Pricing and Hedging Derivative Securities: Theory and Methods

Kerry Back   Steven L. Heston   Hong Liu   Mark Loewenstein

Table of contents

Preface

This is a sequel to *A Course in Derivative Securities: Introduction to Theory and Computation* published by Springer. It is currently a work in progress.

1 Calls and Puts

Financial options are rights to buy and sell assets at pre-specified prices. The rights are traded on exchanges and also as private contracts (called over-the-counter or OTC). A call option is a right to buy an asset. A put option is a right to sell an asset. The pre-specified price is called the exercise price, the strike price, or simply the strike. The asset to which an option pertains is called the underlying asset, or, more briefly, the underlying.

Here is a snapshot of market statistics regarding call options on Apple stock (AAPL) traded on the Chicago Board Options Exchange (CBOE). We are accessing the data courtesy of Yahoo Finance. We will refer to this data throughout the chapter.

```
import pandas as pd
import yfinance as yf
from datetime import datetime
import pytz
from datetime import datetime

est = pytz.timezone('US/Eastern')
fmt = '%Y-%m-%d %H:%M:%S %Z%z'
now = datetime.today().astimezone(est).strftime(fmt)

ticker = "AAPL"         # ticker to pull
kind = "call"           # call or put
maturity = 4            # option maturity in order of maturities trading

tick = yf.Ticker(ticker.upper())

# Pull last stock price
close = tick.history().iloc[-1].Close

# Get maturity date
date = tick.options[maturity]

# Pull options data
df = (
    tick.option_chain(date).calls
```

```python
    if kind == "call"
    else tick.option_chain(date).puts
)

df.lastTradeDate = df.lastTradeDate.map(
    lambda x: x.astimezone(est).strftime(fmt)
)

# Formatting
cols = [
    "strike",
    "bid",
    "ask",
    "lastPrice",
    "change",
    "percentChange",
    "lastTradeDate",
    "volume",
    "openInterest",
    "impliedVolatility",
]
df = df[cols]
df["impliedVolatility"] = df["impliedVolatility"].map("{:.1%}".format)
df["change"] = df["change"].round(2)
df["percentChange"] = (df["percentChange"]/100).map("{:.1%}".format)
df.columns = [
    "Strike",
    "Bid",
    "Ask",
    "Last Price",
    "Change",
    "% Change",
    "Time of Last Trade",
    "Volume",
    "Open Interest",
    "Implied Volatility",
]
df = df.set_index("Strike")
print(f"Code was executed at \t{now}")
print(f"Last {ticker.upper()} price was \t${close:.2f}.")
print(f'Maturity date of options:\t{date}')
pd.set_option('display.max_rows', 500)
```

```
pd.set_option('display.max_columns', 500)
pd.set_option('display.width', 1000)
print(df)
```

```
Code was executed at    2024-09-27 09:56:05 EDT-0400
Last AAPL price was     $229.09.
Maturity date of options:   2024-10-25
        Bid  Ask  Last Price  Change % Change            Time of Last Trade  Volume  Open In
Strike
100.0   0.0  0.0      130.57     0.0     0.0%  2024-09-20 15:55:52 EDT-0400     1.0
140.0   0.0  0.0       85.28     0.0     0.0%  2024-09-25 14:27:29 EDT-0400     1.0
145.0   0.0  0.0       89.11     0.0     0.0%  2024-09-20 15:42:03 EDT-0400     5.0
150.0   0.0  0.0       76.00     0.0     0.0%  2024-09-25 15:28:50 EDT-0400     3.0
160.0   0.0  0.0       63.21     0.0     0.0%  2024-09-12 12:02:27 EDT-0400     NaN
165.0   0.0  0.0       55.91     0.0     0.0%  2024-09-09 10:05:45 EDT-0400     NaN
170.0   0.0  0.0       62.96     0.0     0.0%  2024-09-20 15:52:58 EDT-0400     6.0
175.0   0.0  0.0       55.17     0.0     0.0%  2024-09-19 13:22:14 EDT-0400     2.0
180.0   0.0  0.0       48.14     0.0     0.0%  2024-09-26 10:36:07 EDT-0400     1.0
185.0   0.0  0.0       41.00     0.0     0.0%  2024-09-25 13:47:05 EDT-0400     2.0
190.0   0.0  0.0       38.78     0.0     0.0%  2024-09-26 15:18:33 EDT-0400     1.0
195.0   0.0  0.0       33.36     0.0     0.0%  2024-09-26 14:13:45 EDT-0400     2.0
200.0   0.0  0.0       29.06     0.0     0.0%  2024-09-26 15:59:19 EDT-0400     5.0
205.0   0.0  0.0       22.60     0.0     0.0%  2024-09-26 09:41:51 EDT-0400     2.0
210.0   0.0  0.0       19.50     0.0     0.0%  2024-09-26 11:10:34 EDT-0400   180.0
215.0   0.0  0.0       14.65     0.0     0.0%  2024-09-26 15:54:40 EDT-0400   497.0
220.0   0.0  0.0       10.60     0.0     0.0%  2024-09-26 15:53:31 EDT-0400   163.0
225.0   0.0  0.0        7.15     0.0     0.0%  2024-09-26 15:55:45 EDT-0400   865.0
230.0   0.0  0.0        4.60     0.0     0.0%  2024-09-26 15:59:52 EDT-0400  2324.0
235.0   0.0  0.0        2.53     0.0     0.0%  2024-09-26 15:59:57 EDT-0400  1792.0
240.0   0.0  0.0        1.23     0.0     0.0%  2024-09-26 15:55:10 EDT-0400   885.0
245.0   0.0  0.0        0.66     0.0     0.0%  2024-09-26 15:58:44 EDT-0400   320.0
250.0   0.0  0.0        0.32     0.0     0.0%  2024-09-26 15:59:59 EDT-0400   223.0
255.0   0.0  0.0        0.17     0.0     0.0%  2024-09-26 15:57:48 EDT-0400   110.0
260.0   0.0  0.0        0.10     0.0     0.0%  2024-09-26 15:59:44 EDT-0400   547.0
265.0   0.0  0.0        0.07     0.0     0.0%  2024-09-26 15:28:22 EDT-0400    27.0
270.0   0.0  0.0        0.03     0.0     0.0%  2024-09-25 15:45:22 EDT-0400    21.0
275.0   0.0  0.0        0.02     0.0     0.0%  2024-09-26 10:07:51 EDT-0400    50.0
280.0   0.0  0.0        0.02     0.0     0.0%  2024-09-24 12:13:37 EDT-0400    10.0
285.0   0.0  0.0        0.01     0.0     0.0%  2024-09-18 13:26:05 EDT-0400     1.0
290.0   0.0  0.0        0.02     0.0     0.0%  2024-09-25 14:03:52 EDT-0400     1.0
295.0   0.0  0.0        0.02     0.0     0.0%  2024-09-18 09:30:05 EDT-0400     4.0
300.0   0.0  0.0        0.02     0.0     0.0%  2024-09-25 14:06:17 EDT-0400    12.0
```

1.1 Intrinsic Value and Time Value

1.2 Investing in Options

1.3 Hedging with Options

1.4 Selling Options for Income

1.5 Option Spreads

1.6 Put-Call Parity

1.7 American Options

1.8 Dividends

If the market value of the asset exceeds the exercise price, then we say the call option is in the money. Buying a call option is a way to bet on the upside of the underlying asset.

A put option is the right to sell an asset at a pre-specified (exercise, strike) price. Buying a put is a way to bet on an asset price becoming low (similar to shorting). A put option is in the money if the exercise price exceeds the value of the asset. Both puts and calls are potentially valuable and hence the buyer of a put or call must pay the seller.

A long put option provides insurance to someone who is long the underlying asset, because it guarantees that the asset can always be sold at the strike price of the put (of course, it can be sold at the market price, if that is higher than the strike of the put). Symmetrically, a long call option provides insurance to someone who is short the underlying asset. The terminology in option markets reflects the parallels between options and insurance contracts. In particular, the seller of an option is said to write the option and the compensation (price) he receives from the buyer is called the option premium, just as an insurance company writes insurance contracts in exchange for premium income. Calculating the price at which one should be willing to trade an option is the main topic of this book.

It is important to recognize the different situations of someone who is short a call option and someone who is long a put. Both positions are bets on the downside of the asset. Both the investor who is short a call and the investor who is long a put may eventually sell the underlying asset and receive the exercise price in exchange. However, the investor who is long a put has an option to sell the asset at the exercise price and the investor who is short a call has an *obligation* to sell the asset at the exercise price, should the counterparty choose to exercise the call. Thus, the investor who is long a put will be selling at the exercise price when it is profitable to do so, whereas the investor who is short a call will be selling at the exercise price when it is unprofitable. The buyer of a put must pay the premium to the seller; he then profits if the asset price is low, with his maximum possible profit being quite large (the maximum value is attained when the market value of the underlying asset reaches zero). In contrast, the seller of a call receives premium income, and the premium is his maximum possible profit, whereas his potential losses are unbounded. Thus, these are very different positions.

Individuals who sell calls usually sell out-of-the-money covered calls. Covered means that they own the underlying asset and can therefore deliver the underlying if the call is exercised without incurring any further expense—they experience only an opportunity cost in delivering it for less than the market price.^[In contrast, one who sells a call without owning the underlying is said to sell a naked call. A call being out of the money implies that the price of the underlying must rise before the call would be exercised against the seller; thus, the seller of an out-of-the money covered call still has some potential for profit from the underlying. In addition, of course, the seller receives the premium income from the call. Institutions often follow this strategy also, using the premium income to enhance their return from the underlying. One can hedge a short call without owning a full share of the underlying asset, if one is able to rebalance the hedge over time. Calculating such hedges is another of the principal topics of this book.

In a certain sense, option markets are zero-sum games. The profit earned by one counterparty to an option transaction is a loss suffered by the other. However, options can allow for an increase in the welfare of all investors by improving the allocation of risk. A producer who must purchase a certain input may buy a call option, giving him the right to buy the input at a fixed price. This caps his expense. The seller of the call now bears the risk that the input price will be high—in this case, the option will be exercised and he will be forced to sell at a price below the market price. It may be that the seller is in a better position to bear the risk (for example, he may have less of the risk in his portfolio) and the option transaction may thereby improve the allocation of risks across investors. The similarity to insurance should be apparent.

Quite complex bets or hedges can be created by combining options. For example, a long call and put with the same strike price is called a straddle. Such a portfolio is (almost) always in the money. It is in fact a bet on volatility—a big move in the underlying asset value away from the exercise price will lead to either the call or put having a high value. Another important example of an option portfolio is a collar. A collar consists of a long put and a short call, or a short call and a long put, with the options having the same maturity. As mentioned before, a long put provides insurance to someone who is long the underlying asset. Selling a call provides premium income that can be used to offset the cost of the put (the most popular type of collar is a zero-cost collar: a collar in which the premium of the call is equal to the premium of the put). The cost of selling a call for an owner of the underlying is that it sells off the upside of the underlying asset—if the value of the asset exceeds the strike price of the call, then the call will be exercised and the underlying asset must be delivered for the strike price (rather than the higher market price). Thus, one can purchase the downside insurance provided by a long put by selling part of the upside potential of the asset, rather than paying the cost of the insurance out of pocket. There are many other examples of option portfolios that could be given.

Some puts and calls are traded on exchanges. In this case, the exchange clearinghouse steps between the buyer and seller and becomes the counterparty to both the buyer and seller. This eliminates the risk that the seller might default on his obligation when the buyer chooses to exercise his option. If the owner of an option chooses to exercise, the clearinghouse randomly chooses someone who is short the option to fulfill the obligation. Most exchange traded options are never exercised, because any gain on a long contract can be captured by selling the contract at the market price, thus cancelling the position. Obviously, however, the right to exercise is essential, because it determines the market price. Puts and calls are also transacted over the counter, which means that they are private contracts of the counterparties. Moreover, puts and calls are embedded in many other financial instruments. A prosaic but important example is that most homeowners have the right to pay off their mortgages early. This means they have call options on their mortgages, with exercise price equal to the remaining mortgage principal. Similarly, callable bonds can be redeemed early by the company issuing them, convertible bonds have embedded call options on the company's stock (which are exercised by converting the bonds) and there are many, many other examples. Puts and calls also exist outside financial markets. For example, a company may begin manufacturing a new product at a small scale; if the product is successful, the scale can be expanded. In this case, the company buys a call option on large-scale production with the premium being the cost of launching small-scale production. Adapting the methods developed for financial options to value such real options is an important and growing field.

1.8.1 Exercise Policies for Calls and Puts

It may be rational to exercise a call if the asset value exceeds the exercise price. Thus, denoting the price of the asset by S and the exercise price by K, the owner of a call option can profit by $S-K$ dollars by exercising the option when $S>K$. If $S<K$, exercise would be irrational. Thus, the payoff to the owner of the call option is[1] $\max(0, S-K)$. It has been said that timing is everything, and the timing here should be made clearer. The simplest type of option is called a European option. A European option has a finite lifetime and can only be exercised at its maturity date. For a European call option, the exercise strategy just described is the optimal one, with S representing the asset price at the maturity date of the option. Equally, if not more, important are American options, which can be exercised at any time before maturity.

For an American call option, the exercise strategy just described is the optimal one at the maturity date, but it may also be optimal to exercise prior to maturity. Let K denote the

[1] We use the standard notation: $\max(a, b)$ denotes the larger of a and b and $\min(a, b)$ denotes the smaller.

exercise price, T the date the option matures, and S_T the price of the underlying asset at date $t \leq T$. The intrinsic value of the call option at date t is defined to be $\max(0, S_T - K)$. One would of course never exercise unless the intrinsic value is positive—i.e., unless the option is in the money. Moreover, if the asset does not pay a dividend (or other type of cash flow) prior to the option maturity then one should not exercise in any circumstances prior to maturity. This is captured in the saying: calls are better alive than dead. Exercise being suboptimal is equivalent to the value of the option exceeding the intrinsic value.

The principle that calls on non-dividend-paying assets are better alive than dead follows from two facts: (i) it is generally a good thing (in financial markets as well as in life) to keep one's options open, and (ii) early exercise implies early payment of the exercise price and hence foregone interest. The usual protest that is heard when this statement is made is that one should surely exercise if he expects the stock price to plummet, because by exercising (and then selling the stock acquired) one can lock in the current stock price rather than waiting for it to fall, in which case the option will surely be worth less. This intuition is a reasonable one, but it ignores the fact that the investor could short sell the stock if he expects it to plummet—he doesn't need to exercise the option to lock in the current stock price. In fact, shorting the stock and retaining the option is always better than exercising, assuming the underlying asset does not pay a dividend.

Specifically, suppose an investor considers exercising at date t. As an alternative to exercising early, consider shorting the stock at date t and retaining the option. This is always better than exercising at date t, because the short position can be covered (the stock can be purchased and returned to the lender to cancel the short position) at cost K at date T by exercising the option, and paying K at date T is better than paying it at date t, given that interest rates must be nonnegative. To be more precise, note that exercise at date t produces $S_T - K$ dollars at date t. On the other hand, retaining the option, shorting the stock at date t, and covering the short either by exercising the option or buying the stock in the market (whichever is cheaper) produces S_T dollars at date t and

$$\max(0, S_T - K) - S_T = \max(-S_T, -K) = -\min(S_T, K) \geq -K$$

dollars at date T. If $S_T > K$, one has $-K$ dollars at date T, in which case retaining the option has been superior due to the time value of money. Furthermore, if $S_T < K$, the strategy of retaining the option and shorting the stock produces $-S_T > -K$ dollars at date T, so retaining the option is superior due both to flexibility (waiting until T to decide whether to exercise turns out to be better than committing at date t) and because of the time value of money.[2]

[2] Recall that we are assuming investors earn interest on the proceeds of short sales; otherwise, the S_T dollars earned from exercising the option and selling the stock will be worth more than the S_T dollars earned from shorting the stock. In this case, early exercise could be optimal. However, assuming institutional investors can earn interest on the proceeds of shorts, such investors should prefer owning the option and shorting the stock to exercising. This means they should bid up the price of the option to the point where it exceeds the value $S_T - K$ of exercise. If this is the case, then an investor who cannot earn interest on the proceeds of shorts should simply sell the option in the market rather than exercise it. Thus, a sufficient condition for calls to be better alive than dead is that there be some investors who can earn interest on the proceeds of

Early exercise of a call option can be optimal when the underlying asset pays a dividend. The above analysis does not apply in this case, because paying the dividend to the lender of the stock is an additional cost for the strategy of retaining the option and shorting the stock. If the dividend is so small that it cannot offset the time value of money on the exercise price, then early exercise will not be optimal. In other cases, deriving the optimal exercise strategy is a complicated problem that we will first begin to study in Chapter **??**.

A European put option will be exercised at its maturity T if the price S_T of the underlying asset is below the exercise price K. In general, the value at maturity can be expressed as $\max(0, K - S_T)$. Early exercise of an American put can be optimal, regardless of whether the underlying pays a dividend. While it is valuable to keep one's options open (for puts as well as calls) the time value of money works in the opposite direction for puts. Early exercise of a put option implies early receipt of the exercise price, and it is better to receive cash earlier rather than later. In general, whether early exercise is optimal depends on how deeply the option is in the money—if the underlying asset price is sufficiently low, then it will be fairly certain that exercise will be optimal, whether earlier or late; in this case, one should exercise earlier to earn interest on the exercise price. How low it should be to justify early exercise depends on the interest rate (a higher rate makes the time-value-of-money issue more important, leading to earlier exercise) and the volatility of the underlying asset price (a lower volatility reduces the value of keeping one's options open, leading also to earlier exercise). We will begin to study the valuation of American puts in Chapter **??** also.

1.8.2 Compounding Interest

During most of the first two parts of the book (the only exception being Chapter **??**) we will assume there is a risk-free asset earning a constant rate of return. For simplicity, we will specify the rate of return as a continuously compounded rate. For example, if the annual rate with annual compounding is r_a, then the corresponding continuously compounded rate is r defined as $r = \log(1 + r_a)$, where log denotes the natural logarithm function. This means that the gross return over a year (one plus the rate of return) is $\mathrm{e}^r = 1 + r_a$. More generally, an investment of x dollars for a time period of length T (we measure time in years, so, e.g., a six-month investment would mean $T = 0.5$) will result in the ownership of $x\mathrm{e}^{rT}$ dollars at the end of the time period.

Expressing the interest rate as a continuously compounded rate enables us to avoid having to specify in each instance whether the rate is for annual compounding, semi-annual compounding, monthly compounding, etc. For example, the meaning of an annualized rate r_s for semi-annual compounding is that an investment of x dollars will grow over a year to $x(1 + r_s/2)^2$. The equivalent continuously compounded rate is defined as $r = \log(1 + r_s/2)^2$, and in terms of this rate we can say that the investment will grow in six months to $x\mathrm{e}^{0.5r}$ and that it will grow in one year to $x\mathrm{e}^r$. We can interpret this rate as being continuously compounded because

shorts. This type of reasoning is possible for each situation in this book where the assumption of earning interest on margin deposits is important, and we will not deal with it in this much detail again.

compounding n times per year at an annualized rate of r results in \$1 growing in a year to $(1+r/n)^n$ and

$$\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^n = \mathrm{e}^r \quad .$$

To develop pricing and hedging formulas for derivative securities, it is a great convenience to assume that investors can trade continuously in time. This requires us to assume also that returns are computed continuously. In the case of a risk-free investment of $x(t)$ dollars at any date t at a continuously compounded rate of r, we will say that the interest earned in an instant $\mathrm{d}t$ is $x(t)r\,\mathrm{d}t$ dollars. This is only meaningful when we accumulate the interest over a non-infinitesimal period of time. So consider investing $x(0)$ dollars at time 0 and reinvesting interest in the risk-free asset over a time period of length T. Let $x(t)$ denote the account balance at date t, for $0 \le t \le T$. The change in the account balance in each instant is the interest earned, so we have $\mathrm{d}x(t) = x(t)r\,\mathrm{d}t$. The real meaning of this equation is that $x(t)$ satisfies the differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t)r \ ,$$

and it is well known (and easy to verify) that the solution is

$$x(t) = x(0)\mathrm{e}^{rt} \ ,$$

leading to an account balance at the end of the time period of $x(T) = x(0)\mathrm{e}^{rT}$. Thus, the statement that the interest earned in an instant $\mathrm{d}t$ is $x(t)r\,\mathrm{d}t$ is equivalent to the statement that interest is continuously compounded at the rate r.

In the last part of the book, we will drop the assumption that the risk-free asset earns a constant rate of return. In this case, we will still generally assume that there is a risk-free asset for very short-term investments (i.e., for investments with infinitesimal durations!). We will let $r(t)$ denote the risk-free rate for an instantaneous investment at date t. This means that an investment of $x(t)$ dollars at date t in the risk-free asset earns interest in an instant $\mathrm{d}t$ equal to $x(t)r(t)\,\mathrm{d}t$. Consider again an investment of $x(0)$ dollars at date 0 in this instantaneously risk-free asset with interest reinvested and let $x(t)$ denote the account balance at date t. Then $x(t)$ must satisfy the differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t)r(t) \quad .$$

The solution of this differential equation is

$$x(t) = x(0)\exp\left(\int_0^t r(s)\,ds\right) \quad .$$

The expression $\int_0^t r(s)\,ds$ can be interpreted as a continuous sum over time of the rates of interest $r(s)$ earned at times s between 0 and t. If these rates are all the same, say equal to r, then $\int_0^t r(s)\,ds = rt$ and our compounding factor $\exp\left(\int_0^t r(s)\,ds\right)$ is e^{rt} as before.

2 Binomial Model and Changes of Measure

This chapter introduces the change of measure (or change of numeraire or martingale) method for valuing derivative securities. The method is introduced in a binomial model and then extended to more general (continuum of states) models. Computations in the more general model require the continuous-time mathematics presented in Chapter **??**.

It should be noted that the pricing and hedging results in this book are not tied to any particular currency. However, for specificity (and as a consequence of the author's habit) the discussion will generally be in terms of dollars. Multiple currencies are addressed in Chapter **??**.

2.1 Fundamental Concepts

2.1.1 Longs, Shorts, and Margin

In financial markets, the owner of an asset is said to be long the asset. If person A owes something to person B, the debt is an asset to person B but a liability to person A. One also says that person A is short the asset. For example, if someone borrows money and invests the money in stocks, then the individual is short cash and long stocks.

One must invest some of one's own money when borrowing money to buy stocks. For example, an individual could invest \$600, borrow \$400, and buy \$1000 of stock. The \$600 is called the margin posted by the investor, and buying stocks in this way is called buying on margin. The investor, or the portfolio, is also said to be levered, because buying \$1000 of stock with only a \$600 investment amplifies the risk and return per dollar of investment. On a percentage basis, we would say the account has 60% margin, the 60% being the ratio of the equity (assets minus liabilities = \$1000 of stock minus \$400 debt) to the assets (\$1000 of stock). If the value of the stock drops sufficiently far, then it may become doubtful whether the investor can repay the \$400. In this case, the investor must either sell the stock or invest more of his own funds (i.e., he receives a margin call). In other words, in actual markets there are margin requirements, that specify a minimum percent margin an investor must have initially (when borrowing money) and a minimum percent margin the investor must maintain.

Rather than borrowing money to buy stocks, an investor can do the opposite—he can borrow stocks to buy money. In this case, buying money means selling the borrowed stocks for cash. Such an investor will be short stocks and long cash. This is called short selling (or, more briefly, shorting) stocks. For example, suppose individual A borrows 100 shares of stock from individual B and then sells them to individual C. Both B and C are long the 100 shares and A is short, so the net long position is $2 \times 100 - 100$, which is the original 100 shares that B was long. A short seller of stocks must pay to the lender of the stocks any dividends that are paid on the stock. In our example, both B and C own the 100 shares so both expect to receive dividends. The company will pay dividends only to C, and A must pay the dividends to~B.

Of course, investors always wish to buy low and sell high. The usual method is to buy stocks and hope they rise. An investor who short sells also wishes to buy low and sell high, but he reverses the order—he sells first and then hopes the stocks fall. The risk is that the stocks will instead rise, which will increase the value of his liability (short stock position) without increasing the value of his assets (long cash position), thus putting him under water. To shield the lender of the stocks from this risk, a short seller must also invest some of his own funds, and this amount is again called the investor's margin. For example, an investor might invest \$600, and borrow and sell \$1000 of stock. In this case, the investor will be long \$1600 cash and short \$1000 worth of stock. His equity is \$600 and his percent margin is calculated as \$600/\$1000 = 60%. Again, there are typically both initial and maintenance margin requirements. An additional feature of short selling for small individual investors is that they typically will not earn interest on the proceeds of the short sale (the \$1000 cash obtained from selling stocks in the above example).

In this book, we will assume there is a single risk-free rate at which one can both borrow and lend. Moreover, we will assume that investors earn this rate on margin deposits, including the proceeds of short sales (and including any margin that may be required when buying and selling forward and futures contracts). Thus, investors gain from buying on margin if the asset return is sure to exceed the risk-free rate, and they gain from short selling if the return on an asset is sure to be below the risk-free rate. These assumptions are not reasonable for small individual investors, but they are fairly reasonable for institutional investors. We will assume that no asset has a return that is certain to be above the risk-free rate nor certain to be below the risk-free rate, because institutional investors could arbitrage such guaranteed high-return or guaranteed low-return assets.

2.1.2 Calls and Puts

Call and put options are the basic derivative securities and the building blocks of many others. A derivative security is a security the value of which depends upon another security. A call option is the right to buy an asset at a pre-specified price. The pre-specified price is called the exercise price, the strike price, or simply the strike. We will often call the asset a stock, but there are options on many other types of assets also, and everything we say will be applicable to those as well.[1] The asset to which the call option pertains is called the underlying asset, or, more briefly, the underlying. If the market value of the asset exceeds the exercise price, then

[1]One caveat is that by asset we mean something that can be stored; thus, for example, electricity is, practically speaking, not an asset.

we say the call option is in the money. Buying a call option is a way to bet on the upside of the underlying asset.

A put option is the right to sell an asset at a pre-specified (exercise, strike) price. Buying a put is a way to bet on an asset price becoming low (similar to shorting). A put option is in the money if the exercise price exceeds the value of the asset. Both puts and calls are potentially valuable and hence the buyer of a put or call must pay the seller.

A long put option provides insurance to someone who is long the underlying asset, because it guarantees that the asset can always be sold at the strike price of the put (of course, it can be sold at the market price, if that is higher than the strike of the put). Symmetrically, a long call option provides insurance to someone who is short the underlying asset. The terminology in option markets reflects the parallels between options and insurance contracts. In particular, the seller of an option is said to write the option and the compensation (price) he receives from the buyer is called the option premium, just as an insurance company writes insurance contracts in exchange for premium income. Calculating the price at which one should be willing to trade an option is the main topic of this book.

It is important to recognize the different situations of someone who is short a call option and someone who is long a put. Both positions are bets on the downside of the asset. Both the investor who is short a call and the investor who is long a put may eventually sell the underlying asset and receive the exercise price in exchange. However, the investor who is long a put has an option to sell the asset at the exercise price and the investor who is short a call has an *obligation* to sell the asset at the exercise price, should the counterparty choose to exercise the call. Thus, the investor who is long a put will be selling at the exercise price when it is profitable to do so, whereas the investor who is short a call will be selling at the exercise price when it is unprofitable. The buyer of a put must pay the premium to the seller; he then profits if the asset price is low, with his maximum possible profit being quite large (the maximum value is attained when the market value of the underlying asset reaches zero). In contrast, the seller of a call receives premium income, and the premium is his maximum possible profit, whereas his potential losses are unbounded. Thus, these are very different positions.

Individuals who sell calls usually sell out-of-the-money covered calls. Covered means that they own the underlying asset and can therefore deliver the underlying if the call is exercised without incurring any further expense—they experience only an opportunity cost in delivering it for less than the market price.^[In contrast, one who sells a call without owning the underlying is said to sell a naked call. A call being out of the money implies that the price of the underlying must rise before the call would be exercised against the seller; thus, the seller of an out-of-the money covered call still has some potential for profit from the underlying. In addition, of course, the seller receives the premium income from the call. Institutions often follow this strategy also, using the premium income to enhance their return from the underlying. One can hedge a short call without owning a full share of the underlying asset, if one is able to rebalance the hedge over time. Calculating such hedges is another of the principal topics of this book.

In a certain sense, option markets are zero-sum games. The profit earned by one counterparty to an option transaction is a loss suffered by the other. However, options can allow for an increase in the welfare of all investors by improving the allocation of risk. A producer who must purchase a certain input may buy a call option, giving him the right to buy the input at a fixed price. This caps his expense. The seller of the call now bears the risk that the input price will be high—in this case, the option will be exercised and he will be forced to sell at a price below the market price. It may be that the seller is in a better position to bear the risk (for example, he may have less of the risk in his portfolio) and the option transaction may thereby improve the allocation of risks across investors. The similarity to insurance should be apparent.

Quite complex bets or hedges can be created by combining options. For example, a long call and put with the same strike price is called a straddle. Such a portfolio is (almost) always in the money. It is in fact a bet on volatility—a big move in the underlying asset value away from the exercise price will lead to either the call or put having a high value. Another important example of an option portfolio is a collar. A collar consists of a long put and a short call, or a short call and a long put, with the options having the same maturity. As mentioned before, a long put provides insurance to someone who is long the underlying asset. Selling a call provides premium income that can be used to offset the cost of the put (the most popular type of collar is a zero-cost collar: a collar in which the premium of the call is equal to the premium of the put). The cost of selling a call for an owner of the underlying is that it sells off the upside of the underlying asset—if the value of the asset exceeds the strike price of the call, then the call will be exercised and the underlying asset must be delivered for the strike price (rather than the higher market price). Thus, one can purchase the downside insurance provided by a long put by selling part of the upside potential of the asset, rather than paying the cost of the insurance out of pocket. There are many other examples of option portfolios that could be given.

