STOCHASTIC CALCULUS AND ARBITRAGE-FREE OPTIONS PRICING

ARI WILSON

Abstract. The purpose of this paper is to derive the Black-Scholes-Merton differential equation and formula for pricing options. In order to derive these equations, we must first understand stochastic calculus. Stochastic calculus is essential for the pricing of options using the Black-Scholes-Merton model, since while traditional calculus is sufficient for differentiable functions, random processes are inherently non-differentiable. Therefore, we must extend our understanding of calculus to stochastic calculus in order to operate on these non-differentiable functions. One useful method of doing this is by using Itô calculus. In order to understand Itô calculus, we must first understand Brownian motion. In order to understand Brownian motion, we must first understand random walks and stochastic processes. By the end of this paper, one will posses an in-depth understanding of these concepts through a combination of intuition, theory, and application.

Contents

1. Introduction to Measure Theory and Probability 1
2. Stochastic Processes 3
3. Random Walks and Brownian Motion 4
4. Ito Calculus 6
5. Black-Scholes Equation and Formula 9
Acknowledgments 14
References 14

1. Introduction to Measure Theory and Probability

Definition 1.1. An *outcome*, ω is defined as the result of a random experiment.

Definition 1.2. The *probability* that a given outcome occurs is defined as the fraction of times that we observe that outcome over a large amount of realized outcomes.

Definition 1.3. An event, E can be thought of as the set of outcomes that we want to assign a probability to.

Definition 1.4. A σ*-algebra*, $\mathcal{F}$, on a set, Ω, is defined as a collection of subsets of Ω that satisfy:
•$\Omega \in \mathcal{F}$.
•If $S \in \mathcal{F}$, then $\Omega \setminus S \in \mathcal{F}$.
•If $S_1, S_2, ... \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} S_i \in \mathcal{F}$

A σ-algebra can be thought of as the amount of available information. An event belongs to a σ-algebra if and only if we can determine whether a given outcome belongs to the event. It follows that given an outcome, we can determine whether an event in our σ-algebra has occurred.

Definition 1.5. A *measure* can intuitively be thought of as the way to assign a size to subsets of a set. Something that is measure zero can be thought to be of neglible size. Something with intermediate measure can be thought to have positive measure (non-neglible size) whose complement also has positive measure (non-neglible size). Something with full measure can be thought of as something with positive measure whose complement is measure zero.

Definition 1.6. A *measure space* is an object containing a set, a σ-algebra, and a measure.

Definition 1.7. A *probability space* is a measure space where the measure of the entire space is equal to one. A probability space can be represented by $(\Omega, \mathcal{F}, \mathbb{P})$ where:

- Ω represents a sample space (non-empty set of all outcomes).
- $\mathbb{P}$ represents a probability (measure on $(\Omega, \mathcal{F})$) which can be intuitively thought of as the chance that an event, H occurs. Note that $\mathbb{P}(\Omega) = 1$ and $P(\emptyset) = 0$.
- $\mathcal{F}$ represents a σ-algebra on the sample space (the collection of events, that we care about)

Definition 1.8. A *random variable* is a measurable function $X : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$ such that for every Borel set, B, $X^{-1}(\mathbb{B}) = \{X \in B\} \in \mathcal{F}$. If this is the case, then we will refer to X as $\mathcal{F}$-measurable.

Let $\sigma(X)$ be the smallest σ-algebra such that X is $\mathcal{F}$-measurable.

Note that $\{X \in B\}$ is shorthand for $\{\omega \in \Omega : X(\omega) \in B\}$

Note that a Borel set is any set that can be formed from open sets using countable union, countable intersection, and relative complement.

Definition 1.9. The *distribution* of a random variable is the function
$\mu_X(B) = \mathbb{P}\{X \in B\} = \mathbb{P}[X^{-1}(B)]$ that satisfies:

If μ_X assigns measure one to a countable set of reals, then X is a *discrete random variable.*

If μ_X assigns measure zero to every countable set, then X is a *continuous random variable.*

Definition 1.10. The *distribution function* of a random variable is defined by
$F_X(x) = \mathbb{P}\{X \le x\} = \mu_X(-\infty, x]$ where F_X satisfies:

•$\lim_{x\to-\infty} F(x) = 0$.

•$\lim_{x\to\infty} F(x) = 1$

•F is a nondecreasing function

•F is right continuous. Therefore, for every x, $F(x+) := \lim_{\epsilon\to 0} F(x+\epsilon) = F(x)$.

