Pricing an American Option using Least Squares Monte Carlo Simulation and Simulated Annealing

Ronak Shah
Anurag Dutt
Project supervised by: Dr. Jiaqiao Hu

Student Number: 112757278, 113019209

Department of Applied Mathematics and Statistics
Stony Brook University

December 15, 2020

Abstract

In this project we discuss how to price American option using a geometric Brownian motion model for stock price. We and consider numerical methods for approximating the price of American options. These numerical methods include Least Squares Monte Carlo and Simulated Annealing. Pricing American options on multiple underlying assets is a challenging, high-dimensional problem that is frequently tackled using the Longstaff-Schwartz method, regressing the continuation value over all Monte Carlo paths in order to decide on early exercise. We extend this approach by requiring the interpolated continuation value. We apply this extension to the pricing problem of an American put on a single underlying asset and an American exchange option, while also varying the number and type of basis functions of the interpolation of the continuation value.

Contents

List of Figures

List of Tables

1 Introduction

1.1 The History and Origin of Options

Options, in one form or another, have been around for at least centuries. There are reports that they have been around for millennia, being used as early as the sixth century B.C. in Ancient Greece. Further there are many other instances where we see options crop up in history, clauses in marine cargo contracts, such as those of the Romans and Phoenicians, would now be considered options.

The reason for their wide spread use is, in part, due to the guarantees that they afford to those who buy them. If we were to be traveling to a foreign country to buy something to trade, it is helpful to know when we get back that we are going to be able to sell our product for a price agreed upon prior to the trip. This allows us insurance on the trip, the knowledge of how much we will make when we return. Alternatively, if someone else was to go on a voyage to acquire something we wanted, and we believed that the price of that commodity was to go up, then making sure we could buy it at the current price would allow us to make profit.

The interest then became how to price these contracts. If you want to sell grain for twice its current price in one months' time, it is obvious that the price of grain is unlikely to double in one month, thus the price of this contract would likely be very high. Conversely, if we were wanting to buy grain at twice the price in one months' time, the cost of this contract would be very low. Owing again to the unlikely event that the grain does double in price. Thus someone would happily take on the contract cheaply knowing they can sell their grain to you for what is likely to be much more than it would be worth in a months' time.

1.2 Option Basics

Before we may discuss the pricing of *options* we need to introduce a few terms. The most key of these being the definition of *options* and, we introduce this and some other key ideas here.

Definition 1.1 (Option)**.** *An **option** is a contract that gives the buyer the right, but not the obligation, to buy or sell an underlying asset at a specific price (known as the **strike price**) on or before a certain date.*

By an *underlying asset* we refer to the financial instrument on which the options price is based. This could be futures, stocks, commodities or currency, noting a change in the price of the underlying asset causes a change in the price of the option.

In the introduction we discussed the two types of options in an imprecise way, the buy side and sell side of an option. These concepts are extended within the next definition, to give rise to the two types of options that are used today, which we define here.

Definition 1.2 (Call)**.** *A **call** option is a contract that gives the buyer the right, but not the obligation, to buy an underlying asset at a specific price on or before a certain date (known as the **expiry date**).*

Definition 1.3 (Put)**.** *A **put** option is a contract that gives the buyer the right, but not the obligation, to sell an underlying asset at a specific price on or before a certain date (known as the **expiry date**).*

There are many different variations of puts and calls, the most popular being *European, American* and *Asian* option types. Other types of options are often referred to as *exotic options*, these include; basket options, barrier options, binary options, and down-and-out options. Each of these types has

a different structure in the way the pay-out is calculated or when it may be *exercised*. This term is often used, and before we can differentiate between the types of options we must introduce it.

Definition 1.4 (Exercise)**.** *We say that we* ***exercise*** *an option when we decide to use that option. That is, to buy or sell the underlying asset the option corresponds to.*

The main difference between European and American options is when you can exercise them. The term **European** is given to those options that may be exercised **only** at the date of expiration. Alternatively **American** options may be exercised **at any time** before, or on, the date of expiration.

For both European and American options, as call options allow us to buy the stock at a specific price, the payoff would be given by $\max(S-K,0)$ and for a put $\max(K-S,0)$, where the stock price at the time the option is exercised (sometimes termed when it reaches maturity) is S and the strike price is K.

Example 1.1. As our second example we will consider an American put on a stock. Here we have the right to sell a stock for a specified strike price, let's assume that this is $\$45$. Much like with the call this is a kind of bet that we are placing, here we are "betting that the stock will decrease in price". So let's say this option cost us $\$15$ then if the stock is trading at $\$50$, as with the previous example we either make nothing, reduce our losses or make profit. However this is an American option and as such can be exercised at any time between purchase and expiry dates. Hence, the only way we are guaranteed to lose the cost of the option is if the stock does not ever drop below $\$44.85$.

If it only drops below this value once and we do not exercise we would loose out. This shows the difficulty when choosing to exercise American options.

2 The Black-Scholes Model for Stocks

The trouble with pricing options is without knowing the path the stock is likely to travel, it is difficult to price the option, as clearly the price of an option must be dependent on the stock price. Here we will develop a continuous time stochastic model for stocks and through this derive an equation for the price of options. We may then solve this explicitly for European options to gain the formulae for the pricing of European options. Before we develop this model, there are a number of key ideas and concepts we will need to introduce.

2.1 Factors Affecting the Price of Stock Options

As we expect the price of the option to reflect the price of the underlying asset, it seems logical that a factor that affects the price of the underlying asset would also affect the price of the option. So for stock options we consider the factors that affect stock prices. There are six major factors. These are;

1. The current price the stock is trading at, S_0,
2. The strike price, K,
3. The time until expiration, T,
4. The volatility of the price of the stock, σ,
5. The risk-free interest rate, r,

6. The dividends that are expected to be paid, q.

Here the risk-free interest rate is the theoretical rate of return on a completely risk free investment. The volatility represents how the price varies over time, which we will give a more rigorous definition of this later. How each of these factors affects the price of European and American puts and calls is given in Table 9. We will discuss how each of these factors affects the option. Note that in Table 9 we use + to mean the factor increases the price of the option, a − for a decrease and a ? when the relationship is unknown.

Variable	European Call	European Put	American Call	American Put
Current stock price	+	-	+	-
Strike Price	-	+	-	+
Time to expiration	?	?	+	+
Volatility	+	+	+	+
Risk-free rate	+	-	+	-
Amount of future dividends	-	+	-	+

Table 1: The effect different factors have on the price of different options

Most of the ways each factor affects each type of option is intuitivee. If the price goes up, the price of a call goes up as we are likely to see greater increases in the price of the stock and the price of a put goes down as the decrease will not be as fast as the increase. This is due to the fact that, as we will see later, the path the stock takes is the initial price multiplied by some variables. Hence increases tend to happen faster than decreases.

The risk free interest rate is a more complex idea. Within the economy, as interest rates increase, investors expect more return from the option, however the value of any money earnt in the future decreases, due to these interest rates, termed inflation. This increases the price of stocks slightly resulting in a higher chance of calls paying off and less of puts paying off.

With relation to the dividends, as the ex-dividend date approaches, that is the date at which entitlement to dividends changes, the stock price decreases. Hence, due to the relationships for calls and puts, the price of calls decreases and the price of a put increases.

The most anomalous observations are those of the time to expiration and the volatility. The time to expiration for American options increases for both puts and calls. This is due to that, for two options, if the only difference is that one has a longer time to expiration, then the owner of the option with a longer time has all of the same opportunities to exercise and more. This increases the value of the option. For European options, this is not necessarily the case. If we have two options that straddle the ex-dividend date, the one with the shorter life would be worth more.

Finally we consider volatility, we have not defined volatility until this point, we will discuss definition in later sections. For now we merely remark that volatility is a measure of how uncertain we are of the stocks future price. If volatility increases, the chance the stock will do very well or very poorly increases. As our maximum loss from an option contract is the price but the profit is either infinite or large this benefits the owner hugely.

2.2 Markov and Wiener Processes

The next concept we will need to introduce is that of a *Markov process*. This is defined as a stochastic process with the *Markov property*; where the future value of the variable is dependent only on its current value and not its history. We often say that a Markov process is memoryless due too this property. We will examine a very specific type of Markov process, known as the *Wiener process*.

Definition 2.1 (Wiener process)**.** *Let $W = (W_t)_{t\in[0,\infty)}$ be a continuous process. We say that this process is a* ***Wiener process*** *if the following properties hold:*

Property 1. *We require that if Δt is a small period of time,*

$$W_{t+\Delta t} - W_t = \epsilon\sqrt{\Delta t},$$

where ϵ has a standardised normal distribution (i.e. a normal distribution with a mean of zero and standard deviation of one).

Property 2. *We require that each increment is independent, so that W_t and W_s are independent for $0 \le s < t$,*

Property 3. *The function defined by $t \to W_t$ is almost surely everywhere continuous (i.e. the probability that W_t is everywhere continuous is one).*

We see this is indeed a Markov process as the attribute described in Property 2 is precisely the Markov property. Also note that the first property shows that the Wiener process is normally distributed. An interesting implication of the Wiener process is found when we consider these normally distributed Markovian variables.