Some puts and calls are traded on exchanges. In this case, the exchange clearinghouse steps between the buyer and seller and becomes the counterparty to both the buyer and seller. This eliminates the risk that the seller might default on his obligation when the buyer chooses to exercise his option. If the owner of an option chooses to exercise, the clearinghouse randomly chooses someone who is short the option to fulfill the obligation. Most exchange traded options are never exercised, because any gain on a long contract can be captured by selling the contract at the market price, thus cancelling the position. Obviously, however, the right to exercise is essential, because it determines the market price. Puts and calls are also transacted over the counter, which means that they are private contracts of the counterparties. Moreover, puts and calls are embedded in many other financial instruments. A prosaic but important example is that most homeowners have the right to pay off their mortgages early. This means they have call options on their mortgages, with exercise price equal to the remaining mortgage principal. Similarly, callable bonds can be redeemed early by the company issuing them, convertible bonds have embedded call options on the company's stock (which are exercised by converting the bonds) and there are many, many other examples. Puts and calls also exist outside financial markets. For example, a company may begin manufacturing a new product at a small scale;

if the product is successful, the scale can be expanded. In this case, the company buys a call option on large-scale production with the premium being the cost of launching small-scale production. Adapting the methods developed for financial options to value such real options is an important and growing field.

2.1.3 Exercise Policies for Calls and Puts

It may be rational to exercise a call if the asset value exceeds the exercise price. Thus, denoting the price of the asset by S and the exercise price by K, the owner of a call option can profit by $S-K$ dollars by exercising the option when $S > K$. If $S < K$, exercise would be irrational. Thus, the payoff to the owner of the call option is[2] $\max(0, S-K)$. It has been said that timing is everything, and the timing here should be made clearer. The simplest type of option is called a European option. A European option has a finite lifetime and can only be exercised at its maturity date. For a European call option, the exercise strategy just described is the optimal one, with S representing the asset price at the maturity date of the option. Equally, if not more, important are American options, which can be exercised at any time before maturity.

For an American call option, the exercise strategy just described is the optimal one at the maturity date, but it may also be optimal to exercise prior to maturity. Let K denote the exercise price, T the date the option matures, and $S(t)$ the price of the underlying asset at date $t \leq T$. The intrinsic value of the call option at date t is defined to be $\max(0, S(t) - K)$. One would of course never exercise unless the intrinsic value is positive—i.e., unless the option is in the money. Moreover, if the asset does not pay a dividend (or other type of cash flow) prior to the option maturity then one should not exercise in any circumstances prior to maturity. This is captured in the saying: calls are better alive than dead. Exercise being suboptimal is equivalent to the value of the option exceeding the intrinsic value.

The principle that calls on non-dividend-paying assets are better alive than dead follows from two facts: (i) it is generally a good thing (in financial markets as well as in life) to keep one's options open, and (ii) early exercise implies early payment of the exercise price and hence foregone interest. The usual protest that is heard when this statement is made is that one should surely exercise if he expects the stock price to plummet, because by exercising (and then selling the stock acquired) one can lock in the current stock price rather than waiting for it to fall, in which case the option will surely be worth less. This intuition is a reasonable one, but it ignores the fact that the investor could short sell the stock if he expects it to plummet—he doesn't need to exercise the option to lock in the current stock price. In fact, shorting the stock and retaining the option is always better than exercising, assuming the underlying asset does not pay a dividend.

Specifically, suppose an investor considers exercising at date t. As an alternative to exercising early, consider shorting the stock at date t and retaining the option. This is always better than exercising at date t, because the short position can be covered (the stock can be purchased

[2] We use the standard notation: $\max(a, b)$ denotes the larger of a and b and $\min(a, b)$ denotes the smaller.

and returned to the lender to cancel the short position) at cost K at date T by exercising the option, and paying K at date T is better than paying it at date t, given that interest rates must be nonnegative. To be more precise, note that exercise at date t produces $S(t) - K$ dollars at date t. On the other hand, retaining the option, shorting the stock at date t, and covering the short either by exercising the option or buying the stock in the market (whichever is cheaper) produces $S(t)$ dollars at date t and

$$\max(0, S(T) - K) - S(T) = \max(-S(T), -K) = -\min(S(T), K) \geq -K$$

dollars at date T. If $S(T) > K$, one has $-K$ dollars at date T, in which case retaining the option has been superior due to the time value of money. Furthermore, if $S(T) < K$, the strategy of retaining the option and shorting the stock produces $-S(T) > -K$ dollars at date T, so retaining the option is superior due both to flexibility (waiting until T to decide whether to exercise turns out to be better than committing at date t) and because of the time value of money.[3]

Early exercise of a call option can be optimal when the underlying asset pays a dividend. The above analysis does not apply in this case, because paying the dividend to the lender of the stock is an additional cost for the strategy of retaining the option and shorting the stock. If the dividend is so small that it cannot offset the time value of money on the exercise price, then early exercise will not be optimal. In other cases, deriving the optimal exercise strategy is a complicated problem that we will first begin to study in Chapter **??**.

A European put option will be exercised at its maturity T if the price $S(T)$ of the underlying asset is below the exercise price K. In general, the value at maturity can be expressed as $\max(0, K - S(T))$. Early exercise of an American put can be optimal, regardless of whether the underlying pays a dividend. While it is valuable to keep one's options open (for puts as well as calls) the time value of money works in the opposite direction for puts. Early exercise of a put option implies early receipt of the exercise price, and it is better to receive cash earlier rather than later. In general, whether early exercise is optimal depends on how deeply the option is in the money—if the underlying asset price is sufficiently low, then it will be fairly certain that exercise will be optimal, whether earlier or late; in this case, one should exercise earlier to earn interest on the exercise price. How low it should be to justify early exercise depends on the interest rate (a higher rate makes the time-value-of-money issue more important, leading to earlier exercise) and the volatility of the underlying asset price (a lower

[3] Recall that we are assuming investors earn interest on the proceeds of short sales; otherwise, the $S(t)$ dollars earned from exercising the option and selling the stock will be worth more than the $S(t)$ dollars earned from shorting the stock. In this case, early exercise could be optimal. However, assuming institutional investors can earn interest on the proceeds of shorts, such investors should prefer owning the option and shorting the stock to exercising. This means they should bid up the price of the option to the point where it exceeds the value $S(t) - K$ of exercise. If this is the case, then an investor who cannot earn interest on the proceeds of shorts should simply sell the option in the market rather than exercise it. Thus, a sufficient condition for calls to be better alive than dead is that there be some investors who can earn interest on the proceeds of shorts. This type of reasoning is possible for each situation in this book where the assumption of earning interest on margin deposits is important, and we will not deal with it in this much detail again.

volatility reduces the value of keeping one's options open, leading also to earlier exercise). We will begin to study the valuation of American puts in Chapter **??** also.

2.1.4 Compounding Interest

During most of the first two parts of the book (the only exception being Chapter **??**) we will assume there is a risk-free asset earning a constant rate of return. For simplicity, we will specify the rate of return as a continuously compounded rate. For example, if the annual rate with annual compounding is r_a, then the corresponding continuously compounded rate is r defined as $r = \log(1 + r_a)$, where log denotes the natural logarithm function. This means that the gross return over a year (one plus the rate of return) is $\mathrm{e}^r = 1 + r_a$. More generally, an investment of x dollars for a time period of length T (we measure time in years, so, e.g., a six-month investment would mean $T = 0.5$) will result in the ownership of $x\mathrm{e}^{rT}$ dollars at the end of the time period.

Expressing the interest rate as a continuously compounded rate enables us to avoid having to specify in each instance whether the rate is for annual compounding, semi-annual compounding, monthly compounding, etc. For example, the meaning of an annualized rate r_s for semi-annual compounding is that an investment of x dollars will grow over a year to $x(1 + r_s/2)^2$. The equivalent continuously compounded rate is defined as $r = \log(1+r_s/2)^2$, and in terms of this rate we can say that the investment will grow in six months to $x\mathrm{e}^{0.5r}$ and that it will grow in one year to $x\mathrm{e}^r$. We can interpret this rate as being continuously compounded because compounding n times per year at an annualized rate of r results in \$1 growing in a year to $(1 + r/n)^n$ and

$$\lim_{n\to\infty}\left(1 + \frac{r}{n}\right)^n = \mathrm{e}^r \ .$$

To develop pricing and hedging formulas for derivative securities, it is a great convenience to assume that investors can trade continuously in time. This requires us to assume also that returns are computed continuously. In the case of a risk-free investment of $x(t)$ dollars at any date t at a continuously compounded rate of r, we will say that the interest earned in an instant $\mathrm{d}t$ is $x(t)r\,\mathrm{d}t$ dollars. This is only meaningful when we accumulate the interest over a non-infinitesimal period of time. So consider investing $x(0)$ dollars at time 0 and reinvesting interest in the risk-free asset over a time period of length T. Let $x(t)$ denote the account balance at date t, for $0 \leq t \leq T$. The change in the account balance in each instant is the interest earned, so we have $\mathrm{d}x(t) = x(t)r\,\mathrm{d}t$. The real meaning of this equation is that $x(t)$ satisfies the differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t)r \ ,$$

and it is well known (and easy to verify) that the solution is

$$x(t) = x(0)\mathrm{e}^{rt} \ ,$$

leading to an account balance at the end of the time period of $x(T) = x(0)\mathrm{e}^{rT}$. Thus, the statement that the interest earned in an instant $\mathrm{d}t$ is $x(t)r\,\mathrm{d}t$ is equivalent to the statement that interest is continuously compounded at the rate r.

In the last part of the book, we will drop the assumption that the risk-free asset earns a constant rate of return. In this case, we will still generally assume that there is a risk-free asset for very short-term investments (i.e., for investments with infinitesimal durations!). We will let $r(t)$ denote the risk-free rate for an instantaneous investment at date t. This means that an investment of $x(t)$ dollars at date t in the risk-free asset earns interest in an instant $\mathrm{d}t$ equal to $x(t)r(t)\,\mathrm{d}t$. Consider again an investment of $x(0)$ dollars at date 0 in this instantaneously risk-free asset with interest reinvested and let $x(t)$ denote the account balance at date t. Then $x(t)$ must satisfy the differential equation

$$\frac{\mathrm{d}x(t)}{\mathrm{d}t} = x(t)r(t) \quad .$$

The solution of this differential equation is

$$x(t) = x(0)\exp\left(\int_0^t r(s)\,ds\right) \quad .$$

The expression $\int_0^t r(s)\,ds$ can be interpreted as a continuous sum over time of the rates of interest $r(s)$ earned at times s between 0 and t. If these rates are all the same, say equal to r, then $\int_0^t r(s)\,ds = rt$ and our compounding factor $\exp\left(\int_0^t r(s)\,ds\right)$ is e^{rt} as before.

2.2 State Prices in a One-Period Binomial Model

To introduce the concepts that will be discussed in the remainder of the chapter, we will consider in this and the following section the following very simple framework. There is a stock with price S today (which we will call date~0). At the end of some period of time of length T, the stock price will take one of two values: either S_u or S_d, where $S_u > S_d$. If the stock price equals S_u we say we are in the up state of the world, and if it equals S_d we say we are in the down state. The stock does not pay a dividend. There is also a risk-free asset earning a continuously compounded rate of interest r. Finally we want to consider a European call option on the stock with maturity T and strike K. The value of the call option at the end of the period is $C_u = \max(0, S_u - K)$ in the up state and $C_d = \max(0, S_d - K)$ in the down state.

We will assume

$$\frac{S_u}{S} > \mathrm{e}^{rT} > \frac{S_d}{S} \; . \tag{2.1}$$

This condition means that the rate of return on the stock in the up state is greater than the risk-free rate, and the rate of return on the stock in the down state is less than the risk-free rate. If it were not true, there would be an arbitrage opportunity: if the rate of return on the stock were greater than the risk-free rate in both states, then one should buy an infinite amount of the stock on margin, and conversely if the rate of return on the stock were less than the risk-free rate in both states, then one should short an infinite amount of stock and put the proceeds in the risk-free asset. So what we are assuming is that there are no arbitrage opportunities in the market for the stock and risk-free asset.

The delta of the call option is $\delta = (C_u - C_d)/(S_u - S_d)$. Multiplying by $S_u - S_d$ gives us $\delta(S_u - S_d) = C_u - C_d$ and rearranging yields $\delta S_u - C_u = \delta S_d - C_d$, which is critical to what follows. Consider purchasing δ shares of the stock at date 0 and borrowing

$$\mathrm{e}^{-rT}(\delta S_u - C_u) = \mathrm{e}^{-rT}(\delta S_d - C_d)$$

dollars at date 0. Then you will owe

$$\delta S_u - C_u = \delta S_d - C_d$$

dollars at date T, and hence the value of the portfolio at date T in the up state will be

$$\text{Value of delta shares} - \text{Dollars owed} = \delta S_u - (\delta S_u - C_u) = C_u \ ,$$

and the value of the portfolio at date T in the down state will be

$$\text{Value of delta shares} - \text{Dollars owed} = \delta S_d - (\delta S_d - C_d) = C_d \ .$$

Thus, this portfolio of buying delta shares and borrowing money (i.e., buying delta shares on margin) replicates the call option. Consequently, the value C of the option at date 0 must be the date–0 cost of the portfolio; i.e.,

$$C = \text{Cost of delta shares} - \text{Dollars borrowed} = \delta S - \mathrm{e}^{-rT}(\delta S_u - C_u) \ . \tag{2.2}$$

Because the call option is equivalent to buying the stock on margin, it can be considered a levered investment in the stock.

We will now rewrite the option pricing Equation **??** in terms of state prices. By substituting for δ in Equation **??**, we can rearrange it as

$$C = \frac{S - \mathrm{e}^{-rT}S_d}{S_u - S_d} \times C_u + \frac{\mathrm{e}^{-rT}S_u - S}{S_u - S_d} \times C_d \ . \tag{2.3}$$

A little algebra also shows that

$$S = \frac{S - \mathrm{e}^{-rT}S_d}{S_u - S_d} \times S_u + \frac{\mathrm{e}^{-rT}S_u - S}{S_u - S_d} \times S_d \ , \tag{2.4}$$

and

$$1 = \frac{S - \mathrm{e}^{-rT} S_d}{S_u - S_d} \times \mathrm{e}^{rT} + \frac{\mathrm{e}^{-rT} S_u - S}{S_u - S_d} \times \mathrm{e}^{rT} \ . \tag{2.5}$$

It is convenient to denote the factors appearing in these equations as

$$\pi_u = \frac{S - \mathrm{e}^{-rT} S_d}{S_u - S_d} \quad \text{and} \quad \pi_d = \frac{\mathrm{e}^{-rT} S_u - S}{S_u - S_d} \ . \tag{2.6}$$

The numbers π_u and π_d are called the state prices, for reasons that will be explained below.

With these definitions, we can write Equation **??**–Equation **??** as

$$C = \pi_u C_u + \pi_d C_d \ , \tag{2.7}$$

$$S = \pi_u S_u + \pi_d S_d \ , \tag{2.8}$$

$$1 = \pi_u \mathrm{e}^{rT} + \pi_d \mathrm{e}^{rT} \ . \tag{2.9}$$

These equations have the following interpretation: the value of a security today is its value in the up state times π_u plus its value in the down state times π_d. This applies to Equation **??** by considering an investment of \$1 today in the risk-free asset—it has value 1 today and will have value e^{rT} in both the up and down states at date T. Moreover, this same equation will hold for any other derivative asset. For example, if we considered a put option, then a delta–hedging argument analogous to that we just gave for the call option will lead to a formula for the value P of the put today which can be expressed as $P = \pi_u P_u + \pi_d P_d$ for the same π_u and π_d defined in Equation **??**.

In this model, we can think of any security as a portfolio of what are called Arrow securities (in recognition of the seminal work of Kenneth Arrow (Arrow 1964)). One of the Arrow securities pays \$1 at date T if the up state occurs and the other pays \$1 at date T if the down state occurs. For example, the stock is equivalent to a portfolio consisting of S_u units of the first Arrow security and S_d units of the second, because the stock is worth S_u dollars in the up state and S_d dollars in the down state. Equations Equation **??**–Equation **??** show that~ π_u is the price of the first Arrow security and π_d is the price of the second. For example, the right-hand side of Equation **??** is the value of the stock at date 0 viewed as a portfolio of Arrow securities when the Arrow securities have prices π_u and π_d. Because the stock clearly is such a portfolio, its price today must equal its value as that portfolio, which is what Equation **??** asserts.

As mentioned before, the prices π_u and π_d of the Arrow securities are called the state prices, because they are the prices of receiving \$1 in the two states of the world. The state prices should be positive, because the payoff of each Arrow security is nonnegative in both states and positive in one. A little algebra shows that the conditions $\pi_u > 0$ and $\pi_d > 0$ are exactly equivalent to our no-arbitrage assumption Equation **??**. Thus, we conclude that **in the absence of arbitrage opportunities, there exist positive state prices such that the price of any security is the sum across the states of the world of its payoff multiplied by the state price.**

This conclusion generalizes to other models, including models in which the stock price takes a continuum of possible values. We will discuss more general models later in this chapter. It is a powerful result that tremendously simplifies derivative security pricing.

2.3 Probabilities and Numeraires

In this section, we will continue our analysis of the binomial example. To apply the statement about state prices appearing in boldface type above in the most convenient way, we will manipulate the state prices so we can interpret the sums on the right-hand sides of Equation **??**–Equation **??** in terms of expectations. The expectation (or mean) of a random variable is of course its probability-weighted average value.

In general, there are different expectations that are useful. In this model, there are two that we can define: one corresponding to the risk-free asset and one corresponding to the stock. Many readers will have experience with the first in the form of risk-neutral probabilities.

The risk-neutral probabilities are defined as $\pi_u \mathrm{e}^{rT}$ for the up state and $\pi_d \mathrm{e}^{rT}$ for the down state. Denoting these as p_u and p_d respectively, Equation **??**–Equation **??** can be written as

$$C = \mathrm{e}^{-rT}[p_u C_u + p_d C_d] \;, \tag{2.10}$$

$$S = \mathrm{e}^{-rT}[p_u S_u + p_d S_d] \;, \tag{2.11}$$

$$1 = p_u + p_d \;. \tag{2.12}$$

The numbers p_u and p_d are both positive (because the state prices are positive under our no-arbitrage assumption) and Equation **??** states that they sum to one, so it is indeed sensible to consider them as probabilities. Equations Equation **??** and Equation **??** state that the value of a security today is its expected value at date T (the expectation taken with respect to the risk-neutral probabilities) discounted at the risk-free rate. Thus, these are present value formulas. Unlike the Capital Asset Pricing Model, for example, there is no risk premium in the discount rate. This is the calculation we would do to price assets under the actual probabilities if investors were risk neutral (or for zero-beta assets). So, we can act as if investors are risk neutral by adjusting the probabilities. Of course, we are not really assuming investors are risk neutral. We have simply embedded any risk premia in the probabilities.[4]

Equations Equation **??** and Equation **??** can be written in an equivalent form, which, though somewhat less intuitive, generalizes more readily. First, let's introduce some notation for the price of the risk-free asset. Considering an investment of \$1 today which grows to e^{rT} at date T, it is sensible to take the price today to be $R = 1$ and the price in the up and down states at date T to be $R_u = R_d = \mathrm{e}^{rT}$.[5] In terms of this notation, Equation **??**–Equation **??** can be written as:

$$\frac{C}{R} = p_u \frac{C_u}{R_u} + p_d \frac{C_d}{R_d} \;, \tag{2.13}$$

$$\frac{S}{R} = p_u \frac{S_u}{R_u} + p_d \frac{S_d}{R_d} \;, \tag{2.14}$$

$$1 = p_u + p_d \;. \tag{2.15}$$

Each of equations Equation **??** and Equation **??** states that the price of a security today divided by the price of the risk-free asset equals the expected future value of the same ratio, when we take expectations using the risk-neutral probabilities. In other words, the mean of the date–T value of the ratio is equal to the ratio today. We will discuss the interpretation and significance of these equations further below. First, we consider the other type of expectation in this model, which is based on probabilities corresponding to the stock.

Note that the risk-neutral probabilities are the state prices multiplied by the gross return on the risk-free asset. Analogously, define numbers $q_u = \pi_u S_u/S$ and $q_d = \pi_d S_d/S$. Substituting for π_u and π_d in Equation **??**–Equation **??** and continuing to use the notation R for the price of the risk-free asset, we obtain

$$\frac{C}{S} = q_u \frac{C_u}{S_u} + q_d \frac{C_d}{S_d} \;, \tag{2.16}$$

$$1 = q_u + q_d \;, \tag{2.17}$$

$$\frac{R}{S} = q_u \frac{R_u}{S_u} + q_d \frac{R_d}{S_d} \;. \tag{2.18}$$

Equation Equation **??** establishes that we can view the q's as probabilities (like the risk-neutral probabilities, they are positive because the state prices are positive). Equations Equation **??**

[4] This fundamental idea is due to Cox and Ross (Cox and Ross 1976).

[5] All of the equations appearing below will also be true if instead we take $R = \mathrm{e}^{-rT}$ and $R_u = R_d = 1$.

and Equation **??** both state that the ratio of a security price to the price of the stock today equals the mean value of the same ratio at date T, when we compute expectations using the q's as probabilities.

Here is some useful terminology:

- An assignment of probabilities to events is called a *probability measure*, or simply a *measure* (because it measures the events, in a sense). Thus, we have described two different probability measures in this section.
- The ratio of one price to another is the value of the first (numerator) asset when we are using the second (denominator) asset as the *numeraire*. The term numeraire means a unit of measurement. For example, the ratio C/S is the value of the call when we use the stock as the unit of measurement: it is the number of shares of stock for which one call option can be exchanged (to see this, note that C/S shares is worth $C/S \times S = C$ dollars, so C/S shares is worth the same as one call.)
- A variable that changes randomly over time with the expected future value being always equal to the current value is called a *martingale*.

The right-hand sides of Equation **??**–Equation **??** and Equation **??** and Equation **??** are expectations under different *probability measures* (the p's or q's). The expected future (date–T) value equals the current (date–0) value, so the random variables (C/R and S/R or C/S and R/S) are *martingales*. The values C/R and S/R are the values of the call and stock *using the risk-free asset as numeraire*, and the values C/S and R/S are the values of the call and risk-free asset *using the stock as numeraire*. Thus, we will express Equation **??**–Equation **??** as the call and stock are martingales when we use the risk-free asset as numeraire. Likewise, we will express Equation **??** and Equation **??** as the call and risk-free asset are martingales when we use the stock as numeraire. It should be understood in both cases that using an asset as numeraire means that we also use the corresponding probability measure (i.e., the p's or q's). In general, our conclusion that assets can be priced in terms of positive state prices when there are no arbitrage opportunities can be rephrased as: **if there are no arbitrage opportunities, then for each (non-dividend-paying) asset, there exists a probability measure such that the ratio of any other (non-dividend-paying) asset price to the first (numeraire) asset price is a martingale.**

We have applied this statement to the risk-free asset, which pays dividends (interest). However, the price $R_u = R_d = \mathrm{e}^{rT}$ includes the interest, so no interest has been withdrawn—the interest has been reinvested—prior to the maturity T of the option. This is what we mean by a non-dividend-paying asset. In general, we will apply the formulas developed in this and the following section to dividend-paying assets by considering the portfolios in which dividends are reinvested.

For this exposition, it was convenient to first calculate the state prices and then calculate the various probabilities. However, that is not the most efficient way to proceed in most applications. In a typical application, we would view the prices of the stock and risk-free asset in the various states of the world as given, and we would be attempting to compute the value of the call option. Note that the sets of equations Equation **??**–Equation **??**, Equation **??**–Equation **??**, and Equation **??**–Equation **??** are all equivalent. In each case we would consider that there are three unknowns—the value C of the call option and either two state prices or two probabilities. In each case the state prices or probabilities can be computed from the last two equations in the set of three equations and then the call value C can be computed from the first equation in the set. All three sets of equations produce the same call value.

In fact, as we will see, it will not even be necessary to calculate the probabilities. The fact that ratios of non-dividend paying asset prices to the numeraire asset price are martingales will tell us enough about the probabilities to calculate derivative values without having to calculate the probabilities themselves.

We conclude this section with another reformulation of the pricing relations Equation **??**–Equation **??**. This formulation will generalize more easily to pricing when there are a continuum of states, the subject of the next section.
Let prob_u denote the actual probability of the up state and prob_d denote the probability of the down state. These probabilities are irrelevant for pricing derivatives in the binomial model, but we will use them to write the pricing relations Equation **??**–Equation **??** as expectations with respect to the actual probabilities. To do this, we can define

$$\phi_u = \frac{\pi_u}{\text{prob}_u} ,$$
$$\phi_d = \frac{\pi_d}{\text{prob}_d} .$$

Then Equation **??**–Equation **??** can be written as

$$C = \text{prob}_u \phi_u C_u + \text{prob}_d \phi_d C_d , \tag{2.19}$$

$$S = \text{prob}_u \phi_u S_u + \text{prob}_d \phi_d S_d , \tag{2.20}$$

$$R = \text{prob}_u \phi_u R_u + \text{prob}_d \phi_d R_d . \tag{2.21}$$

The right-hand sides are expectations with respect to the actual probabilities. For example, the right-hand side of Equation **??** is the expectation of the random variable that equals $\phi_u C_u$ in the up state and $\phi_d C_d$ in the down state. The risk-neutral probabilities can be calculated from ϕ_u and ϕ_d as $p_u = \text{prob}_u \phi_u R_u / R$ and $p_d = \text{prob}_d \phi_d R_d / R$. Likewise, the probabilities using the stock as the numeraire can be calculated from ϕ_u and ϕ_d as $q_u = \text{prob}_u \phi_u S_u / S$ and $q_d = \text{prob}_d \phi_d S_d / S$. In the following section, we will assume (which can be shown to be true under some technical conditions) that relations such as Equation **??**–Equation **??** hold in a general (non-binomial) model given the absence of arbitrage opportunities. We will then show, using definitions analogous to the definitions of p_u, p_d, q_u, and q_d in this paragraph, that relations analogous to Equation **??**–Equation **??** and Equation **??**–Equation **??** hold.

2.4 Asset Pricing with a Continuum of States

In this section, we will define the concepts of state prices and probabilities corresponding to different numeraires in a more general framework than that of the preceding section. This leads to what we will call the fundamental pricing equation, namely Equation **??**. There are really no new concepts in this section, only a bit more mathematics.

Consider a non-dividend-paying security having the random price $S(T)$ at date T. We call the contingencies that affect the price $S(T)$ the states of the world. Our principle regarding state prices developed in the preceding section can in general be expressed as:[6] **if there are no arbitrage opportunities, there exists for each date T a positive random variable $\phi(T)$ such that the value at date 0 of a non-dividend-paying security with price S is**

$$S(0) = E[\phi(T)S(T)] \; . \tag{2.22}$$

Here, $E[\phi(T)S(T)]$ denotes the expectation of the random variable $\phi(T)S(T)$. The random variable $\phi(T)$ is called the state price density.[7] In a binomial model (or in any model with only a finite number of states of the world), the concept of an expectation is clear: it is just a weighted average of outcomes, the weights being the probabilities. In the binomial model, the right-hand side of Equation **??** is the same as the right-hand side of Equation **??**.[8]

To convert from state prices to probabilities corresponding to different numeraires, we follow the same procedure as at the end of the previous section: we multiply together (i) the probability of the state, (ii) the value of $\phi(T)$ in the state, and (iii) the gross return of the numeraire in the state. If there is a continuum of states, then the actual probability of any individual state will typically be zero, so this multiplication will produce a zero probability. However, we can nevertheless add up these probabilities to define the probability of any event A, an event being a set of states of the world. To do this, let 1_A denote the random variable that takes the value 1 when A is true and which is zero otherwise. Then the probability of A using S as the numeraire is defined as

$$E\left[1_A \phi(T)\frac{S(T)}{S(0)}\right] \; . \tag{2.23}$$

[6] We have proven this in the binomial model, but we will not prove it in general. As is standard in the literature, we will simply adopt it as an assumption. A general proof is in fact difficult and requires a definition of no arbitrage that is considerably more complicated than the simple assumption Equation **??** that is sufficient in the binomial model.

[7] The term density reflects the fact that in each state of the world $\phi(T)$ can be interpreted as the state price per unit of probability, just as the normal meaning of density is mass per unit of volume.

[8] In general the expectation (or mean) of a random variable is an intuitive concept, and an intuitive understanding will be sufficient for this book, so I will not give a formal definition. It should be understood that we are assuming implicitly, whenever necessary, that the expectation exists (which is not always the case). In this regard, it is useful to note in passing that a product of two random variables XY has a finite mean whenever X and Y have finite variances.

This makes sense as a probability because it is nonnegative and because, if A is the set of all states of the world, then its probability is $E[\phi(T)S(T)/S(0)]$, which equals one by virtue of Equation **??**. From Equation **??** of the probability of any event A, it can be shown that the expectation of any random variable X using S as the numeraire is

$$E\left[X\phi(T)\frac{S(T)}{S(0)}\right] . \tag{2.24}$$

The use of the symbol S to denote the price of the numeraire may be confusing, because S is usually used to denote a stock price. The numeraire here could be any non-dividend-paying asset. For example, we can take $S(t) = \mathrm{e}^{rt}$, the price of the risk-free asset. The definition of probabilities as

$$E[1_A\phi(T)\mathrm{e}^{rT}] \tag{2.25}$$

will be called the risk-neutral probability measure or simply risk-neutral measure as before.