Definition 1.11. The *density* of a random variable is a function $f = f_X : \mathbb{R} \to [0, \infty\}$ such that $\mathbb{P}\{a \le X \le b\} = \int_a^b f(x)dx$

Definition 1.12. The *indicator function* of event E is the random variable

$$1_E(\omega) = \begin{cases} 1 & \omega \in E \\ 0 & \omega \notin E \end{cases}$$

Definition 1.13. A random variable has a *normal distribution* if it has density $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$ where μ represents the mean and σ^2 represents the variance.

Note that if $\mu = 0$ and $\sigma^2 = 1$, then X is said to have a *standard normal distribution.*

The distribution function of the standard normal distribution is denoted Φ where:
$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$

Definition 1.14. The *expectation* of a discrete random variable is $\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i)$

The expectation of a continuous random variable is:
$\mathbb{E}[X] = \int X d\mathbb{P}$

Definition 1.15. The *variance* of a random variable, is expressed:
$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
Note *variance* may also be expressed by: $\mathbb{E}[X^2] - \mathbb{E}[X]^2$.
We will now demonstrate the equality of these two expressions:
$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$
$= \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2]$
$= \mathbb{E}[X^2] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}[X]^2$
$= \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X^2]$
$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$
The square root of the variance is referred to as the *standard deviation.*

Definition 1.16. Two events, A, B are *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Theorem 1.17. *(Central Limit Theorem) If $X_1, X_2, ..., X_n$ are independent, identically distributed random variables with expectation μ and variance $\sigma^2 < \infty$, then* $\lim_{n\to\infty} \mathbb{P}(a \leq Q \leq b) = \Phi(b) - \Phi(a)$ *where*
$Q = \frac{X_1 + X_2 + ... + X_n - n\mu}{\sigma\sqrt{n}}$.
For the proof, see Section 5 of [3]

The central limit theorem implies that the distribution of Q approaches a standard normal distribution.

2. Stochastic Processes

Definition 2.1. A *stochastic process* X_t is a collection of random variables indexed by time $t \in T \subset \mathbb{R}$. If T is a countable set, then time is discrete. If T is an interval, then time is continuous.

Definition 2.2. A *filtration* $\{\mathcal{F}_t\}$ is a collection of increasing σ-algebras for stochastic process X_t on $(\Omega, \mathcal{F}, \mathbb{P})$ such that if $s < r$, then $\mathcal{F}_s \subset \mathcal{F}_r$ where $\mathcal{F}_s, \mathcal{F}_r \in \mathcal{F}$.

Note that the definition of a filtration intuitively implies that information is not lost over time.

Definition 2.3. The *natural filtration* $\{\mathcal{F}_t\}$ is the filtration where any σ-algebra in the collection is the smallest sigma algebra that contains all of the information in a stochastic process up to time t.

Only the natural filtration is relevant for this paper. Therefore, throughout the remainder of this paper, when we use filtration, we will be referencing the natural filtration.

Definition 2.4. A stochastic process X_t is *adapted* to a filtration if it is $\mathcal{F}_t$ measurable for all t.

Note that every stochastic process is adapted to a filtration since we are only referencing the natural filtration in this paper, however, this does not hold for every filtration.

Definition 2.5. The *conditional expectation* of a random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is the unique random variable $E[X|\mathcal{F}_n]$ such that:
•$E[X|\mathcal{F}_n]$ is $\mathcal{F}_n$ measurable
•For all $\mathcal{F}_n$ measurable events, A: $\mathbb{E}[E[X|\mathcal{F}_n]1_A] = \mathbb{E}[X1_A]$.

We can think of $\mathbb{E}[X]$ as the best guess for the random variable given no information about the outcome of the random experiment which produces X; however, when we have some, but not all information about an event, we may utilize conditional expectation.

Consider the example of a fair coin flip. Let $S_n = X_1 + ... + X_n$ be the total number of heads in the first n flips. We know that $\mathbb{E}[S_n] = n/2$. Our best guess for S_3 depends on S_1. We may say $E[S_3|X_1 = 1] = 2$ and $E[S_3|X_1 = 0] = 1$.

