$

where Z follows a Wiener process.

We use this somewhat informal equation to mean that at a certain time t, X changes according to a deterministic component given by $a(X,t)$, and a stochastic/noise component given by $b(X,t)$. In the stochastic differential equation, the function $a(X,t)$ determines the *drift rate* of the Ito process, while the function $b(X,t)$ determines the magnitude of the stochastic component. Intuitively, over a small time interval Δt, we may treat a and b as effectively constant, so the expected change in X is $E[\Delta X] = a\Delta t$, and the variance of the change is $\mathrm{Var}[\Delta X] = b^2\Delta t$ (meaning that the standard deviation is $b\sqrt{\Delta t}$). The definition of an Ito process can be made formal, although we will not discuss the theory here, as the preceding informal discussion will be enough for our purposes. The curious reader may consult any of the standard texts on stochastic calculus.

2.2. **Ito's Lemma.** We now turn our attention to an indispensable tool in stochastic calculus, Ito's Lemma. It is analogous to the chain rule in basic calculus. Ito's Lemma gives an SDE for any process that is a function of another process given by an SDE.

Theorem 2.6. *(Ito's Lemma). Consider a function $f(X,t)$, where X follows an Ito process of the form*

$$dX = a(X,t)dt + b(X,t)dZ$$

where Z is a Wiener process. Then $Y = f(X,t)$ must also follow an Ito process of the form

$$dY = \left(\frac{\partial f}{\partial x}a + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}b^2\right)dt + \frac{\partial f}{\partial x}b\,dZ. \tag{2.7}$$

While a rigorous proof of Theorem 2.6 is outside of the scope of this paper, this result will prove pivotal in the derivation of the Black-Scholes Model.

3. The Stochastic Nature of Stock Prices

The Efficient Market Hypothesis (EMH) posits that in a reasonably liquid market, all available information about the value of a stock is almost instantly reflected in its price. The high volume of traders in such markets ensures that any opportunity to profit from a systematic discrepancy between a stock's intrinsic value and its market price is rapidly exploited. Consequently, the most relevant predictor of a stock's future price behavior is its current price, as any information embedded in its price history has already been assimilated by market participants. However, due to the inherent uncertainty of the real world, it is impossible to precisely determine a stock's intrinsic value, allowing for differing opinions on its "fair" price among market actors. Nevertheless, these varying predictions typically offset each other, resulting in the market price following a random walk around its intrinsic value. These properties suggest that a stock's price is well modeled by a stochastic process.

A reasonable mathematical model for the behavior of stock prices is an Ito process where the drift rate and magnitude of the stochastic component (known as *volatility*) are proportional to the current price of the stock. Intuitively, this makes sense because stocks themselves represent a fraction of a company's perceived value. Therefore, we expect the stochastic component of the stock price to be a fraction of the stock price, rather than a flat amount. If we let S_t be the price of a stock at time t, then its corresponding Ito process is

$$dS_t = \mu S_t dt + \sigma S_t dZ \tag{3.1}$$

where μ is the expected growth rate and σ is the volatility of the stock price. We say the distribution followed by S_t for any fixed t is lognormal.

4. An Introduction to Options

We will introduce the concept of options through a classic motivating example. Suppose a crude oil company "A" situated in the Middle East is deciding whether or not to send a vessel of petroleum to Europe, where they expect to make a profit off of the higher market price in the region. However, the trip takes many months, and the risk of the market price of petroleum decreasing during that time is significant. How can company A increase the chances of realizing a profit on your investment?

To solve this problem, modern markets make use of options. In the case of the crude oil company, an appropriate way to hedge their risk is to purchase a "put option." This is a contract with a buyer "B" which gives A the *right*, but *not the obligation*, to sell their shipment of petroleum at a certain *strike price*. By purchasing this contract, the crude oil company ensures that, should the market price of petroleum decrease when their shipment arrives, they will still be able to offload their shipment at the predetermined price. Alternatively, if the market price of petroleum happens to be above the strike price, the company can simply choose not to exercise the option, and make a greater profit by selling their shipment at market value.