Proposition 2.1. *Let X and Y be two independent normally distributed variables with means μ_X and μ_Y respectively and variances σ_X^2 and σ_Y^2 respectively. Then the variable defined as $Z = X + Y$ is normally distributed with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$. We will use the notation $A \sim N(\mu, \sigma^2)$ to mean "A variable A is normally distributed with mean μ and variance σ^2".*

Proof. The characteristic function of X and Y are by definition,

$$G_X(s) = \mathbb{E}\left[e^{isX}\right] \text{ and } G_Y(s) = \mathbb{E}\left[e^{isY}\right].$$

We sum these two variables to generate a new variable $Z = X + Y$, which has characteristic function given by,

$$G_Z(s) = \mathbb{E}\left[e^{isZ}\right] = \mathbb{E}\left[e^{isX+isY)}\right] = \mathbb{E}\left[e^{isY}\right]\mathbb{E}\left[e^{isX}\right] = G_X(s)G_Y(s),$$

as these variables are independent. Now using the general formula for the characteristic function of a normal distribution from section **??** we have that,

$$\begin{aligned} G_z(s) = G_X(s)G_Y(s) &= \exp\left(it\mu_x - \frac{\sigma_X^2 s^2}{2}\right)\exp\left(it\mu_Y - \frac{\sigma_Y^2 s^2}{2}\right) \\ &= \exp\left(it(\mu_X + \mu_Y) - \frac{(\sigma_x^2 + \sigma_Y^2)s^2}{2}\right). \end{aligned}$$

Which is precisely the characteristic function of a normally distributed variable with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$. Hence, $W \sim (\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$ mean W has normal distribution with mean $\mu_X + \mu_Y$ and variance $\sigma_X^2 + \sigma_Y^2$. □

Consider a variable X that follows a Markov process. If we know that the change in value of a single day is a normal distribution with mean zero and variance one, then we may find the distribution for two days. Due to the Markov property the above proposition applies, and we have that the change in the variable over two days is normally distributed with mean zero and variance two.

Note that we may apply this as many times as we please to find the change in the variable over any period but the variance increases showing the uncertainty of these predictions.

It follows that we may consider the change in the variable W over a relatively large period of time W. We denote this $W_T - W_0$ and it can be thought of as the sum of changes in W in N small time intervals of length $\Delta t = T/N$. Hence we have that,

$$W_T - W_0 = \sum_{i=1}^{N} \epsilon_i \sqrt{\Delta t}, \tag{1}$$

where each of the ϵ_i for $i = 1, 2, \ldots, N$ are normally distributed with mean zero and variance one.

It follows from (1) that, the mean of $W_T - W_0$ is zero and has variance $N\Delta t = T$. In normal calculus we consider the limit of a discrete process as the changes head toward zero. Similar notations and conventions exist in stochastic calculus, here we will use the notation $\mathrm{d}W$ to refer to the Wiener process $W_{\Delta t+t} - W_t$ in the limit $\Delta t \to 0$.

2.3 Generalisation of the Wiener Process and Itôs Process

Insofar we have discussed the Wiener process, however there is a problem with this. We can see this by defining the *drift rate* and *variance rate*, these are given as the mean change per unit time and variance change per unit time. In our prior discussion of the Wiener process it is obvious that these have been zero and one respectively. This leads to a small issue, the expected value of W at any given time t, is equal to its current value. This is clearly an issue as stocks trend upward or downward. We may capture this aspect of stocks by generalising the Wiener process. The *generalised Wiener process* for a variable, here denoted as x is defined as,

$$\mathrm{d}x = a\mathrm{d}t + b\mathrm{d}W, \tag{2}$$

where a and b are constants and $\mathrm{d}W$ is the Wiener process. Here if the $\mathrm{d}W$ term where removed from (2), then it would have solution $x = x_0 + at$ for some x_0 specified by an initial condition. This is the way we assume the stock will grow, moving at a constant rate a. The $\mathrm{d}W$ term adds "noise" to this. As stock movement is assumed to be random, this term introduces the randomness in the form of a Wiener process.

Expressing this in discrete terms we have that $\Delta x = a\Delta t + b\epsilon\sqrt{\Delta t}$, with ϵ as before. Hence we have that the mean is now $a\Delta t$ and variance is $b^2\Delta t$. Again following similar arguments as with the Wiener process, we see that for the continuous process that the mean and variance are now aT and b^2T respectively.

It is easy to see that even this model which allows us to drift the stock either up or down in a given direction is not particularly favourable. It assumes that the drift is constant. This is, in practice, is not the case, a stock may rise at a constant rate but crash the next day. Here we introduce an *Itô process* to compensate for this.

Definition 2.2 (Itô processes)**.** *An Itô process is a is a type of generalised Wiener process where the constants a and b are now dependent on the value of the underlying variable x and time t. In*

mathematical terms,

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

Itô processes address the issues discussed. Following the same structure as previous arguments this has drift rate $a(x,t)$ and variance rate $b(x,t)^2$.

2.4 Itôs Lemma

Before we may prove Itôs lemma we must first prove a very important result, often regarded as the corner stone of Itô calculus, $dW_t^2 = dt$. This form of expressing this is merely short hand, the exact statement is given and proven below.

Theorem 2.1 ($dW_t^2 = dt$)**.** *Given a function $g(X_t)$ we have that,*

$$\int_{t_0}^{t} g(X_t)dW_t^2 = \int_{t_0}^{t} g(X_t)dt.$$

Proof. To begin first let us define a new variable Y as,

$$\begin{aligned} Y &= \int_{t_0}^{t} g(X_t)dW_t - \int_{t_0}^{t} g(X_t)dt \\ &= \sum g(X_{i-1})\Delta W_{t_i}^2 - \sum g(X_{i-1})\Delta t_i \\ &= \sum g(X_{i-1})(\Delta W_{t_i}^2 - \Delta t_i) \end{aligned}$$

We may then calculate the moments of this new variable.

$$\begin{aligned} \mathbb{E}[Y] &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\Big[g(X_{i-1})(\Delta W_{t_i}^2 - \Delta t_i)\Big] \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[g(X_{i-1})]\,\mathbb{E}\Big[(\Delta W_{t_i}^2 - \Delta t_i)\Big], \end{aligned}$$

as $g(X_{i-1})$ and $(\Delta W_{t_i}^2 - \Delta t_i)$ are independent, so the mean of the product is the product of the mean. Furthermore, as Δt is deterministic and merely a constant, from the properties discussed in the preliminaries,

$$\mathbb{E}[Y] = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}[g(X_{i-1})]\,(\mathbb{E}\Big[\Delta W_{t_i}^2\Big] - \Delta t_i).$$

Now note that, as $\mathbb{E}\Big[W_{t_i}W_{t_{i-1}}\Big] = \mathbb{E}\Big[(W_{t_i} - W_{t_{i-1}})W_{t_{i-1}} + W_{t_{i-1}}^2\Big]$. Splitting this up we have that,

$$\mathbb{E}\Big[(W_{t_i} - W_{t_{i-1}})W_{t_{i-1}} + W_{t_{i-1}}^2\Big] = \mathbb{E}\Big[(W_{t_i} - W_{t_{i-1}})W_{t_{i-1}}\Big] + \mathbb{E}\Big[W_{t_{i-1}}^2\Big]$$

The first term consists of two independent variables, thus,

$$\mathbb{E}\Big[(W_{t_i} - W_{t_{i-1}})W_{t_{i-1}}\Big] = \mathbb{E}\Big[(W_{t_i} - W_{t_{i-1}})\Big]\mathbb{E}\Big[W_{t_{i-1}}\Big] = 0,$$

as the average of the Wiener process is zero. We then notice the second term may be expressed as,

$$\begin{aligned}\mathbb{E}\left[W_{t_{i-1}}^2\right] &= \mathbb{E}\left[W_{t_{i-1}}\right]^2 + \mathbb{E}\left[W_{t_{i-1}}^2\right] - 2\mathbb{E}\left[W_{t_{i-1}}\right]^2 + \mathbb{E}\left[W_{t_{i-1}}\right]^2 \\ &= \mathrm{Var}[W_{t_{i-1}}] + \mathbb{E}\left[W_{t_{i-1}}\right]^2.\end{aligned}$$

The average and variance of the Wiener process are known. Hence, we know that $\mathbb{E}\left[W_{t_{i-1}}^2\right] = t_{i-1}$. Then we have,

$$\mathbb{E}\left[\Delta W_{t_i}^2\right] = \mathbb{E}\left[(W_{t_i} - W_{t_{i-1}})^2\right] = \mathbb{E}\left[W_{t_i}^2\right] + \mathbb{E}\left[W_{t_{i-1}}^2\right] - 2\mathbb{E}\left[W_{t_i}W_{t_{i-1}}\right] = |t_i - t_{i-1}| = \Delta t_i,$$

hence the average of our variable Y is zero. Now note that,

$$\mathrm{Var}[Y] = \mathbb{E}\left[Y^2\right] - \mathbb{E}[Y]^2 = \mathbb{E}\left[Y^2\right].$$

So the second moment is precisely the variance. Considering the second moment we have that,

$$\begin{aligned}\mathbb{E}\left[Y^2\right] &= \lim_{n\to\infty} \mathbb{E}\left[\left(\sum_{i=1}^{n} g(X_{t_{i-1}})(\Delta W_{t_i}^2 - \Delta t)\right)^2\right] \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[g(X_{t_{i-1}})^2\right]\mathbb{E}\left[(\Delta W_{t_i} - \Delta t)^2\right] \\ &\quad + 2\sum_{i=1}^{n}\sum_{j<i} \mathbb{E}\left[\Delta W_{t_i}^2 - \Delta t\right]\mathbb{E}\left[g(X_{t_{i-1}})g(X_{t_{j-1}})(\Delta W_{t_j}^2 - \Delta t)\right].\end{aligned}$$

Note that in the second term in the above we have that $\mathbb{E}\left[\Delta W_{t_i}^2 - \Delta t\right]$ is independent of all the other terms, so we may split up the angular brackets. However $\mathbb{E}\left[\Delta W_{t_i}^2 - \Delta t\right] = 0$ hence there is no contribution from the second term. Now consider $\mathbb{E}\left[(\Delta W_{t_i} - \Delta t)^2\right]$,

$$\begin{aligned}\mathbb{E}\left[(\Delta W_{t_i} - \Delta t)^2\right] &= \mathbb{E}\left[W_{t_i}^4\right] - 2\Delta t\mathbb{E}\left[W_{t_i}^2\right] + \Delta t^2 \\ &= 3\Delta t^2 - 2\Delta t^2 + \Delta t^2 = 2\Delta t^2.\end{aligned}$$