Proposition 2.6. *Let X_t be a stochastic process on filtration $\{\mathcal{F}_t\}$. Let X be a random variable. Then the conditional expectation $E[X|\mathcal{F}_t]$ satisfies:*

- *If X is $\mathcal{F}_t$-measurable, then $E[X|\mathcal{F}_t] = X$. Note that from this, we may say $\mathbb{E}[E[X|\mathcal{F}_t]] = \mathbb{E}[X]$ based on the construction of conditional expectation.*
- *If a collection of random variables, $X_1, ..., X_n$ are independent of X, then $E[X|\mathcal{F}_t] = \mathbb{E}[X]$. (since $\mathcal{F}_t$ does not contain any useful information about X)*
- *Linearity: If X, Y are random variables and a, b are constants, then $E[aX + bY|\mathcal{F}_t] = aE[X|\mathcal{F}_t] + bE[Y|\mathcal{F}_t]$*
- *Tower Property: If $s < r$, then $E[E[X|\mathcal{F}_r]|\mathcal{F}_s] = E[X|\mathcal{F}_s]$*
- *If Y is a $\mathcal{F}_t$-measurable random variable, then $E[XY|\mathcal{F}_t] = YE[X|\mathcal{F}_t]$*

Definition 2.7. A *martingale* M_t is a stochastic process that satisfies:

- $\mathbb{E}[|M_t|] < \infty$.
- If $s < t$, then $E[M_t|\mathcal{F}_s] = M_s$.

Martingales can be intuitively thought of as a model of a fair game, since regardless of previous values, the expected change from time s to time t is zero.

3. Random Walks and Brownian Motion

Definition 3.1. A *random walk* is a stochastic process defined by $S_n = \sum_{i=1}^{n} X_i$ where $S_0 = 0$ and each X_i is an independent random variable and all X_i are identically distributed.

The fact that each random variable is independent means that the information provided at one step in the random walk does not effect the outcome of any of the other steps in the random walk.

Definition 3.2. A random walk is *simple* if $X_i = 1$ or $X_i = -1$ with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = 1 - p$.

Consider stochastic process $S_n = \sum_{i=1}^{n} X_i$ where $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$. We also know that $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 = (1)^2 \cdot \frac{1}{2} + (-1)^2 \cdot \frac{1}{2} - 0 = 1$.

Proposition 3.3. *If S_n is a simple random walk where $\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}$, then S_n is a martingale.*

Proof. We will first demonstrate that $\mathbb{E}[|S_n|] < \infty$. We know $\mathbb{E}[|S_n|] = \mathbb{E}[|X_1 + ...X_n|] = \mathbb{E}[|X_1|] + ... + \mathbb{E}[|X_n|]$ and we know that each $\mathbb{E}[|X_i|] \leq 1$ since $|X_i| = 1$ and $\mathbb{P}(X_i = 1)$ and $\mathbb{P}(X_i = -1) \leq 1$. Therefore, $\mathbb{E}[|S_n|]$ is at most n and is therefore finite. Thus, S_n satisfies property one of a Martingale.

We will now demonstrate that if $m < n$, then $E[S_n|\mathcal{F}_m] = S_m$.

In order to take the conditional expectation of S_n, we must first ensure that S_n is $\mathcal{F}_n$ measurable for all n, however, this just means that S_n is adapted to a filtration which we know to be true for every stochastic process. Therefore, we are now ready to take the conditional expectation of S_n.

We know that $E[S_n|\mathcal{F}_m] = E[S_m|\mathcal{F}_m] + E[X_{m+1}|\mathcal{F}_m] + ... + E[X_n|\mathcal{F}_m]$ by linearity. We also know that $E[S_m|\mathcal{F}_m] = S_m$. It follows that $E[S_n|\mathcal{F}_m] = S_m + E[X_{m+1}|\mathcal{F}_m] + ... + E[X_n|\mathcal{F}_m]$.

We know that each X_i is independent of $\mathcal{F}_m$ where $i > m$ by the definition of a random walk. Therefore, each $E[X_i|\mathcal{F}_m] = \mathbb{E}[X_i]$ since each random variable is independent. We know that $\mathbb{E}[X_i] = 1 \cdot \frac{1}{2} + (-1) \cdot \frac{1}{2} = 0$. It follows that $E[S_n|\mathcal{F}_m] = S_m + 0 \cdot (n - m)$. Thus, $E[S_n|\mathcal{F}_m] = S_m$ and S_m is a martingale. ☐

Definition 3.4. A one-dimensional *Brownian motion* B_t is a continuous stochastic process with *drift rate* m and *variance rate* σ^2 that satsfies:

- $B_0 = 0$ with probability one.
- Let $0 \leq s \leq t$. Then the distribution of $B_t - B_s$ is normal with mean $\mu(t - s)$ and variance $\sigma^2(t - s)$.
- Let $0 \leq s_1 \leq t_1 \leq ... \leq s_n \leq t_n$. Then the random variables $B_{t1} - B_{s1}, ..., B_{tn} - B_{sn}$ are independent.
- The function $t \mapsto B_t$ is a continuous function of t with probability one.