To make this contract a fair deal, however, the option writer must be compensated for the risk that they are taking on. Therefore, the crude oil company pays an upfront cost, known as a *premium*, for the option. If the crude oil company decides not to exercise their option, then the option writer earns a profit from the premium without paying any additional costs.

This concept of buying and selling options is widely used in the stock market today. Traders use options to hedge against risks posed by assets and make speculative bets on the future movement of stock prices. Put options are primarily purchased when the buyer believes that the underlying stock's price will go down. Its counterpart, the *call option*, is purchased when the buyer believes that the stock price will go up.

Call and put options fall under a broader classification of contracts known as *derivatives*. These contracts derive their value from the value of an *underlying asset*, which is often a stock. In the previous example, the underlying asset was crude oil. There are four positions that a trader can take using call and put options: They can buy a call option, buy a put option, sell a call option, or sell a put option. The profit from each position is shown below.

Proposition 4.1. *Payouts of Options. In the following scenarios, let K be the strike price of an option, T be the time of expiration, P be the premium, and S_T be the price of the underlying asset at time T.*

(1) The return from buying a call option is $\max[0, (S_T - K) - P]$.
(2) The return from buying a put option is $\max[0, (K - S_T) - P]$.
(3) The return from selling a call option is $\min[0, P - (S_T - K)]$.
(4) The return from selling a put option is $\min[0, P - (K - S_T)]$.

An important distinction to make here is between American options and European options. American options allow the buyer to exercise the option at any point on or before the expiration date, while European options can only be exercised on the expiration date. Despite the regional affiliations suggested by their names, both options are traded on stock exchanges around the world. For the purposes of this paper, we assume all options to be European.

From the formulae above, it is clear that the price of the underlying stock throughout an option's lifetime plays a role in determining the value of the option. We would like to express the value of the option as a function of the underlying stock, which itself is a stochastic process. Our framework for doing this will be Arbitrage Pricing Theory and portfolio replication.

5. Arbitrage Pricing Theory and Portfolio Replication

The modern financial market operates off of a concept known as Arbitrage Pricing Theory. Arbitrage Pricing Theory states that two investments which produce the same returns regardless of future developments in the market must require the same initial investment. If there is a price mismatch, then traders could sell the more expensive investment and purchase the less expensive one to reach a market-neutral position while making a riskless profit. An example of this is the *put-call parity*, which dictates the price of a put option based on a call option with the same parameters.

Consider a world with two financial assets: a stock and a fixed-rate bond. The stock's price at time t is S_t, dependent on various market factors. The bond's price at time t is B_t, determined solely by its interest rate r, which is fixed. We will assume that interest is continuously compounded, so that the value of the bond follows an exponential growth model. Bonds are an example of a *risk-free asset* because their return at a future time is guaranteed and independent of market factors. Once a bond has been written, its payout is essentially guaranteed. A common example of this is a US government treasury bill (T-bill).

We will set up two portfolios with the same payout. Portfolio 1 consists of a call option on S with expiration time T and strike price K, as well as a bond with a payout of K at time T. Portfolio 2 consists of a put option with the same parameters as the call option, and a share of stock S.

Since both options have a strike price of K, there are two disjoint scenarios which can occur at time T: either $S_T > K$ or $S_T \leq K$. If $S_T > K$, then the call option is exercised. Portfolio 1's return is $(S_T - K) + K = S_T$. The put option is not exercised, so portfolio 2's return is S_T, which is identical to portfolio 1.

If $S_T \leq K$, then the call option is not exercised, while the put option is. The payout of portfolio 1 is K, while portfolio 2's payout is $(K - S_T) + S_T = K$.