Here we used that as ΔW is a Gaussian variable with zero mean this means that $\mathbb{E}\left[W^4\right] = 3\mathbb{E}\left[W^2\right]$, this can be seen by direct integration of the Gaussian distribution and is quoted but not derived here. Hence, we now have,

$$\begin{aligned}\mathbb{E}\left[Y^2\right] &= \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{E}\left[g(X_{t_{i-1}})^2 2\Delta t^2\right] \\ &= \lim_{n\to\infty} 2\Delta t \sum_{i=1}^{n} \mathbb{E}\left[g(X_{t_{i-1}})^2 2\Delta t\right].\end{aligned}$$

Note that taking the limit $n \to \infty$ is equivalent to taking $\Delta t \to 0$, as $\Delta t = (t - t_0)/n$. Hence, $\mathbb{E}\left[Y^2\right] = 0$. We have shown Y has mean and variance that are zero. The argument also holds for higher moments as $\Delta t \to 0$ and these can be shown to be zero as well. We now have a variable with all moments equal to zero and hence $Y \equiv 0$. This proves our result. □

With this corner stone of stochastic calculus we may now prove Itô's lemma, sometimes referred to as Itô's equation or Itô's differentiation rule. Proving this will then allow us to derive the Black-Scholes formula.

Lemma 2.1 (Itôs Lemma)**.** *Consider an Itô process described by,*

$$\mathrm{d}X_t = A(t, X_t)\mathrm{d}t + B(t, X_t)\mathrm{d}W_t, \tag{3}$$

where W_t is the Wiener process. Then, if $g(t, x)$ is a twice-differentiable scalar function of two variables $x, t \in \mathbb{R}$, then,

$$\mathrm{d}g(t, X_t) = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial X}A + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}B^2\right)\mathrm{d}t + \frac{\partial g}{\partial X}B\mathrm{d}W.$$

Proof. We do not give a full rigorous proof here as it is beyond the scope of this project, however we may derive this result using results from Riemann calculus. Consider $g(t, X_t)$, then from Taylor's Theorem we know that an approximation for the derivative of $g$of order $\mathcal{O}(dt)$ is given by,

$$\mathrm{d}g(t, X_t) = \frac{\partial g}{\partial t}\mathrm{d}t + \frac{\partial g}{\partial X}\mathrm{d}X_t + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}\mathrm{d}X_t^2.$$

We include the final term as when we substitute in (3) we will obtain a $\mathrm{d}W^2$ term which we know to be dt. Substituting (3) into the above we obtain,

$$\begin{aligned}\mathrm{d}g(t, X_t) = \frac{\partial g}{\partial t}\mathrm{d}t + \frac{\partial g}{\partial X}(A(t, X_t)\mathrm{d}t + B(t, X_t)\mathrm{d}W_t) \\ + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}(A(t, X_t)\mathrm{d}t + B(t, X_t)\mathrm{d}W_t)^2.\end{aligned}$$

Recall that $\left\langle \Delta W_t^2 \right\rangle = \Delta t$. Now taking limits we see that $\left\langle \mathrm{d}W_t^2 \right\rangle = \mathrm{d}t$, so $\mathrm{d}W_t$ can be thought of as $\mathcal{O}(\mathrm{d}t)$. Expanding and removing terms using this rule of order greater than $\mathrm{d}t$ we have that,

$$\begin{aligned}\mathrm{d}g(t, X_t) &= \frac{\partial g}{\partial t}\mathrm{d}t + \frac{\partial g}{\partial X}(A(t, X_t)\mathrm{d}t + B(t, X_t)\mathrm{d}W_t) + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}\big[A(t, X_t)^2\mathrm{d}t^2 \\ &\quad + B(t, X_t)^2\mathrm{d}W_t^2 + 2A(t, X_t)B(t, X_t)\mathrm{d}t\mathrm{d}W_t^2\big] \\ &= \frac{\partial g}{\partial t}\mathrm{d}t + \frac{\partial g}{\partial X}(A(t, X_t)\mathrm{d}t + B(t, X_t)\mathrm{d}W_t) + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}\big[B(t, X_t)^2\mathrm{d}W_t^2\big].\end{aligned}$$

Replacing $\mathrm{d}W_t^2$ with $\mathrm{d}t$ we have,

$$\mathrm{d}g(t, X_t) = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial X}A(t, X_t) + \frac{1}{2}\frac{\partial^2 g}{\partial X^2}B(t, X_t)^2\right)\mathrm{d}t + \frac{\partial g}{\partial X}B(t, X_t)\mathrm{d}W_t,$$

as required. □

2.5 Black-Scholes Model

As discussed before it is incorrect to assume that the stock follows a generalized Wiener process. We then introduced an Itô process to counter this. It remains to be determined what the functions $a(x,t)$ and $b(x,t)$ need to be in our Itô process. The appropriate assumption is that the expected return (the drift over the stock price) is constant. Thus we must have that the expected return is μS for some $\mu \in \mathbb{R}$.

Furthermore, we assume that the percentage return's variability over a small time interval, Δt, is constant and independent of stock price. This means that a buyer is as uncertain of the return (as a percentage) when the stock costs $\$1$ as when the stock costs $\$1000$. This leads to the fact that the the stock price should be proportional to the standard deviation over a small period of time Δt. This leads to the following final model,

$$\mathrm{d}S = \mu S \mathrm{d}t + \sigma S \mathrm{d}W, \tag{4}$$

where the variable σ is the volatility of the stock per year and μ is the expected rate of return on the stock per year. This is the most widely used model for stock behavior. Now using (4) and applying Itô's lemma we obtain, for a function $G(S,t)$ we have the process G follows is given by,

$$\mathrm{d}G = \left(\frac{\partial G}{\partial S} + \frac{\partial G}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 G}{\partial S^2}\right)\mathrm{d}t + \frac{\partial G}{\partial S}\sigma S \mathrm{d}W. \tag{5}$$

We see from (4) that the volatility is a measure of how unsure we are about the path the stock will take. This is because it is multiplying the random component. It can also be viewed as the standard deviation of the lognormal distribution of S_T, as we will see in the next section.

2.6 The Lognormal Property

Consider the equation as described in (4). We may use Itôs lemma to derive the process. Let $G = \log(S)$ by applying Itôs lemma, (2.1), with $X_t = S$ and $g(x,t) = \log(S)$ we obtain,

$$\mathrm{d}G = \left(\mu - \frac{\sigma^2}{2}\right)\mathrm{d}t + \sigma \mathrm{d}W.$$

With μ as, the expected rate of return, and σ being the volatility of the stock, these are constant. Thus G follows a generalised Wiener process, with drift rate $\mu - \frac{\sigma^2}{2}$ and variance rate σ. Therefore, by Proposition 2.1, between time 0 and T we see that G has mean $(\mu - \frac{\sigma^2}{2})T$ and variance $\sigma^2 T$. Hence,

$$\log(S_T) - \log(S_0) \sim \phi\left[\left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right]$$
$$\Rightarrow \log(S_T) \sim \phi\left[\log(S_0) + \left(\mu - \frac{\sigma^2}{2}\right)T, \sigma\sqrt{T}\right].$$

Where S_t is the stock price at time t. Then by Definition **??** we have that,

$$\mathbb{E}(S_T) = S_0 \mathrm{e}^{\mu T}$$
$$\mathrm{Var}(S_T) = S_0^2 \mathrm{e}^{2\mu T}\left(\mathrm{e}^{\sigma^2 T} - 1\right).$$

2.7 The Black-Scholes Differential Equation

Using this model for stock prices, we may derive the Blakc-Scholes differential equation. Given that f is an option subject to S, then f must be some function of S and t. Hence, from (5),

$$\mathrm{d}f = \left(\frac{\partial f}{\partial S} + \frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right)\mathrm{d}t + \frac{\partial f}{\partial S}\sigma S \mathrm{d}W.$$

The equations (4) and the above have discretized versions,

$$\Delta S = \mu S \Delta t + \sigma S \Delta W. \tag{6}$$

$$\Delta f = \left(\frac{\partial f}{\partial S} + \frac{\partial f}{\partial t} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2}\right)\Delta t + \frac{\partial f}{\partial S}\sigma S \Delta W, \tag{7}$$

over a time interval Δt. As the Wiener processes contained in Δf and ΔS are the same, it follows that we may construct a portfolio to eliminate this. Such a portfolio should sell an option and buy $\frac{\partial f}{\partial S}$ shares. Then by definition our portfolio, Π, is,

$$\Pi = -f + \frac{\partial f}{\partial S} S. \tag{8}$$

The change in this over Δt is,

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

By substituting in (6) and (7) we obtain,

$$\Delta \Pi = \left(\frac{\partial f}{\partial t} - \frac{1}{2}\frac{\partial^2 f}{\partial S^2}\sigma^2 S^2\right)\Delta t \tag{9}$$

Over the time period Δt we have eliminated ΔW, so the portfolio must be riskless in this time period and must therefore make the riskfree interest rate. Thus,

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

substituting (8) and (9) into the above, we yield,

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

so that,

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

The above is known as the Black-Scholes differential equation. It is solvable for some boundary conditions and unsolvable analytically for others. In particular this is solvable for European options, which we investigate in the next section.

2.8 Black-Scholes Formula for European Options

Theorem 2.2 (Black-Scholes Pricing Formula)**.** *The value of a call option, f_c, and a put option, f_p, are given by the following formulae;*

$$f_c = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2),$$
$$f_p = Ke^{-rT}\Phi(-d_2) - S_0\Phi(-d_1),$$

where

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}, \tag{10}$$
$$d_2 = \frac{\log(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma.\sqrt{T}. \tag{11}$$

Before proving this we prove the following claim.