Brownian motion can intuitively be thought of as reducing the size of time increments of a random walk; however, we must be careful when we do this, otherwise, the Brownian motion would simply cover the entire number line which is the reason that the additional properties are required.

Note that *drift rate* refers to the mean change per unit time of a stochastic process and that *variance* refers to the variance per unit time.

Note that Brownian motion is also referred to using the name *Wiener process.*

Definition 3.5. A *standard Brownian motion* is a Brownian motion with drift rate 0 and variance rate 1.

Proposition 3.6. *Suppose that a standard Brownian motion B_t is adapted to filtration $\{\mathcal{F}_t\}$ and that $\mathbb{E}[|B_n|] < \infty$. Then B_t is a martingale.*

Proof. Let $s < t$. It follows that $E[B_t|\mathcal{F}_s] = E[B_s|\mathcal{F}_s] + E[B_t - B_s|\mathcal{F}_s]$. Therefore, $E[B_t|\mathcal{F}_s] = B_s + \mathbb{E}[B_t - B_s] = B_s + 0$. Thus, $E[B_t|\mathcal{F}_s] = B_s$. ☐

Theorem 3.7. *With probability one, the function $t \mapsto B_t$ is nowhere differentiable.*

For the proof, see Section 2.6 of [4]

This can be intuitively understood since for any interval, the probability that the difference between sets of random variables is strictly increasing or strictly decreasing on that interval is vanishingly small since the differences between any two random variables are independent. Since any interval can be made into any amount of smaller intervals, brownian motion will not be differentiable anywhere.

Since Brownian motion is not differentiable, standard calculus will not be sufficient for continuous stochastic processes.

4. Ito Calculus

An essential breakthrough associated with typical differential and integral calculus is that one can determine values of a function if they know the rate of change of that function. Typical differential equations take the form

$$df(t) = C(t, f(t))dt.$$

This can be manipulated into the form in which it is typically expressed:

$$\frac{df}{dt} = f'(t) = C(t, f(t)).$$

Differential equations can be intuitively understood as: at time t, the f moves infinitesimally along a straight line with slope $C(t, f(t))$. A solution to this differential equation with initial condition $f(0) = x_0$ is

$$f(t) = x_0 + \int_0^t C(s, f(s))ds.$$

Sometimes, one can do the integration to determine the exact function. When this is not possible, one can use a computer to approximate the solution. This can be accomplished using a technique such as *Euler's method* where one takes a small increment Δt and evaluate:

$$f((k+1)\Delta t) = f(k\Delta t) + \Delta t C(k\Delta t, f(k\Delta t)).$$

Stochastic calculus examines stochastic differential equations in order to get around the non-differentiability of Brownian motion.

Definition 4.1. A *stochastic differential equation* is a differential equation in the form:

$$dX_t = m(t, X_t)dt + \sigma(t, X_t)dB_t$$

A solution to this equation takes the form:

$$X_t = X_0 + \int_0^t m(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s.$$

The ds integral is standard in typical integral calculus, however, the dB_s term is referred to as the *Itô integral*. The *Itô integral* is extremely similar to the Riemann integral in typical integral calculus.

Stochastic differential equations are often solved using numerical methods, such as the *stochastic Euler method* which takes the form:

$X((k+1)\Delta t) = X(k\Delta t) + \Delta t m(k\Delta t, X(k\Delta t)) + \sqrt{\Delta t}\sigma(k\Delta t, X(k\Delta t))N_k$ where N_k is a $N(0,1)$ random variable.

Definition 4.2. A *Markov process* is a type of stochastic process where only the current value of a variable is relevant for predicting the future. Therefore, past history has no affect on the future value. Stock prices are typically assumed to follow a Markov process.

Definition 4.3. An *Itô process* is a generalized Wiener process in which the parameters a and b are functions of the value of the underlying variable x and time t in the form:

$$dx = a(x,t)dt + b(x,t)dB_t.$$

Note that an Itô process is a Markov process since the change in x at time t only depends on the value of x at time t and not at any earlier time.