Since portfolio 1 and 2 have the same payout regardless of circumstance, Arbitrage Pricing Theory dictates that both portfolios must require the same initial investment. The cost of portfolio 1 is the cost of the call option, c, in addition to the price of the fixed-rate bond, B_0. For a bond with interest rate r to have a payoff of K at time T, its value at $t = 0$ must be $\frac{K}{e^{rT}} = Ke^{-rT}$. The cost of portfolio 2 is the cost of the put option, p, and the cost of the stock at $t = 0$, S_0. Equating the two gives us the famous put-call parity equation:

$$c + Ke^{-rT} = p + S_0 \tag{5.1}$$

The method of analysis used in deriving the put-call parity is known as *portfolio replication*. More generally, given an asset whose no-arbitrage price is hard to determine, we attempt to construct a replication portfolio with the same financial consequences as the asset. If the replication portfolio's price can be found, then we can find the price of the asset using Arbitrage Pricing Theory. As we shall now see, portfolio replication is a powerful method of analysis when pricing options in reasonably efficient markets.

6. The Black-Scholes Model

The Black-Scholes Model provides a theoretical framework for determining the price of an option on an underlying asset. The model is derived by constructing a replicating portfolio which fluctuates with the price of the underlying asset. By continuously rebalancing the portfolio's holdings in response to the current value of the asset, we can ensure that the portfolio's value consistently matches the value of the option. Consequently, if we can determine the initial investment needed to establish this replicating portfolio, we can accurately calculate the option's price.

As in the example of put-call parity, we assume a simplified version of the market with only two financial assets, a stock and a bond, in addition to the option whose price we are attempting to find. We assume that the bond is a risk-free asset. Some additional assumptions are presented below. We will then state the theorem and then present a proof.

Assumptions 6.1. (Conditions for the Black-Scholes PDE)

(1) The price of the underlying stock follows the lognormal distribution presented in (4.1).
(2) The short-term risk-free interest rate r is known and remains constant.
(3) The volatility σ of the underlying stock remains constant throughout the lifetime of the option.
(4) The market is frictionless, that is, there are no transaction costs and transactions are instantaneous.
(5) The stock pays no dividends.
(6) Short selling is permitted without requiring collateral.
(7) The purchase of fractional shares and derivatives is possible.

Theorem 6.1. *(Black-Scholes PDE). Let $f(S, t)$ be the price of a derivative, and S_t be the price of its underlying asset. Let r be the constant risk-free rate and σ be the volatility of the underlying asset. Then the price of the derivative must obey*

$$rf = \frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} \tag{6.2}$$

Proof. From the assumptions above, as well as Section 3, we know that the price of the stock obeys

$$S_t = \mu S_t dt + \sigma S_t dZ \tag{6.3}$$

and the price of the bond satisfies the exponential growth model given by

$$dB_t = rB_t dt \tag{6.4}$$

The payout of a call option with strike price K and expiration T is

$$\max(0, S_T - K). \tag{6.5}$$

We would like to construct a replication portfolio with a known initial investment which has the same payout as the call option. We approach this by considering a portfolio with a continuously changing amount of bonds and stocks. Let a_t denote the amount of stock and b_t denote the amount of bonds in the replication portfolio at time t, so that the total value of the portfolio is

$$\Pi(t) = a_t S_t + b_t B_t. \tag{6.6}$$

Since the call option requires only an initial investment in the form of its premium, our replication portfolio must also not require additional investments throughout its lifetime. Therefore, any purchases of stocks or bonds throughout the lifetime of the portfolio must be offset by liquidation of the other asset. This is known as the *self-financing condition*, which can be expressed as

$$d\Pi = a_t dS_t + b_t dB_t. \tag{6.7}$$

At time T, we want the value in the portfolio, $\Pi(T)$ to equal the payout of the call option, $\max(0, S_T - K)$. Now we want to find functions which satisfy the constraints listed above. Since the bond price model is deterministic, the only stochastic source in the value of the portfolio is from the dZ term in the stock price model. Thus, we may assume that $\Pi(t)$ can be represented by a function $f(S_t, t)$. Substituting (6.3) and (6.4) into (6.7) gives

$$d\Pi(t) = a_t S_t[\mu dt + \sigma dZ] + b_t r B_t dt = [a_t \mu S_t + b_t r B_t]dt + a_t \sigma S_t dZ. \tag{6.8}$$