Claim:. If V is lognormally distributed and the standard deviation of $\log(V)$ is ω, then,

$$\mathbb{E}(\max(V-K,0)) = \mathbb{E}(V)N(d_1) - KN(d_2),$$

where $\mathbb{E}$ denotes the expected value, and we have that,

$$d_1 = \frac{\log\left(\frac{\mathbb{E}(V)}{K}\right) + \frac{\omega^2}{2}}{\omega}$$
$$d_2 = \frac{\log\left(\frac{\mathbb{E}(V)}{K}\right) - \frac{\omega^2}{2}}{\omega}.$$

Proof of claim. Define $h(V)$ to be the probability density function of V. Then we must have that,

$$\mathbb{E}(\max(V-K,0)) = \int_0^\infty \max(V-K,0)h(V)\mathrm{d}V \tag{12}$$
$$= \int_K^\infty (V-K)h(V)\mathrm{d}V. \tag{13}$$

By assumption the variable $\log(V)$ is normally distributed with standard deviation ω. Then from Definition **??** we have that the mean, m, is given by,

$$m = \log(\mathbb{E}(V)) - \frac{\omega^2}{2}.$$

We further define a new variable, W, by the following,

$$W = \frac{\log(V) - m}{\omega}. \tag{14}$$

This is the transformation that turns the distribution of $\log(V)$ to the standard normal distribution. Let the probability distribution function of W be $g(W)$, so that,

$$g(W) = \frac{1}{\sqrt{2\pi}}e^{\frac{-W^2}{2}}.$$

Using (14) as a change of variable for (13) we obtain that,

$$\mathbb{E}(\max(V-K,0)) = \int_{\frac{\log(K)-m}{\omega}}^{\infty} \left(e^{Q\omega+m} - K\right) h(Q)\mathrm{d}Q \tag{15}$$

$$= \int_{\frac{\log(K)-m}{\omega}}^{\infty} e^{Q\omega+m} h(Q)\mathrm{d}x - K\int_{\frac{\log(K)-m}{\omega}}^{\infty} h(Q)\mathrm{d}Q. \tag{16}$$

To solve this first consider,

$$e^{Q\omega+m}h(Q) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(Q-\omega)^2+2m+\omega^2}{2}\right)$$
$$= e^{m+\frac{\omega^2}{2}}h(Q-\omega),$$

thus we now have that (16) is,

$$\mathbb{E}(\max(V-K,0)) = e^{m+\frac{\omega^2}{2}}\int_{\frac{\log(K)-m}{\omega}}^{\infty} h(Q-\omega)\mathrm{d}Q - K\int_{\frac{\log(K)-m}{\omega}}^{\infty} h(Q)\mathrm{d}Q. \tag{17}$$

In (17) the first integral, the integrand is a normal distribution with a shifted mean. Thus it must be a function of cumulative normal distribution as we are summing the area under the probability distribution function, giving that,

$$\int_{\frac{\log(K)-m}{\omega}}^{\infty} h(Q-\omega)\mathrm{d}Q = [\,\Phi(Q-\omega)]_{\frac{\log(K)-m}{\omega}}^{\infty}$$
$$= 1 - \Phi\left(\frac{\log(K)-m}{\omega} - \omega\right)$$
$$= \Phi\left(\frac{-\log(K)+m}{\omega} + \omega\right).$$

Substituting for m we obtain,

$$\Phi\left(\frac{-\log(K)+m}{\omega} + \omega\right) = \Phi\left(\frac{\log\left(\frac{\mathbb{E}(V)}{K}\right) + \frac{\omega^2}{2}}{\omega}\right) = \Phi(d_1).$$

Similarly we obtain that the second integral in (17) is $\Phi(d_2)$. Thus we have that,

$$\mathbb{E}(\max(V-K,0)) = e^{m+\frac{\omega^2}{2}}\Phi(d_1) - K\Phi(d_2).$$

Substituting for m and we obtain our required result. □

Proof of Black-Scholes Formula. We will prove this for a call option. The proof for a put follows similarly using a similarly proved claim and a similar strategy for this proof.

Consider a call option on a non-dividend paying stock with time of expiry T (initial time $t = 0$), strike price K, risk-free interest rate r, current stock price S_0 and volatility σ. The value of such a call at T in a risk-neutral world would be,

$$\mathbb{E}(\max(S_T - K, 0)). \tag{18}$$

Note here this is the expected value in a risk neutral world and not necessarily the real world. For the rest of this proof all expected values will be as such.

Then as this is its value at time T, its value, c, would be (18) discounted to the initial time,

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

Then as we have shown the stochastic process underlying the stock is log normally distributed. Then at time $t = T$, we have S_T is log normally distributed, and from Section 2.6 $\mathbb{E}(S_T) = S_0 e^{rS_T}$, the standard deviation of $\log(S_T)$ is $\sigma\sqrt{T}$. Using our claim,

$$\begin{aligned} c &= e^{-rT}\left(S_0 e^{rT}\Phi(d_1) - \Phi(d_2)\right) \\ &= S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2). \end{aligned}$$

Further in this case,

$$\begin{aligned} d_1 &= \frac{\log\left(\frac{\mathbb{E}(S_T)}{K}\right) + \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}} = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \\ d_2 &= \frac{\log\left(\frac{\mathbb{E}(S_T)}{K}\right) - \frac{\sigma^2 T}{2}}{\sigma\sqrt{T}} = \frac{\log(S_0/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}}. \end{aligned}$$

□

3 Monte Carlo

Monte Carlo is the first numerical method we will look at for finding the price of options. The method is incredibly simple and relies on some of the ideas we built in the previous section. We shall develop these ideas. The concept is as follows;

1. Attempt to predict the path of the price of the underlying stock from $t = 0$ to $t = T$ where T is the expiry date of the option,
2. Evaluate the stock price at $t = T$,
3. Using the previous step, calculate the option price at maturity ($t = T$),
4. Repeat (1) - (3) a statistically significant number of times,
5. Calculate the average option price at maturity,
6. Discount the average option price by the interest rate to obtain the option price at $t = 0$.

For the first step we use the model for stock behavior we developed in (4). We may solve this with an application of Itôs lemma along with the $dW_t^2 = dt$ formula. Firstly note that by Itôs lemma we have that,

$$d(\ln(S_t)) = \frac{1}{S_t}dS_t - \frac{1}{2S_t^2}dS_t^2. \tag{19}$$

Then by using (4) with (19) and substituting in for dS_t we have that,

$$d(\ln(S_t)) = \frac{1}{S_t} S_t(\mu dt + \sigma W_t) - \frac{1}{2S_t^2} S_t^2(\sigma^2 dW_t^2)$$
$$= \mu dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt.$$

Then through exponentiation we have that,

$$S_t = S_0 \left[\exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma W_t \right) \right], \tag{20}$$

where S_0 is our initial stock price as before. Now note that as the Wiener process is normally distributed with mean zero and standard deviation $\sqrt{T}$ we may rewrite (20) as

$$S_t = S_0 \left[\exp\left((\mu - \frac{1}{2}\sigma^2)t + \sigma\sqrt{T}N \right) \right], \tag{21}$$

where N is a standardized normal random variable. Note here that by taking logarithms of the above we may again see that $\ln(S_t)$ is normally distributed.

Using (21) we may perform our first step by sampling for N to generate our path. This works well for European options however for Asian options as we are considering the average this may not generate a realistic average.

To compensate for this we split our interval into many different time points $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$. We sample at each of these time points to generate a path between $[t_i, t_{i+1}]$ for $i = 0, \ldots, n-1$. This allows us to generate more realistic paths.

The next step in the algorithm that was described above is very simple. However the third step changes and varies depending on what kind of option we are considering so we shall see how we adapt these methods for European, American and Asian options.

3.1 Monte Carlo for American Options

It is possible to adapt Monte Carlo to American options however it is very difficult. The problem arises from the possibility of early exercise, therefore we would need to find an optimal exercise rule. As we will develop other methods to price American options we will not consider this here.

At each timestep, the option holder has to make a (rational) choice – to exercise the option at the current time step or to hold on to the option. In fact, if the price of the American option (that would be obtained by holding onto the option) is smaller than the payoff, the option holder should exercise immediately. A generic algorithm for pricing an American option on an asset or underlying assets using a Monte Carlo approach would be:

1. Create random paths $S_j\ (t_i)$, $i = t_0 \ldots\ nstep$, $j = 1 \ldots npath$, $t_i = t_0 + i.\delta$t
2. For each path, set price at maturity to the payoff $V_j(T) = h(S_j(T))$
3. Start at maturity $i = n_{timestep}$
4. Go one step backwards in time $i = i$ - 1

5. For each path j, discount the price $V_j(t_i) = e^{\int_{t_i}^{t_{i+1}} r(s)ds} V_j(t_{i+1})$
6. For each path j, compute the continuation value c_j
7. For each path j with h_j ¿ c_j, exercise, i.e. $V_j(t_i) = h(S_j(t_i))$
8. If last timestep or timestep i = 1:
   (a) For each path j, discount the price: $V_j(t_0) = e^{\int_{t_0}^{t_1} r(s)ds} V_j(t_1)$
   (b) The price today is the average over all paths:

$$V(S(t_0)) = \frac{1}{n_{path}} \sum_{j=1}^{n_{path}} V_j t_0 \tag{22}$$

The basic idea is that after creating the random paths, the algorithm goes backwards in time: First, create the ensemble of random paths $S_j(t_i)$ with path label$S_j = 1...n_{path}$ and discrete time step t_i ,i= 0...T of length dt. Perform the following steps for each path: At maturity, set the value of the option to the payoff V_j(T) = $h(S_j(T))$. Now iterate backwards in time by first discounting the price along each path and compute the continuation value – i.e. the value of the option that the holder would achieve if they held on to the option in the current time step. Compare the former with the latter and decide for each path whether or not to exercise. Modify the price at this time step accordingly. Iterate these steps until the first time step. As it makes no sense to have the option holder exercise the option in the moment of buying it, compute the price of the option along each path for the initial time t_0 by simple discounting from the first time step t_1 to t_0 without comparing it to the continuation value. The price of the option at t_0 is then simply given by the average over all n_{path} paths. In principle, computing the continuation value c = V($S_j(t_i)$,ti)in step 6 requires a full revaluation of the option price at each time step with a new set of random paths. This strategy scales with the square of the number of paths, making it prohibitively expensive to compute.