Note that B_t represents a standard Brownian motion (normal distribution with drift 0 and variance 1).

We will now discretize this process. Discretization can be thought of as allowing a time interval of Δt to pass by stretching the standard normal distribution of standard Brownian motion by a factor of Δt (where Δt represents the difference between two random variables in our Brownian motion). This can be accomplished by multiplying both the drift and the variance by a factor of Δt; however, the variance term is comprised of a standard Brownian motion that possesses a standard normal distribution (with mean 0 and variance 1). Therefore, if we multiply this distribution by a factor of Δt, then the variance would be influenced by a factor of Δt^2 since each value is squared in the variance formula included below for convenience (note that the expectation of the standard normal distribution is zero).

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Therefore, in order to influence the variance term by a factor of Δt, we must multiply the standard normal distribution (with mean 0 and variance 1) by a factor of $\sqrt{\Delta t}$. Let ϵ represent the standard normal distribution. It follows that the discretized variance term after time Δt has passed would be expressed by $b(x,t)\epsilon\sqrt{\Delta t}$ or equivalently $b(x,t)\phi(0,\Delta t)$ where ϕ is a normal distribution with mean 0 and variance Δt. After combining this term with the drift term, the resulting change in Δx can concisely be expressed:

$$\Delta x = a\Delta t + b\epsilon\sqrt{\Delta t}.$$

Note that the variance rate has a coefficient of $\sqrt{\Delta t}$ while the drift term has a coefficient of Δt. Therefore, as Δt gets very small, $\sqrt{\Delta t}$ will be much larger than Δt causing the variance term to significantly influence the value of Δx. For this reason the path of an Itô process is quite jagged. This jaggedness should provide more intuition behind the reason that standard calculus is not sufficient for continuous stochastic processes.

Proposition 4.4. *(Itô's Lemma) A function G of x and t follows the process:*

$$dG = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}b dB_t.$$

Note that G also follows an Itô process with drift

$$\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2$$

and variance

$$\left(\frac{\partial G}{\partial x}\right)^2 b^2$$

The reason for this can be considered an extension of typical differential calculus. Consider a function G of variables x and y where G is continuous and differentiable. If Δx and Δy are small changes in x and y and ΔG is the resulting change in G which can be expressed:

$$\Delta G \approx \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial y}\Delta y.$$

If greater precision is desired, one may use a Taylor series expansion of ΔG:

$$\Delta G = \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial y}\Delta y + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\Delta x^2 + \frac{\partial^2 G}{\partial x \partial y}\Delta x \Delta y + \frac{1}{2}\frac{\partial^2 G}{\partial y^2}\Delta y^2 + ...$$

As Δx and Δy approach zero, this becomes

$$\Delta G = \frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial y}dy.$$

Note that the second order terms are disregarded since they are negligible as Δx and Δy approach zero.

We will now investigate the analog of this idea for Itô processes.

Remember that an Itô process can be expressed in the form:

$$dx = a(x,t) + b(x,t)dB_t.$$

After discretizing and removing arguments, we can see that

$$\Delta x = a\Delta t + b\epsilon\sqrt{\Delta t}.$$

Let G be a function of x and of time t. We may evaluate ΔG using a Taylor expansion:

$$\Delta G = \frac{\partial G}{\partial x}\Delta x + \frac{\partial G}{\partial t}\Delta t + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}\Delta x^2 + \frac{\partial^2 G}{\partial x \partial t}\Delta x \Delta t + \frac{1}{2}\frac{\partial^2 G}{\partial t^2}\Delta t^2 + ...$$

Since $\Delta x = a\Delta t + b\epsilon\sqrt{\Delta t}$, we know $\Delta x^2 = b^2\epsilon^2\Delta t +$ higher order terms. Therefore, Δx^2 contains a first order term that cannot be ignored.

We know that the variance of the standard normal distribution is 1. It follows that

$$\mathbb{E}[\epsilon^2] - \mathbb{E}[\epsilon]^2 = 1.$$

Since the expectation of the standard normal distribution is 0, we know that $\mathbb{E}[\epsilon^2] = 1$. It follows that the expectation of $\epsilon^2\Delta t = \Delta t$.

Note that if you square the standard normal distribution, the resulting variance is 2. It follows that the variance of $\epsilon^2\Delta t = 2\Delta t^2$. Since this is a second order term, it is considered negligible as Δt approaches zero. Therefore, the variance is disregarded and Δx^2 is solely represented by its expectation meaning that in the limit, the result is an exact value rather than a distribution. Thus, $\Delta x^2 = b^2\Delta t$.