Different numeraires lead to different probability measures and hence to different expectations. To keep this straight, we will use the numeraire as a superscript on the expectation symbol: for example, E^S will denote expectation with respect to the probability measure that corresponds to S being the numeraire. Also, we will use the symbol $\text{prob}^S(A)$ to denote the probability of an event A when we use S as the numeraire. So, Equation **??** and Equation **??** will be written as

$$\text{prob}^S(A) = E\left[1_A\phi(T)\frac{S(T)}{S(0)}\right] , \tag{2.26}$$

$$E^S[X] = E\left[X\phi(T)\frac{S(T)}{S(0)}\right] . \tag{2.27}$$

Our key result in the preceding section was that the ratio of the price of any non-dividend paying asset to the price of the numeraire asset is not expected to change when we use the probability measure corresponding to the numeraire. We will demonstrate the same result in this more general model. Recall that T denotes an arbitrary but fixed date at which we have defined the probabilities using S as the numeraire in Equation **??**. At each date $t < T$, let E^S_t denote the expectation given information at time t and using S as the numeraire (we will continue to write the expectation at date 0 without a subscript; i.e., E^S has the same meaning as E^S_0). Let Y denote the price of another non-dividend-paying asset. We will show that

$$\frac{Y(t)}{S(t)} = E^S_t\left[\frac{Y(T)}{S(T)}\right] . \tag{2.28}$$

Thus, the expected future (date–T) value of the ratio Y/S always equals the current (date–t) value when we use S as the numeraire. As discussed in the preceding section, the mathematical term for a random variable whose expected future value always equals its current value is martingale. Thus, we can express Equation **??** as stating that the ratio Y/S is a martingale when we compute expectations using the probability measure that corresponds to S being the numeraire.

The usefulness of Equation **??** is that it gives us a formula for the asset price $Y(t)$ at any time t—and recall that this formula holds for every non-dividend paying asset. The formula is obtained from Equation **??** by multiplying through by $S(t)$:

$$Y(t) = S(t)E_t^S\left[\frac{Y(T)}{S(T)}\right] . \tag{2.29}$$

We will call Equation ?? the fundamental pricing formula. It is at the heart of modern pricing of derivative securities. It is a present value relation: the value at time t of the asset is the expectation of its value $Y(T)$ at time T discounted by the (possibly random) factor $S(t)/S(T)$. To emphasize that the numeraire can be any non-dividend-paying asset (and not necessarily a stock price, as the symbol S might suggest), we can write Equation **??** in the equivalent form

$$Y(t) = \text{num}(t)E_t^{\text{num}}\left[\frac{Y(T)}{\text{num}(T)}\right] , \tag{2.30}$$

where now $\text{num}(t)$ denotes the price of the (non-dividend-paying) numeraire asset at time t.

For example, letting $R(t)$ denote the value e^{rt} of the risk-free asset and using it as the numeraire, Equation **??** becomes

$$Y(t) = \mathrm{e}^{rt}E_t^R\left[\frac{Y(T)}{\mathrm{e}^{rT}}\right] = \mathrm{e}^{-r(T-t)}E_t^R[Y(T)] , \tag{2.31}$$

which means that the value $Y(t)$ is the expected value of $Y(T)$ discounted at the risk-free rate for the remaining time $T-t$, when the expectation is computed under the risk-neutral probability measure.

We end this section with a proof of Equation **??**, a proof that the reader may skip if desired.[9]

Note

Consider any time t and any event A that is distinguishable by time t. Consider the trading strategy of buying one share of the asset with price Y at time t when A has happened and financing this purchase by short selling $Y(t)/S(t)$ shares of the asset with price S. Each share of this asset that you short brings in $S(t)$ dollars, so shorting $Y(t)/S(t)$ shares brings in $Y(t)$ dollars, exactly enough to purchase the desired share of

[9] The proof is due to Harrison and Kreps (Harrison and Kreps 1979). See also Geman, El Karoui and Rochet (Geman, El Karoui, and Rochet 1995). We omit here technical assumptions regarding the existence of expectations.

the first asset. Hold this portfolio until time T and then liquidate it. Liquidating it will generate

$$1_A\left(Y(T)-\frac{Y(t)}{S(t)}S(T)\right)$$

dollars. The multiplication by the random variable 1_A is because we only implement this strategy when A occurs (i.e., when $1_A=1$). Consider the security that pays this number of dollars at time T. Because we obtained it with a trading strategy that required no investment at time t, its price at time 0 must be 0. We already observed that we can represent the price in terms of state prices, so we conclude that

$$E\left[\phi(T)1_A\left(Y(T)-\frac{Y(t)}{S(t)}S(T)\right)\right]=0\ .$$

When we divide by $S(0)$, this will still equal zero. Factoring $S(T)$ outside the parentheses gives

$$E\left[1_A\frac{S(T)}{S(0)}\phi(T)\left(\frac{Y(T)}{S(T)}-\frac{Y(t)}{S(t)}\right)\right]=0\ .$$

We see from Equation **??** for expectations using S as the numeraire that we can write this as

$$E^S\left[1_A\left(\frac{Y(T)}{S(T)}-\frac{Y(t)}{S(t)}\right)\right]=0\ .$$

This is true for any event A distinguishable at time t, so the expectation of $Y(T)/S(T)-Y(t)/S(t)$ must be zero given any information at time t when we use S as the numeraire; i.e.,

$$E_t^S\left[\frac{Y(T)}{S(T)}-\frac{Y(t)}{S(t)}\right]=0\ ,$$

or, equivalently

$$E_t^S\left[\frac{Y(T)}{S(T)}\right]=\frac{Y(t)}{S(t)}\ .$$

2.5 Introduction to Option Pricing

A complete development of derivative pricing requires the continuous-time mathematics to be covered in the next chapter. However, we can present the basic ideas using the tools already developed. Consider the problem of pricing a European call option. Let T denote the maturity of the option and K its strike price, and let S denote the price of the underlying. We will assume for now that the underlying does not pay dividends, but we will make no assumptions about the distribution of its price $S(T)$ at the maturity of the option. Assume there is a risk-free asset with constant interest rate r.

Our convention will be that date 0 denotes the date at which we are attempting to value a derivative. The value of the option at maturity is $\max(0, S(T)-K)$. Consider a contract that pays $S(T)$ at date T when $S(T) \geq K$ and that pays zero when $S(T) < K$, and consider another contract that pays K at date T when $S(T) \geq K$ and zero when $S(T) < K$. In Chapter **??**, we will call the first contract a share digital and the second contract a digital. The call option is equivalent to a portfolio that is long the first contract and short the second, because the value of the call at maturity is $S(T) - K$ when $S(T) \geq K$ and it is zero otherwise. So, we can value the call if we can value the share digital and the digital. This splitting up of complex payoffs into simpler contracts is a key to analyzing many types of derivatives.

2.5.1 Pricing Share Digitals

Consider first the problem of valuing the share digital. Let $Y(t)$ denote its value at each date $t \leq T$. We seek to find $Y(0)$. Our fundamental pricing Equation **??** tells us that

$$Y(0) = \mathrm{num}(0) E^{\mathrm{num}} \left[\frac{Y(T)}{\mathrm{num}(T)} \right] ,$$

for any numeraire with price $\mathrm{num}(t)$. We want to choose the numeraire to simplify the calculation of the expectation. The expectation only involves the states of the world in which $S(T) \geq K$, because $Y(T) = 0$ when $S(T) < K$. In the states of the world in which $S(T) \geq K$, the value of the share digital is $S(T)$. The calculation of the expectation would be simplified if the value were a constant when it was nonzero, because, if you are to receive a constant amount in a certain event, your expected payoff is the constant times the probability of the event (e.g., the expected payoff of a gamble that pays $1 when a fair die rolls a 6 is 1/6). This suggests we should use the stock as the numeraire, because then we will have

$$\frac{Y(T)}{\mathrm{num}(T)} = \frac{S(T)}{S(T)} = 1$$

when $S(T) \geq K$, implying that

$$E^{\mathrm{num}} \left[\frac{Y(T)}{\mathrm{num}(T)} \right] = \mathrm{prob}^S(S(T) \geq K) ,$$

where prob^S denotes the probability using S as the numeraire. This implies that the value of the share digital is

$$S(0) \times \mathrm{prob}^S(S(T) \geq K) .$$

The remaining question is obviously how to compute the probability. We will **not** use Equation **??** which expresses the probability in terms of an expectation involving state prices. To attempt to do so would simply raise the question of how to compute the state prices. Instead, we use the fundamental pricing formula again, this time replacing the derivative value Y with the value of the risk-free asset. This is exactly analogous to computing the q probabilities from

Equation **??** and Equation **??** in Section **??**. Recall that the fundamental formula holds for any non-dividend-paying asset, so it holds for $R(t) = \mathrm{e}^{rt}$, telling us that the ratio $R(t)/S(t)$ is a martingale when we use S as the numeraire. In a continuous-time model (at least until we introduce stochastic volatility) this will give us exactly the information we need to compute the distribution of $S(T)$ when we use S as the numeraire, and from the distribution of $S(T)$ we can easily compute $\text{prob}^S(S(T) \geq K)$. This calculation will be covered in Chapter **??** for the Black-Scholes model.

2.5.2 Pricing Digitals

Now consider the problem of pricing the digital. We will change notation to let $Y(t)$ denote now the value of the digital at time t. Again we want to compute

$$Y(0) = \text{num}(0) E^{\text{num}} \left[\frac{Y(T)}{\text{num}(T)} \right] ,$$

and again this expectation only involves the states of the world in which $S(T) \geq K$. In these states of the world, the value of the digital is already a constant K, so we should take the numeraire to have a constant value at T, so that the ratio $Y(T)/\text{num}(T)$ will be constant in the states in which $S(T) \geq K$. This means that we should take the numeraire to be the risk-free asset. For this numeraire, the pricing formula is

$$Y(0) = \mathrm{e}^{-rT} E^R[Y(T)] = \mathrm{e}^{-rT} K \times \text{prob}^R(S(T) \geq K) ,$$

so we need to compute the risk-neutral probability that $S(T) \geq K$. We will do this by using the fact that $S(t)/R(t) = \mathrm{e}^{-rt} S(t)$ is a martingale under the risk-neutral probability measure. This is analogous to computing the risk-neutral probabilities from Equation **??** and Equation **??** in Section **??**. This calculation will also be covered in Chapter **??** for the Black-Scholes model.

Readers familiar with the Black-Scholes formula may already have surmised that, under the Black-Scholes assumptions,

$$\text{prob}^S(S(T) \geq K) = \mathrm{N}(d_1) \quad \text{and} \quad \text{prob}^R(S(T) \geq K) = \mathrm{N}(d_2) ,$$

where N denotes the cumulative normal distribution function . The numbers d_1 and d_2 are different, and hence these are different probabilities, even though they are both probabilities of the option finishing in the money ($S(T) \geq K$). They are different probabilities because they are computed under different numeraires.

2.5.3 A Remark

It seems worthwhile here to step back a bit from the calculations and try to offer some perspectives on the methods developed in this chapter. The change of numeraire technique probably seems mysterious. Even though one may agree that it works after following the steps in the chapter, there is probably a lingering question about why it works. The author's opinion is that it may be best to regard it simply as a computational trick. Fundamentally it works because valuation is linear. Linearity simply means that the value of a cash flow $X = X_1 + X_2$ is the sum of the values of the cash flows X_1 and X_2 and the value of the cash flow aX is a times the value of X, for any constant a. This linearity is manifested in the statement that the value of a cash flow is the sum across states of the world of the state prices multiplied by the size of the cash flow in each state. The change of numeraire technique exploits the linearity to further simplify the valuation exercise. There are other ways the linearity can be used (for example, it produces solvable partial differential equations) but the particular trick we have developed in this chapter seems the most useful to the author (and to others, though perhaps not to everyone). After enough practice with it, it will seem as natural as other computational tricks one might have learned.

2.6 An Incomplete Markets Example

In this section, we consider a more difficult valuation problem than the binomial model and discuss the general implications of this example. We only need to make the problem slightly more difficult to see the issues. Consider a trinomial model, in which the asset price takes three possible values: $S_u > S_m > S_d$ (m for middle, medium, median, …). We continue to make the no arbitrage assumption Equation **??**. State prices π_u, π_m and π_m must satisfy equations analogous to Equation **??**–Equation **??**; specifically,

$$S = \pi_u S_u + \pi_m S_m + \pi_d S_d \ , \tag{2.32}$$

$$1 = \pi_u \mathrm{e}^{rT} + \pi_m \mathrm{e}^{rT} + \pi_d \mathrm{e}^{rT} \ . \tag{2.33}$$

In the binomial case, these equations can be solved for π_u and π_d, as shown in Equation **??**. However, in the trinomial case, we have only two equations in three unknowns. Thus, there exist many solutions.

Given any particular solution (π_u, π_m, π_d) of Equation **??** - Equation **??**, we can define the risk-neutral probabilities p_u, p_m and p_d as before—e.g., $p_u = \pi_u \mathrm{e}^{rT}$. Likewise, we can define the probabilities using the stock as numeraire. Thus, we can value calls and puts and other derivative securities. However, the values we obtain will depend on the particular solution (π_u, π_m, π_d). There are many arbitrage-free values for a call option, one for each solution of Equation **??** - Equation **??**.

The reason that there are many arbitrage-free values for a call (or put) is that a call cannot be replicated in a trinomial model using the stock and risk-free asset; we can say equivalently that there is no delta hedge for a call option. Recall that we first found the value of a call in the binomial model by finding the replicating portfolio and calculating its cost. A similar analysis is impossible in the trinomial model. To see this, consider a portfolio of a dollars invested in the risk free asset and b dollars invested in the stock. The value of the portfolio at date T will be $a\mathrm{e}^{rT} + bS_x/S$, where $x \in \{u, m, d\}$. To replicate the call, we need a and b to satisfy

$$\begin{aligned}
a\mathrm{e}^{rT} + bS_u/S &= \max(0, S_u - K)\ , \\
a\mathrm{e}^{rT} + bS_m/S &= \max(0, S_m - K)\ , \\
a\mathrm{e}^{rT} + bS_d/S &= \max(0, S_d - K)\ .
\end{aligned}$$

These are three linear equations in the two unknowns a and b. For any strike price K between S_d and S_u, none of the equations is redundant, and the system has no solution. When there are state-contingent claims (such as the call option payoff) that cannot be replicated by trading in the marketed assets (the stock and risk-free asset in this case), one says that the market is incomplete. Thus, the trinomial model is an example of an incomplete market.

To value derivative securities in this situation, we have to select some particular solution (π_u, π_m, π_d) of Equation **??** - Equation **??** and assume that the market uses that solution for valuation. Equivalently, we can assume the market uses a particular set of risk-neutral probabilities (p_u, p_m, p_d). This type of valuation is often called equilibrium valuation, as opposed to arbitrage valuation, because to give a foundation for our particular choice of risk-neutral probabilities, we would have to assume something about the preferences and endowments of investors and the production possibilities. We will encounter incomplete markets when we consider stochastic volatility in Chapter **??**.

2.7 Exercises

Exercise 2.1. Create an Excel worksheet in which the user inputs S, S_d, S_u, K, r and T. Check that the no-arbitrage condition Equation **??** is satisfied. Compute the value of a call option in each of the following ways:

1. Compute the delta and use Equation **??**.
2. Compute the state prices and use Equation **??**.
3. Compute the risk-neutral probabilities and use Equation **??**.
4. Compute the probabilities using the stock as numeraire and use Equation **??**.

Verify that all of these methods produce the same answer.

Exercise 2.2. In a binomial model, a put option is equivalent to δ_p shares of the stock, where $\delta_p = (P_u - P_d)/(S_u - S_d)$ (this will be negative, meaning a short position) and some money invested in the risk-free asset. Derive the amount of money x that should be invested in the risk-free asset to replicate the put option. The value of the put at date~0 must be $x + \delta_p S$.

Exercise 2.3. Using the result of the previous exercise, repeat Problem~**??** for a put option.

Exercise 2.4. Here is a chance to apply option pricing theory to real life. Suppose you have a significant other who would marry you if you ask.

1. What type of option do you have on marriage? Can you tell when it is in the money?
2. Under what circumstances should you exercise this option early?

3. What is the put option in a marriage contract called?

3 Continuous-Time Models

3.1 From Discrete to Continuous Time

Our main point of departure is knowing how a given quantity evolves through time. For example, $X(t)$ might denote how much wealth you have at time t. Partition the interval $[0, T]$ into N intervals of length $\Delta t = \frac{T}{N}$, in other words

$$[0, \Delta t = \frac{T}{N}, 2\Delta t, ..., (N-1)\Delta t, N\Delta t = T].$$

If you invest your wealth $X((i-1)\Delta t)$ at a fixed rate of interest $r\Delta t$ over a period of length $\Delta t = \frac{T}{N}$, you will have $X((i-1)\Delta t)(1 + r\Delta t)$ at the end of the period; the change of your wealth from $(i-1)\Delta t$ to $t = i\Delta t$ is

$$\Delta X(t) = X(i\Delta t) - X((i-1)\Delta t) = (1 + r\Delta t)X((i-1)\Delta t) - X(t) = rX(t)\Delta t.$$

The limiting form of this equation is

$$X(t) = rX(t)dt \quad \text{or} \quad \frac{dX(t)}{dt} = rX(t)$$

which is a differential equation. Notice this can also be written as a sum

$$X(N\Delta t) = X(0) + \sum_{i=0}^{N}(X(i\Delta t) - X((i-1)\Delta t)) = X(0) + \sum_{i=0}^{N} rX((i-1)\Delta t)\Delta t;$$

in the limit we get an integral equation

$$X(T) = X(0) + \int_0^T rX(t)dt.$$

The solution is $X(T) = X(0)e^{rT}$ which is easily verified by taking the derivative $\frac{d}{dt}e^{rt}X(t) = rX(t)$ or by inserting the solution into the integral equation. This is also the limiting value of the discretely compunded interest as the number of compounding periods n goes to infinity:

$$X(T) = \lim_{t \to \infty} \left(1 + r\frac{T}{N}\right)^N.$$

This can be verified by taking the log of $\left(1 + r\frac{T}{N}\right)^N$ and using l'Hospital's rule. Given the initial wealth, $X(0)$ and the rate of change $rX(t)$ (which, if the unit of account is dollars, is in units dollars / time), we can we solve for the value of our wealth at all times.

```
import plotly.graph_objects as go
import numpy as np

trace1= go.Scatter(
    x=np.arange(31),
    y=1.08**np.arange(31),
    mode="lines",
    name="8%"
)

trace2 = go.Scatter(
    x=np.arange(31),
    y=1.04**np.arange(31),
    mode="lines",
    name="4%"
)

fig = go.Figure()
fig.add_trace(trace1)
fig.add_trace(trace2)
string = "year %{x}<br>balance = $%{y:.2f}"
fig.update_traces(hovertemplate=string)
fig.update_layout(
    template="none",
    xaxis_title="Year",
    yaxis_title="Account Balance",
    yaxis_tickprefix="$",
    yaxis_tickformat=",.0f",
    legend=dict(
        yanchor="top",
        y=0.99,
        xanchor="left",
        x=0.01
        )
```

```
)
fig.show()
```

Figure 3.1

(a) Growth of an investment account at different rates of return.

```
Unable to display output for mime type(s): text/html
```

(b)

```
Unable to display output for mime type(s): text/html
```

However, most investments involve risk. We might think that part of the return is more or less known while a portion of the investment is random or risky. Therefore we would like a model which can be written as

$$dX(t) = \mu X(t)dt + \text{`noise'} \times X(t).$$

It makes sense that the units of the noise should be in the unit of account (dollars) and it should have zero mean but it is not obvious how to make sense of a noisy change over an instant. In the next section we examine the most popular choice for this model, namely Brownian motion. Loosely speaking, we can take expectations to get

$$dE[X(t)] = \mu E[X(t)]dt$$

which gives a solution

$$E[X(T)] = X(0)e^{\mu T}.$$

So we can think of this model as a known expected continuously compunded investment at a rate μ plus a noisy return.

Historically, these deterministic mathematics trace back to Issac Newton who developed the classical laws of motion. The model of noise was developed by Ito to account for random disturbances, such as unpredictable wind gusts, to a deterministic model, for example, in a model predicting the position of a falling object.

This chapter has three objectives. The first is to introduce the concept of a Brownian motion. A Brownian motion is a random process (a variable that changes randomly over time) that evolves continuously in time and has the property that its change over any time period is normally distributed with mean zero and variance equal to the length of the time period. The mean zero feature means that a Brownian motion is a martingale. We will also give a different characterization (Levy's theorem) emphasizing the quadratic variation process, which is a property of the paths (how the variable evolves over time, in a given state of the world) of the process.

The second objective is to explain Ito's formula, which is the chain rule for stochastic calculus. In the Black-Scholes model, the stock price is assumed to satisfy

$$\frac{\mathrm{d}S}{S} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B\ ,$$

where B is a Brownian motion. In the case that the stock pays no dividend, the rate of return is its price change $\mathrm{d}S$ divided by the initial price S, so the model states that the expected rate of return in each instant $\mathrm{d}t$ is $\mu\,\mathrm{d}t$ (of course, t denotes time, so $\mathrm{d}t$ is the change in time). The variance of the rate of return depends on σ. This model can be equivalently written in terms of the natural logarithm of S, which we will write as $\log S$. The above equation for the rate of return is equivalent to

$$\mathrm{d}\ \log S = \left(\mu - \frac{1}{2}\sigma^2\right)\ \mathrm{d}t + \sigma\,\mathrm{d}B\ .$$

We will explain this equivalence and other similar calculations that are useful for pricing derivatives.

The third objective is to explain how, when we change numeraires, as described in the previous chapter, we can calculate the expectation in the fundamental pricing Equation **??**. The question is what effect does changing the numeraire (and hence the probability measure) have on the distribution of an asset price.

Everything in the remainder of the book is based on the mathematics presented in this chapter. For easy reference, the essential formulas have been highlighted in boxes.

3.2 Simulating a Brownian Motion

We begin with the fact that changes in the value of a Brownian motion are normally distributed with mean zero and variance equal to the length of the time period. Let $B(t)$ denote the value of a Brownian motion at time t. Then for any date $u > t$, given the information at time t, the random variable $B(u) - B(t)$ is normally distributed with mean zero and variance equal to $u - t$. Unless stated otherwise, our convention will be that a Brownian motion starts at $B(0) = 0$.

We can generate an approximate Brownian motion in Python. To do so, we take a small time period Δt and define the value at the end of the period to be the value of the Brownian motion at the beginning plus a normally distributed variable with mean 0 and variance Δt. In the following procedure, we input the length $T = 0.5$ of the entire time period over which the

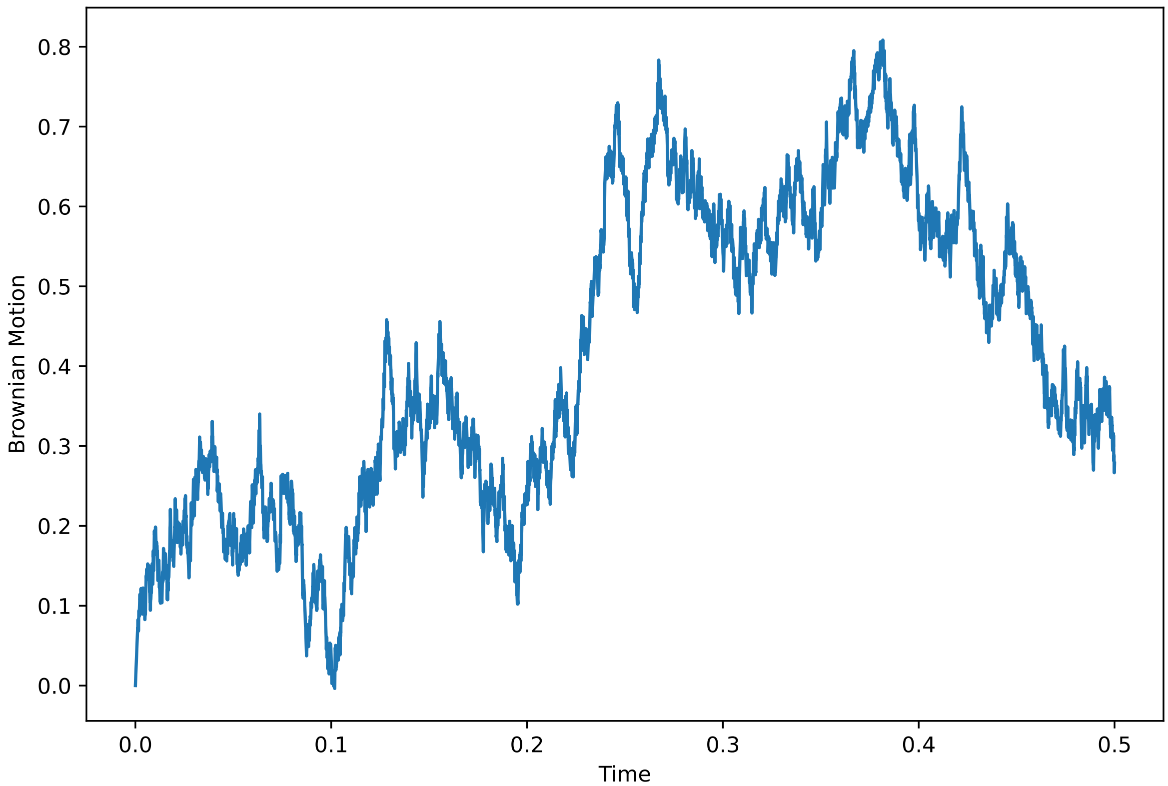
Brownian motion is to be simulated. One input the is number $n = 10000$ of time periods of length Δt within the full interval $[0, T]$. The length Δt of each individual time period is then calculated as T/n. The quality of the approximation of this simulation to a true Brownian motion will be always be improved by increasing the number n. Plotting the output of the procedure creates a picture of what we call a path of the Brownian motion, which means that it shows the value taken at each time in one state of the world. The procedure generates $m = 2$ paths which can be interpreted as the values of the Brownian motion in another state of the world. In other words, the path of the Brownian motion is itself random, depending in this approximation on the numbers produced by Python's random number generating function. The random number generator np.random.normal(loc = 0, scale = vol,size = (n,m)) is an algorithm which produces numbers which mimic $n \times m$ normally distributed random variables with mean 0 (loc=0) and standard deviation $\sqrt{\Delta t}$ (scale=vol). By setting the seed=1234 the generator is initialized and we will always get the same output whenever we run the code. The Brownian path is just the cumulative sum of n of the normal increments.

```
import numpy as np
import matplotlib.pyplot as plt

# number of subdivisions
n = 10000
# number of paths
m = 2
# last date
T = 0.5
# delta t
dt = T/n
# volatility is standard deviation
vol = np.sqrt(dt)
# seed for random generator
seed= 1234
# define a random generator
np.random.seed(seed)

# generate "dB" for each date on each path distributed N(0,vol)
inc = np.random.normal(loc = 0, scale = vol,size = (n,m))
Bt = np.zeros(shape = (n + 1, m))

# Brownian path starts at 0 and is cumulative sum of the dB
Bt[1:] = inc.cumsum(axis = 0)
# Could do previous step in a loop
for i in range(1, n + 1):
    Bt[i] = Bt[i - 1] + inc[i - 1]
```

```
# plot one path
t = np.array(range(0, n + 1, 1)) * dt
plt.figure(figsize=(9,6))
plt.plot(t,Bt[:,0])
plt.xlabel("Time")
plt.ylabel("Brownian Motion")
```

```
Text(0, 0.5, 'Brownian Motion')
```

Simulated Brownian Motion



–>

3.3 Quadratic Variation

If we take a large number n of time steps in the simulation of the preceding section, we will see the distinctive characteristic of a Brownian motion: it jiggles rapidly, moving up and down

in a very erratic way. The name Brownian motion derives from the botanist Robert Brown's observations of the erratic behavior of particles suspended in a fluid. This has long been thought to be a reasonable model for the behavior of a stock price.
The plot of other functions with which we may be familiar will be much smoother. This is captured in the concept of quadratic variation.