3.2 Antithetic Variates

We will here explore a variance reduction technique, antithetic variates. In this method for each Y_i generated we create a corresponding $\hat{Y}_i$. Each pair $(Y_i, \hat{Y}_i)$ must be i.i.d.. Define the antithetic estimator as the average of these $2n$ replications, we then have the value of our estimator is of this is given by,

$$\begin{aligned} \overline{Y}^* &= \frac{1}{2n}\left(\sum_{i=1}^{n} Y_i + \sum_{i=1}^{n} \hat{Y}_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i + \hat{Y}_i}{2}\right). \end{aligned}$$

There are several tricks we may employ here to reduce computing time, one such trick is choosing $\hat{N}_i = -N_i$ for Gaussian variables (note that this is non-zero as we apply the payoff to obtain Y_i). These tricks are not always readily available thus in the worst case scenario to do this method we will require $2n$ replications. Thus we need to compare the new variance after the variance reduction

technique to the variance of if we used a normal method with $2n$ replications. It follows that this method is reduces variance if,

$$\text{Var}[\overline{Y}^*] < \text{Var}\left[\frac{1}{2n}\sum_{i=1}^{2n} Y_i\right]$$

thus the above implies that,

$$\text{Var}[Y_i + \hat{Y}_i] < 2\text{Var}[Y_i]. \tag{23}$$

We note that the left hand side can be written as,

$$\text{Var}[Y_i + \hat{Y}_i] = \text{Var}[Y_i] + \text{Var}[\hat{Y}_i] + 2\text{Covar}[Y_i, \hat{Y}_i] \tag{24}$$

$$= 2\text{Var}[Y_i] + 2\text{Covar}[Y_i, \hat{Y}_i], \tag{25}$$

as both Y_i and $\hat{Y}_i$ have the same distribution and therefore must have the same variance. So using (23) with (25) we see that the condition for antithetic variance reduction to be effective is,

$$\text{Covar}[Y_i, \hat{Y}_i] < 0,$$

for each path.

Now we may apply this to the field of option pricing. To do this we will use the trick we discussed earlier. The algorithm is as follows,

1. Simulate the price to generate the paths in the usual way as given in Section 3 and calculate their payoff,
2. Use the same random components and choose, for the new paths, that $N_i = -N_i$ where N_i is the ith random variable we simulated in step 1 and calculate the pay off of these new paths,
3. For each path generated in step 1 average it with the new path generated in step 2,
4. Take the average of these paths and divide by n to find the approximate value of the option.

3.3 The Longstaff-Schwartz Algorithm for American Options

The previous method described uses regression methods to compute the continuation value in step 6. As mentioned previously, computing the continuation value of an option is computationally rather expensive. Hence, we suggest approximating the continuation value at each timestep using a least squares regression over all n_{path} paths. At each time step t_i, expand the continuation value c(S) (as a function of the underlying asset price) in terms of a function basis ψ_j.

$$c(x, t_i) = E[V(S(t_{i+1}))S(t_i) = x] = \sum_{k=0}^{n_{order}} \beta_k \psi_k(x) \tag{26}$$

where the expansion coefficients β_k are obtained by a least squares fit to the (discounted)values of the option at the next time step:

$$\beta = (B_{\psi\psi})^{-1} B_{V\psi}, \tag{27}$$

where $B_{\psi\psi}$, $B_{V\psi}$ at time step t_i are given by:

$$(B_{V\psi})_l = E[V(S(t_{i+1})\psi_l(S(t_i))] \tag{28}$$

$$(B_{\psi\psi})kl) = E[\psi_k(S(t_i))\psi_l(S(t_i))] \tag{29}$$

Using this least squares regression for the continuation value results in the algorithm, proved almost sure convergence of this algorithm in the limit of infinitely many sample paths.In contrast to simulating the early exercise decision by setting the price:

$$V_j(t_i) = h(S_j(t_i)) \tag{30}$$

for all paths j where $h(S_j(t_i))$ ¿ $e^{\int_{t_i}^{t_{i+1}}} r(s)ds\ V_j(t_{i+1})$ in step 7, [26] assume that the non-exercised value is the continuation value:

$$V_j(t_i) = max[c(S_j(t_i)), h(S_j(t_i))] \tag{31}$$

which has the drawback of accumulating the sampling error from the difference between c. and V. Hence we keep using (27) instead Deciding on early exercise by comparing $c(S(t_i), t_i)$ to the immediate payoff $h(S(t_i), t_i)$ typically results in a sub-optimal exercise pattern. Hence the resulting price of the option today is a lower bound on the option's true price. This approach is trading the complexity of running a full Monte Carlo valuation for each path at each time step for a least squares fit to a set of basis functions with $n_{asset} \times n_{orderterms}$. The order of expansion n_{order} and the set of basis functions$\psi_j$should be tuned to each option under investigation, incorporating any prior knowledge about the option at hand.Employing the Longstaff-Schwartz algorithm as described above, the continuation value can violate no-arbitrage bounds. A generic algorithm for pricing an American option on an asset or underlying assets using a Monte Carlo approach would be:

1. Create random paths $S_j\ (t_i)$, $i = t_0 ...\ nstep$, $j = 1 ... npath$, $t_i = t_0 + i.\delta$t
2. For each path, set price at maturity to the payoff $V_j(T) = h(S_j(T))$
3. Start at maturity $i = n_{timestep}$
4. Go one step backwards in time $i = i$ - 1
5. For each path j, discount the price $V_j(t_i) = e^{\int_{t_i}^{t_{i+1}} r(s)ds}\ V_j\ (t_{i+1})$
6. For each path j, compute the continuation value c_j
   - Solve for the regression coefficients β

     $$\beta = (B_{\psi\psi})^{-1} B_{V\psi}, \tag{32}$$

   - For each path j, compute the continuation value c_j=$\sum_{k=0}^{n_{order}} \beta_k\ \psi_k\ (S_j(t_i))$
7. For each path j with h_j ¿ c_j, exercise, i.e. $V_j(t_i) = h(S_j(t_i))$
8. If last timestep or timestep i = 1:
   (a) For each path j, discount the price: $V_j(t_0) = e^{\int_{t_0}^{t_1} r(s)ds}\ V_j\ (t_1)$
   (b) The price today is the average over all paths:

$$V(S(t_0)) = \frac{1}{n_{path}} \sum_{j=1}^{n_{path}} V_j t_0 \tag{33}$$

3.4 Basis functions and methodology

Every minute prior to the expiration date of an American option, the option holder must choose the optimal outcome between exercising the option or keeping it for an-other term. The choice is based on the immediate exercise payoff and the future payoff expected. If the former is greater than the latter, then the option is exercised. The idea of the Least-Squares Monte Carlo Method is to approximate the continuation value by using least-squares regression at every moment in which it is possible to exercise the option. The regression is done based on the data obtained for the state variables via Monte Carlo simulation and by choosing the trajectories where the option is in the money.

Within this context, whenever it is possible to exercise the option, the continuation value can be expressed as a linear combination of orthogonal basis functions,such as Power, Chebychev Legendre and Laguerre polynomials. This follows from the Finance Literature that considers the payoff functions that belong to the function space of finite variance.Since this is a Hilbert space,any function F that belongs to this space can be written as a linear combination of orthogonal basis functions. Thus, the same function F can be rewritten as:

$$c(x, t_i) = \sum_{k=0}^{n_{order}} \beta_k \psi_k(x) \tag{34}$$

where the polynomial basis, represented by ψ_j, is a function of one of the state variables. Note that the coefficients of the basis, β_j, are not previously known, but they can be estimated by linear regression. When considering the number of basis to be used, it is important to point out a problem that the use of a large number of basis to conclude that several degrees of the chosen polynomial basis and its respective crossed products, can increase the accuracy of the estimation. In some cases,however, an excessive increase in the number of basis can reduce the precision of the method, making it computationally expensive. The polynomials used in this study are Power, Legendre, Laguerre, Chebychev and Hermite. All of them can be alternatively expressed by Rodrigues'formula, explicit form or by the recurrence law:

$$f_n(x) = \frac{1}{a_n.\rho(x)} \frac{\partial^n}{\partial x^n}[\rho(x).(g(x))^n] \tag{35}$$

Where n is the polynomial degree (n≥0). The coefficients and functions of Rodrigues'formula for each of the polynomials are detailed in the table 2:

Polynomial	$f_n(x)$	a_n	$\rho(x)$	g(x)
Power	$W_n(x)$	$\frac{(2n)!}{n!}$	$x^(2n)$	1
Legendre	$P_n(x)$	$(-1)^n.2^n.n!$	1	$1-x^2$
Laguerre	$L_n(x)$	$n!$	e^{-x}	x
Hermite	$H_n(x)$	$(-1)^n$	e^{-x^2}	1
Chebychev 1	$C_{n1}(x)$	$\frac{1}{\sqrt{\pi}}(-1)^n.2^n.(n-\frac{1}{2})!$	$(1-x^2)^{-\frac{1}{2}}$	$(1-x^2)$
Chebychev 2	$C_{n2}(x)$	$\frac{1}{(n+1)\sqrt{\pi}}(-1)^n.2^{n+1}.(n+\frac{1}{2})!$	$(1-x^2)^{\frac{1}{2}}$	$(1-x^2)$