Now, we can find another SDE relating f to its derivatives using Ito's Lemma. This will allow us to solve for a_t and b_t. Applying Ito's Lemma shows that $\Pi(t) = f(S_t, t)$ must satisfy

$$d\Pi = \left(\frac{\partial f}{\partial S}\mu S_t + \frac{df}{dt} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S_t^2\right) dt + \frac{\partial f}{\partial S}\sigma S_t \, dZ. \tag{6.9}$$

for all t. We can match the coefficients for the dt and dZ terms to find expressions for a_t and b_t. From the dZ term, we find that

$$a_t = \frac{\partial f}{\partial S} \tag{6.10}$$

and the dt term gives

$$\mu S_t \frac{\partial f}{\partial S} + r b_t B_t = \mu S_t \frac{\partial f}{\partial S} + \frac{df}{dt} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S_t^2 \tag{6.11}$$

Solving for b_t in (6.11), we get

$$b_t = \frac{1}{rB_t}\left(\frac{df}{dt} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S_t^2\right) \tag{6.12}$$

Finally, we substitute (6.10) and (6.12) into (6.3) to get

$$f(S_t, t) = \Pi(t) = \frac{\partial f}{\partial S} S_t + \frac{1}{rB_t}\left(\frac{df}{dt} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S_t^2\right) B_t. \tag{6.13}$$

Canceling out the B_t term on the RHS gives

$$f(S,t) = \frac{\partial f}{\partial S} S_t + \frac{1}{r}\left(\frac{df}{dt} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S_t^2\right) \tag{6.14}$$

which simplfies to the *Black-Scholes PDE*:

$$rf = \frac{\partial f}{\partial t} + rS\frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} \tag{6.15}$$

□

Using the boundary conditions in Proposition 4.1, we can find unique solutions to the Black-Scholes PDE corresponding to the no-arbitrage price of a call or put option. We present the formulae below, omitting the detailed solution process, which is beyond the scope of this paper. The interested reader may refer to external resources for the general solution techniques of PDEs.

In the following theorem, we use $\Phi(x)$ to denote the cumulative distribution function of the standard normal distribution given by

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,dt.$$

Theorem 6.16. *(Black-Scholes(BSM) Formula). Let $\Phi(x)$ denote the CDF of the normal distribution. If S_0 is the current price of a stock, K is the strike price, T is the expiration date, and r is the risk-free rate, then the no-arbitrage price of a European call option is*

$$C = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2) \tag{6.17}$$

and the price of a European put option is

$$P = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1) \tag{6.18}$$

where

$$d_1 = \frac{\ln\left(\frac{S_0}{K}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \qquad d_2 = d_1 - \sigma\sqrt{T}$$

7. Empirical Testing of the Black-Scholes Formula

Our derivation of the Black-Scholes Model involved making a variety of assumptions and simplifications. In the real world, there are rarely such ideal conditions. We would like to quantify the accuracy of our model, and analyze its shortcomings.

To test our model empirically, we implemented the Black-Scholes Formula using a computer program. Our dataset is a list of approximately 15 million options which were listed on the New York Stock Exchange (NYSE) and NASDAQ from 2021 until 2024. For each option, we have the corresponding parameters needed to calculate the input variables of the Black-Scholes formula, including strike price, expiration date, current stock price, and the risk free rate, as well as several metrics for historical volatility.

For each option, the program calculates the theoretical price given by the Black-Scholes Formula and compares it to the actual bid and ask prices of the option listing. The results from our tests on several subsets of the data are presented below.