Consider a discrete partition

$$0 = t_0 < t_1 < t_2 < \cdots < t_N = T$$

of the time interval $[0, T]$. Let B be a Brownian motion and calculate the sum of squared changes

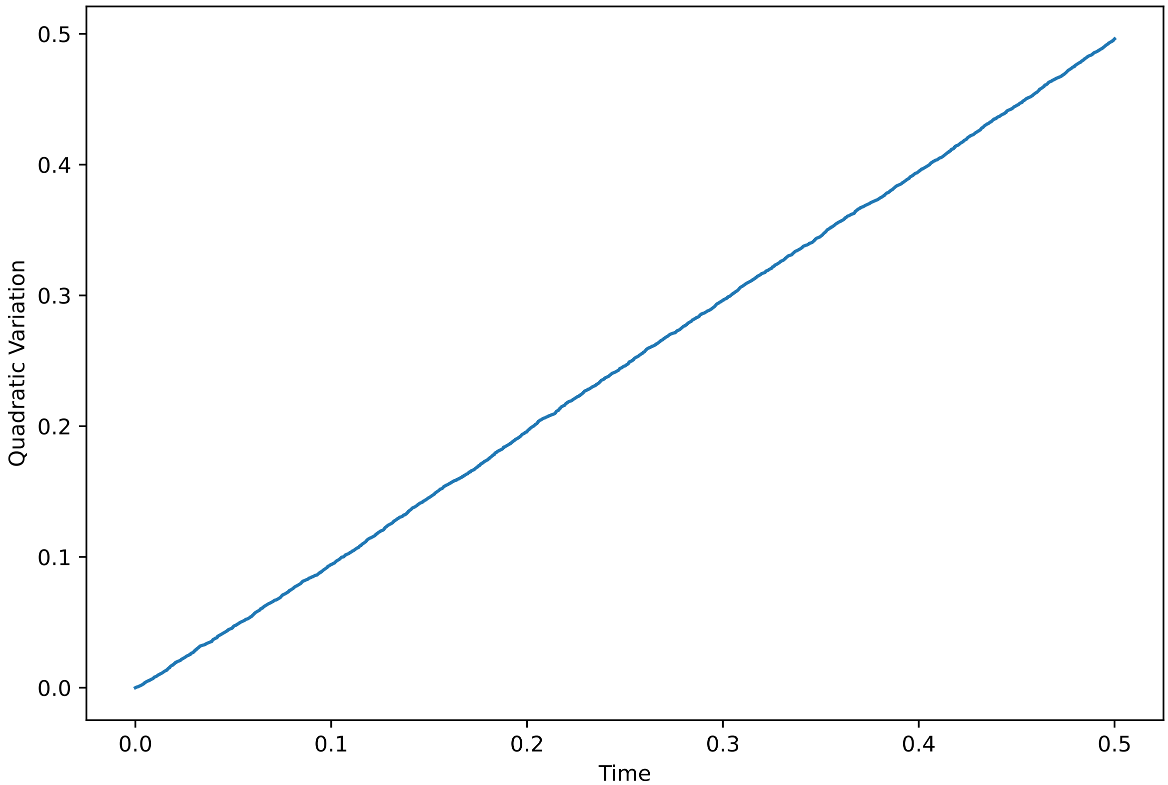
$$\sum_{i=1}^{N} [\Delta B(t_i)]^2 ,$$

where $\Delta B(t_i)$ denotes the change $B(t_i) - B(t_{i-1})$. If we consider finer partitions with the length of each time interval $t_i - t_{i-1}$ going to zero, the limit of the sum is called the quadratic variation of the process. For a Brownian motion, the quadratic variation over an interval $[0, T]$ is equal to T with probability one. Here is a plot of the quadratic variation. For large values of n, it is equal to t.

```
# quadratic variation
Q = np.zeros(shape = (n, m))
# Generate dB^2
Q = (Bt[1:n + 1] - Bt[0:n])**2

# Quadratic variation is the cumulative sum of dB^2
QV = np.zeros(shape = (n + 1, m))
QV[1:] = Q.cumsum(axis = 0)
plt.figure(figsize=(9,6))
plt.plot(t,QV[:,0])
plt.xlabel("Time")
plt.ylabel("Quadratic Variation")
```

```
Text(0, 0.5, 'Quadratic Variation')
```

Quadratic Variation



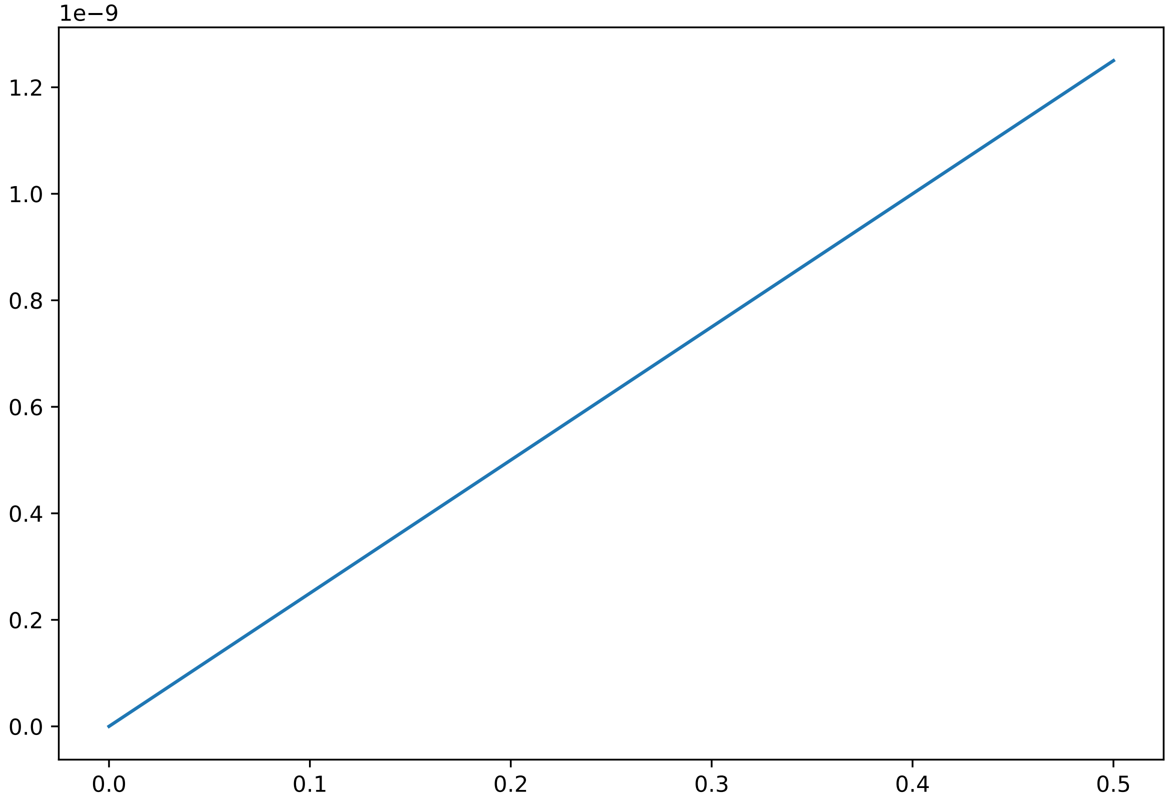
The functions with which we are normally familiar are continuously differentiable. If X is a continuously differentiable function of time (in each state of the world), then the quadratic variation of X will be zero. A simple example is a linear function: $X(t) = at$ for some constant a. Then, taking $t_i - t_{i-1} = \Delta t = T/N$ for each i, the sum of squared changes is

$$\sum_{i=1}^{N} [\Delta X(t_i)]^2 = \sum_{i=1}^{N} [a\,\Delta t]^2 = Na^2(\Delta t)^2 = Na^2 \left(\frac{T}{N}\right)^2 = \frac{a^2T^2}{N} \to 0$$

as $N \to \infty$. Essentially the same argument shows that the quadratic variation of any continuously differentiable function is zero, because such a function is approximately linear at each point. Below is a plot of the quadratic variation for the function $X(t) = t$. Look carefully at the scale of the y-axis

```
# Plot dt^2
plt.figure(figsize=(9,6))
plt.plot(t,t * dt **2)
```

Figure 3.2: Approximate Squared Variation of $f(t) = t$



Thus, the jiggling of a Brownian motion, which leads to the nonzero quadratic variation, is quite unusual. To explain exactly how unusual it is, it is helpful to introduce the concept of total variation, which is defined in the same way as quadratic variation but with the squared changes $[\Delta B(t_i)]^2$ replaced by the absolute value of the changes $|\Delta B(t_i)|$. If the quadratic variation of a continuous function is nonzero, then its total variation is necessarily infinite, so each path of a Brownian motion has infinite total variation (with probability one). It was mentioned above that, with a large number of time steps in the simulation of the preceding section, one could see the distinctive jiggling property of a Brownian motion. This is not quite right. Any plot drawn by a pencil (or a laser printer, for that matter) must have finite total variation, because the total variation is the total distance traveled by the pencil. Hence, no matter how many time steps one uses, one will never create a continuous plot with the nonzero quadratic variation (and infinite total variation) that a Brownian path has. Another way to understand this is to consider focusing on a small segment of a plot and viewing it with a magnifying glass. If the segment is small enough, and excluding the finite number of kinks that a pencil can draw in the plot of a function, it will look approximately like a straight line under the magnifying glass (with slope equal to the derivative of the function). However, if one could view a segment of a path of a true Brownian motion under a magnifying glass, it would look much the same as the entire picture does to the naked eye—no matter how small the segment, one would still see the characteristic jiggling.

One may well question why we should be interested in this curious mathematical object. The reason is that asset pricing inherently involves martingales (variables that evolve randomly over time in such a way that their expected changes are always zero), as our fundamental pricing Equation **??** establishes. Furthermore, continuous processes (variables whose paths are continuous functions of time) are much more tractable mathematically than are processes that can jump at some instants. More importantly, it is possible in a mathematical model with continuous processes to define perfect hedges much more readily than it is in a model involving jump processes. So, we are led to a study of continuous martingales. An important fact is that any non-constant continuous martingale must have infinite total variation! So, the normal functions with which we are familiar are left behind once we enter the study of continuous martingales.

There remains perhaps the question of why we focus on Brownian motion within the world of continuous martingales. The answer here is that any continuous martingale is really just a transformation of a Brownian motion. This is a consequence of the following important fact, which is known as Levy's theorem:

> Tip
>
> A continuous martingale is a Brownian motion if and only if its quadratic variation over each interval $[0, T]$ equals T.

Thus, among continuous martingales, a Brownian motion is defined by the condition that the quadratic variation over each interval $[0, T]$ is equal to T. This is really just a normalization. A different continuous martingale may have a different quadratic variation, but it can be converted to a Brownian motion just by deforming the time scale. Furthermore, many continuous martingales can be constructed as stochastic integrals with respect to a Brownian motion. We take up this topic in the next section.

3.4 Ito Processes

An Ito process is a variable X that changes over time as

$$\mathrm{d}X(t) = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B(t)\;, \tag{3.1}$$

where B is a Brownian motion, and μ and σ can also be random processes. Some regularity conditions are needed on μ and σ which we will omit, except for noting that $\mu(t)$ and $\sigma(t)$

should be known at time t. In particular, constant μ and σ are certainly acceptable. When we add the changes over time, we get

$$X(T) = X(0) + \int_0^T \mu(t)\,\mathrm{d}t + \int_0^T \sigma(t)\,\mathrm{d}B(t)$$

for any $T > 0$. There are other types of random processes, in particular, processes that can jump, but we will not consider them in this book.

We will not formally define the integral $\int_0^T \sigma(t)\,\mathrm{d}B(t)$, but it should be understood as being approximately equal to a discrete sum of the form

$$\sum_{i=1}^{N} \sigma(t_{i-1})\,\Delta B(t_i)\ ,$$

where $0 = t_0 < \cdots t_N = T$ and the time periods $t_i - t_{i-1}$ are small. Given that we can simulate the changes $\Delta B(t_i)$ as random normals, we can approximately simulate the random variable $\int_0^T \sigma(t)\,\mathrm{d}B(t)$ and hence we can approximately simulate $X(T)$.

An Ito process evolves continuously over time. We interpret $\mu(t)\,\mathrm{d}t$ as the expected change in X in an instant $\mathrm{d}t$. The quantity $\mu(t)$ is also called the drift of the process X at time t. The coefficient $\sigma(t)$ is called the diffusion coefficient of X at time t.

If μ and σ are constant, it is standard to refer to an Ito process X as a (μ, σ)–Brownian motion. In this case we have

$$X(t) = \mu t + \sigma B_t$$

Of course, it is not a martingale when $\mu \neq 0$. For example, when $\mu > 0$, X tends to increase over time. However, it has the jiggling property of a Brownian motion, scaled by the diffusion coefficient σ.

A very important fact is that an Ito process such as Equation **??** can be a martingale only if $\mu = 0$. This should seem sensible, because $\mu\,\mathrm{d}t$ is the expected change in X, and a process is a martingale only if its expected change is zero.[1] This observation plays a fundamental role in deriving asset pricing formulas, as we will begin to see in Section **??**. Conversely, if $\mu = 0$ and

$$E\left[\int_0^T \sigma^2(t)\,\mathrm{d}t\right] < \infty \tag{3.2}$$

for each T,
then the Ito process is a continuous martingale and the variance of its date–T value, calculated

[1] If the sources of uncertainty in the market can be modeled as Brownian motions, then in fact every martingale is an Ito process with $\mu = 0$. This is some justification for the assumption we will make in this book, when studying continuous-time models, that all martingales are Ito processes.

with the information available at date 0, is:

$$\operatorname{var}[X(T)] = E\left[\int_0^T \sigma^2(t)\,\mathrm{d}t\right] .$$

Whether μ is zero or not, and independently of the assumption Equation **??**, the quadratic variation of the Ito process X is

$$\lim_{N\to\infty}\sum_{i=1}^{N}[\Delta X(t_i)]^2 = \int_0^T \sigma^2(t)\,\mathrm{d}t \tag{3.3}$$

with probability one. Thus we obtain (when $\mu = 0$ and Equation **??** holds) a continuous martingale with a different quadratic variation than a Brownian motion via the diffusion function σ. In fact, when Equation **??** holds, a somewhat more precise definition of the stochastic integral is the (unique) martingale with quadratic variation given by Equation **??**.

To compute the quadratic variation of an Ito process, we use the following simple and important rules (for the sake of brevity, we drop the (t) notation from $B(t)$ here and sometimes later):

> Tip
>
> $$(\mathrm{d}t)^2 = 0\ , \tag{3.4}$$
>
> $$(\mathrm{d}t)(\mathrm{d}B) = 0\ , \tag{3.5}$$
>
> $$(\mathrm{d}B)^2 = \mathrm{d}t\ . \tag{3.6}$$

We apply these rules to compute the quadratic variation of X as follows:

> Tip
>
> If $\mathrm{d}X = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B$ for a Brownian motion B, then
>
> $$\begin{aligned}(\mathrm{d}X)^2 &= (\mu\,\mathrm{d}t + \sigma\,\mathrm{d}B)^2\\ &= \mu^2(\mathrm{d}t)^2 + 2\mu\sigma(\mathrm{d}t)(\mathrm{d}B) + \sigma^2(\mathrm{d}B)^2\\ &= 0 + 0 + \sigma^2\,\mathrm{d}t\ .\end{aligned}$$

We integrate this from 0 to T to obtain the quadratic variation Equation **??** over that time period:[2]

$$\int_0^T (\mathrm{d}X(t))^2 = \int_0^T \sigma^2(t)\,\mathrm{d}t\ . \tag{3.7}$$

[2] In a more formal mathematical presentation, one normally writes $\mathrm{d}\langle X, X\rangle$ for what we are writing here as $(\mathrm{d}X)^2$. This is the differential of the quadratic variation process, and the quadratic variation through date

3.5 Ito's Formula

First we recall some facts of the ordinary calculus. If $y = g(x)$ and $x = f(t)$ with f and g being continuously differentiable functions, then

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{\mathrm{d}y}{\mathrm{d}x} \times \frac{\mathrm{d}x}{\mathrm{d}t} = g'(x(t))f'(t) \;.$$

Over a time period $[0, T]$, this implies that

$$y(T) = y(0) + \int_0^T \frac{\mathrm{d}y}{\mathrm{d}t}\,\mathrm{d}t = y(0) + \int_0^T g'(x(t))f'(t)\,\mathrm{d}t \;.$$

Substituting $\mathrm{d}x(t) = f'(t)\,\mathrm{d}t$, we can also write this as

$$y(T) = y(0) + \int_0^T g'(x(t))\,\mathrm{d}x(t) \;. \tag{3.8}$$

We can contrast Equation **??** with a special case of Ito's formula for the calculus of Ito processes (the more general formula will be discussed in the next section). If B is a Brownian motion and $Y = g(B)$ for a twice-continuously differentiable function g, then

$$Y(T) = Y(0) + \int_0^T g'(B(t))\,\mathrm{d}B(t) + \frac{1}{2}\int_0^T g''(B(t))\,\mathrm{d}t \;. \tag{3.9}$$

Thus, relative to the ordinary calculus, Ito's formula has an extra term involving the second derivative g''. We can write Equation **??** in differential form as

$$dY(t) = \frac{1}{2}g''(B(t))\,\mathrm{d}t + g'(B(t))\,\mathrm{d}B(t).$$

Thus, $Y = g(B)$ is an Ito process with drift $g''(B(t))/2$ and diffusion coefficient $g'(B(t))$.

To gain some intuition for the extra term in Ito's formula, we return to the ordinary calculus. Given dates $t < u$, the derivative defines a linear approximation of the change in y over this time period; i.e., setting $\Delta x = x(u) - x(t)$ and $\Delta y = y(u) - y(t)$, we have the approximation

$$\Delta y \approx g'(x(t))\,\Delta x \;.$$

A better approximation is given by the second-order Taylor series expansion

$$\Delta y \approx g'(x(t))\,\Delta x + \frac{1}{2}g''(x(t))\,[\Delta x]^2 \;.$$

T is

$$\langle X, X\rangle(T) = \int_0^T \mathrm{d}\langle X, X\rangle(t) = \int_0^T \sigma^2(t)\,\mathrm{d}t \;.$$

An interpretation of Equation **??** is that the linear approximation works perfectly for infinitesimal time periods $\mathrm{d}t$, because we can compute the change in y over the time period $[0, T]$ by summing up the infinitesimal changes $g'(x(t))\,\mathrm{d}x(t)$. In other words, the second-order term $\frac{1}{2}g''(x(t))\,[\Delta x]^2$ vanishes when we consider very small time periods.

The second-order Taylor series expansion in the case of $Y = g(B)$ is

$$\Delta Y \approx g'(B(t))\,\Delta B + \frac{1}{2}g''(B(t))\,[\Delta B]^2 \ .$$

For example, given a partition $0 = t_0 < t_1 < \cdots < t_N = T$ of the time interval $[0, T]$, we have, with the same notation we have used earlier,

$$\begin{aligned} Y(T) &= Y(0) + \sum_{i=1}^{N} \Delta Y(t_i) \\ &\approx Y(0) + \sum_{i=1}^{N} g'(B(t_{i-1}))\,\Delta B(t_i) + \frac{1}{2}\sum_{i=1}^{N} g''(B(t_{i-1}))\,[\Delta B(t_i)]^2 \ . \end{aligned} \tag{3.10}$$

If we make the time intervals $t_i - t_{i-1}$ shorter, letting $N \to \infty$, we cannot expect that the extra term here will disappear, leading to the result Equation **??** of the ordinary calculus, because we know that

$$\lim_{N\to\infty} \sum_{i=1}^{N} [\Delta B(t_i)]^2 = T \ ,$$

whereas for the continuously differentiable function $x(t) = f(t)$, the same limit is zero. In fact it seems sensible to interpret the limit of $[\Delta B]^2$ as $(\mathrm{d}B)^2 = \mathrm{d}t$. This is perfectly consistent with Ito's formula: if we take the limit in Equation **??**, replacing the limit of $[\Delta B(t_i)]^2$ with $(\mathrm{d}B)^2 = \mathrm{d}t$, we obtain Equation **??**.

The code below defines a function $g(x) = e^x$ (g'(x)=e^x$ and g''(x) = e^x$) and simulates the value e^{B_t} in two ways. The first way is to simulate the Ito expansion

$$e^{B_t} = 1 + \int_0^t e^{B_s} dB_s + \frac{1}{2}\int_0^t e^{B_s} ds$$

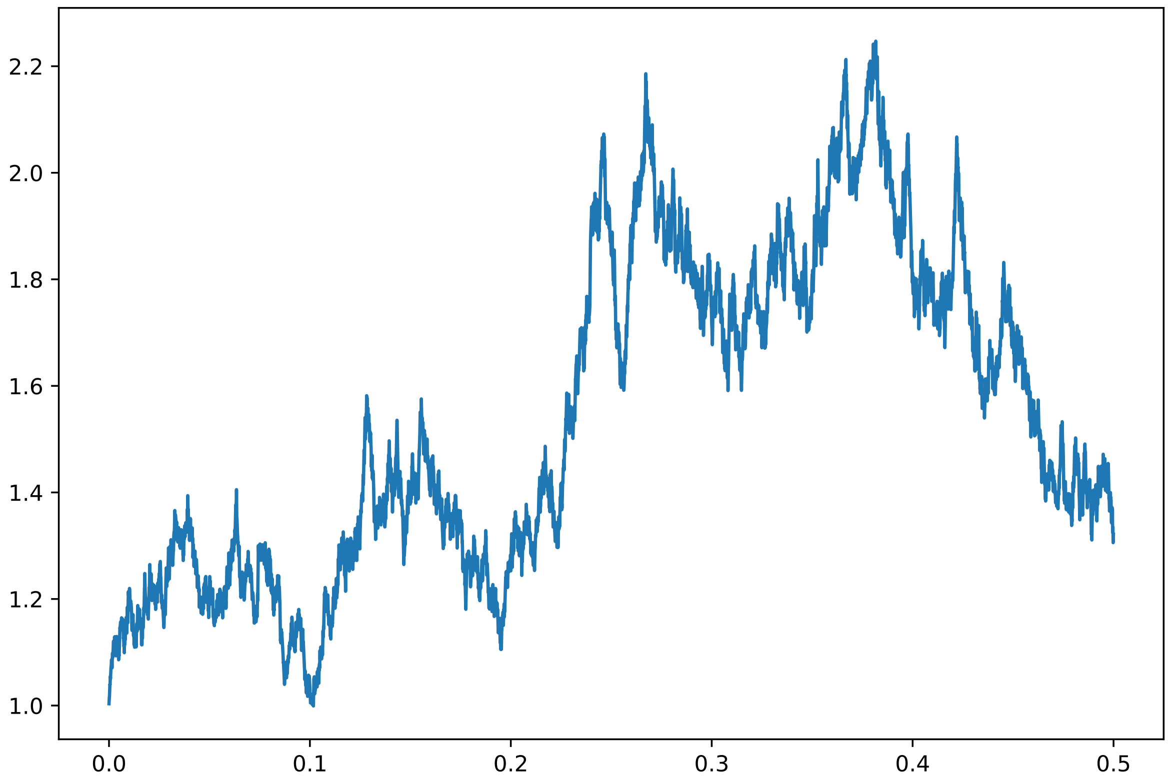
using the discretization

$$\Delta e^{B_t} = e^{B_t}\Delta B_t + \frac{1}{2}e^{B_t}\Delta t$$

```
# Define a function and its first and second derivative
G = lambda x: np.exp(x)
DG = lambda x: np.exp(x)
DDG = lambda x: np.exp(x)
# Build G'(x)dB
```

```
GdB = np.zeros(shape = (n, m))
GdB[0] = np.repeat(DG(0), m) * inc[0]
GdB[1:] = DG(Bt[0:n - 1]) * inc[1:]
SI = np.zeros(shape = (n, m))
# Stochastic Integral is cumulative sum of G'(B)dB plus initial
SI = GdB.cumsum(axis = 0) + 0.5 * (DDG(Bt[0:n])*Q).cumsum(axis =0) + G(0)
# Compare Ito's Lemma
plt.figure(figsize=(9,6))
plt.plot(t[0:n], SI[:,0])
```

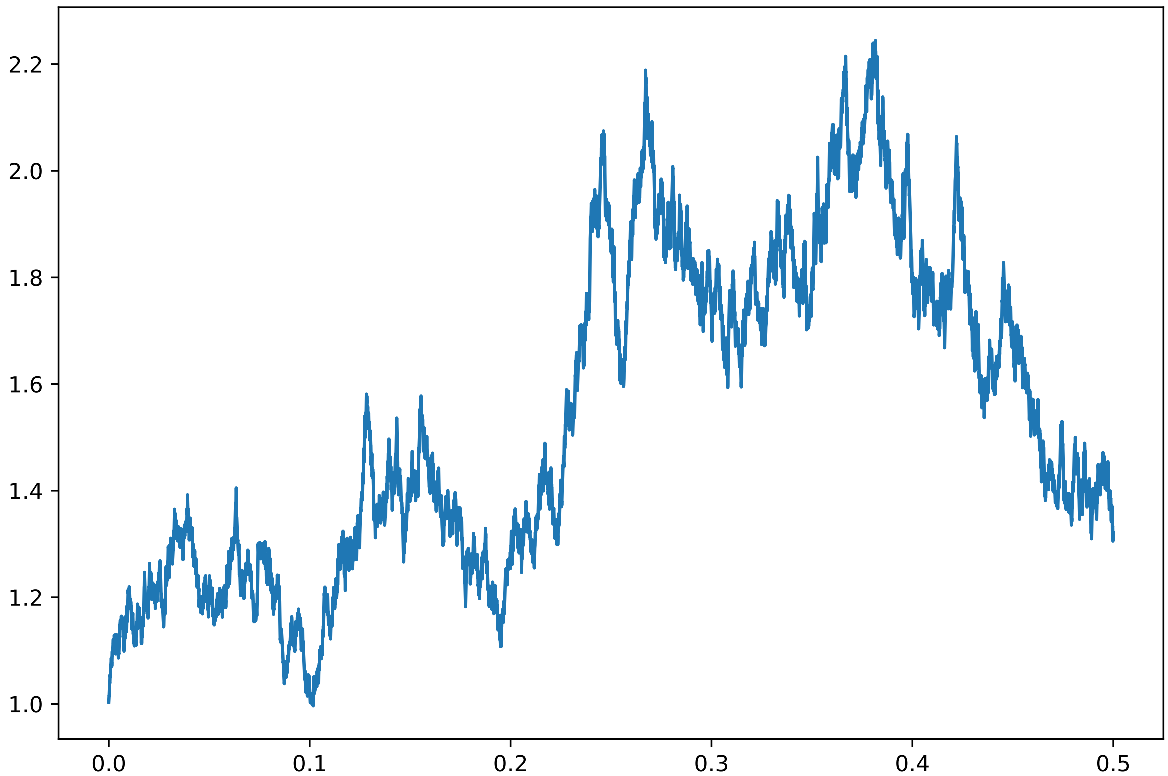
Figure 3.3: Simulated Ito Formula



Below is the exact simulated solution e^{B_t}. It is almost impossible to see a difference.

```
#exact solution
plt.figure(figsize=(9,6))
plt.plot(t[0:n], G(Bt[1:n + 1, 0]))
```

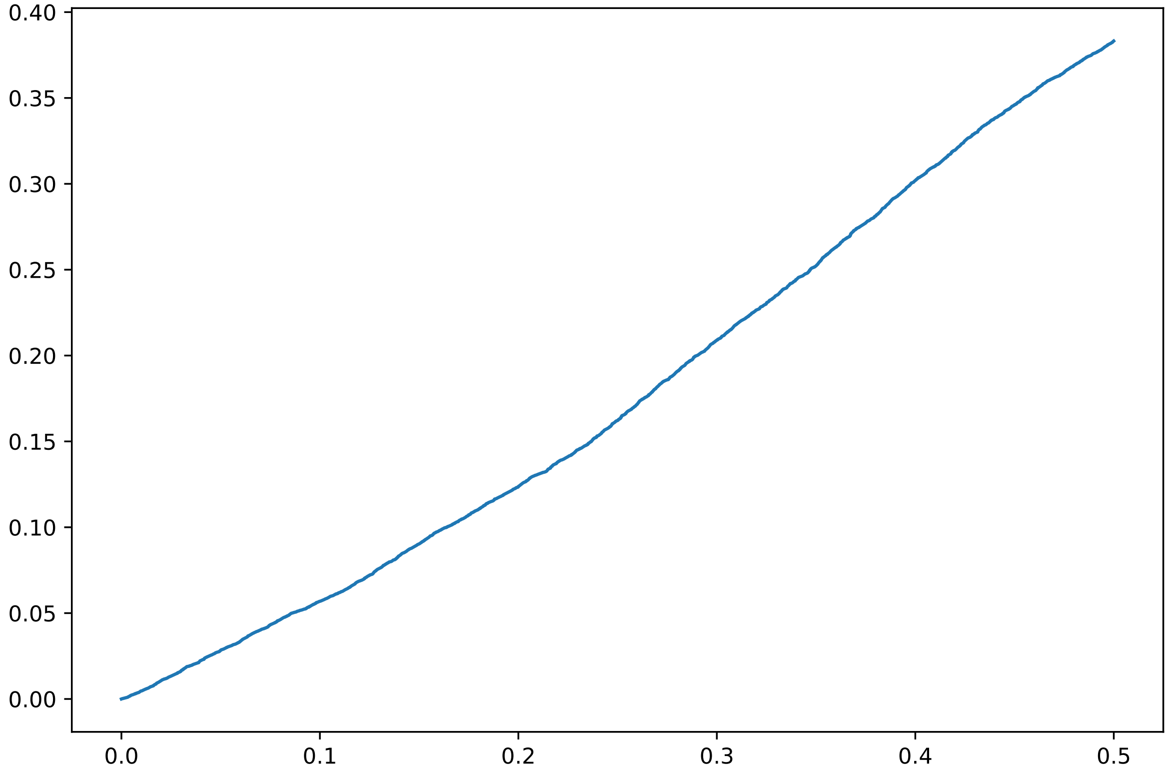
Figure 3.4: Simulated Exact Solution



Below we plot the $\int_0^t \frac{1}{2} g''(B_s)ds$ term. Without this term the two plots above will not match.

```
# Correction term in Ito formula
QVV = np.zeros(shape = (n, m))
QVV = 0.5*(DDG(Bt[0:n]) * Q).cumsum(axis = 0)
plt.figure(figsize=(9,6))
plt.plot(t[0:n], QVV[:,0])
```

Figure 3.5: Second Derivative Term in Ito's Formula



3.6 Multiple Ito Processes

Now consider two Ito processes

$$\mathrm{d}X(t) = \mu_x(t)\,\mathrm{d}t + \sigma_x(t)\,\mathrm{d}B_x(t)\ , \tag{3.11}$$

$$\mathrm{d}Y(t) = \mu_y(t)\,\mathrm{d}t + \sigma_y(t)\,\mathrm{d}B_y(t)\ , \tag{3.12}$$

where B_x and B_y can be different Brownian motions. The relation between the two Brownian motions is determined by their covariance or correlation. Given dates $t < u$, we know that both changes $B_x(u) - B_x(t)$ and $B_y(u) - B_y(t)$ are normally distributed with mean 0 and variance equal to $u - t$. There will exist a (possibly random) process ρ such that the covariance of these two normally distributed random variables, given the information at date t, is

$$E_t\left[\int_t^u \rho(s)\,\mathrm{d}s\right]\ .$$

The process ρ is called the correlation coefficient of the two Brownian motions, because when it is constant the correlation of the changes $B_x(u) - B_x(t)$ and $B_y(u) - B_y(t)$ is

$$\frac{\text{covariance}}{\text{product of standard deviations}} = \frac{\int_t^u \rho \, \mathrm{d}s}{\sqrt{u-t}\sqrt{u-t}} = \frac{(u-t)\rho}{u-t} = \rho \, .$$

Moreover, given increasingly fine partitions $0 = t_0 < \cdots < t_N = T$ of an interval $[0, T]$ as before, we will have

$$\sum_{i=1}^{N} \Delta B_x(t_i) \times \Delta B_y(t_i) \to \int_0^T \rho(t) \, \mathrm{d}t$$

as $N \to \infty$, with probability one.