Table 2: Coefficients and functions of the basis functions using Rodrigues'formula

Alternatively, it is possible to use the explicit form, whose terms are specified in Table 3, to

represent the polynomials:

$$f_n = d_n \sum_{m=0}^{N} c_m . g_m(x) \tag{36}$$

Polynomial	$f_n(x)$	N	d_n	c_m	$g_m(x)$
Power	$W_n(x)$	0	1	1	x^n
Legendre	$P_n(x)$	[n/2]	2^{-n}	$(-1)^m . \binom{n}{m} . \binom{(2.n-2.m)}{n}$	x^{n-2m}
Laguerre	$L_n(x)$	n	1	$\frac{(-1)^m}{m!} \cdot \binom{n}{(n-m)}$	x^m
Hermite	$H_n(x)$	$\frac{n}{2}$	$n!$	$(-1)^m \frac{1}{m!(n-2m)!}$	$(2x)^{n-2m}$
Chebychev 1	$C_{n2}(x)$	[n/2]	1	$\frac{(-1)^m}{(n-m)} \binom{(n-m)}{m}$	$(2x)^{n-2m}$
Chebychev 2	$C_{n2}(x)$	[n/2]	1	$(-1)^m \binom{(n-m)}{m}$	$(2x)^{n-2m}$

Table 3: Explicit expressions of basis functions

The recurrence law can also be used to express the polynomials and it is described as:

$$a_{n+1} . f_{n+1} = (a_n + b_n . x) . f_n(x) - a_{n-1} . f_{n-1}(x) \tag{37}$$

Table 3 specifies every term of the recurrence law. The format used to write a polynomial is chosen in such a way that the pricing procedure becomes operationally more practical. This method proved to be effective throughout the whole process, excluding the possibility that the rank of the coefficient matrix of the least square regression is insufficient

3.5 Why Least Squares - Pros and Cons

We use ordinary least squares to estimate the conditional expectation function. In some cases, however, it may be more efficient to use other techniques such as weighted least squares. generalized least squares, or even Gaussian Mixture Models in estimating the conditional expectation function. For example, if the process for the state variables has state dependent volatility, the residuals from the regression may be heteroskedastic and these alternative least squares techniques may have advantages. In estimating the least squares regressions, it may be noted that the R^2s from the regressions are often somewhat low. The reason for this is simply the volatility of realized cash flows around their conditional expected values. The R^2 from the regression measures the percentage of the total variation in the ex post cash flows explained by the variation in the conditional expectation function; a low R^2 simply means that the volatility of unexpected cash flows is large relative to the volatility of expected cash flows. Thus low R^2 are to be expected when unexpected cash flows are highly volatile. In general, since the LSM algorithm is based on conditional first moments rather than second moments, the R^2s from the regression should have little impact on the quality of the LSM approximation to the American option value.

3.6 Choice of basis functions

Extensive numerical tests indicate that the results from the LSM algorithm are remarkably robust to the choice of basis functions. For example, we use the first Laguerre or Chebychev polynomials as basis functions in the American put, we obtain results that are virtually identical to those each other

Similarly, when we use S, S', and S3 as basis functions, when we use the Hermite polynomials as basis functions the results are identical as well. As reported earlier, few basis functions are needed to closely approximate the conditional expectation function over the relevant range where early exercise may be optimal. While the results are robust to the choice of basis functions, it is important to be aware of the numerical implications of the choice. For example, the weighted Laguerre polynomials used in the American put include an exponential term in the stock price S. Thus, directly applying the weighted Laguerre polynomials to the problem could result in computational underflows. To avoid this problem, we can renormalized the American put example by dividing all cash flows and prices by the strike price, and estimating the conditional expectation function in the renormalized space. Note, that this is only for numerical convenience; the option value is unaffected since we discount back the un-normalized value of the cash flows along each path to obtain its value. We recommend normalizing appropriately to avoid numerical errors resulting from scaling problems. Finally, the choice of basis functions also has implications for the statistical significance of individual basis functions in the regression. In particular, some choices of basis functions are highly correlated with each other, resulting in estimation difficulties for individual regression coefficients akin to the multicolinearity problem in econometrics. This difficulty does not affect the LSM algorithm since the focus is on the fitted value of the regression rather than on individual coefficients; the fitted value of the regression is unaffected by the degree of correlation among the explanatory variables. However, if the choice of basis functions leads to a cross-moment matrix that is nearly singular, then numerical inaccuracies in some least squares algorithms may affect the functional form of estimated conditional expectation function. To avoid these types of numerical problems, the regressions are estimated using the double-precision algorithm in LSM which estimates least squares via an iterative-refinement algorithm. We also cross-checked the results by estimating the regressions using a variety of alternative procedures such as Cholesky-decomposition and QR-algorithm least squares techniques.

4 Simulated Annealing

Simulated annealing provides a methodology for using Monte Carlo where instead of treating the problem of finding of exercise boundary as a free boundary-problem (like in Black-Scholes method), we use Monte-Carlo via a genetic algorithm approach such as Simulated Annealing) to find the solution of the maximization problem itself. We give here the outline of a simulated annealing algorithm. A state of the system in this case is a 6–component vector of positive real numbers:

$$x = (a_1, b_1, c_1, a_2, b_2, c_2) \tag{38}$$

hese components are used in the following equation to form the exercise boundary.

$$B(\tau) = \frac{K - S_\tau}{K} = a_1 ln(b_1.\tau^{c_1} + 1) + a_2 ln(b_2\tau^{c_2} + 1), 0 \leq \tau \leq T \tag{39}$$

Here τ is the time remaining to expiration. The energy of the system is a function of x and can be, for example:

$$E_x = \frac{1}{option\_price} \tag{40}$$

The option price will be maximized by minimizing the system's energy. An important parameter of an anneal is the system temperature T. The temperature cools from some high value at the start

of the algorithm to a low value at the end. Thus T is a function of algorithm time t as measured by iterations in the algorithm's main loop. Two common cooling schedules are geometric:

$$T = ab^t, a > 0, 0 < b < 1 \tag{41}$$

and logarithmic:

$$T = \frac{c}{ln(1+t)}, c > 0 \tag{42}$$

The parameters a and b or c are parameters of the anneal. Usually preliminary trials of the algorithm are performed in order to get an idea as to the size of and, working from there together with knowing how many trials will be allocated, appropriate values for them are determined. For example, theoretically c should be equal to the largest value can have. Knowing the number of trials determines the size of p near the end of the run; p should be less than 5 % at the end. (If geometric cooling is used, b should be quite close to 1, e.g. b=0.9999 or larger, otherwise p tends to zero too quickly.)

In addition, each state x must define a neighborhood $N(x)$ about itself defining what it means to be close to x. New trials of the algorithm use points close to the current trial. An appropriate neighborhood for this problem is the set of vectors $y = (a_1', b_1', c_1', a_2', b_2', c_2')$ such that$||yx|| < \epsilon$ for some ϵ and some vector norm, for example L_1

$$|a_1' - a_1| + |b_1' - b_1| + ... + |c_2' - c_2| < \epsilon \tag{43}$$

Hence the algorithm is given as:

4.1 Simulated Annealing algorithm

Algorithm 1 Simulated Annealing algorithm

1: **procedure** SIMULATED_ANNEALING(a,b, neighborhoods $N(x)$, and the number of iterations to perform, n)
2: initialize x randomly and evaluate E_x
3: initialize temperature T
4: **for** $t = 1, ..., n$ **do**
5: from $N(x)$choose y at random and evaluate E_y
6: put $\Delta E = E_y$ - E_x
7: with probability $p = e^{\Delta E/T}$ replace x by y
8: otherwise keep x
9: update the temperature, $T = ab^t$
10: report the option price $\frac{1}{E_x}$
11: (and the maximizing state x if desired)

The key step in the algorithm is the line in which the old state x is replaced by the new one y with probability p(as achieved by drawing a computer random number r and replacing if (r¡p). If the new energy E_y is less than the old, then p as calculated is greater than 1 and the new state is accepted. But the new state can also be accepted even if the new state is worse. And this will happen with greater probability at high temperature than low. In this way the algorithm does not become trapped in local minima.

5 Results

5.1 Testing configurations

Initially we started our testing for the following configuration:-

Parameter	Value
S_0	40
r	0.06
σ	0.5
T	1 year
n_{paths}	100,000
δT	$\frac{1}{250}$
K	40

Table 4: Initial Conditions (We assumed 250 trading days in an year)

For the aforementioned testing configuration, we got the following results, for each of the following polynomial basis functions:

Polynomial	Prices	Variance
Power	6.8034	0.0229
Legendre	6.8036	0.0226
Laguerre	6.8051	0.0225
Hermite	6.8109	0.0229
Chebychev 1	6.8272	0.0227
Chebychev 2	6.8112	0.0226

Table 5: Option prices for initial testing configuration

We use the simualated annealing methodology to set our benchmarks, and the stock price obtained is 5.8711. We use this as our lower bound and proceed to perform tests in multiple configurations, which is listed in our Appendix section.