After substituting in for Δx^2, taking limits as Δx and Δt approach zero, and removing negligible terms, our Taylor expansion becomes:

$$dG = \frac{\partial G}{\partial x}dx + \frac{\partial G}{\partial t}dt + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2 dt.$$

After substituting in for dx from our Itô process, we arrive at our desired result of

$$dG = \left(\frac{\partial G}{\partial x}a + \frac{\partial G}{\partial t} + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b^2\right)dt + \frac{\partial G}{\partial x}bdB_t.$$

Definition 4.5. A process satisfies *Geometric Brownian motion* with drift μ and volatility σ if it satisfies:

$$dX_t = \mu X_t dt + \sigma X_t dB_t.$$

Note that Geometric Brownian motion is widely used in finance since it models how the price of an asset changes as a percentage rather than an absolute change which is often more relevant.

5. Black-Scholes Equation and Formula

The Black-Scholes model was a significant breakthrough in financial theory since it rigorously evaluated the theoretical price of financial derivatives (financial instruments whose prices are derived from that of an underlying asset). In this paper, we will focus on European call options and European put options whose underlying assets are stocks. A European option is an option that can only be exercised at expiry. A call option provides the right but not the obligation to buy a stock at a strike price K at an expiry time T. A put option provides the right but not the obligation to sell a stock at a strike price K at an expiry time T. Throughout the remainder of this paper, we will investigate the Black-Scholes model with respect to European call options; however, all of the analysis that we perform may be easily modified for European put options. Since stock options provide a right, not an obligation, they will only be exercised if the stock price is above the strike price at expiry (the option is in the money). Otherwise the option's owner will simply do nothing.

The Black-Scholes-Merton Model requires several assumptions about the underlying stock, a risk-free asset, and the overall market:

- The underlying stock's price follows a Geometric Brownian motion with non-random drift and non-random volatility. This is a significant assumption that does not hold in reality since volatility is actually random with fat tails and skew in its distribution.
- The underlying stock does not pay dividends between now and expiry. Note that alterations of Black Scholes attempt to account for dividends, however, this process is complex since the timing and amount of an upcoming dividend is not always known ahead of time.
- The risk-free rate is non-random. This assumption is not as significant since the risk-free rate tends to be relatively stable during the duration of an option contract.
- Arbitrage is impossible. Arbitrage refers to an opportunity for a risk-free profit. This assumption allows for the analysis of an option's price. The Black Scholes Model seeks to price an option so that it is impossible to

arbitrage. If the actual price of the option diverged from the no-arbitrage price, then the option would provide a risk-free profit. Therefore, everyone would buy the option until the option's price was back at the no-arbitrage price.
- Any amount of the risk-free asset may be borrowed or lent.
- Any amount of stock may be bought or sold (including fractional shares)
- There are no transaction costs.
- Transactions occur immediately.

The Black-Scholes-Merton model is often altered for modern-day applications, however, the overall intuition behind the original model is still extremely useful.

Definition 5.1. The payoff of a call option on a stock S_T with strike price K at expiry time T is:

$$f(S_T) = \max(S_T - K, 0)$$

Proposition 5.2. *The Black-Scholes-Merton model assumes that the price of an underlying stock S follows Geometric Brownian motion and can be modeled by:*

$$dS = \mu S dt + \sigma S dB_t.$$

where μ represents the stock's annual expected return and σ represents the stock's annual volatility.

Note that we have assumed that μ and σ are constant. This assumption is used in the derivation of the Black Scholes model, however, this is a big assumption, since volatility is considered random.

We will now provide some intuition behind this equation. While some may think that a stock's price follows a generalized Wiener process (constant expected drift rate and variance rate), that model would not account for the fact that the expected return of a given stock is independent of a stock's current price (in theory, the expected percentage return on a \$50 stock should be the same as on a \$20 stock). In order for this to occur, higher priced stocks would be expected to change by greater absolute amounts. Therefore, constant expected drift should be replaced by constant expected return (drift divided by stock price). Thus, Geometric Brownian motion is used to account for this. If S is the stock price at time t, we would expect the drift rate of stock S with annual expected return μ to be μS. It can also be assumed that the variability of a stock's return is constant (an investor is just as uncertain of the return of a \$50 stock as a \$20 stock) and is therefore represented by σS. Combining these terms results in our desired equation.