We know that

$$\sum_{i=1}^{N} [\Delta X(t_i)]^2 \to \int_0^T \sigma_x^2(t) \, \mathrm{d}t \quad \text{and} \quad \sum_{i=1}^{N} [\Delta Y(t_i)]^2 \to \int_0^T \sigma_y^2(t) \, \mathrm{d}t \, . \tag{3.13}$$

Furthermore, it can be shown that the sum of products satisfies

$$\sum_{i=1}^{N} \Delta X(t_i) \times \Delta Y(t_i) \to \int_0^T \sigma_x(t)\sigma_y(t)\rho(t) \, \mathrm{d}t \, . \tag{3.14}$$

Tip

By adding the rule

$$(\mathrm{d}B_x)(\mathrm{d}B_y) = \rho \, \mathrm{d}t \tag{3.15}$$

to the rules Equation **??**–Equation **??**, we can compute the limit in Equation **??** as

$$\begin{aligned} \lim_{N\to\infty} \sum_{i=1}^{N} \Delta X(t_i) \times \Delta Y(t_i) &= \int_0^T (\mathrm{d}X)(\mathrm{d}Y) \\ &= \int_0^T (\mu_x \, \mathrm{d}t + \sigma_x \, \mathrm{d}B_x)(\mu_y \, \mathrm{d}t + \sigma_y \, \mathrm{d}B_y) \\ &= \int_0^T \sigma_x(t)\sigma_y(t)\rho(t) \, \mathrm{d}t \, . \end{aligned} \tag{3.16}$$

The most general case of Ito's formula that we will need is for a function $Z(t) = g(t, X(t), Y(t))$ where X and Y are Ito processes as in Equation **??** - Equation **??**. In this case, Ito's formula is[3]

[3] We need to assume $g(t, x, y)$ is continuously differentiable in t and twice continuously differentiable in (x, y) for Equation **??** and Equation **??** to be valid. Note also that we are using a short-hand notation here. The partial derivatives of g will generally depend on t, $X(t)$ and $Y(t)$ just as g does.

$$
\begin{aligned}
Z(T) = Z(0) &+ \int_0^T \frac{\partial g}{\partial t}\,\mathrm{d}t + \int_0^T \frac{\partial g}{\partial x}\,\mathrm{d}X(t) + \int_0^T \frac{\partial g}{\partial y}\,\mathrm{d}Y(t) \\
&+ \frac{1}{2}\int_0^T \frac{\partial^2 g}{\partial x^2}\,(\mathrm{d}X(t))^2 + \frac{1}{2}\int_0^T \frac{\partial^2 g}{\partial y^2}\,(\mathrm{d}Y(t))^2 \\
&+ \int_0^T \frac{\partial^2 g}{\partial x \partial y}\,(\mathrm{d}X(t))(\mathrm{d}Y(t))\;. \qquad (3.17)
\end{aligned}
$$

In this equation, we apply the rules Equation **??**–Equation **??** to compute

$$
\begin{aligned}
(\mathrm{d}X(t))^2 &= \sigma_x^2(t)\,\mathrm{d}t\;, \\
(\mathrm{d}Y(t))^2 &= \sigma_y^2(t)\,\mathrm{d}t\;, \\
(\mathrm{d}X(t))(\mathrm{d}Y(t)) &= \sigma_x(t)\sigma_y(t)\rho(t)\,\mathrm{d}t\;.
\end{aligned}
$$

Ito's Equation **??** appears a bit simpler (and easier to remember) if we write it in differential form." We have:

> Tip
>
> If $Z(t) = g(t, X(t), Y(t))$ where X and Y are Ito processes as in Equation **??** - Equation **??**, then
>
> $$
> \begin{aligned}
> \mathrm{d}Z = \frac{\partial g}{\partial t}\,\mathrm{d}t + \frac{\partial g}{\partial x}\,\mathrm{d}X + \frac{\partial g}{\partial y}\,\mathrm{d}Y + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\,(\mathrm{d}X)^2 + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}\,(\mathrm{d}Y)^2 \\
> + \frac{\partial^2 g}{\partial x \partial y}\,(\mathrm{d}X)(\mathrm{d}Y)\;. \qquad (3.18)
> \end{aligned}
> $$

3.7 Examples of Ito's Formula

The following are the applications of Ito's formula that will be used most frequently in the book. They follow from the boxed formula at the end of the previous section by taking $g(x, y) = xy$ or $g(x, y) = y/x$ or $g(x) = e^x$ or $g(x) = \log x$.

> Tip
>
> **Products.**; If $Z = XY$, then $\mathrm{d}Z = X\,\mathrm{d}Y + Y\,\mathrm{d}X + (\mathrm{d}X)(\mathrm{d}Y)$. We can write this as
>
> $$
> \frac{\mathrm{d}Z}{Z} = \frac{\mathrm{d}X}{X} + \frac{\mathrm{d}Y}{Y} + \left(\frac{\mathrm{d}X}{X}\right)\left(\frac{\mathrm{d}Y}{Y}\right)\;. \qquad (3.19)
> $$

Tip

Ratios.; If $Z = Y/X$, then

$$\frac{\mathrm{d}Z}{Z} = \frac{\mathrm{d}Y}{Y} - \frac{\mathrm{d}X}{X} - \left(\frac{\mathrm{d}Y}{Y}\right)\left(\frac{\mathrm{d}X}{X}\right) + \left(\frac{\mathrm{d}X}{X}\right)^2 . \tag{3.20}$$

Tip

Exponentials.; If $Z = \mathrm{e}^X$, then

$$\frac{\mathrm{d}Z}{Z} = \mathrm{d}X + \frac{(\mathrm{d}X)^2}{2} . \tag{3.21}$$

Tip

Logarithms.; If $Z = \log X$, then

$$\mathrm{d}Z = \frac{\mathrm{d}X}{X} - \frac{1}{2}\left(\frac{\mathrm{d}X}{X}\right)^2 . \tag{3.22}$$

Tip

Compounding/Discounting. Let

$$Y(t) = \exp\left(\int_0^t q(s)\,\mathrm{d}s\right)$$

for some (possibly random) process q and define $Z = XY$ for any Ito process X. The usual calculus gives us $\mathrm{d}Y(t) = q(t)Y(t)\,\mathrm{d}t$, and the product rule above implies

$$\frac{\mathrm{d}Z}{Z} = q\,\mathrm{d}t + \frac{\mathrm{d}X}{X} . \tag{3.23}$$

This is the same as in the usual calculus.

3.8 Reinvesting Dividends

Frequently, we will assume that the asset underlying a derivative security pays a constant dividend yield, which we will denote by q. This means, for an asset with price $S(t)$, that the dividend in an instant $\mathrm{d}t$ is $qS(t)\,\mathrm{d}t$. If the dividends are reinvested in new shares, the number of shares will grow exponentially at rate q. To see this, consider the portfolio starting with a single share of the asset and reinvesting dividends until some date T. Let $X(t)$ denote the number of shares resulting from this strategy at any time $t \leq T$. Then the dividend received at date t is $qS(t)X(t)\,\mathrm{d}t$, which can be used to purchase $qX(t)\,\mathrm{d}t$ new shares. This implies that $\mathrm{d}X(t) = qX(t)\,\mathrm{d}t$, or $\mathrm{d}X(t)/\mathrm{d}t = qX(t)$, and it is easy to check (and very well known) that this equation is solved by $X(t) = \mathrm{e}^{qt}X(0)$. In our case, with $X(0) = 1$, we have $X(t) = \mathrm{e}^{qt}$.

The dollar value of the trading strategy just described will be $X(t)S(t) = \mathrm{e}^{qt}S(t)$. Denote this by $V(t)$. This is the value of a non-dividend-paying portfolio, because all dividends are reinvested. From the Compounding/Discounting example in Section **??**, we know that

$$\frac{\mathrm{d}V}{V} = q\,\mathrm{d}t + \frac{\mathrm{d}S}{S}\ . \tag{3.24}$$

This means that the rate of return on the portfolio is the dividend yield $q\,\mathrm{d}t$ plus the return $\mathrm{d}S/S$ due to capital gains.

3.9 Geometric Brownian Motion

A random variable is lognormally distributed if it can be written as $\tilde{y} = e^{\tilde{x}}$ where $\tilde{x}$ is distributed according to a normal distribution with mean m and standard deviation s. The expected value of $\tilde{y}$ is given by $E[\tilde{y}] = e^{m+\frac{s^2}{2}}$.

Tip

Lognormal Random Variable.; If $\tilde{x}$ is normally distributed with mean m and standard deviation s, then $e^{\tilde{x}}$ is lognormally distributed and

$$E[e^{\tilde{x}}] = e^{m+\frac{1}{2}s^2};. \tag{3.25}$$

An important stochastic process is the Geometric Brownian Motion given by

$$S(t) = S(0)\exp\left(\mu t - \sigma^2 t/2 + \sigma B(t)\right) \tag{3.26}$$

for constants μ and σ, where B is a Brownian motion. Note that for each time t, geometric Brownian motion is a lognormal random variable. Using the product rule and the rule for exponentials, we obtain

$$\frac{\mathrm{d}S}{S} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B\ . \tag{3.27}$$

When we see an equation of the form Equation **??**, we should recognize Equation **??** as the solution.

The process S is called a geometric Brownian motion. In keeping with the discussion of Section **??**, we interpret Equation **??** as stating that $\mu\,\mathrm{d}t$ is the expected rate of change of

S and $\sigma^2\,\mathrm{d}t$ is the variance of the rate of change in an instant $\mathrm{d}t$. We call μ the drift and σ the volatility. The geometric Brownian motion will grow at the average rate of μ, in the sense that $E[S(t)] = \mathrm{e}^{\mu t}S(0)$; one way to verify this uses the formula for the mean of a lognormal random variable.

Taking the natural logarithm of Equation **??** gives an equivalent form of the solution:

$$\log S(t) = \log S(0) + \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma B(t)\ . \tag{3.28}$$

This shows that $\log S(t) - \log S(0)$ is a $(\mu - \sigma^2/2, \sigma)$–Brownian motion. Given information at time t, the logarithm of $S(u)$ for $u > t$ is normally distributed with mean $(u-t)(\mu-\sigma^2/2)$ and variance $(u-t)\sigma^2$. Because S is the exponential of its logarithm, S can never be negative. For this reason, a geometric Brownian motion is a better model for stock prices than is a Brownian motion.

The differential of Equation **??** is

$$\mathrm{d}\log S(t) = \left(\mu - \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B(t)\ . \tag{3.29}$$

We conclude:

> Tip
>
> The equation
>
> $$\frac{\mathrm{d}S}{S} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B$$
>
> is equivalent to the equation
>
> $$\mathrm{d}\log S(t) = \left(\mu - \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B(t)\ .$$
>
> The solution of both equations is Equation **??** or the equivalent Equation **??**.

Over a discrete time interval Δt, Equation **??** implies that the change in the logarithm of S is

$$\Delta\log S = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\,\Delta B\ . \tag{3.30}$$

If S is the price of a non-dividend-paying asset, then over the time period t_{i-1} to t_i, with $t_i - t_{i-1} = \Delta t$, we have

$$\Delta\log S = r_i\,\Delta t\ , \tag{3.31}$$

where r_i is the continuously compounded annualized rate of return during the period Δt. This follows from the definition of the continuously compounded rate of return as the constant rate over the time period Δt that would cause S to grow (or fall) from $S(t_{i-1})$ to $S(t_i)$. To be precise, r_i is defined by

$$\frac{S(t_i)}{S(t_{i-1})} = \mathrm{e}^{r_i \Delta t} \, ,$$

which is equivalent to Equation **??**. Thus, the geometric Brownian motion model Equation **??** implies that the continuously compounded annualized rate of return over a period of length Δt is given by

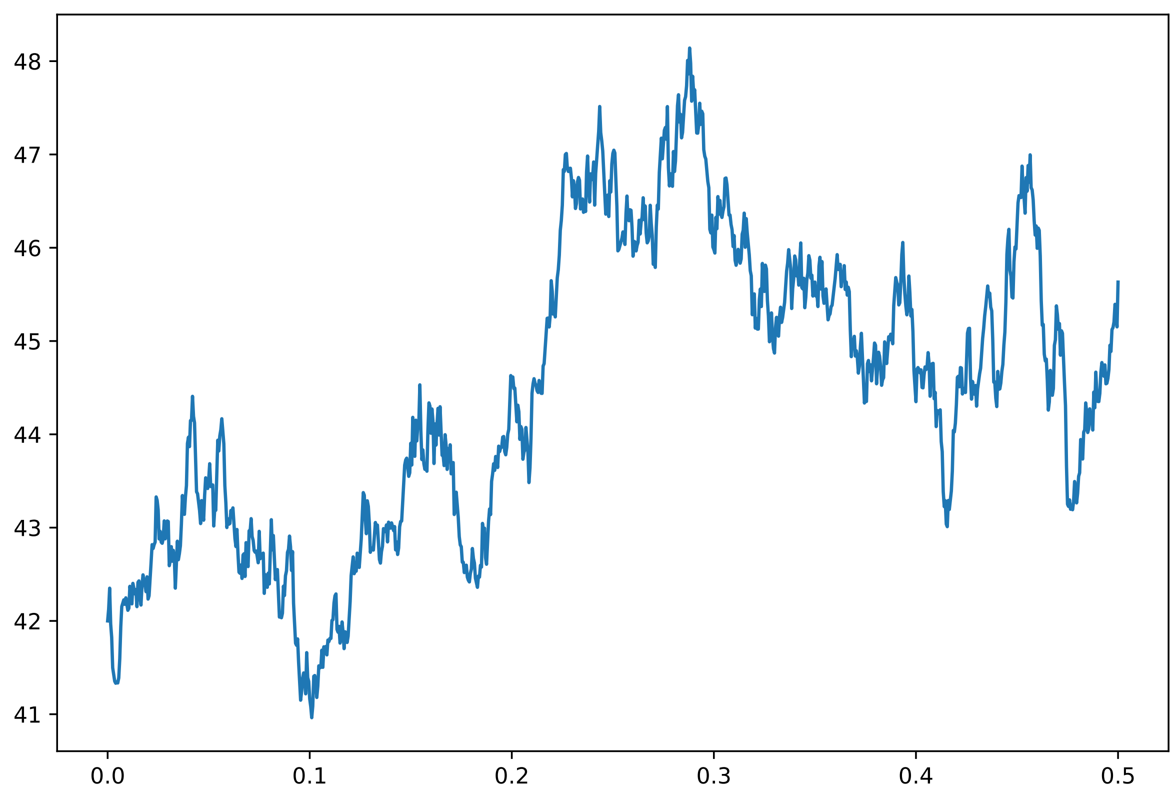
$$r_i = \mu - \frac{1}{2}\sigma^2 + \frac{\sigma \Delta B}{\Delta t} \, .$$

This means that r_i is normally distributed with mean $\mu - \sigma^2/2$ and variance $\sigma^2/\Delta t$. Given historical data on the rates of return, the parameters μ and σ can be estimated by standard methods (see Chapter **??**).

We can simulate a path of S by simulating the changes $\Delta \log S$. The random variable $\sigma\Delta B$ in Equation **??** has a normal distribution with zero mean and variance equal to $\sigma^2\Delta t$. We simulate it as $\sigma\sqrt{\Delta t}$ multiplied by a standard normal. The code below simulates $n = 10000$ paths with $m = 1000$ time steps. There are some features of the simulation which will prove useful later. The drift μ is labelled the interest rate $r = 0.1$. Other parameters are $\sigma = 0.2$, and $T = 0.5$. The drift of the log stock price is labelled $drift = r - \frac{\sigma^2}{2}$. The plot output is one of the simulated sample paths. In practice, if we are only interested in the terminal value of the stock price we would use many fewer subdivisions, $n = 1$. Given a simulated mean zero normal random variable, z, changing the sign to $-z$ is also a simulated normal random variable with zero mean and the same standard deviation. As a result, we have two simulations for the stock price labelled St and $St1$, but we only plot one sample path for St.

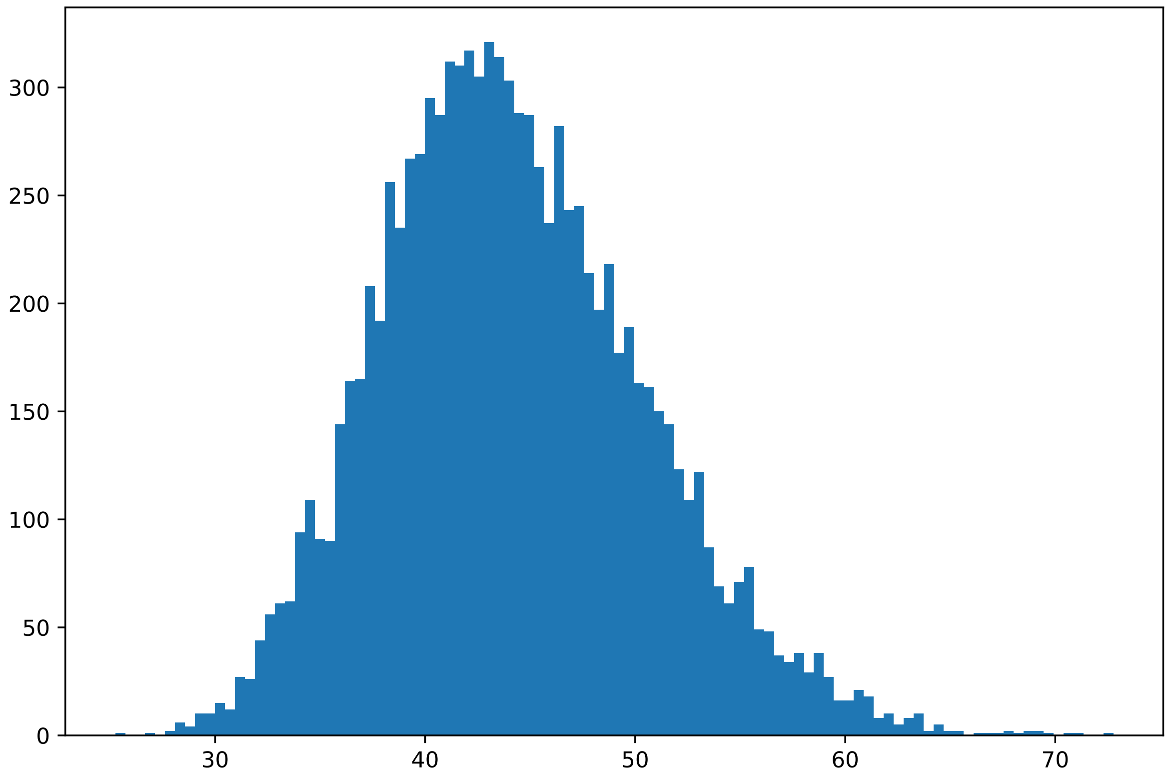
```
# Simulate Geometric Brownian Motion
import numpy as np
import matplotlib.pyplot as plt
# number of paths
n = 10000
#number of divisions
m = 1000
# Interest rate (We set the drift equal to the interest rate)
r = 0.1
# Volatility
sig = 0.2
# Initial Stock Price
S0 = 42
# Maturity
T = 0.5
# Delta t
dt = T/m
# Drift
```

```
drift = (r-0.5*sig**2)
# Volatility
vol = sig * np.sqrt(dt)

t = np.array(range(0,m + 1,1)) * dt

# seed for random generator
seed= 2020
# define a random generator
np.random.seed(seed)
inc = np.zeros(shape = (m + 1, n))
inc[1:] = np.transpose(np.random.normal(loc = 0, scale = vol,size = (n,m)))
St = np.zeros(shape = (m + 1, n))
St = S0 * np.exp(np.cumsum(inc,axis=0) + (drift * t[0:m + 1])[:,None])
St1 = S0 * np.exp(-np.cumsum(inc,axis=0) + (drift * t[0:m + 1])[:,None])
plt.figure(figsize=(9,6))
plt.plot(t,St[:,1])
```



The plot below is the distribution of the simulated stock price at T.

```
plt.figure(figsize=(9,6))
a=plt.hist(St[m,:], bins=100)
```



While Geometric Brownian Motion is an important stochastic process to model stock prices, the process with drift equal to zero given by

$$X(t) = \exp\left(-\frac{\kappa^2}{2} + \kappa B_t\right)$$

satisfies $E[X(t)] = 1$ and $E[X(t)|X(s)] = X(s)$ and is an important example of a strictly positive martingale. Again, these facts can be verified using the formula for the expected value of a lognormal random variable. Notice we can write

$$dX(t) = \kappa X(t) dB_t$$

which agrees with the martingale characterization of $\int_0^t \sigma(X_t, t) dB_t$.

3.10 Numeraires and Probabilities

When we change probability measures, we cannot expect a process B that was a Brownian motion to remain a Brownian motion. The expected change in a Brownian motion must always be zero, but when we change probabilities, the expected change of B is likely to become nonzero. (Likewise, a martingale is unlikely to remain a martingale when we change probabilities.) However, the Brownian motion B will still be an Ito process under the new probability measure. In fact, every Ito process under one probability measure will still be an Ito process under the new probability measure, and the diffusion coefficient of the Ito process will be unaffected by the change in probabilities.[4] Changing probabilities only changes the drift of an Ito process.

In a sense, this should not be surprising. It was noted in Section **??** that a Brownian motion B can be defined as a continuous martingale with paths that jiggle in such a way that the quadratic variation over any interval $[0, T]$ is equal to T. Changing the probabilities will change the probabilities of the various paths (so it may affect the expected change in B) but it will not affect how each path jiggles. So, under the new probability measure, B should still be like a Brownian motion but it may have a nonzero drift. If we consider a general Ito process, the reasoning is the same. The diffusion coefficient σ determines how much each path jiggles, and this is unaffected by changing the probability measure. Furthermore, instantaneous covariances—the $(\mathrm{d}X)(\mathrm{d}Y)$ terms—between Ito processes are unaffected by changing the probability measure. Only the drifts are affected.

As explained in Section **??**, we need to know the distribution of the underlying under probability measures corresponding to different numeraires. Let S be the price of an asset that has a constant dividend yield q, and, as in Section **??**, let $V(t) = \mathrm{e}^{qt}S(t)$. This is the price of the portfolio in which all dividends are reinvested, and we have

$$\frac{\mathrm{d}V}{V} = q\,\mathrm{d}t + \frac{\mathrm{d}S}{S} \ .$$

Let Y be the price of another another asset that does not pay dividends. Let $r(t)$ denote the instantaneous risk-free rate at date t and let $R(t) = \exp\left(\int_0^t r(s)\,\mathrm{d}s\right)$. Assume

$$\frac{\mathrm{d}S}{S} = \mu_s\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s \ ,$$
$$\frac{\mathrm{d}Y}{Y} = \mu_y\,\mathrm{d}t + \sigma_y\,\mathrm{d}B_y \ ,$$

where B_s and B_y are Brownian motions under the actual probability measure with correlation ρ, and where $\mu_s, \mu_y, \sigma_s, \sigma_y$ and ρ can be quite general random pro-cesses.
We consider the dynamics of the asset price S under three different probability measures. In each case, we follow the same steps: (i) we note that the ratio of an asset price to the numeraire

[4] To be a little more precise, this is true provided sets of states of the world having zero probability continue to have zero probability when the probabilities are changed. Because of the way we change probability measures when we change numeraires (cf. Equation **??**) this will always be true for us.

asset price must be a martingale, (ii) we use Ito's formula to calculate the drift of this ratio, and (iii) we use the fact that the drift of a martingale must be zero to compute the drift of $\mathrm{d}S/S$.

3.10.1 Risk-Neutral Probabilities

Under the risk-neutral measure, $Z(t)$ defined as

$$Z(t) = \frac{V(t)}{R(t)} = \exp\left(-\int_0^t r(s)\,\mathrm{d}s\right) V(t)$$

is a martingale. Using the compounding/discounting rule, we have

$$\frac{\mathrm{d}Z}{Z} = -r\,\mathrm{d}t + \frac{\mathrm{d}V}{V} = (q-r)\,\mathrm{d}t + \frac{\mathrm{d}S}{S}\ .$$

For Z to be a martingale, the drift ($\mathrm{d}t$ part) of $\mathrm{d}Z/Z$ must be zero. Therefore, the drift of $\mathrm{d}S/S$ must be $(r-q)\,\mathrm{d}t$ under the risk-neutral measure. Because the change of measure does not affect the volatility, this implies:

Tip

$$\frac{\mathrm{d}S}{S} = (r-q)\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s^*\ , \tag{3.32}$$

where B_s^* is a Brownian motion under the risk-neutral measure.

3.10.2 Underlying as the Numeraire

When V is the numeraire, the process $Z(t)$ defined as

$$Z(t) = \frac{R(t)}{V(t)} = \frac{\exp\left(\int_0^t r(s)\,\mathrm{d}s\right)}{V(t)}$$

is a martingale. Using the rule for ratios, we have

$$\frac{\mathrm{d}Z}{Z} = r\,\mathrm{d}t - \frac{\mathrm{d}V}{V} + \left(\frac{\mathrm{d}V}{V}\right)^2 = (r-q+\sigma_s^2)\,\mathrm{d}t - \frac{\mathrm{d}S}{S}\ .$$

Because the drift of $\mathrm{d}Z/Z$ must be zero, this implies that the drift of $\mathrm{d}S/S$ is $(r-q+\sigma_s^2)\,\mathrm{d}t$. We conclude that:

Tip

$$\frac{\mathrm{d}S}{S} = (r - q + \sigma_s^2)\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s^*\,, \tag{3.33}$$

where now B_s^* denotes a Brownian motion when $V(t) = \mathrm{e}^{qt}S(t)$ is the numeraire.

3.10.3 Another Risky Asset as the Numeraire

When Y is the numeraire, $Z(t)$ defined as

$$Z(t) = \frac{V(t)}{Y(t)}$$

must be a martingale. Using again the rule for ratios, we have

$$\begin{aligned}
\frac{\mathrm{d}Z}{Z} &= \frac{\mathrm{d}V}{V} - \frac{\mathrm{d}Y}{Y} - \left(\frac{\mathrm{d}V}{V}\right)\left(\frac{\mathrm{d}Y}{Y}\right) + \left(\frac{\mathrm{d}Y}{Y}\right)^2 \\
&= \frac{\mathrm{d}V}{V} - \frac{\mathrm{d}Y}{Y} - \rho\sigma_s\sigma_y\,\mathrm{d}t + \sigma_y^2\,\mathrm{d}t \\
&= \frac{\mathrm{d}S}{S} - \frac{\mathrm{d}Y}{Y} + (q - \rho\sigma_s\sigma_y\,\mathrm{d}t + \sigma_y^2)\,\mathrm{d}t\,.
\end{aligned}$$

We can apply our previous example to compute the dynamics of Y when Y is the numeraire. This shows that the drift of $\mathrm{d}Y/Y$ is $(r + \sigma_y^2)\,\mathrm{d}t$. Because the drift of $\mathrm{d}Z/Z$ must be zero, it follows that the drift of $\mathrm{d}S/S$ is $(r - q + \rho\sigma_s\sigma_y)\,\mathrm{d}t$. We conclude that:

Tip

$$\frac{\mathrm{d}S}{S} = (r - q + \rho\sigma_s\sigma_y)\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s^*\,, \tag{3.34}$$

where B_s^* denotes a Brownian motion under the probability measure corresponding to the non-dividend-paying risky asset Y being the numeraire, and where ρ is the correlation of S and Y.

Notice that Equation **??**, while more complicated, is also more general than the others. In fact, it includes the formulas Equation **??** and Equation **??** as special cases: (i) if Y is the price of the instantaneously risk-free asset, then $\sigma_y = 0$ and Equation **??** simplifies to Equation **??**, and (ii) if $Y = V$, then $\sigma_y = \sigma_s$ and $\rho = 1$, so Equation **??** simplifies to Equation **??**.

3.10.4 Further Discussion

It would be natural for one to ask at this point: what is the Brownian motion B_s^* and where did it come from? We have argued that once we know the drift, and the fact that the volatility does not change, we can immediately write down, for example,

$$\frac{\mathrm{d}S}{S} = (r - q)\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s^*$$

for a Brownian motion B_s^* under the risk-neutral measure. To answer this question, we will give here the definition of B_s^* under the risk-neutral measure. The definition shows that we are justified in writing down Equation **??**–Equation **??**, but we will not repeat the definition each time we make a statement of this sort.