7.1. **A Survey of Large-Cap Stock Options.** We first examine options on the top 500 stocks by daily trading volume. The parameters we used for filtering the data are:

(1) Option listed between 1 January 2021 and 5 July 2024.
(2) Daily trading volume of underlying stock $\geq$ 10.5 million units per day, which filters down to the approximately 500 largest stocks by trade volume.
(3) Option premium $\geq$ \$5.00. This eliminates options whose intrinsic value is too low for the BSM formula to reasonably calculate, given the precision of the data.

We calculate the percent error of the Black Scholes Model's prediction for each option. The summary stats are presented below, along with a histogram to visualize the data.

# data points	1.011446e+06
mean	-0.2330984
std	8.791468
min	-127.7954
Q1	-4.448919
median	0.4942469
Q3	3.483092
max	199.8684

Table 1. Summary statistics of the percent error between the BSM Formula's prediction and market price of options for large-cap stocks.

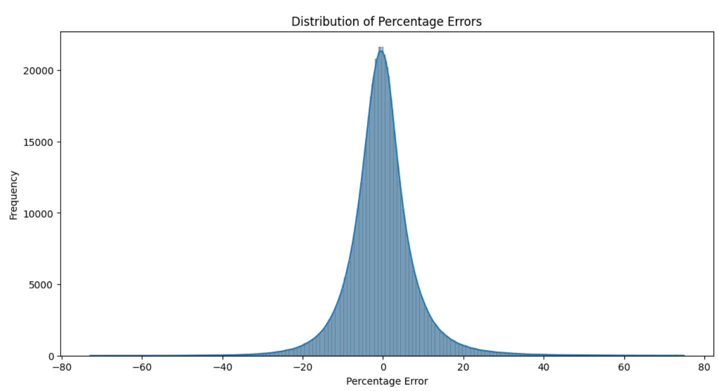

Figure 1. Distribution curve of the percent error between the BSM Formula's prediction and market price of options for large-cap stocks.

At first glance, the results are extremely promising. The mean error of the model is less than 1 percent, and the distribution curve formed by the percent error plot is almost perfectly symmetrical. This suggests that the model itself is good at representing the behavior of the market. Strangely, the mean and median, although both less than a percent away from 0, fall on different sides. This prompts us to

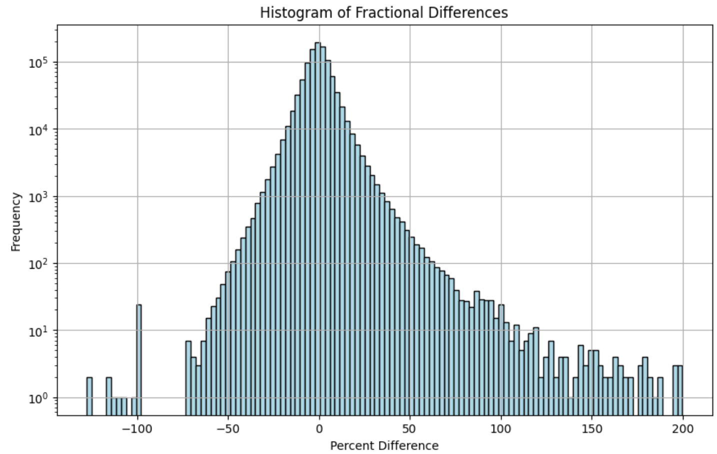

FIGURE 2. Log-scaled distribution curve of the percent error between the BSM Formula's prediction and market price of options for large-cap stocks.

investigate the tail behavior of the distribution graph. Examining the logarithmically scaled distribution curve shown in Figure 2 reveals that the data points are skewed slightly to the right, suggesting that the model tends to price options higher than the market when it deviates significantly from the real-life data.