5.2 Performance

We ran the Monte-Carlo simulation for 100,000 runs over 1000 times and averaged the execution times to a get a measure of performance. As expected, Hermite and Power polynomials perform the best as the polynomial terms in these two polynomials require the least number of mathematical operations. Also as expected the Chebychev polynomials (first and second kind) very similar times as the polynomial have very similar mathematical forms for both:

Polynomial	Results
Power	24.1780
Legendre	29.1276
Laguerre	27.8132
Hermite	25.8838
Chebychev 1	29.1822
Chebychev 2	29.1097

Table 6: Performance times averaged over 100,000 runs for 1000 times

5.3 Graphs

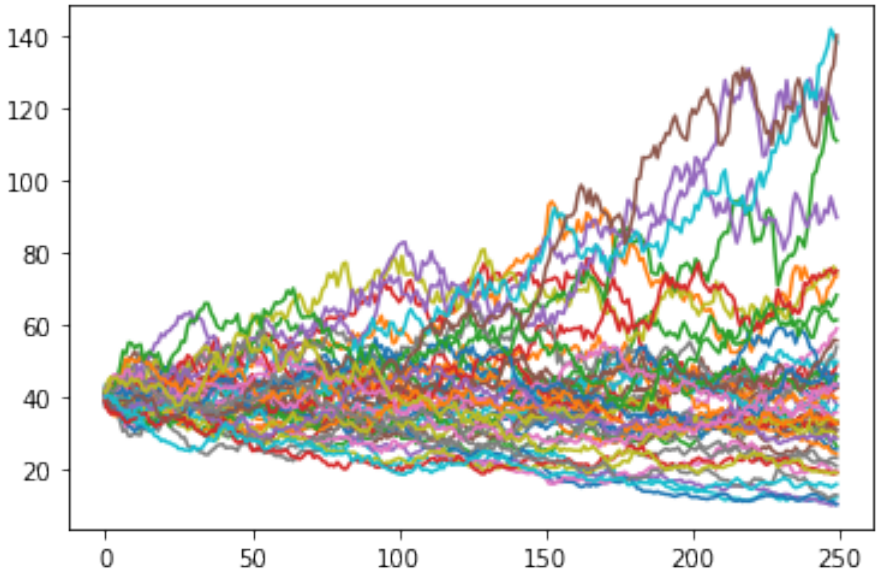
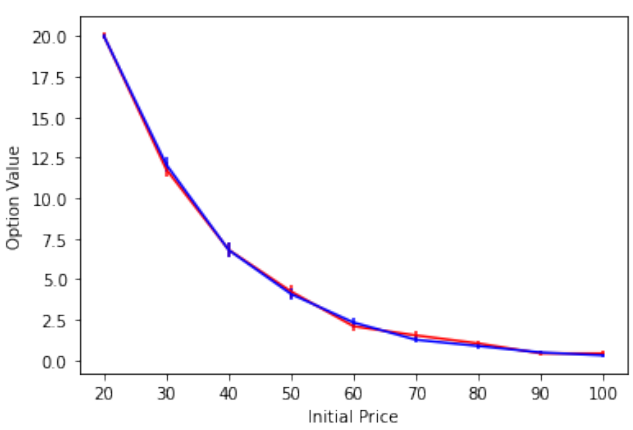
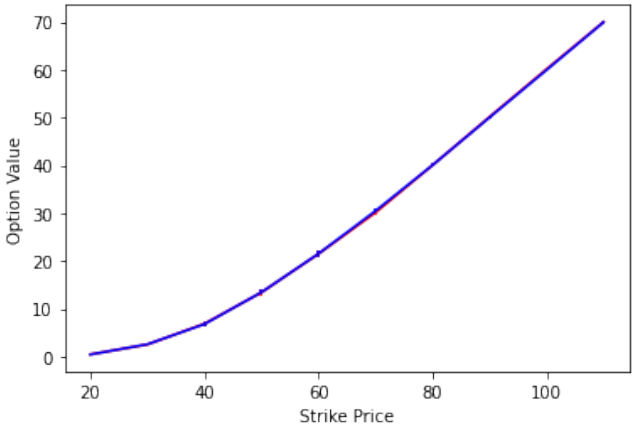


Figure 1: 100 replications of Stock price movements

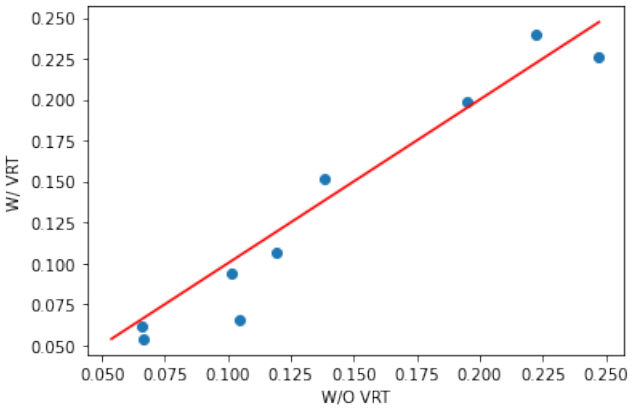


(a) Intitial Stock Price S_0 Vs. the Option Price after 100k runs

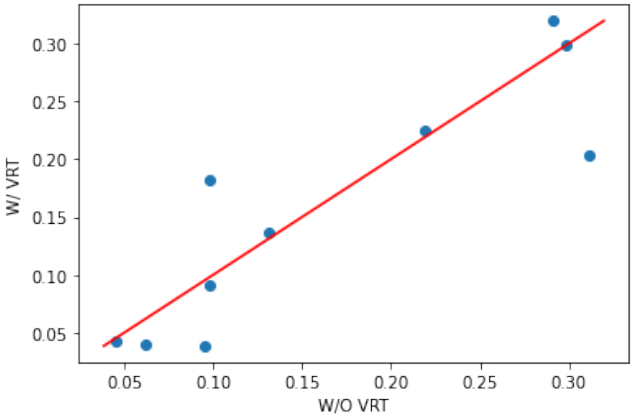


(b) Strike Price of the option S_0 Vs. the Option Price after 100k runs

Figure 2: Comparison Plots: Initial Price and Strike Price

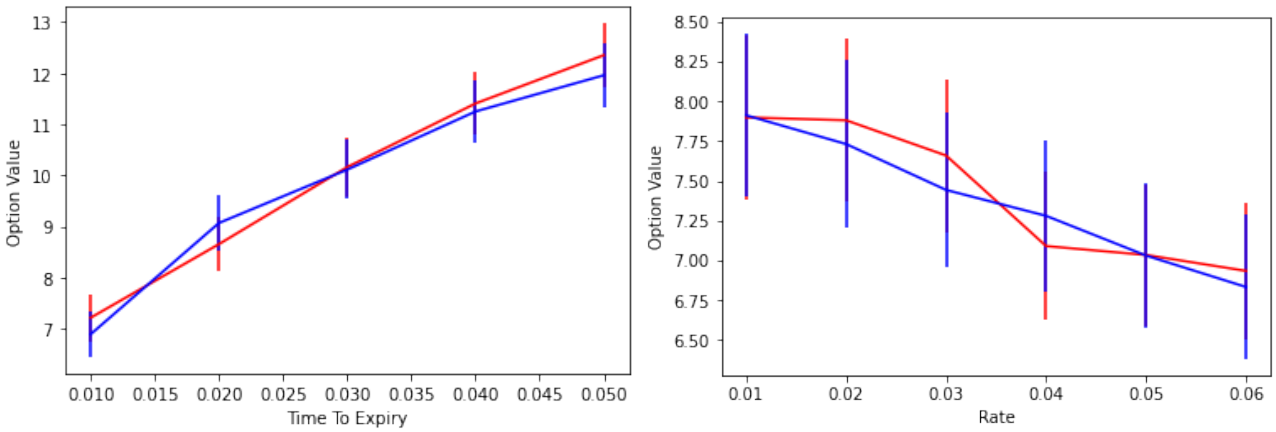


(a) Use of VRT for initial Stock Price vs Option Price



(b) Use of VRT for strike Price vs Option Price

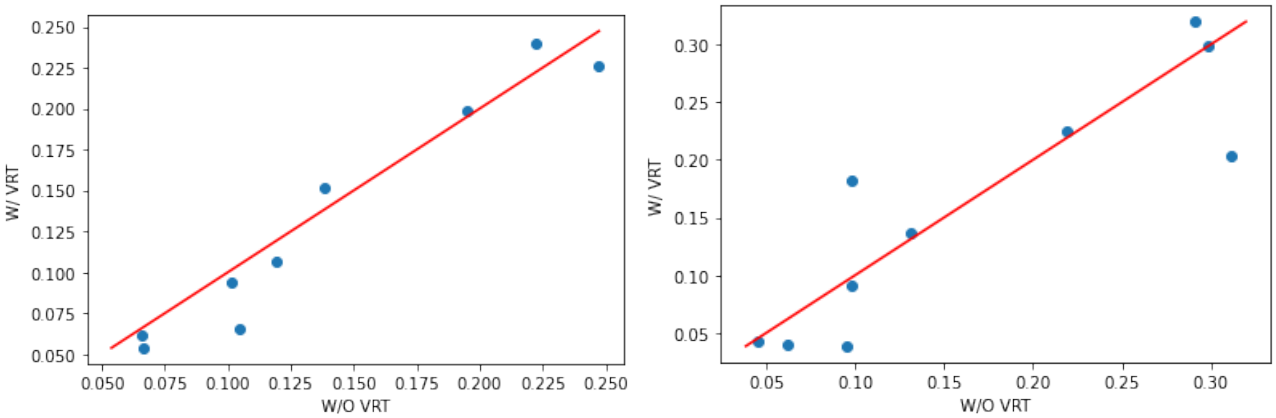
Figure 3: VRT comparison: Initial Stock Price and Strike Price