We showed that Z is a martingale under the risk-neutral measure, where Z satisfies

$$\frac{\mathrm{d}Z}{Z} = (q - r)\,\mathrm{d}t + \frac{\mathrm{d}S}{S} = (q - r + \mu_s)\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s\,. \tag{3.35}$$

Define $B_s^*(0) = 0$ and

$$\mathrm{d}B_s^* = \left(\frac{q - r + \mu_s}{\sigma_s}\right)\,\mathrm{d}t + \mathrm{d}B_s\,. \tag{3.36}$$

Then

$$\mathrm{d}B_s^* = \frac{1}{\sigma_s}\left(\frac{\mathrm{d}Z}{Z}\right)$$

and hence is a continuous martingale under the risk-neutral measure. We can compute its quadratic variation as

$$(\mathrm{d}B_s^*)^2 = \left(\frac{q - r + \mu_s}{\sigma_s}\right)^2(\mathrm{d}t)^2 + 2\left(\frac{q - r + \mu_s}{\sigma_s}\right)(\mathrm{d}t)(\mathrm{d}B_s) + (\mathrm{d}B_s)^2 = \mathrm{d}t\,.$$

Therefore, by Levy's theorem (Section **??**), B_s^* is a Brownian motion under the risk-neutral measure. From Equation **??** and Equation **??** we have

$$(q - r)\,\mathrm{d}t + \frac{\mathrm{d}S}{S} = \sigma_s\,\mathrm{d}B_s^* \quad\Longleftrightarrow\quad \frac{\mathrm{d}S}{S} = (r - q)\,\mathrm{d}t + \sigma_s\,\mathrm{d}B_s^*\,,$$

as in Equation **??**.

3.11 Tail Probabilities of Geometric Brownian Motions

For each of the numeraires discussed in the previous section, we have

$$\mathrm{d}\log S = \alpha\,\mathrm{d}t + \sigma\,\mathrm{d}B\,, \tag{3.37}$$

for some α and σ, where B is a Brownian motion under the probability measure associated with the numeraire. Specifically, $\sigma = \sigma_s$, $B = B_s^*$, and

1. for the risk-neutral measure, $\alpha = r - q - \sigma_s^2/2$,
2. when $e^{qt}S(t)$ is the numeraire, $\alpha = r - q + \sigma_s^2/2$,
3. when another risky asset price Y is the numeraire, $\alpha = r - q + \rho\sigma_s\sigma_y - \sigma_s^2/2$.

We will assume in this section that α and σ are constants. The essential calculation in pricing options, as we will see in the next chapter and in Chapter **??**, is to compute $\text{prob}(S(T) > K)$ and $\text{prob}(S(T) < K)$ for a constant K (the strike price of an option), where prob denotes the probabilities at date 0 (the date we are pricing an option) associated with a particular numeraire.

Equation Equation **??** gives us

$$\log S(T) = \log S(0) + \alpha T + \sigma B(T)\ .$$

Given this, we deduce

$$\begin{aligned}
S(T) > K \quad &\iff \quad \log S(T) > \log K \\
&\iff \quad \sigma B(T) > \log K - \log S(0) - \alpha T \\
&\iff \quad \frac{B(T)}{\sqrt{T}} > \frac{\log K - \log S(0) - \alpha T}{\sigma\sqrt{T}} \\
&\iff \quad -\frac{B(T)}{\sqrt{T}} < \frac{\log S(0) - \log K + \alpha T}{\sigma\sqrt{T}} \\
&\iff \quad -\frac{B(T)}{\sqrt{T}} < \frac{\log\left(\frac{S(0)}{K}\right) + \alpha T}{\sigma\sqrt{T}}\ .
\end{aligned} \tag{3.38}$$

The random variable on the left-hand side of Equation **??** has the standard normal distribution—it is normally distributed with mean equal to zero and variance equal to one. As is customary, we will denote the probability that a standard normal is less than some number d as $\mathrm{N}(d)$. We conclude:

Tip

Assume $\mathrm{d}\log S = \alpha\,\mathrm{d}t + \sigma\,\mathrm{d}B$, where B is a Brownian motion. Then, for any number K,

$$\text{prob}(S(T) > K) = \mathrm{N}(d)\ , \tag{3.39}$$

where

$$d = \frac{\log\left(\frac{S(0)}{K}\right) + \alpha T}{\sigma\sqrt{T}}\ . \tag{3.40}$$

The probability $\text{prob}(S(T) < K)$ can be calculated similarly, but the simplest way to derive it is to note that the events $S(T) > K$ and $S(T) < K$ are complementary—their probabilities sum to one (the event $S(T) = K$ having zero probability). Therefore $\text{prob}(S(T) < K) = 1 - \text{N}(d)$. This is the probability that a standard normal is greater than d, and by virtue of the symmetry of the standard normal distribution, it equals the probability that a standard normal is less than $-d$. Therefore, we have:

Tip

Assume $\mathrm{d}\log S = \alpha\,\mathrm{d}t + \sigma\,\mathrm{d}B$, where B is a Brownian motion. Then, for any number K,

$$\text{prob}(S(T) < K) = \text{N}(-d)\ , \tag{3.41}$$

where d is defined in Equation **??**.

3.12 Volatilities

As mentioned in Section **??**, when we encounter an equation of the form

$$\frac{\mathrm{d}S}{S} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B$$

where B is a Brownian motion, we will say "σ is the volatility of S. For example, in the Black-Scholes model, the most important assumption is that the volatility of the underlying asset price is constant. We will occasionally need to compute the volatilities of products or ratios of random processes. These computations follow directly from Ito's formula.

Suppose

$$\frac{\mathrm{d}X}{X} = \mu_x\,\mathrm{d}t + \sigma_x\,\mathrm{d}B_x \qquad \text{and} \qquad \frac{\mathrm{d}Y}{Y} = \mu_y\,\mathrm{d}t + \sigma_y\,\mathrm{d}B_y\ ,$$

where B_x and B_y are Brownian motions with correlation ρ, and μ_x, μ_y, σ_x, σ_y, and ρ may be quite general random processes.

3.12.1 Products

If $Z = XY$, then Equation **??** gives us

$$\frac{\mathrm{d}Z}{Z} = (\mu_x + \mu_y + \rho\sigma_x\sigma_y)\,\mathrm{d}t + \sigma_x\,\mathrm{d}B_x + \sigma_y\,\mathrm{d}B_y\ . \tag{3.42}$$

The instantaneous variance of $\mathrm{d}Z/Z$ is calculated, using the rules for products of differentials, as

$$\begin{aligned}\left(\frac{\mathrm{d}Z}{Z}\right)^2 &= (\sigma_x\,\mathrm{d}B_x + \sigma_y\,\mathrm{d}B_y)^2 \\ &= (\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y)\,\mathrm{d}t\ .\end{aligned}$$

As will be explained below, the volatility is the square root of the instantaneous variance (dropping the $\mathrm{d}t$). This implies:

Tip

The volatility of XY is

$$\sqrt{\sigma_x^2 + \sigma_y^2 + 2\rho\sigma_x\sigma_y}\,. \tag{3.43}$$

3.12.2 Ratios

If $Z = Y/X$, then Equation **??** gives us

$$\frac{\mathrm{d}Z}{Z} = (\mu_y - \mu_x - \rho\sigma_x\sigma_y + \sigma_x^2)\,\mathrm{d}t + \sigma_y\,\mathrm{d}B_y - \sigma_x\,\mathrm{d}B_x\,. \tag{3.44}$$

The instantaneous variance of $\mathrm{d}Z/Z$ is therefore

$$\begin{aligned}\left(\frac{\mathrm{d}Z}{Z}\right)^2 &= (\sigma_y\,\mathrm{d}B_y - \sigma_x\,\mathrm{d}B_x)^2 \\ &= (\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y)\,\mathrm{d}t\,.\end{aligned}$$

This implies:

Tip

The volatility of Y/X is

$$\sqrt{\sigma_x^2 + \sigma_y^2 - 2\rho\sigma_x\sigma_y}\,. \tag{3.45}$$

3.12.3 Further Discussion

To understand why taking the square root of $(\mathrm{d}Z/Z)^2$ (dropping the $\mathrm{d}t$) gives the volatility, consider for example the product case $Z = XY$. Define a random process B by $B(0) = 0$ and

$$\mathrm{d}B = \frac{\sigma_x}{\sigma}\,\mathrm{d}B_x + \frac{\sigma_y}{\sigma}\,\mathrm{d}B_y\,, \tag{3.46}$$

where σ is the volatility defined in Equation **??**. Then we can write Equation **??** as

$$\frac{\mathrm{d}Z}{Z} = \left(\mu_x + \mu_y + \rho\sigma_x\sigma_y\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B\,. \tag{3.47}$$

From the discussion in Section **??**, we know that B is a continuous martingale. We can compute its quadratic variation from

$$\begin{aligned}(\mathrm{d}B)^2 &= \left(\frac{\sigma_x \,\mathrm{d}B_x + \sigma_s \,\mathrm{d}B_s}{\sigma}\right)^2 \\ &= \frac{(\sigma_x^2 + \sigma_s^2 + 2\rho\sigma_x\sigma_s)\,\mathrm{d}t}{\sigma^2}\ , \\ &= \mathrm{d}t\ .\end{aligned}$$

By Levy's theorem (see Section **??**), any continuous martingale with this quadratic variation is necessarily a Brownian motion. Therefore, Equation **??** shows that σ is the volatility of Z as defined at the beginning of the section.

3.13 Exercises

Exercise 3.1. Consider a discrete partition $0 = t_0 < t_1 < \cdots t_N = T$ of the time interval $[0, T]$ with $t_i - t_{i-1} = \Delta t = T/N$ for each i. Consider the function

$$X(t) = \mathrm{e}^t\ .$$

Write a code, which computes and plots $\sum_{i=1}^N [\Delta X(t_i)]^2$, where

$$\Delta X(t_i) = X(t_i) - X(t_{i-1}) = \mathrm{e}^{t_i} - \mathrm{e}^{t_{i-1}}\ .$$

Exercise 3.2. Repeat the previous problem for the function $X(t) = t^3$. In both this and the previous problem, what happens to $\sum_{i=1}^N [\Delta X(t_i)]^2$ as $N \to \infty$?

Exercise 3.3. Either use the code provided or write a code to compute $\sum_{i=1}^N [\Delta B(t_i)]^2$, where B is a simulated Brownian motion. For a given T, what happens to the sum as $N \to \infty$?

Exercise 3.4. Repeat the previous problem to compute $\sum_{i=1}^N [\Delta B(t_i)]^3$, where B is a simulated Brownian motion. For a given T, what happens to the sum as $N \to \infty$?

Exercise 3.5. Repeat the previous problem, computing instead $\sum_{i=1}^N |\Delta B(t_i)|$ where $|\cdot|$ denotes the absolute value. What happens to this sum as $N \to \infty$?

Exercise 3.6. Use Ito's Lemma to derive the stochastic differential equation for $S(t)^2$. Argue that $S(t)^2$ is Geometric Brownian Motion and find $E[S(t)^2]$.

Exercise 3.7. Ito's Lemma can be used in different ways to get the same answer. For example, let $X(t) = at + bB_t$ and use Ito's lemma on the function $e^{X(t)}$. Alternatively, let $f(t, B_t) = e^{at+bB_t}$. Use Ito's lemma on $f(,)$.

Exercise 3.8. Use the facts $e^{x+y} = e^x \times e^y$ and $\frac{e^x}{e^y} = e^{x-y}$ to deduce the drift and volatility of the product and ratio of two Geometric Brownian motions.

Exercise 3.9. Consider a discrete partition $0 = t_0 < t_1 < \cdots t_N = T$ of the time interval $[0, T]$ with $t_i - t_{i-1} = \Delta t = T/N$ for each i. Consider a geometric Brownian motion

$$\frac{\mathrm{d}Z}{Z} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B\,.$$

An approximate path $\tilde{Z}(t)$ of the geometric Brownian motion can be simulated as

$$\Delta\tilde{Z}(t_i) = \tilde{Z}(t_{i-1})\left[\mu\,\Delta t + \sigma\,\Delta B\right]\,. \tag{3.48}$$

Modify the code to generate both a path $Z(t)$ and an approximate path $\tilde{Z}(t)$ according to Equation **??**, using the same ΔB for both paths and taking $\tilde{Z}(0) = Z(0)$. Plot both paths in the same figure. How well does the approximation work for large N? Warning:
For N larger than about $100T$, the approximation will look perfect—you won't be able to tell that there are two plots in the figure. One reason this is true is an exact formula is

$$Z(t_i) = Z(t_{i-1})\exp\left[\left(\mu - \frac{\sigma^2}{2}\right)\,\Delta t + \sigma\,\Delta B\right]\,. \tag{3.49}$$

and using Taylor's Theorem for small Δt, $e^{\left(\mu-\frac{\sigma^2}{2}\right)\Delta t} \approx 1 + \left(\mu - \frac{\sigma^2}{2}\right)\Delta t$ and $e^{\sigma\Delta B_t} \approx 1 + \sigma\Delta B_t + \frac{1}{2}\sigma^2(\Delta B_t)^2$ and $(dB_t)^2 = \Delta t$.

Exercise 3.10. Use simulation to find $E^*[e^{-rT}\mathbf{1}_{\{S(T)\geq K\}}]$ in the risk neutral measure where

$$dS(t) = rS(t)dt + \sigma S(t)dB^*_t$$

Verify that $S(t)/e^{rt}$ is a martingale in the $*$ measure where B^* is a Brownian motion. Then use simulation to find $S(0)E^S[\frac{1}{S(T)}\mathbf{1}_{\{S(T)\geq K\}}]$ in the pricing measure which uses the share as numeraire where

$$dS(t) = (r + \sigma^2)S(t)dt + \sigma S(t)dB^S_t$$

so the log satisfies

$$\log(S(t)) = \log(S(0)) + (r + \frac{\sigma^2}{2})t + \sigma B^S_t$$

You should verify $e^{rt}/S(t)$ is a martingale in the S measure where B^S is a Brownian motion. Both estimates should be the same up to simulation error and give the time zero value of the random payoff $\mathbf{1}_{\{S(T)\geq K\}}$ which is a random variable equal to 1 if $S(T) \geq K$ and 0 otherwise. You should choose the values for r, σ, T, and K.

4 Black-Scholes

In this chapter, we will study the value of European digital and share digital options and standard European puts and calls under the Black-Scholes assumptions. We will also explain how to calculate implied volatilities and the option Greeks. The Black-Scholes assumptions are that the underlying asset pays a constant dividend yield q and has price S satisfying

$$\frac{\mathrm{d}S}{S} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B \tag{4.1}$$

for a Brownian motion B. Here σ is assumed to be constant (though we will allow it to vary in a non-random way at the end of the chapter) and μ can be a quite general random process. It is also assumed that there is a constant continuously-compounded risk-free rate r.

Under these assumptions, we will complete the discussion of Section **??** to derive option pricing formulas. Recall that, to price a European call option, all that remains to be done is to calculate the probabilities of the option finishing in the money when we use the risk-free asset and the underlying asset as numeraires. We will do this using the results of Section **??**. As in Section **??**, we will approach the pricing of call and put options by first considering their basic building blocks: digitals and share digitals.

4.1 Digital Options

A digital (or binary) option pays a fixed amount in a certain event and zero otherwise. Consider a digital that pays \$1 at date T if $S(T) > K$, where K is a number that is fixed by the contract. This means that the digital pays x dollars at date T where x is defined as

$$x = \begin{cases} 1 & \text{if } S(T) > K\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

Using the risk-neutral pricing Equation **??**, the value of the digital at date~0 is $\mathrm{e}^{-rT}E^R[x]$. Note that

$$\begin{aligned} E^R[x] &= 1 \times \mathrm{prob}^R(x=1) \;+\; 0 \times \mathrm{prob}^R(x{=}0) \\ &= \mathrm{prob}^R(x=1) \\ &= \mathrm{prob}^R\big(S(T) > K\big)\ . \end{aligned}$$

So we need to calculate this probability of the digital finishing in the money.

In Section **??**—see Equation **??**—we learned that under the Black-Scholes assumption Equation **??** we have

$$\frac{\mathrm{d}S}{S} = (r-q)\,\mathrm{d}t + \sigma\,\mathrm{d}B^* \,,$$

where B^* is a Brownian motion under the risk-neutral measure.[1] on the volatility coefficients and on B and B^* to distinguish the Brownian motion driving S from the Brownian motion driving Y and to distinguish their volatilities are not needed here. In Section **??**, we observed that this is equivalent to

$$\mathrm{d}\log S = \left(r - q - \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B^* \,.$$

Now using the formulas Equation **??**–Equation **??**, with $\alpha = r - q - \sigma^2/2$, we have $\mathrm{prob}^R(S(T) > K) = \mathrm{N}(d_2)$ where

$$d_2 = \frac{\log\left(\frac{S(0)}{K}\right) + \left(r - q - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} \,. \tag{4.2}$$

The notation d_2 is standard notation from the Black-Scholes formula, and we use it—rather than a simple d—to distinguish the number Equation **??** from a similar number—to be called d_1 of course—that we will see in the next section.
We conclude:

> Tip
>
> The value of a digital option that pays \$1 when $S(T) > K$ is $\mathrm{e}^{-rT}\mathrm{N}(d_2)$, where d_2 is defined in Equation **??**.

Consider now a digital that pays when the underlying asset price is low; i.e., consider a security that pays y dollars at date T where

$$y = \begin{cases} 1 & \text{if } S(T) < K \,, \\ 0 & \text{otherwise} \,. \end{cases}$$

Using risk-neutral pricing again, the value of this digital at date 0 is

$$\mathrm{e}^{-rT}E^R[y] = \mathrm{e}^{-rT}\mathrm{prob}^R(y=1) = \mathrm{e}^{-rT}\mathrm{prob}^R(S(T) < K) \,.$$

From this fact and Equation **??**, we conclude:

[1]There is no other risky asset price Y in this model, so the subscripts we used in Section **??**

Tip

The value of a digital option that pays \$1 when $S(T) < K$ is $\mathrm{e}^{-rT}\mathrm{N}(-d_2)$, where d_2 is defined in Equation **??**.

4.2 Share Digitals

Consider a derivative security that pays one share of the underlying asset at date T if $S(T) > K$ and pays zero otherwise. This is called a share digital. As before, let

$$x = \begin{cases} 1 & \text{if } S(T) > K \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

Then the payoff of the share digital at date T is $xS(T)$. Let $Y(t)$ denote the value of this claim for $0 \le t \le T$. We have $Y(T) = xS(T)$ and we want to find $Y(0)$.

From Section **??**, we know that $V(t) = \mathrm{e}^{qt}S(t)$ is the price of a non-dividend-paying portfolio. From our fundamental pricing Equation **??**, using V as the numeraire, we have

$$\begin{aligned} Y(0) &= S(0)E^V\left[\frac{Y(T)}{\mathrm{e}^{qT}S(T)}\right] \\ &= \mathrm{e}^{-qT}S(0)E^V[x] \ . \end{aligned}$$

As in the previous section, $E^V[x] = \mathrm{prob}^V(x = 1)$, so we need to compute this probability of the option finishing in the money.

We follow the same steps as in the previous section. From Equation **??** we have

$$\frac{\mathrm{d}S}{S} = (r - q + \sigma^2)\,\mathrm{d}t + \sigma\,\mathrm{d}B^*,$$

where now B^* denotes a Brownian motion when V is the numeraire. This is equivalent to

$$\mathrm{d}\log S = \left(r - q + \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B^* \ . \tag{4.3}$$

Thus, from the formulas Equation **??**–Equation **??**, with $\alpha = r - q + \sigma^2/2$, we have

$$\mathrm{prob}^V(S(T) > K) = \mathrm{N}(d_1) \ ,$$

where

$$d_1 = \frac{\log\left(\frac{S(0)}{K}\right) + \left(r - q + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}} \ . \tag{4.4}$$

This implies:

Tip

The value of a share digital that pays one share when $S(T) > K$ is $\mathrm{e}^{-qT}S(0)\mathrm{N}(d_1)$, where d_1 is defined in Equation **??**.

Consider now a share digital that pays one share of the stock at date T if $S(T) < K$. Letting

$$y = \begin{cases} 1 & \text{if } S(T) < K \;, \\ 0 & \text{otherwise} \;, \end{cases}$$

the payoff of this option is $yS(T)$. Its value at date 0 is

$$\begin{aligned} \mathrm{e}^{-qT}S(0)E^V[y] &= \mathrm{e}^{-qT}S(0) \times \mathrm{prob}^V(y=1) \\ &= \mathrm{e}^{-qT}S(0) \times \mathrm{prob}^V\big(S(T) < K\big) \;, \end{aligned}$$

and from Equation **??** we have

$$\mathrm{prob}^V\big(S(T) < K\big) = \mathrm{N}(-d_1) \;.$$

We conclude:

Tip

The value of a share digital that pays one share when $S(T) < K$ is $\mathrm{e}^{-qT}S(0)\mathrm{N}(-d_1)$, where d_1 is defined in Equation **??**.

4.3 Puts and Calls

A European call option pays $S(T)-K$ at date T if $S(T) > K$ and 0 otherwise. Again letting

$$x = \begin{cases} 1 & \text{if } S(T) > K \;, \\ 0 & \text{otherwise} \;, \end{cases}$$

the payoff of the call can be written as $xS(T) - xK$. This is equivalent to one share digital minus K digitals, with the digitals paying in the event that $S(T) > K$. The share digital is worth $\mathrm{e}^{-qT}S(0)\mathrm{N}(d_1)$ at date 0 and each digital is worth $\mathrm{e}^{-rT}\mathrm{N}(d_2)$. Note that equations Equation **??** and Equation **??** for d_1 and d_2 imply $d_2 = d_1 - \sigma\sqrt{T}$. Therefore, combining the results of the previous two sections yields the Black-Scholes formula:

Tip

The value of a European call option at date 0 is

$$\mathrm{e}^{-qT}S(0)\mathrm{N}(d_1) - \mathrm{e}^{-rT}K\mathrm{N}(d_2)\ , \tag{4.5}$$

where d_1 is defined in Equation **??** and $d_2 = d_1 - \sigma\sqrt{T}$.

A European put option pays $K - S(T)$ at date T if $S(T) < K$ and 0 otherwise. As before, let

$$y = \begin{cases} 1 & \text{if } S(T) < K\ , \\ 0 & \text{otherwise}\ . \end{cases}$$

The payoff of the put option is $yK - yS(T)$. This is equivalent to K digitals minus one share digital, all of the digitals paying when $S(T) < K$. Thus, we have:

Tip

The value of a European put option at date 0 is

$$\mathrm{e}^{-rT}K\mathrm{N}(-d_2) - \mathrm{e}^{-qT}S(0)\mathrm{N}(-d_1)\ , \tag{4.6}$$

where d_1 is defined in Equation **??** and $d_2 = d_1 - \sigma\sqrt{T}$.

Again, this is the Black-Scholes formula.

The values of the European put and call satisfy put-call parity, and we can also find one from the other by^[The put-call parity relation follows from the fact that both the left and the right-hand sides are the prices of portfolios that have value $\max(S(T), K)$ at the maturity of the option. To see this for the left-hand side, note that $\mathrm{e}^{-rT}K$ is sufficient cash to accumulate to K at date T, allowing exercise of the call when it is in the money and retention of the cash K otherwise. For the right-hand side, note that $\mathrm{e}^{-qT}S(0)$ is enough cash to buy e^{-qT} shares of the stock at date 0 which, with reinvestment of dividends, will accumulate to one share at date T, enabling exercise of the put if it is in the money or retention of the share otherwise.}

$$\mathrm{e}^{-rT}K + \text{Call Price} = \mathrm{e}^{-qT}S(0) + \text{Put Price}\ . \tag{4.7}$$

4.4 Greeks

The derivatives (calculus derivatives, not financial derivatives!) of an option pricing formula with respect to the inputs are commonly called Greeks. The most important Greek is the option delta. This measures the sensitivity of the option value to changes in the value of the underlying asset. The following table shows the standard Greeks, with reference to the Black-Scholes pricing formula.

Table 4.1: Black-Scholes Greeks

Input	**Input Symbol**	**Greek**	**Greek Symbol**
Stock price	S	delta	δ
delta	δ	gamma	Γ
- Time to maturity	$-T$	theta	Θ
Volatility	σ	vega	V
Interest rate	r	rho	ρ

Input	**Input Symbol**	**Greek**	**Greek Symbol**
Stock price	S	delta	δ
delta	δ	gamma	Γ
- Time to maturity	$-T$	theta	Θ
Volatility	σ	vega	V
Interest rate	r	rho	ρ

The second line of the above shows δ as an input.[2] Of course, it is not an input but instead is calculated. Gamma, the derivative of δ, is the second derivative of the option price with respect to the underlying asset price. The reason for calculating Θ as the derivative with respect to $-T$ instead of T is that the time-to-maturity T decreasing ($-T$ increasing) is equivalent to time passing, so Θ measures the change in the option value when time passes.

We can calculate these from the Black-Scholes formula using the chain rule from differential calculus. The derivative of the normal distribution function N is the normal density function nd defined as

$$\mathrm{n}d(d) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-d^2/2} \; .$$

One can easily verify directly that

$$\mathrm{e}^{-qT} S \mathrm{n}d(d_1) = \mathrm{e}^{-rT} K \mathrm{n}d(d_2) \; , \tag{4.8}$$

[2] The delta is frequently denoted by the upper case Δ, but we will use the lower case, reserving the upper case for discrete changes, e.g., Δt. One may have noticed also that the symbol for vega is a little different from the others; this reflects the fact that vega is not actually a Greek letter.

which simplifies the calculations for the Black-Scholes call option pricing formula. For this formula, the Greeks are as follows:

$$\begin{aligned}
\delta &= \mathrm{e}^{-qT}\mathrm{N}(d_1) + \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\frac{\partial d_1}{\partial S} - \mathrm{e}^{-rT} K\mathrm{nd}(d_2)\frac{\partial d_2}{\partial S} \\
&= \mathrm{e}^{-qT}\mathrm{N}(d_1) + \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\left(\frac{\partial d_1}{\partial S} - \frac{\partial d_2}{\partial S}\right) \\
&= \mathrm{e}^{-qT}\mathrm{N}(d_1)\ , \\
\Gamma &= \mathrm{e}^{-qT}\mathrm{nd}(d_1)\frac{\partial d_1}{\partial S} = \mathrm{e}^{-qT}\mathrm{nd}(d_1)\frac{1}{S\sigma\sqrt{T}}\ ,
\end{aligned}$$

$$\begin{aligned}
\Theta &= -\mathrm{e}^{-qT} S\mathrm{nd}(d_1)\frac{\partial d_1}{\partial T} + q\mathrm{e}^{-qT} S\mathrm{N}(d_1) \\
&\quad + \mathrm{e}^{-rT} K\mathrm{nd}(d_2)\frac{\partial d_2}{\partial T} - r\mathrm{e}^{-rT} K\mathrm{N}(d_2) \\
&= \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\left(\frac{\partial d_2}{\partial T} - \frac{\partial d_1}{\partial T}\right) \\
&\quad + q\mathrm{e}^{-qT} S\mathrm{N}(d_1) - r\mathrm{e}^{-rT} K\mathrm{N}(d_2) \\
&= -\mathrm{e}^{-qT} S\mathrm{nd}(d_1)\frac{\sigma}{2\sqrt{T}} + q\mathrm{e}^{-qT} S\mathrm{N}(d_1) - r\mathrm{e}^{-rT} K\mathrm{N}(d_2)\ , \\
V &= \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\frac{\partial d_1}{\partial \sigma} - \mathrm{e}^{-rT} K\mathrm{nd}(d_2)\frac{\partial d_2}{\partial \sigma} \\
&= \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\left(\frac{\partial d_1}{\partial \sigma} - \frac{\partial d_2}{\partial \sigma}\right) \\
&= \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\sqrt{T}\ , \\
\rho &= \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\frac{\partial d_1}{\partial r} - \mathrm{e}^{-rT} K\mathrm{nd}(d_2)\frac{\partial d_2}{\partial r} + T\mathrm{e}^{-rT} K\mathrm{N}(d_2) \\
&= \mathrm{e}^{-qT} S\mathrm{nd}(d_1)\left(\frac{\partial d_1}{\partial r} - \frac{\partial d_2}{\partial r}\right) + T\mathrm{e}^{-rT} K\mathrm{N}(d_2) \\
&= T\mathrm{e}^{-rT} K\mathrm{N}(d_2)\ .
\end{aligned}$$

We can calculate the Greeks of a European put option from the call option Greeks and put-call parity:

$$\text{Put Price} = \text{Call Price} + \mathrm{e}^{-rT}K - \mathrm{e}^{-qT}S(0)\ .$$

For example, the delta of a put is the delta of a call (with the same strike and maturity) minus e^{-qT}, and the gamma of a put is the same as the gamma of the corresponding call.

4.5 Delta Hedging

The ability to create a fully hedged (risk-free) portfolio of the stock and an option is the essence of the arbitrage argument underlying the Black-Scholes formula, as we saw in Chapter **??** for the binomial model. For a call option, such a portfolio consists of delta shares of the underlying asset and a short call option, or a short position of delta shares of the underlying and a long call option. These portfolios have no instantaneous exposure to the price of the underlying. To create a perfect hedge, the portfolio must be adjusted continuously, because the delta changes when the price of the underlying changes and when time passes. In practice, any hedge will therefore be imperfect, even if the assumptions of the model are satisfied.