It is very difficult to pinpoint a reason for this skew. A possible explanation of the skew is that some traders will use options to speculate on future market movements, relying on factors not reflected in current market conditions. These riskier speculations may be more frequent in markets during times of high volatility, when stock prices have more potential to chart massive gains or losses. If the factors that the traders are relying on turn out to be biased or otherwise incorrect, it can lead to many options becoming near-worthless. The buyer of these options might be tempted to resell them on the exchange in hopes of salvaging a portion of the option premium.

While this prediction is hard to ascertain, it does lead us to wonder if the market tends to deviate more from the Black-Scholes Model for stocks with higher volatility.

7.2. **High vs. Low Volatility.** We now take the dataset from Trial 1 and stratify it by volatility. The high volatility subgroup consists of the top quartile of options from the previous dataset by volatility, while the low volatility subgroup consists of the bottom quartile. Figure 3 shows log-scaled distribution curves of both subgroups overlaid on each other.

A visual inspection of the distribution graph reveals that the high volatility subgroup is much more skewed than the low volatility subgroup, which matches our expectations. The skewness of the high volatility subgroup is 6.03, while the skewness of the low volatility subgroup is -1.33. The summary stats for the high volatility subgroup suggests that traders tend to deviate more from the Black-Scholes Model during times of high volatility. This is not surprising, since periods of high volatility tend to encompass major events for the underlying stock, such as quarterly earnings announcements or a change of leadership in the company.

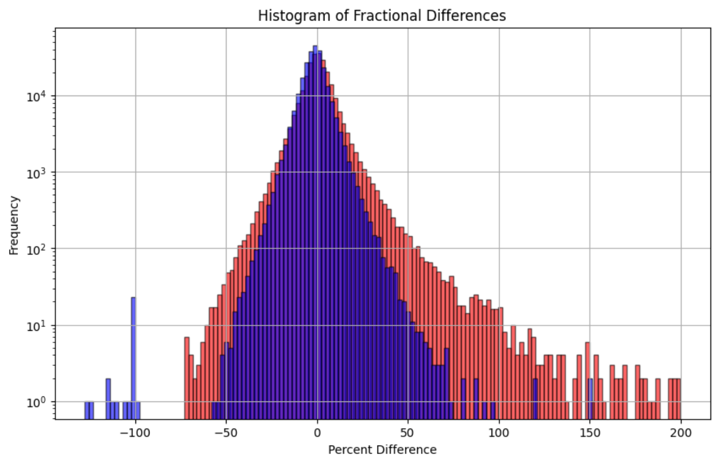

FIGURE 3. Log-scaled distribution curve of the percent error between BSM Model's prediction and market prices. The options from the low volatility group are in blue and the options from the high volatility group are in red.

# data points	252840
mean	0.870750
std	11.071532
min	-72.945531
Q1	-4.120367
median	0.305213
Q3	4.928789
max	199.868386

TABLE 2. Summary Stats: Percent error between BSM prediction and market in high volatility group.

What is more surprising is the trend revealed by the summary stats of the low volatility subgroup. We can see from Table 3 that the mean and median of the group has shifted significantly to the left compared to the full dataset and the high volatility subgroup. This implies that traders tend to transact at a higher price relative to the Black-Scholes Model's prediction during periods of lower volatility. This shift in the dataset is apparent in Figure 4, which shows linearly scaled distribution curves of the two subgroups.

# data points	252748
mean	-1.222670
std	7.514144
min	-127.795441
Q1	-5.018839
median	-1.236977
Q3	2.296168
max	151.454063

Table 3. Summary Stats: Percent error between BSM prediction and market in low volatility group.