(a) Time to Expiry S_0 Vs. the Option Price after 100k runs

(b) Risk Free Rate S_0 Vs. the Option Price after 100k runs

Figure 4: Comparison Plots: Time To expiry and Risk Free Rate



(a) Use of VRT for Time to expiry vs Option Price

(b) Use of VRT for Risk Free vs Option Price

Figure 5: VRT comparison: Time To expiry and Risk Free Rate

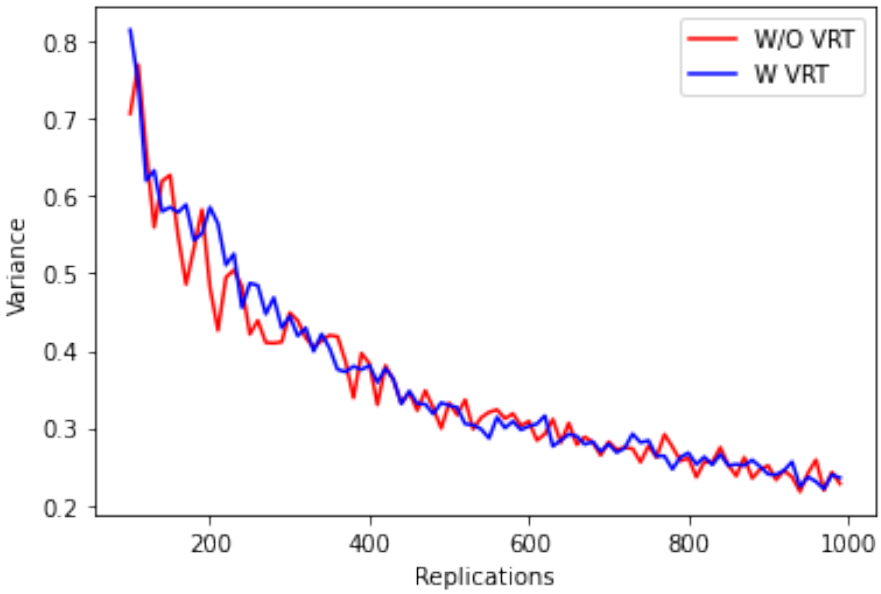
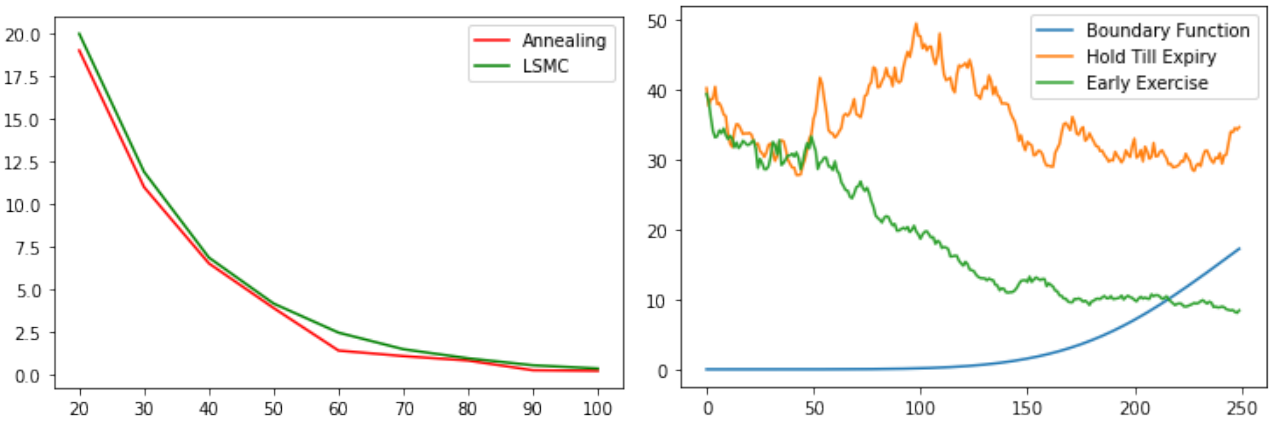


Figure 6: Variance vs Replications



(a) LSM algorithm compared to simulated annealing

(b) Approximate boundary function

Figure 7: Simulated Annealing Benchmark

Figure 1 shows the sample paths generated for a single monte-carlo execution. Figure 2 (a.) shows the plot of Comparison of Initial Stock Price vs option value which is to be expected as we are pricing an American Put. The lower the stock price, the higher should be the option value due to greater chance of execution and vice versa. Figure 2 (b.) shows the plot of Comparison of Strike Price vs option value which is to be again expected as we are pricing an American Put. The higher the strike price for a given value of S_0, the higher should be the option value due to greater chance of execution (As the value $K - S_0$ increases when K increases and vice versa. Figure 3 (a.) shows the plot to compare the variance(error) in the option price with and without the use of antithetic variables plotted against initial stock price. Figure 3 (b.) studies the variance in the option price with and without the use

of antithetic variable against the strike price of the option. In both the cases we see the results are fluctuating and this could be attributed to relatively high precision of the LSMC algorithm. Figure 4 (a.) (b.) studies the option price against Time to Expiry and the Risk Free rate. Figure 5 (a.) (b.) study the variance with and without the use of antithetic variable plotted against Time to Expiry and Risk Free rate. The fluctuations in variance reduction is because of the high precision of the LSMC algorithm. Figure 6 studies the variance of the option price as the number of replications increase. Figure 7 (a.) compares the results of the Simulated annealing against the LSMC algorithm. LSMC gives a good estimate of the option price in a highly computationally efficient manner. Figure 7 (b.) shows the exercise boundary as estimated using the simulated annealing. Whenever the stock price crosses the boundary the option is exercised.

6 Conclusion

Our experiments using the six chosen polynomial basis ratified one of the method's elements, i.e., that any orthogonal polynomial basis can be used for pricing American put options, since in all of the simulated paths the estimated prices of the option sare virtually the same. For all of the American puts priced, and for each of the basis used, the convergence of the price estimate results from the decrease of the standard deviation while the number of simulated paths increases. Considering 100,000 simulated trajectories, our study suggests that it is possible to empirically choose one specific polynomial basis for pricing an American put.

The results show that one polynomial basis is marginally more accurate than others, by providing a lower standard deviation when we compared it other it delivers a better performance to the algorithm when pricing a complex option as American options. Such empirical outcome is theoretically unpredictable, since in principle all basis can be indistinctly used when pricing the derivative. Therefore, for practical purposes, if a trader is looking for a faster and more accurate way to valuing an American put option, the polynomial basis suggested to perform this procedure is using Hermite or Power polynomials. Due to the considerably less mathematically less operations it is much faster to execute Hermite or Power polynomials without much loss in accuracy.

As a framework for valuing and risk managing derivatives, simulation has many important advantages. With the ability to value American options, the applicability of simulation techniques becomes much broader and more promising, particularly in markets with multiple factors. Furthermore, simulation techniques make it much easier to implement advanced models such as Heath, Jarrow, and Morton or Santa-Clara and Sornette in practice.

7 Appendix

Following are tables for different configuration that we ran the Least squares monte carlo algorithm for:-

For T = 0.5 (125 trading days):

	$\sigma_1 = 0.4$	$\sigma_1 = 0.5$	$\sigma_1 = 0.6$
$S_0 = 20$	19.99043	19.99041	19.98759
$S_0 = 40$	3.97847	5.059122	6.15192
$S_0 = 60$	0.36627	0.90201	1.61503

Table 7: Initial Conditions (We assumed 250 trading days in an year)

For T = 1 (250 trading days):

	$\sigma_1 = 0.4$	$\sigma_1 = 0.5$	$\sigma_1 = 0.6$
$S_0 = 20$	19.99038	19.98611	20.06997
$S_0 = 40$	5.30142	6.80505	8.27332
$S_0 = 60$	1.17150	2.26312	3.56218

Table 8: Initial Conditions (We assumed 250 trading days in an year)

For T = 1.5 (375 trading days):

	$\sigma_1 = 0.4$	$\sigma_1 = 0.5$	$\sigma_1 = 0.6$
$S_0 = 20$	19.99039	19.96093	20.34277
$S_0 = 40$	6.18723	7.98582	9.75557
$S_0 = 60$	1.92530	3.43873	5.11212

Table 9: Initial Conditions (We assumed 250 trading days in an year)

8 References

The following papers/books were used as references: [6], [2], [3], [10], [11], [1], [7], [4], [5], [8], [9], [12]

9 Bibliography

[1] Mckean h. (1965) appendix a free boundary problem for the heat equation arising from a problem in mathematical economics. industrial management review 6 32-39. - references - scientific research publishing.

[2] Clément Emmanuelle, Lamberton Damien, and Protter Philip. An analysis of a least squares regression method for american option pricing. *Finance and Stochastics*, 6:449–471, 10 2002.

[3] Paul Glasserman. Monte carlo methods in financial engineering — paul glasserman — springer, 2019.

[4] Peskir Goran and Shiryaev Albert N. *Optimal Stopping and Free-Boundary Problems*. Birkhäuser Basel, 2006.

[5] Stentoft Lars. Convergence of the least squares monte carlo approach to american option valuation. *Management Science*, 50:1193–1203, 2004.

[6] Francis A. Longstaff and Eduardo S. Schwartz. Valuing american options by simulation: A simple least-squares approach. *Review of Financial Studies*, pages 113–147, 2001.

[7] Moreno Manuel and Navas Javier F. On the robustness of least-squaresmonte carlo (lsm) for pricing american options. *Review of Derivatives Research*, 6:107–128, 2003.

[8] Mostovyi Oleksii. On the stability the least squares monte carlo. *arXiv:1102.3218 [q-fin]*, 07 2011.

[9] Boyle Phelim P. Options: A monte carlo approach. *Journal of Financial Economics*, 4:323–338, 05 1977.

[10] Glasserman Paul and Yu Bin. Number of paths versus number of basis functions in american option pricing 1. *The Annals of Applied Probability*, 14:2090–2119, 2004.

[11] S Shreve and I Karatzas. Brownian motion and stochastic calculus, 2019.

[12] Shonkwiler Ronald W. *Finance with Monte Carlo*. Springer-Verlag, 2013.