Using this model, the expectation of a stock's price after the passage of time T can be expressed:

$$\mathbb{E}[S_T] = S_0 e^{\mu T}$$

where μ is the stock's expected rate of return and S_0 is the initial stock price.

Definition 5.3. A variable follows a *lognormal distribution* if its natural logarithm is normally distributed.

Proposition 5.4. *The model of a stock S_T at time T used by the Black-Scholes-Merton model follows a log-normal distribution expressed:*

$$lnS_T \sim \phi\left[\ln S_0\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right].$$

We will demonstrate this using Itô's Lemma. Let $G = \ln S$. We know that

$$\frac{\partial G}{\partial S} = \frac{1}{S}, \frac{\partial^2 G}{\partial S^2} = -\frac{1}{S^2}, \frac{\partial G}{\partial t} = 0.$$

Therefore, the process followed by G is:

$$dG = \left(\mu - \frac{\sigma^2}{2}\right)dt + \sigma dB_t.$$

Since we are assuming that μ and σ are constant, we know that G follows a generalized Wiener process with constant drift rate $\mu - \frac{\sigma^2}{2}$ and constant variance rate σ^2. Therefore, the change in $\ln S$ between time 0 and some future time T is normally distributed. This may be expressed in the form:

$$lnS_T - lnS_0 \sim \phi\left[\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right].$$

or

$$lnS_T \sim \phi\left[\ln S_0\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma^2 T\right].$$

Proposition 5.5. *(Black-Scholes-Merton differential equation):*

Suppose that f is the price of a derivative contingent on underlying asset dS and on time t. From Itô Lemma, we may express:

$$df = \left(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)dt + \frac{\partial f}{\partial S}\sigma S dB_t.$$

The equations for dS and df can be discretized by:

$$\Delta S = \mu S \Delta t + \sigma S \Delta B_t.$$

$$\Delta f = \left(\frac{\partial f}{\partial S}\mu S + \frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t + \frac{\partial f}{\partial S}\sigma S \Delta B_t.$$

Note that ΔB_t (the Brownian motion term) appears in both equations. Therefore, a portfolio of both a stock and its derivative can be constructed to eliminate this term. This portfolio may contain -1 derivative and $\frac{\partial f}{\partial S}$ shares of stock. Let Π represent the value of the portfolio. Therefore

$$\Pi = -f + \frac{\partial f}{\partial S}S.$$

It follows that the change in the value of the portfolio in time interval Δt is given by:

$$\Delta\Pi = -\Delta f + \frac{\partial f}{\partial S}\Delta S.$$

Substituting our previous results into this equation yields:

$$\Delta\Pi = \left(-\frac{\partial f}{\partial t} - \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t.$$

Notice that this equation does not involve ΔB_t. Therefore, this portfolio must be riskless during time Δt. This idea is known in finance as delta hedging where a portfolio is hedged against changes in the underlying asset by shorting $\frac{\partial f}{\partial S}$ (referred to as delta) shares of stock for every one option purchased.

Remember that an assumption of the Black Scholes formula is that there are no arbitrage opportunities. Since the portfolio is riskless, it must earn the same rate of return as other risk-free assets (which we will refer to as the risk-free rate). If this was not the case there would be an arbitrage opportunity.

Since the portfolio earns the risk-free rate over time, we may express:

$$\Delta\Pi = r\Pi\Delta t.$$

After substituting prior results into this equation, we obtain:

$$\left(\frac{\partial f}{\partial t} + \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t = r\left(f - \frac{\partial f}{\partial S}S\right)\Delta t.$$

Thus:

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

This is the Black-Scholes-Merton differential equation.

Notice that all variables in this equation (current stock price, time, stock volatility, and risk-free interest rate) are unaffected by investors' risk tolerance, whereas μ (which does not appear in the Black-Scholes-Merton differential equation) is affected by risk-preferences. Therefore, this equation will still hold if we let $\mu = r$. This assumes a risk-neutral world (that all investors are risk neutral and do not require premium to take on additional risk). The real world is not risk neutral (most people would require premium to take on additional risk), however, since risk-preferences do not enter the equation, they cannot affect the solution to the equation. Thus, the option price in a risk-neutral world also applies to the real world, however, assuming a risk-neutral world will help facilitate solving this equation. Note that when moving from a risk-neutral world to our current world, the expected growth rate in the stock price changes and the discount rate that must be used for any payoffs of the derivative changes, and these changes exactly offset.