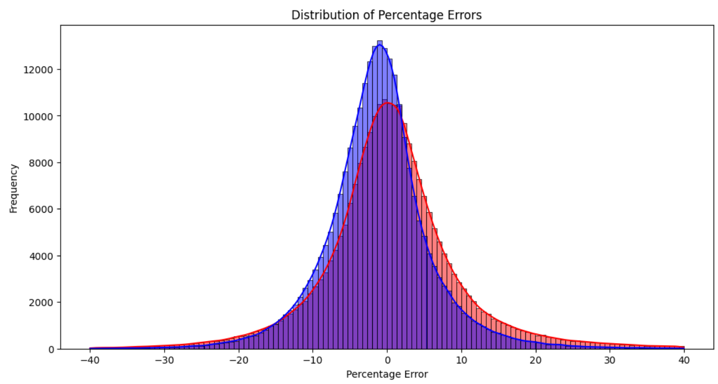

Figure 4. Distribution curve of percent difference between BSM model prediction and market in high and low volatility subgroups.

7.3. **The Super Micro Computer Inc. Anomaly.** We now turn our attention to an anomaly in the recent large-cap options market. The parameters for our data are presented below:

(1) Option listing dates between 28 December 2023 and 5 July 2024.
(2) Daily trading volume of underlying $\geq$ 5.05 million shares.
(3) Option premium $\geq$ \$5.00.

The summary stats of the percent error of the model's prediction is shown below, along with a histogram.

# data points	317981
mean	0.811951
std	8.639830
min	-72.438899
Q1	-3.164326
median	0.265585
Q3	3.861148
max	180.210774

Table 4. Summary Stats: Percent error between BSM Model's prediction and market price for options.

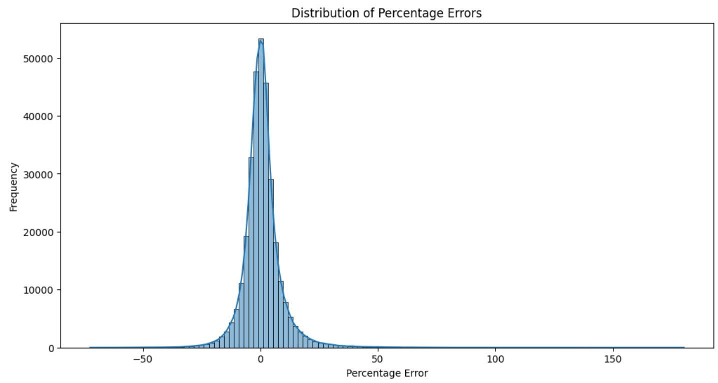

FIGURE 5. Distribution curve of percent error between BSM Model prediction and market price of options.

At a glance, it appears that the BSM formula has predicted the real-life options quite well. However, when we plot the actual price of each option against the corresponding prediction of the BSM formula, we find an anomaly.

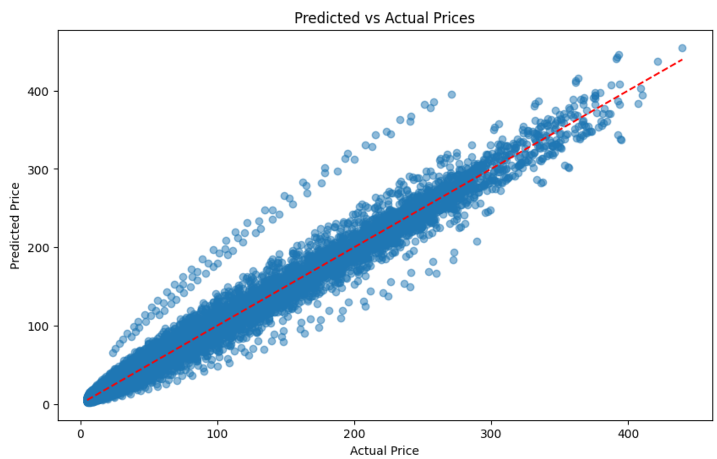

FIGURE 6. Scatter plot of BSM model's prediction vs. market price of options. Line of best fit: $y = 1.00x + 0.14$. Coefficient of determination (R^2): 0.990

Despite a near-perfect one-to-one correlation and a R^2 value of 0.99, there is a distinct set of points which fall significantly above the line, indicating a trend of options being sold for far less than their theoretical value. Extracting the individual data points reveals that out of the 81 outliers, 67 were options on SMCI (Super Micro Computer Inc.), with AVGO (Broadcom), NVDA (Nvidia), and ADBE (Adobe) making up the remaining 14 points. Moreover, 60 out of the 67 SMCI options were calls.