We first consider the continuous-time hedging argument. Consider a European call option with maturity T, and let $C(S,t)$ denote the value of the option at date $t<T$ when the stock price is S at date t. Consider a portfolio that is short one call option and long δ shares of the underlying asset and that has a (short) cash position equal to $C-\delta S$. This portfolio has zero value at date t.

The change in the value of the portfolio in an instant $\mathrm{d}t$ is

$$-\mathrm{d}C+\delta\,\mathrm{d}S+q\delta S\,\mathrm{d}t+(C-\delta S)r\,\mathrm{d}t\;. \tag{4.9}$$

The first term reflects the change in the value of the option, the second term is the capital gain or loss on δ shares of stock, the third term is the dividends received on δ shares of stock, and the fourth term is the interest expense on the short cash position.

On the other hand, we know from Ito's formula that

$$\begin{aligned}\mathrm{d}C&=\frac{\partial C}{\partial S}\,\mathrm{d}S+\frac{\partial C}{\partial t}\,\mathrm{d}t+\frac{1}{2}\frac{\partial^2 C}{\partial S^2}(\mathrm{d}S)^2\\&=\delta\,\mathrm{d}S+\Theta\,\mathrm{d}t+\frac{1}{2}\Gamma\sigma^2S^2\,\mathrm{d}t\;.\end{aligned} \tag{4.10}$$

Substituting Equation **??** into Equation **??** shows that the change in the value of the portfolio is

$$-\Theta\,\mathrm{d}t-\frac{1}{2}\Gamma\sigma^2S^2\,\mathrm{d}t+q\delta S\,\mathrm{d}t+(C-\delta S)r\,\mathrm{d}t\;. \tag{4.11}$$

Several aspects of this are noteworthy. First, as noted earlier, the delta hedge (being long δ shares of the underlying) eliminates the exposure to changes in the price of the underlying—there is no $\mathrm{d}S$ term in Equation **??**. Second, Θ will be negative, because it captures the time decay in the option value; being short the option means the portfolio will profit from time decay at rate $-\Theta$. Third, this portfolio is short gamma. We can also say it is short convexity, the term convexity referring to the convex shape of the option value as a function of the price of the underlying, which translates mathematically to a positive second derivative (gamma). The volatility in the stock makes convexity valuable, and a portfolio that is short convexity will suffer losses. Finally, the portfolio is earning dividends but paying interest.

It is straightforward to check, from the definitions of Θ, Γ and δ in the preceding section, that the sum of the terms in Equation **??** is zero. The time decay in the option value and dividends received on the shares of the underlying exactly offset the losses due to convexity and interest. Therefore, the delta hedge is a perfect hedge. The portfolio, which has a zero cost, neither earns nor loses money. This is true not only on average but for every possible change in the stock price.

To see how well this works with only discrete adjustments to the hedge, one can simulate the changes in S over time and sum the gains and losses over discrete rebalancing periods. One should input the actual (not risk-neutral) expected rate of return on the asset to compute the actual distribution of gains and losses. This is discussed further in **?@sec-blackscholes_python**.

4.6 Gamma Hedging

To attempt to improve the performance of a discretely rebalanced delta hedge, one can use another option to create a portfolio that is both delta and gamma neutral. Being delta neutral means hedged as in the previous section—the portfolio value has no exposure to changes in the underlying asset price. In other words, it means that the derivative of the portfolio value with respect to the price of the underlying (the portfolio delta) is zero. Being gamma neutral means that the delta of the portfolio has no exposure to changes in the underlying price, which is equivalent to the second derivative of the portfolio value with respect to the price of the underlying (the portfolio gamma) being zero. If the delta truly did not change, then there would be no need to rebalance continuously, and hence no hedging error introduced by only adjusting the portfolio at discrete times rather than continuously. However, there is certainly no guarantee that a discretely-rebalanced delta/gamma hedge will perform better than a discretely rebalanced delta hedge.

A delta/gamma hedge can be constructed as follows. Suppose we have written (shorted) a call option and we want to hedge both the delta and gamma using the underlying asset and another option, for example, another call option with a different strike. In practice, one would want to use a liquid option for this purpose, which typically means that the strike of the option will be near the current value of the underlying (i.e., the option used to hedge would be approximately at the money).

Let δ and Γ denote the delta and gamma of the written option and let δ' and Γ' denote the delta and gamma of the option used to hedge. Consider holding a of shares of the stock and b units of the option used to hedge in conjunction with the short option. The delta of the stock is one ($\mathrm{d}S/\mathrm{d}S = 1$), so to obtain a zero portfolio delta we need

$$0 = -\delta + a + b\delta'. \tag{4.12}$$

The gamma of the stock is zero ($\mathrm{d}^2S/\mathrm{d}S^2 = d\,1/\mathrm{d}S = 0$), so to obtain a zero portfolio gamma we need

$$0 = -\Gamma + b\Gamma' \ . \tag{4.13}$$

Equation Equation **??** shows that we should hold enough of the second option to neutralize the gamma of the option we have shorted; i.e.,

$$b = \frac{\Gamma}{\Gamma'}$$

Equation Equation **??** shows that we should use the stock to delta hedge the portfolio of options; i.e.,

$$a = \delta - \frac{\Gamma}{\Gamma'}\delta' \ .$$

4.7 Implied Volatilities

All of the inputs into the option pricing formulas are in theory observable, except for the volatility coefficient σ. We can estimate σ from historical data (see Chapter **??**), or estimate it from the prices of other options. The latter method exploits the fact that there is a one-to-one relationship between the price given by the Black-Scholes formula and the σ that is input, so one can take the price as given and infer σ from the formula. The σ computed in this way is called the implied volatility. The implied volatility from one option can be used to price another (perhaps non-traded or less actively traded) option. The calculation of implied volatilities is discussed in **?@sec-blackscholes_implied**.

Even if we acknowledge that the model is not correct, the computation of implied volatilities is still useful for characterizing market prices, because we can quickly describe an option as expensive or cheap depending on whether its implied volatility is large or small. Somewhat paradoxically, it is less easy to see if an option is expensive or cheap by looking at its price, because one must consider the price in the context of the exercise price and maturity. To some extent, the implied volatility normalizes the price relative to the exercise price and maturity. Of course, it does not always pay to sell expensive options or buy cheap options, unless they are expensive or cheap relative to an accurate model!

4.8 Term Structure of Volatility

The option pricing formulas in this chapter are derived from the fact that the natural logarithm of the stock price at maturity is normally distributed with a certain mean (depending on the numeraire) and variance equal to $\sigma^2 T$. It is not actually necessary that the volatility be constant. The formulas are still valid if

$$\frac{\mathrm{d}S(t)}{S(t)} = \mu(t)\,\mathrm{d}t + \sigma(t)\,\mathrm{d}B(t)$$

where $\sigma(t)$ is some non-random function of time (and again μ can be a quite general random process). In this case, the variance of $\log S(T)$ will be

$$\int_0^T \sigma^2(t)\,\mathrm{d}t\ , \tag{4.14}$$

which is essentially the sum of the instantaneous variances $\sigma^2(t)\,\mathrm{d}t$. In the d_1's and d_2's in the option pricing formulas, $\sigma^2 T$ should be replaced by Equation **??**. A convenient way of expressing this is as follows. Let σ_{avg} be the positive number such that

$$\sigma_{\text{avg}}^2 = \frac{1}{T}\int_0^T \sigma^2(t)\,\mathrm{d}t\ . \tag{4.15}$$

Then we simply need to input σ_{avg} as `sigma` in our option pricing functions. We will call σ_{avg} the average volatility, though note that it is not really the average of $\sigma(t)$ but instead is the square root of the average of $\sigma^2(t)$.

It is important to recognize that, throughout this chapter, date 0 means the date at which the option is being valued. It is not necessarily the date at which the option was first bought or sold. So σ_{avg} is the average (in a sense) volatility during the remaining lifetime of the option, which need not be the same as the average during the option's entire lifetime. It is this remaining volatility that is important for pricing and hedging. Moreover, it is a mistake at date 0 to use $\sigma(0)$ as the volatility to compute prices and hedges. Instead, prices and hedges should be based on σ_{avg}.

These considerations provide a way to address the following situation. If we compute implied volatilities for options with different maturities, we will normally get different numbers. For example, consider two at-the-money options with maturities T_1 and T_2 where $T_2 > T_1$. Denote the implied volatilities by $\hat\sigma_1$ and $\hat\sigma_2$. We want to interpret these as average volatilities for the time periods $[0, T_1]$ and $[0, T_2]$ respectively. This requires the existence of a function $\sigma(t)$ such that

$$\hat\sigma_1^2 = \frac{1}{T_1}\int_0^{T_1} \sigma^2(t)\,\mathrm{d}t \quad \text{and} \quad \hat\sigma_2^2 = \frac{1}{T_2}\int_0^{T_2} \sigma^2(t)\,\mathrm{d}t\ .$$

This would imply

$$\hat\sigma_2^2 T_2 - \hat\sigma_1^2 T_1 = \int_{T_1}^{T_2} \sigma^2(t)\,\mathrm{d}t\ ,$$

which requires

$$\hat\sigma_2^2 T_2 - \hat\sigma_1^2 T_1 \geq 0\ .$$

Equivalently,

$$\hat\sigma_2 \geq \sqrt{\frac{T_1}{T_2}}\hat\sigma_1\ .$$

Provided this last inequality is satisfied, we can easily construct the function $\sigma(t)$ as

$$\sigma(t) = \begin{cases} \hat{\sigma}_1 & \text{for } t \leq T_1 \\ \sqrt{\frac{\hat{\sigma}_2^2 T_2 - \hat{\sigma}_1^2 T_1}{T_2 - T_1}} & \text{for } T_1 < t \leq T_2. \end{cases}$$

More generally, given a sequence of at-the-money options with maturities $T_1 < T_2 < \cdots T_N$ and implied volatilities $\hat{\sigma}_1, \dots, \hat{\sigma}_N$, we define

$$\sigma(t) = \sqrt{\frac{\hat{\sigma}_{i+1}^2 T_{i+1} - \hat{\sigma}_i^2 T_i}{T_{i+1} - T_i}}$$

for $T_i < t \leq T_{i+1}$, provided the expression inside the square root symbol is positive. This $\sigma(t)$ is often called the term structure of (implied) volatilities. Generally, we may expect $\sigma(t)$ to be a decreasing function of time t when the current market is especially volatile and to be an increasing function when the current market is especially quiet.

4.9 Smiles and Smirks

If we compute implied volatilities for options with the same maturity but different strikes, we will again obtain different implied volatilities for different options. If we plot implied volatility against the strike, the pattern one normally sees for equities and equity indices is the implied volatility declining as the strike increases until the strike is somewhere near the current value of the underlying (so the option is at the money). The implied volatility will then generally flatten out or increase slightly at higher strikes. The graph looks like a twisted smile (smirk). This pattern has been very pronounced in equity index option prices since the crash of 1987. In contrast to the term structure of implied volatilities, this moneyness structure of implied volatilities is simply inconsistent with the model. It suggests that the risk-neutral return distribution is not lognormal but instead exhibits a higher likelihood of extreme returns than the lognormal distribution (i.e., it has fat tails) with the likelihood of extreme negative returns being higher than the likelihood of extreme positive returns (i.e., it is skewed). We will return to this subject in Section **??**.

4.10 Calculations in Python

The following calculates the Black Scholes call, put, call delta, call gamma, and impliied volatility.

```python
import numpy as np
from scipy.stats import norm
import scipy.optimize as optimize

def black_scholes_call(S, K, r, sigma, q, T):
    """
    Inputs:
    S = initial stock price
    K = strike price
    r = risk-free rate
    sigma = volatility
    q = dividend yield
    T = time to maturity
    """
    if sigma == 0:
        return max(0, np.exp(-q * T) * S - np.exp(-r * T) * K)
    else:
        d1 = (np.log(S / K) + (r - q + 0.5 * sigma ** 2) * T) / (sigma * np.sqrt(T))
        d2 = d1 - sigma * np.sqrt(T)
        N1 = norm.cdf(d1)
        N2 = norm.cdf(d2)
        return np.exp(-q * T) * S * N1 - np.exp(-r * T) * K * N2

def black_scholes_put(S, K, r, sigma, q, T):
    """
    Inputs:
    S = initial stock price
    K = strike price
    r = risk-free rate
    sigma = volatility
    q = dividend yield
    T = time to maturity
    """
    if sigma == 0:
        return max(0, np.exp(-r * T) * K - np.exp(-q * T) * S)
    else:
        d1 = (np.log(S / K) + (r - q + 0.5 * sigma ** 2) * T) / (sigma * np.sqrt(T))
        d2 = d1 - sigma * np.sqrt(T)
        N1 = norm.cdf(-d1)
        N2 = norm.cdf(-d2)
        return np.exp(-r * T) * K * N2 - np.exp(-q * T) * S * N1
```

```python
def black_scholes_call_delta(S, K, r, sigma, q, T):
    """
    Inputs:
    S = initial stock price
    K = strike price
    r = risk-free rate
    sigma = volatility
    q = dividend yield
    T = time to maturity
    """
    d1 = (np.log(S / K) + (r - q + 0.5 * sigma ** 2) * T) / (sigma * np.sqrt(T))
    return np.exp(-q * T) * norm.cdf(d1)

def black_scholes_call_gamma(S, K, r, sigma, q, T):
    """
    Inputs:
    S = initial stock price
    K = strike price
    r = risk-free rate
    sigma = volatility
    q = dividend yield
    T = time to maturity
    """
    d1 = (np.log(S / K) + (r - q + 0.5 * sigma ** 2) * T) / (sigma * np.sqrt(T))
    nd1 = np.exp(-d1 ** 2 / 2) / np.sqrt(2 * np.pi)
    return np.exp(-q * T) * nd1 / (S * sigma * np.sqrt(T))

def black_scholes_call_implied_vol(S, K, r, q, T, CallPrice):
    """
    Inputs:
    S = initial stock price
    K = strike price
    r = risk-free rate
    q = dividend yield
    T = time to maturity
    CallPrice = call price
    """
    def objective(sigma):
        return black_scholes_call(S, K, r, sigma, q, T) - CallPrice

    if CallPrice < np.exp(-q * T) * S - np.exp(-r * T) * K:
        raise ValueError("Option price violates the arbitrage bound.")
```

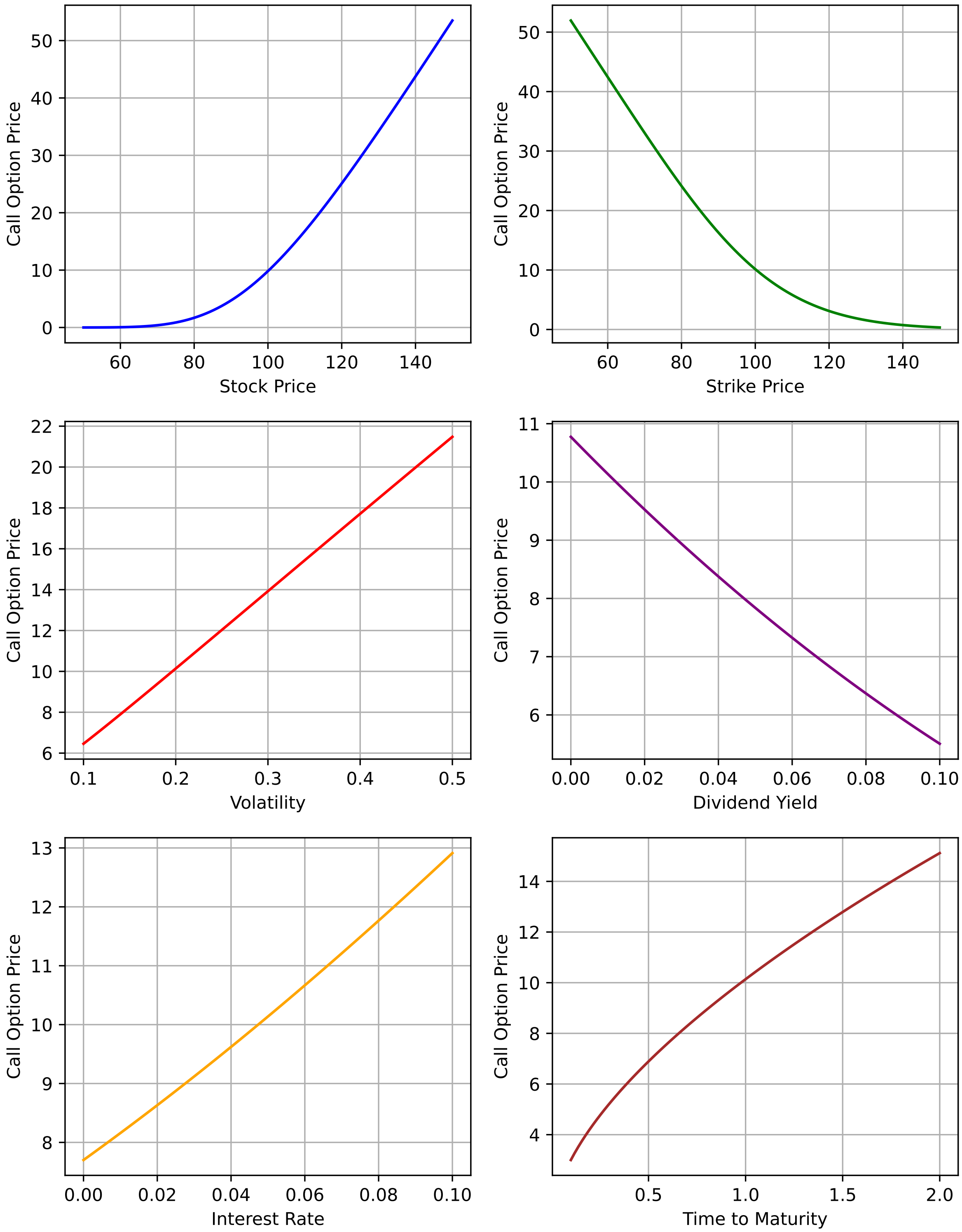
```python
    tol = 1e-6
    lower = 0
    upper = 1
    fupper = objective(upper)

    while fupper < 0:
        upper *= 2
        fupper = objective(upper)

    implied_vol = optimize.bisect(objective, lower, upper, xtol=tol)
    return implied_vol


# Example usage (you can replace these with input values)
S = 100  # Initial stock price
K = 100  # Strike price
r = 0.05  # Risk-free rate
sigma = 0.2  # Volatility
q = 0.02  # Dividend yield
T = 1  # Time to maturity in years
CallPrice = 10  # Call price for implied volatility calculation

# Calculate Black-Scholes call and put prices
call_price = black_scholes_call(S, K, r, sigma, q, T)
put_price = black_scholes_put(S, K, r, sigma, q, T)
print(f"Call Price: {call_price}, Put Price: {put_price}")

# Calculate Delta and Gamma for the call option
call_delta = black_scholes_call_delta(S, K, r, sigma, q, T)
call_gamma = black_scholes_call_gamma(S, K, r, sigma, q, T)
print(f"Call Delta: {call_delta}, Call Gamma: {call_gamma}")

# Calculate implied volatility for a given call price
implied_vol = black_scholes_call_implied_vol(S, K, r, q, T, CallPrice)
print(f"Implied Volatility: {implied_vol}")
```

```
Call Price: 9.227005508154036, Put Price: 6.330080627549918
Call Delta: 0.586851146134764, Call Gamma: 0.018950578755008718
Implied Volatility: 0.2203836441040039
```

The following plot the Black Scholes call price against the stock price, the strike price, volatility, dividend yield, interest rate, and time to maturity.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

black_scholes_call(S, K, r, sigma, q, T)

# Parameters
S = np.linspace(50, 150, 100)
K = 100
T = 1
r = 0.05
sigma = 0.2
q = 0.01

# Plot Black-Scholes call price against various parameters
fig, axs = plt.subplots(3, 2, figsize=(8, 10))

# Plot against stock price
call_prices_S = [black_scholes_call(s, K, r, sigma, q, T) for s in S]
axs[0, 0].plot(S, call_prices_S, label='Call Option Price', color='blue')
axs[0, 0].set_title('')
axs[0, 0].set_xlabel('Stock Price')
axs[0, 0].set_ylabel('Call Option Price')
axs[0, 0].grid(True)

# Plot against strike price
strike_prices = np.linspace(50, 150, 100)
call_prices_K = [black_scholes_call(S[50],k, r, sigma, q, T) for k in strike_prices]
axs[0, 1].plot(strike_prices, call_prices_K, label='Call Option Price', color='green')
axs[0, 1].set_title('')
axs[0, 1].set_xlabel('Strike Price')
axs[0, 1].set_ylabel('Call Option Price')
axs[0, 1].grid(True)

# Plot against volatility
volatilities = np.linspace(0.1, 0.5, 100)
call_prices_sigma = [black_scholes_call(S[50],K, r, sigma, q, T) for sigma in volatilities]
axs[1, 0].plot(volatilities, call_prices_sigma, label='Call Option Price', color='red')
axs[1, 0].set_title('')
axs[1, 0].set_xlabel('Volatility')
axs[1, 0].set_ylabel('Call Option Price')
```

```python
axs[1, 0].grid(True)

# Plot against dividend yield
dividend_yields = np.linspace(0, 0.1, 100)
call_prices_q = [black_scholes_call(S[50],K, r, sigma, q, T) for q in dividend_yields]
axs[1, 1].plot(dividend_yields, call_prices_q, label='Call Option Price', color='purple')
axs[1, 1].set_title('')
axs[1, 1].set_xlabel('Dividend Yield')
axs[1, 1].set_ylabel('Call Option Price')
axs[1, 1].grid(True)

# Plot against interest rate
interest_rates = np.linspace(0, 0.1, 100)
call_prices_r = [black_scholes_call(S[50],K, r, sigma, q, T) for r in interest_rates]
axs[2, 0].plot(interest_rates, call_prices_r, label='Call Option Price', color='orange')
axs[2, 0].set_title('')
axs[2, 0].set_xlabel('Interest Rate')
axs[2, 0].set_ylabel('Call Option Price')
axs[2, 0].grid(True)

# Plot against time to maturity
times_to_maturity = np.linspace(0.1, 2, 100)
call_prices_T = [black_scholes_call(S[50],K, r, sigma, q, T) for T in times_to_maturity]
axs[2, 1].plot(times_to_maturity, call_prices_T, label='Call Option Price', color='brown')
axs[2, 1].set_title('')
axs[2, 1].set_xlabel('Time to Maturity')
axs[2, 1].set_ylabel('Call Option Price')
axs[2, 1].grid(True)

plt.tight_layout()
plt.show()
```


Call Option Price
Stock Price
Call Option Price
Strike Price
Call Option Price
Volatility
Call Option Price
Dividend Yield
Call Option Price
Interest Rate
Call Option Price
Time to Maturity

4.11 Example: Replicating Portfolios and Simulating Portfolio Insurance

Another derivation of the Black Scholes formula is provided by Merton. He asked the question whether by trading the stock and the risk free asset whether the payoff to a European call option can be replicated. Let θ_t be the number of shares of the stock held at time t and α_t the number of shares of an initial investment of one dollar in the risk free asset. Then the portfolio is worth $\alpha_t R_t + \theta_t S_t$ where $R_t = e^{rt}$ is the time t value of an initial time 0 investment of one dollar in the risk free asset. The portfolio should start with an initial value, should not have any cash inflows or outflows and have a terminal value equal to a call payoff so the changes in value are completely dictated by the changes in the value of the assets. That is, assuming continuous trading,

$$dW_t = \theta_t dS_t + \alpha_t dR_t = \theta_t \left(\mu S_t dt + q S_t dt + \sigma S_t dB_t\right) + \alpha_t r R_t dt$$

with terminal condition

$$W_T = \alpha_T R_T + \theta_T S_T = (S_T - K)^+$$

The problem is to find θ_t and α_t for all times and states. If we can accomplish this, then by 'no-arbitrage' the call price must be the value of the initial investment. Assume the call price is a function of the stock price and time: $C(t, S_t)$. Then by Ito's Lemma

$$dC(t, S_t) = \left(\frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}(\mu - q)S_t + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S_t^2\right) dt + \frac{\partial C}{\partial S}\sigma S_t dB_t$$

It should be apparent that we want to hold $\theta = \frac{\partial C}{\partial S}$, which is the delta of the call option. By doing so, we match the diffusion term in thw change in wealth and the change in the call option. Then matching the drift terms in both expressions

$$\frac{\partial C}{\partial S}\mu S_t + \alpha_t r R_t = \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}(\mu - q)S_t + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S_t^2$$

which can be solved to give

$$\alpha_t r R_t = r\left(W_t - \frac{\partial C}{\partial S}S_t\right) = \frac{\partial C}{\partial t} - \frac{\partial C}{\partial S}q S_t + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S_t^2$$

which gives the equation

$$rW_t = \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}(r - q)S_t + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S_t^2$$

with a boundary condition $W_T = (S_T - K)^+$. However, no-arbitrage suggests $W_t = C(t, S_t)$ which gives us the partial differential equation

$$rC = \frac{\partial C}{\partial t} + \frac{\partial C}{\partial S}(r - q)S + \frac{1}{2}\frac{\partial^2 C}{\partial S^2}\sigma^2 S^2$$

with a boundary condition $C(T, S_T) = (S_T - K)^+$. This is a partial differential equation and a fairly tedious set of calcuations show the Black Scholes formula is a solution (in fact it is the only positive solution). Close observation of the right hand side we see this is the drift term of Ito expansion for C if we work in the risk neutral measure. The right hand side then says in the risk neutral measure, the call option earns the risk free return.

However, there is nothing special about a call option. The same argument will apply for any European style option. The only difference is the boundary condition. This procedure allows us to replicate the payoff of any European option even for those which might not be traded. This observation had a profound effect on practice. A particularly popular example is portfolio insurance.

Recall, that a protective put position buys a put and buys a share and the payoff at the expiration of the put is given by $\max(K, S_T)$. The reason for the name protective put is apparent since the position can pay off no less than K. The cost of this insurance is the price of the put. However, if the put is not traded, we can synthetically replicate this payoff using the prodedure above assuming we can trade continuously. The basic recipe is to start with inital wealth equal to that for a protective put position: $W_0 = P(0, S_0) + S_0$. The delta of the protective put position can be calcuated to be the delta of the put plus 1 which is $N(d_1)$, where d_1 is calculated at each point in time. However, in practice we cannot trade continuously. A simple discrete strategy would rebalance at intervals Δt. The strategy calculates $N(d_1)$ at time 0 and holds $P(0, S_0) + S_0 - N(d_1)S_0$ dollars in the risk free asset and $N(d_1)$ shares of the asset. Thereafter these holdings are adjusted. The change in portfolio value over the interval Δt is

$$\Delta W = W_{i\Delta t} - W_{(i-1)\Delta t}$$

$$= (P((i{-}1)\Delta t, S_{(i-1)\Delta t}) + S_{(i-1)\Delta t} - N(d_1{-})S_{(i-1)\Delta t})(R_{i\Delta t} - R_{(i-1)\Delta t}) + N(d1{-})(S_{i\Delta t} - S_{(i-1)\Delta t})$$

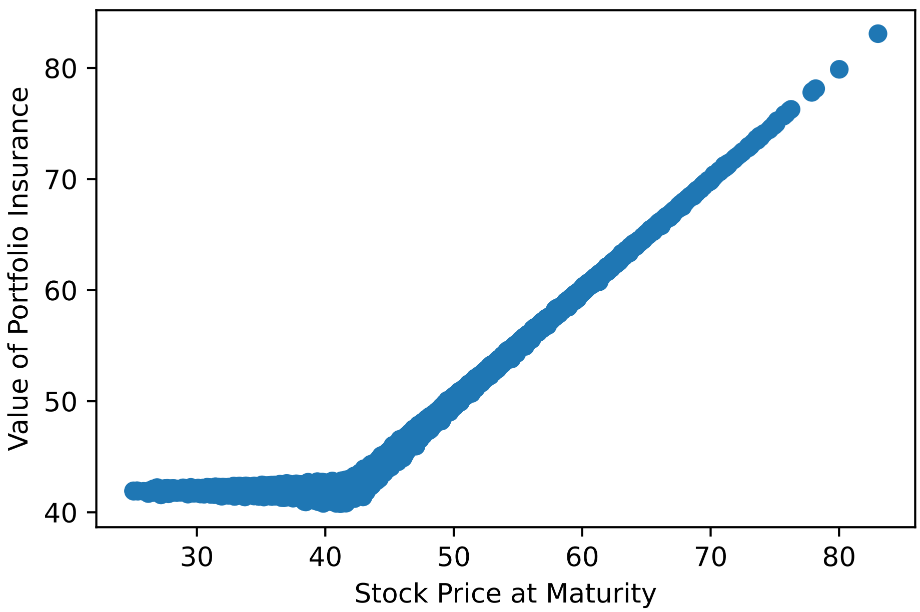
where $N(d1-)$ is the delta chosen at time $(i-1)\Delta t$. The question is if the Black Scoles model is correct, how accurate can a discrete rebalancing scheme be? This is simulated in the following code:

```
import numpy as np
# from bsfunctions import *
import matplotlib.pyplot as plt
import time
from math import pow, exp, sqrt
from scipy import stats
# incs = np.genfromtxt('incs.csv',delimiter=",",skip_header=1)
def blackscholes(S0, K, r, q, sig, T, call = True):
    '''Calculate option price using B-S formula.

    Args:
        S0 (num): initial price of underlying asset.
        K (num): strick price.
```

```
        r (num): risk free rate.
        q (num): dividend yield
        sig (num): Black-Scholes volatility.
        T (num): maturity.
        call (bool): True returns call price, False returns put price.