Lemma 5.6. *If V is lognormally distributed and the variance of $\ln V$ is w^2, then*

$$\mathbb{E}[\max(V-K,0)] = E(V)\Phi\left(\frac{\ln\left[\frac{E(V)}{K}\right] + \frac{w^2}{2}}{w}\right) - K\Phi\left(\frac{\ln\left[\frac{E(V)}{K}\right] - \frac{w^2}{2}}{w}\right).$$

For the proof, see pages 352 and 353 of [5].

Remember that Φ is the distribution function of the standard normal distribution.

Theorem 5.7. *(Black-Scholes-Merton Formula) The price of a European call option c can be expressed:*

$c = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$ *where*

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

We know that the expected value of a European call option at maturity is

$$E[\max(S_T - K, 0)].$$

It follows that the price of a European call option in a risk neutral world can be modeled by this expectation discounted by the risk-free rate:

$$c = e^{-rT} E[\max(S_T - K, 0)].$$

It follows that

$$c = e^{-rT}\left[E(S_T)\Phi\left(\frac{\ln\left[\frac{E(S_T)}{K}\right] + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right) - K\Phi\left(\frac{\ln\left[\frac{E(S_T)}{K}\right] - \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right)\right]$$

by the Lemma stated above (remember that $\ln S_T$ is normally distributed with variance $\sigma^2 T$ by Proposition 6.4).

We know that the expectation of a stock's price at time T can be modeled by:

$$\mathbb{E}[S_T] = S_0 e^{\mu T}$$

and is assumed by the Black-Scholes-Merton model to follow a log-normal distribution.

Since we are able to assume a risk-neutral world, we may set $\mu = r$. Therefore,

$$\mathbb{E}[S_T] = S_0 e^{rT}.$$

It follows that

$$c = e^{-rT} S_0 e^{rT}\Phi\left(\frac{\ln\left[\frac{S_0 e^{rT}}{K}\right] + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right) - Ke^{-rT}\Phi\left(\frac{\ln\left[\frac{S_0 e^{rT}}{K}\right] - \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right).$$

$$= S_0\Phi\left(\frac{\ln\left[\frac{S_0}{K}\right] + \ln(e^{rT}) + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right) - Ke^{-rT}\Phi\left(\frac{\ln\left[\frac{S_0}{K}\right] + \ln(e^{rT}) - \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right).$$

$$= S_0\Phi\left(\frac{\ln(S_0/K) + rT + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right) - Ke^{-rT}\Phi\left(\frac{\ln(S_0/K) + rT - \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}}\right).$$

$$= S_0\Phi\left(\frac{\ln(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}\right) - Ke^{-rT}\Phi\left(\frac{\ln(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}\right).$$

This is commonly expressed in the form: $c = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$ where

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

This is the Black-Scholes-Merton formula for the price of an arbitrage-free European call option.

Note that the Black-Scholes-Merton formula may also be expressed in the form $c = e^{-rT}N(d_2)[S_0 e^{rT}N(d_1)/N(d_2) - K]$ where:

e^{-rT} : Discount factor

$N(d_2)$: Probability of exercise

$e^{rT}N(d_1)/N(d_2)$: Expected percentage increase in stock price in risk-neutral world if option is exercise

K: Strike price paid if option is exercised

This form provides further intuition behind the factors that contribute to an option's price.

Acknowledgments

It is a pleasure to thank my mentor, Jake Fiedler, for his insightful guidance throughout the program. I am also extremely grateful for Professor Peter May for arranging such an incredible program.

References

[1] Erhan Cinlar. Probability and Stochastics. Springer-Verlag New York, 2014.
[2] Fischer Black and Myron Scholes. The Pricing of Options and Corporate Liabilities. https://www.cs.princeton.edu/courses/archive/fall09/cos323/papers/black scholes73.pdf.
[3] Gregory F. Lawler. Notes On Probability. https://www.math.uchicago.edu/ lawler/probnotes.pdf.
[4] Gregory F. Lawler. Stochastic Calculus: An Introduction with Applications. https://www.math.uchicago.edu/ lawler/finbook.pdf.
[5] John C. Hull. Options, Futures, and Other Derivatives. Pearson. 1989.
[6] Sheldon Natenberg. Option Volatility and Pricing: Advanced Trading Strategies and Techniques. McGraw-Hill Education, 2015.