A moment's reflection on macroeconomic trends during the time period provides vital context for the anomaly. All four stocks benefited greatly from massive developments in the artificial intelligence sector, which led to unprecedented increases for the stock price of many established companies within the industry. For example, SMCI's stock price rose fourfold in the first three months of 2024. However, the speculative nature of the "AI bubble" also meant that traders may have deviated from standard metrics for pricing options, instead opting to make risky bets on the continuation of the stocks' historic runs. Along with the period of sell-off that SMCI experienced in the second quarter of 2024, an increase in speculation could provide a plausible reason for the significant number of call options which were selling for well below their theoretical value.

This case study highlights one of the major shortcomings of the Black-Scholes Model: an inability to factor in macroeconomic trends such as disruptive technological advancements. It also provides some context for the results of Section 7.2, where we observed a tendency for the distribution curve to skew right for high-volatility stock options.

Our investigation into the empirical accuracy of the Black-Scholes Model highlights both the remarkable consistency of the model and some of its limitations. In the first trial, the model demonstrated a strong alignment with market prices for options on large-cap stocks, suggesting that the Black-Scholes Model performs well for highly liquid and relatively stable markets. In Section 7.2, we found that the model fails to capture a tendency for markets to price options lower than their theoretical value in high-volatility environments. We also observed a tendency for the model to systematically undervalue options in low-volatility environments relative to the market. This demonstrates the model's shortcomings in accounting for speculative trading or major market events, which is underscored by the findings in Section 7.3. Our results suggest that while there may be room for improvement in the Black-Scholes Model when dealing with volatile market conditions, the model provides a stable foundation for understanding the pricing of options in reasonably efficient markets.

Acknowledgments

It is my great pleasure to thank my advisor, Max Ovsiankin, for going above and beyond in his support of every aspect of my project. Max provided invaluable guidance as I explored potential research topics and meticulous feedback throughout the process of writing my paper. I'm deeply grateful for his dedication to this project and my growth as a researcher.

I would also like to thank Professor Daniil Rudenko for his captivating lectures on advanced mathematics, as well as Professor Greg Lawler for his talks on stochastic processes. Finally, I would like to thank Professor Peter May for organizing the UChicago Math REU program which has provided me, along with many other students, the opportunity to explore the fascinating world of mathematics.

References

[1] Eugene F. Fama. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, Vol. 25, No. 2, pp. 383-417, 1970. DOI: 10.2307/3003046

[2] Eugene F. Fama. Random Walks in Stock Market Prices. Financial Analysts Journal, vol. 21, no. 5, 1965, pp. 55–59.

[3] Fischer Black and Myron Scholes. The Pricing of Options and Corporate Liabilities. Journal of Political Economy, vol. 81, no. 3, 1973. DOI: 10.1086/260062.

[4] Gregory F. Lawler. Random Walk and the Heat Equation. University of Chicago, 2002.

[5] J. Michael Steele. Stochastic Calculus and Financial Applications. Springer, 2001.

[6] John C. Hull. Options, Futures, and Other Derivatives, 10th edition. Pearson, 2017.
[7] Steven P. Lalley. Brownian Motion. University of Chicago, 2018
[8] Wiener Process. Wikipedia. https://en.wikipedia.org/wiki/Wiener_process.
[9] Options Database, Dolthub (2024). Available at https://www.dolthub.com/repositories/post-no-preference/options.
[10] Stocks Database, Dolthub (2024). Available at https://www.dolthub.com/repositories/post-no-preference/stocks.
[11] Federal Interest Rates Database, Dolthub (2024). Available at https://www.dolthub.com/repositories/post-no-preference/rates.