    Returns:
        num
    '''
    d1 = (np.log(S0/K) + (r -q + sig**2/2) * T)/(sig*np.sqrt(T))
    d2 = d1 - sig*np.sqrt(T)
#     norm = sp.stats.norm
    norm = stats.norm
    if call:
        return np.exp(-q*T)*S0 * norm.cdf(d1,0,1) - K * np.exp(-r * T) * norm.cdf(d2,0, 1)
    else:
        return np.exp(-q*T)*S0 * -norm.cdf(-d1,0,1) + K * np.exp(-r * T) * norm.cdf(-d2,0, 1

def blackscholes_delta(S0, K, r, q, sig, T, call = True):
    '''Calculate option price using B-S formula.

    Args:
        S0 (num): initial price of underlying asset.
        K (num): strick price.
        r (num): risk free rate.
        q (num): dividend yield
        sig (num): Black-Scholes volatility.
        T (num): maturity.
        call (bool): True returns call price, False returns put price.

    Returns:
        num
    '''
    d1 = (np.log(S0/K) + (r -q + sig**2/2) * T)/(sig*np.sqrt(T))
    d2 = d1 - sig*np.sqrt(T)
#     norm = sp.stats.norm
    norm = stats.norm
    if type(call) == bool:
        if call:
            return np.exp(-q*T)*norm.cdf(d1,0,1)
        else:
            return np.exp(-q*T)*norm.cdf(-d1,0,1)
```

```
    else:
        print("Not a valid value for call")

# parameters
# number of paths
# n = incs.shape[1]
n = 100000
# number of divisions
# m = incs.shape[0]
m = 100
# interest rate
r = .1
# dividend yield
q=0.0
# true drift
mu = .15
# volatility
sig = .2
# Initial Stock Price
S0 = 42
# Strike Price
K = 42
# Maturity
T = 0.5


# seed for random generator
seed= 1234
# define a random generator
rg = np.random.RandomState(seed)
# initialize


# generate normal random vairables
dt= T/m
vol=sig*np.sqrt(dt)
incs = rg.normal(0,vol,[m,n])


tline = np.linspace(0,T,m+1)
```

```
St = np.zeros((m+1,n))
#St1 = np.zeros((m+1,n))

V_vec = np.zeros((m+1,n))

delta = np.zeros((m,n))

put= blackscholes(S0,K,r, q, sig,T,call=False)

incs_cumsum =  np.concatenate((np.zeros((1,n)),incs),axis=0).cumsum(axis=0)
V_vec = np.zeros((m+1,n))
t_mat =  np.repeat(tline.reshape((m+1,1)), n, axis=1)
drift_cumsum = (mu -q -0.5*sig**2) * t_mat

St = S0 * np.exp(incs_cumsum + drift_cumsum)

delta = blackscholes_delta(St[:-1,:],K,r, q, sig,T-t_mat[:-1,:])

V_vec[0,:] = S0 + put

for i in range(1,m+1):
    V_vec[i,:] = V_vec[i-1,:] + (np.exp(r*dt)-1) * (V_vec[i-1,:] - delta[i-1,:] * St[i-1,:])

# Uses actual simulated changes in riskfree and stock price not the dt and dB approximations
# plot ST versus VT
plt.scatter(St[m,:],V_vec[m,:])
plt.xlabel('Stock Price at Maturity')
plt.ylabel('Value of Portfolio Insurance')
plt.show()
```



With $m = 100$ rebalancing dates over $T = 0.5$ for the parameters chosen the repliacting strategy does a pretty good job. The hedging errors occur when the stock price is close to the strike price. This is not surprising since the delta changes (measured by the gamma) fastest around this point. A gamma hedge would potentially improve the performance.

The portfolio insurance rebalancing scheme involves sell stock and buying bonds when the stock price goes down and buying stocks and selling bonds when the stock price goes up. This can be destabilizing and was identified as a contributor to the 1987 stock market crash.

4.11.1 Discretely-Rebalanced Delta Hedges

To compute the real-world distribution of gains and losses from a discretely-rebalanced delta hedge, we input the expected rate of return μ. We consider adjusting the hedge at dates $0 = t_0 < t_1 < \cdots < t_N = T$, with $t_i - t_{i-1} = \Delta t = T/N$ for each i. The changes in the natural logarithm of the stock price between successive dates t_{i-1} and t_i are simulated as

$$\Delta \log S = \left(\mu - q - \frac{1}{2}\sigma^2\right) \Delta t + \sigma \, \Delta B \; ,$$

where ΔB is normally distributed with mean zero and variance Δt. The random variables ΔB are simulated as standard normals multiplied by $\sqrt{\Delta t}$. We begin with the portfolio that is short a call, long δ shares of the underlying, and short $\delta S - C$ in cash. After the stock price changes, say from S to S', we compute the new delta δ'. The cash flow from adjusting the hedge is $(\delta - \delta')S'$. Accumulation (or payment) of interest on the cash position is captured by the factor $e^{r\Delta t}$. Continuous payment of dividends is modelled similarly: the dividends earned during the period Δt is taken to be $\delta S\left(e^{q\Delta t} - 1\right)$. The cash position is adjusted due to interest, dividends, and the cash flow from adjusting the hedge. At date T, the value of the portfolio is the cash position less the intrinsic value of the option.

To describe the distribution of gains and losses, we compute percentiles of the distribution. You should see that the hedge becomes more nearly perfect as the number of periods N is increased. Note that this is true regardless of the μ that is input, which reaffirms the point that option values and hedges do not depend on the expected rate of return of the underlying. The percentile is calculated with the Excel `Percentile` function.[3]

```python
import numpy as np
from scipy.stats import norm
import scipy.optimize as optimize

def simulated_delta_hedge_profit(S0, K, r, sigma, q, T, mu, M, N, pct):
    """
    Inputs:
    S0 = initial stock price
    K = strike price
    r = risk-free rate
    sigma = volatility
    q = dividend yield
    T = time to maturity
    mu = expected rate of return
    N = number of time periods
    M = number of simulations
    pct = percentile to be returned
    """
    dt = T / N
    SigSqrdt = sigma * np.sqrt(dt)
    drift = (mu - q - 0.5 * sigma ** 2) * dt
    Comp = np.exp(r * dt)
    Div = np.exp(q * dt) - 1
    LogS0 = np.log(S0)
    Call0 = black_scholes_call(S0, K, r, sigma, q, T)
    Delta0 = black_scholes_call_delta(S0, K, r, sigma, q, T)
    Cash0 = Call0 - Delta0 * S0
    Profit = np.zeros(M)
```

[3] If numsims = 11 and pct =0.1, the percentile function returns the second lowest element in the series. The logic is that 10% of the numbers, excluding the number returned, are below the number returned—i.e., 1 out of the other 10 are below—and 90% of the others are above. In particular, if pct = 0.5, the percentile function returns the median. When necessary, the function interpolates; for example, if numsims = 10 and pct=0.1, then the number returned is an interpolation between the lowest and second lowest numbers.

```python
    for i in range(M):
        LogS = LogS0
        Cash = Cash0
        S = S0
        Delta = Delta0

        for j in range(1, N):
            LogS += drift + SigSqrdt * np.random.randn()
            NewS = np.exp(LogS)
            NewDelta = black_scholes_call_delta(NewS, K, r, sigma, q, T - j * dt)
            Cash = Comp * Cash + Delta * S * Div - (NewDelta - Delta) * NewS
            S = NewS
            Delta = NewDelta

        LogS += drift + SigSqrdt * np.random.randn()
        NewS = np.exp(LogS)
        HedgeValue = Comp * Cash + Delta * S * Div + Delta * NewS
        Profit[i] = HedgeValue - max(NewS - K, 0)

    return np.percentile(Profit, pct * 100)

# Example usage (you can replace these with input values)
S = 100  # Initial stock price
K = 100  # Strike price
r = 0.05  # Risk-free rate
sigma = 0.2  # Volatility
q = 0.02  # Dividend yield
T = 1  # Time to maturity in years
CallPrice = 10  # Call price for implied volatility calculation

# Simulate delta hedging profit
S0 = 100  # Initial stock price
mu = 0.1  # Expected rate of return
M = 1000  # Number of simulations
N = 252  # Number of time periods
pct = 0.95  # Percentile to be returned

delta_hedge_profit = simulated_delta_hedge_profit(S0, K, r, sigma, q, T, mu, M, N, pct)
print(f"Delta Hedge Profit (95th percentile): {delta_hedge_profit}")
```

```
Delta Hedge Profit (95th percentile): 0.6283250482023715
```

4.12 Exercises

Exercise 4.1. Create a Python code which inputs K, r, σ, q and T. Compute the delta of a call option for stock prices $S = .01K$, $.02K$, ..., $1.99K$, $2K$ (i.e., $S = iK/100$ for $i = 1, \ldots 200$) and plot the delta against the stock price.

Exercise 4.2. The delta of a digital option that pays \$1 when $S(T) > K$ is

$$\frac{\mathrm{e}^{-rT}\mathrm{n}d(d_2)}{\sigma S\sqrt{T}}\,.$$

Repeat the previous problem for the delta of this digital. Given that in reality it is costly to trade (due to commissions, the bid-ask spread and possible adverse price impacts for large trades), do you see any problems with delta hedging a short digital near maturity if it is close to being at the money?

Exercise 4.3. Modify the Python code for replicating portfolio insurance to simulate a discrete replication of a digital option using the delta in the previous problem. Run the code for $10, 20, 100, 1000$ rebalancing dates. When does the strategy do a good job and when does it fail?

Exercise 4.4. Repeat Exercise **??** for the gamma of a call option.

Exercise 4.5. Consider delta and gamma hedging a short call option, using the underlying and a put with the same strike and maturity as the call. Calculate the position in the underlying and the put that you should take, using the analysis in Section **??**. Will you ever need to adjust this hedge? Relate your result to put-call parity.

Exercise 4.6. The delta of a share digital that pays one share when $S(T) > K$ is

$$\mathrm{e}^{-qT}\mathrm{N}(d_1) + \frac{\mathrm{e}^{-qT}\mathrm{n}d(d_1)}{\sigma\sqrt{T}}\,.$$

Repeat Exercise **??** for the delta of this share digital.

Exercise 4.7. Compute the value of an at-the-money call option ($S = K$) using the Python code for volatilities $\sigma = .01, .02, \ldots, 1.0$. Plot the call value against the volatility.

Exercise 4.8. Repeat the previous problem for $S = 1.2K$ (an example of an in-the-money call option).

Exercise 4.9. The file CBOEQuotes.txt (available at `www.kerryback.net`) contains price data for call options on the S&P 500 index. The options expired in February, 2003, and the prices were obtained on January 22, 2003. The first column lists various exercise prices. The second column gives the bid price and the third column the ask price. Import this data into an Excel worksheet and compute and plot the implied volatility against the exercise price using this data. Use the ask price as the market price for the option. The options have 30 days to maturity (so $T = 30/365$). At the time the quotes were downloaded, the S&P 500 was at 884.25. According to the CBOE, the dividend yield on the S&P 500 was 1.76%. Use 1.25% for the risk-free interest rate.

Exercise 4.10. Attempt to repeat the previous problem using the bid price as the market price of the option. If this doesn't work, what is wrong? Does this indicate there is an arbitrage opportunity?

::: Suppose an investor invests in a portfolio with price S and constant dividend yield q. Assume the investor is charged a constant expense ratio α (which acts as a negative dividend) and at date T receives either his portfolio value or his initial investment, whichever is higher. This is similar to a popular type of variable annuity. Letting D denote the number of dollars invested in the contract, the contract pays

$$\max\left(D, \frac{D\mathrm{e}^{(q-\alpha)T}S(T)}{S(0)}\right) \tag{4.16}$$

at date T.
We can rearrange the expression Equation **??** as

$$\begin{aligned}\max\left(D, \frac{D\mathrm{e}^{(q-\alpha)T}S(T)}{S(0)}\right) &= D + \max\left(0, \frac{D\mathrm{e}^{(q-\alpha)T}S(T)}{S(0)} - D\right) \\ &= D + \mathrm{e}^{-\alpha T}D\max\left(0, \frac{\mathrm{e}^{qT}S(T)}{S(0)} - \mathrm{e}^{\alpha T}\right) \,. \end{aligned} \tag{4.17}$$

Thus, the contract payoff is equivalent to the amount invested plus a certain number of call options written on the gross holding period return $\mathrm{e}^{qT}S(T)/S(0)$. Note that $Z(t) = \mathrm{e}^{qt}S(t)/S(0)$ is the date–t value of the portfolio that starts with $1/S(0)$ units of the asset (i.e., with a \$1 investment) and reinvests dividends. Thus, the call options are call options on a non-dividend paying portfolio with the same volatility as S and initial price of \$1. This implies that the date–0 value of the contract to the investor is $\mathrm{e}^{-rT}D$ plus

$e^{-\alpha*T} * D*\texttt{Black\_Scholes\_Call}(1, e^{-\alpha*T}, r, sigma, q, T)$

1. Create a Python function to compute the fair expense ratio; i.e., find α such that the date–0 value of the contract is equal to D. Hint: Modify the

`Black_Scholes_Call_Implied_Vol`

function. You can use $\alpha = 0$ as a lower bound. Because the value of the contract is decreasing as α increases, you can find an upper bound by iterating until the value of the contract is less than D. 2. How does the fair expense ratio vary with the maturity T? Why?

::: ::: {#exr-nolabel} Modify the function `Simulated_Delta_Hedge_Profit` to compute percentiles of gains and losses for an investor who writes a call option and constructs a delta and gamma hedge using the underlying asset and another call option. Include the exercise price of the call option used to hedge as an input, and assume it has the same time to maturity as the option that is written. Hint: In each period `j = 1 to N-1`, the updated cash position can be calculated as

```
Cash = exp(r*dt)*Cash + a*S*(exp(q*dt)-1) - (Newa-a)*NewS _
     - (Newb-b)*PriceHedge ,
```

where `a` denotes the number of shares of the stock held, `b` denotes the number of units held of the option that is used for hedging, and `PriceHedge` denotes the price of the option used for hedging (computed from the Black-Scholes formula each period). This expression embodies the interest earned (paid) on the cash position, the dividends received on the shares of stock and the cash inflows (outflows) from adjusting the hedge. At the final date `N`, the value of the hedge is

```
exp(r*dt)*Cash + a*S*(exp(q*dt)-1) + a*NewS _
     + b*Application.Max(NewS-KHedge,0) ,
```

and the value of the overall portfolio is the value of the hedge less

```
Application.Max(NewS-KWritten,0) ,
```

where `KHedge` denotes the strike price of the option used to hedge and `KWritten` denotes the strike of the option that was written. :::

5 Estimating and Modeling Volatility

Thus far, we have assumed that the volatility of the underlying asset is constant or varying in a non-random way during the lifetime of the derivative. In this chapter we will look at models that relax this assumption and allow the volatility to change randomly. This is very important, because there is plenty of evidence that volatilities do change over time in a random way.

In the first three sections, we will consider the problem of estimating the volatility. The discussion of estimation methods leads naturally into the discussion of modeling a changing volatility.

5.1 Statistics Review

We begin with a brief review of basic statistics. Given a random sample $\{x_1, \ldots, x_N\}$ of size N from a population with mean μ and variance σ^2, the best estimate of μ is of course the sample mean

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \ .$$

The variance is the expected value of $(x-\mu)^2$, so an obvious estimate of the variance is the sample average of $(x_i-\mu)^2$, replacing μ with its estimate $\bar{x}$. This would be

$$\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2$$

However, because $\bar{x}$ is computed from the x_i, the x_i will deviate less on average from $\bar{x}$ than they do from the true mean μ. Hence the estimate proposed above will on average be less than σ^2. To eliminate this bias, it suffices just to scale the estimate up by a factor of $N/(N-1)$. This leads to the estimate

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2 \ ,$$

and the best estimate of σ is the square root

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2} \ .$$

To calculate s^2, notice that

$$\begin{aligned}\sum_{i=1}^N (x_i - \bar{x})^2 &= \sum_{i=1}^N (x_i^2 - 2x_i\bar{x} + \bar{x}^2) \\ &= \sum_{i=1}^N x_i^2 - 2\bar{x}\sum_{i=1}^N x_i + \sum_{i=1}^N \bar{x}^2 \\ &= \sum_{i=1}^N x_i^2 - 2\bar{x}(N\bar{x}) + N\bar{x}^2 \\ &= \sum_{i=1}^N x_i^2 - N\bar{x}^2 \; .\end{aligned}$$

Therefore

$$s = \sqrt{\frac{1}{N-1}\left(\sum_{i=1}^N x_i^2 - N\bar{x}^2\right)} \; .$$

It is important to know how much variation there would be in $\bar{x}$ if one had access to multiple random samples. More variation means that an $\bar{x}$ computed from a single sample will be a less reliable estimate of μ. The variance of $\bar{x}$ in repeated samples is σ^2/N,[1] and our best estimate of this variance is s^2/N. The standard deviation of $\bar{x}$ in repeated samples, which is called the standard error of $\bar{x}$, is $\sigma/\sqrt{N}$, and we estimate this by $s/\sqrt{N}$, which equals

$$\sqrt{\frac{1}{N(N-1)}\left(\sum_{i=1}^N x_i^2 - N\bar{x}^2\right)} \; .$$

If the population from which x is sampled has a normal distribution, then a 95% confidence interval for μ will be $\bar{x}$ plus or minus 1.96 standard errors. Even if x does not have a normal distribution, by the Central Limit Theorem, $\bar{x}/\sqrt{N}$ will be approximately normally distributed if the sample size N is large enough, and plus or minus 1.96 standard errors will still be approximately a 95% confidence interval for μ.

5.2 Estimating a Constant Volatility and Mean

Consider an asset price that is a geometric Brownian motion under the actual probability measure:

$$\frac{\mathrm{d}S}{S} = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}B \; ,$$

[1] The variance of $\bar{x} = (1/N)(x_1 + \cdots + x_N)$ is, by independence of the x_i, equal to $(1/N)^2(\mathrm{var} x_1 + \cdots + \mathrm{var} x_N)$, and, because the x_i all have the same variance σ^2, this is equal to $(1/N)^2 \times N\sigma^2 = \sigma^2/N$.

where μ and σ are unknown constants and B is a Brownian motion. We can as usual write this in log form as

$$\mathrm{d}\log S = \left(\mu - \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B\ .$$

Over a discrete time period of length Δt, this implies

$$\Delta \log S = \left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\Delta B\ . \tag{5.1}$$

Suppose we have observed the asset price S at dates $0 = t_0 < t_1 < \cdots < t_N = T$, where $t_i - t_{i-1} = \Delta t$. If the asset pays dividends, we will take S to be the value of the portfolio in which the dividends are reinvested in new shares. Thus, in general, $S(t_i)/S(t_{i-1})$ denotes the gross return (one plus the rate of return) between dates t_{i-1} and t_i. This return is measured on a non-compounded and non-annualized basis. The annualized continuously-compounded rate of return is the rate r_i defined by

$$\frac{S(t_i)}{S(t_{i-1})} = \mathrm{e}^{r_i\Delta t}\ .$$

This implies that

$$r_i = \frac{\log S(t_i) - \log S(t_{i-1})}{\Delta t} = \mu - \frac{1}{2}\sigma^2 + \sigma\frac{B(t_i) - B(t_{i-1})}{\Delta t}\ . \tag{5.2}$$

Because $B(t_i) - B(t_{i-1})$ is normally distributed with mean zero and variance Δt, the sample $\{r_1, \ldots, r_N\}$ is a sample of independent random variables each of which is normally distributed with mean $\mu - \sigma^2/2$ and variance $\sigma^2/\Delta t$. We are focused on estimating σ^2, so it will simplify things to define

$$y_i = r_i\sqrt{\Delta t} = \frac{\log S(t_i) - \log S(t_{i-1})}{\sqrt{\Delta t}}\ . \tag{5.3}$$

The sample $\{y_1, \ldots, y_N\}$ is a sample of independent random variables each of which is normally distributed with mean $(\mu - \sigma^2/2)\sqrt{\Delta t}$ and variance σ^2. As was discussed in the previous section, the best estimate of the mean of y is the sample mean

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i\ ,$$

and the best estimate of σ^2 is

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2\ .$$

This means that we estimate μ as

$$\hat{\mu} = \frac{\bar{y}}{\sqrt{\Delta t}} + \frac{1}{2}\hat{\sigma}^2 = \bar{r} + \frac{1}{2}\hat{\sigma}^2\ .$$

Let us digress for a moment to discuss the reliability of $\hat{\mu}$ as an estimate of μ. Notice that

$$\begin{aligned}
\bar{r} &= \frac{\sum_{i=1}^{N} \log S(t_i) - \log S(t_{i-1})}{N\Delta t} \\
&= \frac{\log S(T) - \log S(0)}{N\Delta t} \\
&= \frac{\log S(T) - \log S(0)}{T} \; . \qquad (5.4)
\end{aligned}$$

Therefore the first component $\bar{r}$ of the estimate of μ depends only on the total change in S over the time period. Hence, the reliability of this component cannot depend on how frequently we observe S within the time period $[0, T]$. The standard deviation of $\bar{r}$ in repeated samples is the standard deviation of $[\log S(T) - \log S(0)]/T$, which is $\sigma/\sqrt{T}$. This is likely to be quite large. For example, with $\sigma = 0.3$ and ten years of data ($T = 10$), the standard deviation of $\bar{r}$ is 9.5%, which means that a 95% confidence interval will be a band of roughly 38%. Given that μ itself should be of the order of magnitude of 10%, such a wide confidence interval is useless for all practical purposes.

Fortunately, it is easier to estimate σ. We observed in the previous section that the $\hat{\sigma}^2$ defined above can be calculated as

$$\frac{1}{N-1} \sum_{i=1}^{N} y_i^2 - \frac{N\bar{y}^2}{N-1} \; . \qquad (5.5)$$

From Equation **??** of y_i and Equation **??**, we have

$$\bar{y} = \frac{\sqrt{\Delta t}}{T} [\log S(T) - \log S(0)] \; .$$

Hence, the second term in Equation **??** is

$$\frac{N}{N-1} \left(\frac{\Delta t}{T^2} \right) [\log S(T) - \log S(0)]^2 \; .$$

If we observe the stock price sufficiently frequently, so that Δt is very small, this term will be negligible. In this circumstance, $\hat{\sigma}^2$ is approximately

$$\begin{aligned}
\frac{1}{N-1} \sum_{i=1}^{N} y_i^2 &= \frac{1}{N-1} \sum_{i=1}^{N} \frac{[\log S(t_i) - \log S(t_{i-1})]^2}{\Delta t} \\
&= \frac{N}{N-1} \times \frac{1}{T} \times \sum_{i=1}^{N} [\log S(t_i) - \log S(t_{i-1})]^2 \; . \qquad (5.6)
\end{aligned}$$

If we observe S more and more frequently, letting $\Delta t \to 0$ and $N \to \infty$, the sum

$$\sum_{i=1}^{N} [\log S(t_i) - \log S(t_{i-1})]^2$$

will converge with probability one to $\sigma^2 T$, as explained in Section **??**. This implies that $\hat{\sigma}^2$ will converge to σ^2. Thus, in theory, we can estimate σ^2 with any desired degree of precision by simply observing S sufficiently frequently. This is true no matter how short the overall time period $[0, T]$ may be.

In practice, this doesn't work out quite so well. If we observe minute-by-minute data, or we observe each transaction, much of the variation in the price S will be due to bouncing back and forth between the bid price and the ask price. This is not really what we want to estimate, and this source of variation will be much less important if we look at weekly or even daily data. So, there are practical limits to how frequently we should observe S. Nevertheless, it is still true that, if σ^2 were truly constant, we could estimate it with a very high degree of precision. In fact, we can estimate the volatility of a stock with enough precision to determine that it really isn't constant! The real problem that we face is to estimate and model a changing volatility.

5.3 Estimating a Changing Volatility

Without attempting yet to model how the volatility may change, we can say a few things about how we might estimate a changing volatility. In this and following sections, we will take the observation interval Δt to be fixed. We assume it is small (say, a day or a week) and focus on the estimate Equation **??**. Recall from Section **??** that the reason we are dividing by $N-1$ rather than N is that the sample standard deviation usually underestimates the actual standard deviation, because it uses the sample mean, which will be closer to the points x_i than will be the true mean. However, Equation **??** does not employ the sample mean (it replaces it with zero), so there is no reason to make this correction. So, we take as our point of departure the estimate

$$\frac{1}{T} \sum_{i=1}^{N} [\log S(t_i) - \log S(t_{i-1})]^2 = \frac{1}{N} \sum_{i=1}^{N} y_i^2 \ .$$

An obvious response to the volatility changing over time is simply to avoid using data from the distant past. Such data is not likely to be informative about the current value of the volatility. What distant should mean in this context is not entirely clear, but, for example, we might want to use only the last 60 observations. If we are using daily data, this would mean that at the end of each day we would add that day's observation and drop the observation from 61 days past. This leads to a somewhat abruptly varying estimate. For example, a very large movement in the price on a particular day increases the volatility estimate for the next 60 days. On the 61st day, this observation would drop from the sample, leading to an abrupt drop in the estimate (presuming that there is not an equally large change in S on the 61st day). This seems unreasonable. An estimate in which the impact of each observation decays smoothly over time is more attractive.

We can construct such an estimate as

$$\hat{\sigma}_{i+1}^2 = (1-\lambda)y_i^2 + \lambda\hat{\sigma}_i^2 \tag{5.7}$$

for any constant $0 < \lambda < 1$. Here, $\hat{\sigma}_{i+1}^2$ denotes the estimate of the volatility from date t_i to date t_{i+1}. The estimate Equation **??** is a weighted average of the estimate $\hat{\sigma}_i^2$ for the previous time period and the most recently observed squared change y_i^2. Following the same procedure, the next estimate will be

$$\begin{aligned}\hat{\sigma}_{i+2}^2 &= (1-\lambda)y_{i+1}^2 + \lambda\hat{\sigma}_{i+1}^2 \\ &= (1-\lambda)y_{i+1}^2 + \lambda(1-\lambda)y_i^2 + \lambda^2\hat{\sigma}_i^2\ .\end{aligned}$$

Likewise, the estimate at the following date will be

$$\hat{\sigma}_{i+3}^2 = (1-\lambda)y_{i+2}^2 + \lambda(1-\lambda)y_{i+1}^2 + \lambda^2(1-\lambda)^2y_i^2 + \lambda^3\hat{\sigma}_i^2\ .$$

This demonstrates the declining importance of the squared deviation y_i^2 for future estimates. At each date, y_i^2 enters with a weight that is lower by a factor of λ, compared to the previous date. If λ is small, the decay in the importance of each squared deviation will be fast. In fact, Equation **??** shows that, if λ is close to zero, the estimate $\hat{\sigma}_{i+1}^2$ is approximately equal to the squared deviation y_i^2—previous squared deviations are relatively unimportant. On the other hand, if λ is close to one, the decay will be slow; i.e., the importance of y_i^2 for the estimate $\hat{\sigma}_{i+2}^2$ will be nearly the same as for $\hat{\sigma}_{i+1}^2$, and nearly the same for $\hat{\sigma}_{i+3}^2$ as for $\hat{\sigma}_{i+2}^2$, etc. This will lead to a smooth (slowly varying) volatility estimate. The slowly varying nature of the estimate in this case is also clear from Equation **??**, because it shows that if λ is close to one, then $\hat{\sigma}_{i+1}^2$ will be approximately the same as $\hat{\sigma}_i^2$.

This method can also be used to estimate covariances, simply by replacing the squared deviations y_i^2 by the product of deviations for two different assets. And, of course, given covariance and variance estimates, we can construct estimates of correlations. To ensure that an estimated correlation is between -1 and $+1$, we will need to use the same λ to estimate each of the variances and the covariance. This is the method used by RiskMetrics.^[See Mina and Xiao (Mina and Xiao 2001), available online at *www.riskmetrics.com].*

5.4 GARCH Models

We are going to adopt a subtle but important change of perspective now. Instead of considering Equation **??** as simply an estimation procedure, we are going to assume that the actual volatility evolves according to Equation **??**, or a generalization thereof. We are also going to reintroduce the expected change in $\log S$, which we dropped in going from Equation **??** to

Equation **??**. Specifically, we return to Equation **??**, but we operate under the risk-neutral measure, so $\mu = r - q$, and we have

$$\log S(t_{i+1}) - \log S(t_i) = \left(r - q - \frac{1}{2}\sigma_{i+1}^2\right)\Delta t + \sigma_{i+1}\Delta B\,. \tag{5.8}$$

We assume the volatility σ_{i+1} between dates t_i and t_{i+1} is given by

$$\sigma_{i+1}^2 = a + b y_i^2 + c\sigma_i^2\,, \tag{5.9}$$

for some constants $a > 0$, $b \geq 0$ and $c \geq 0$, with y_i now defined by

$$y_i = \frac{\log S(t_i) - \log S(t_{i-1}) - \left(r - q - \frac{1}{2}\sigma_i^2\right)\Delta t}{\sqrt{\Delta t}}\,.$$

From Equation **??**, applied to the period from t_{i-1} to t_i, this implies that y_i is normally distributed with mean zero and variance σ_i^2, and of course y_{i+1} has variance σ_{i+1}^2, etc.
Under these assumptions, the random process $\log S$ is called a GARCH(1,1) process.[2] There are many varieties of GARCH processes that have been proposed in the literature, but we will only consider GARCH(1,1), which is the simplest.

We assume $b + c < 1$, in which case we can write the variance equation as a generalization of Equation **??**. Namely, %

$$\begin{aligned} \sigma_{i+1}^2 &= (1-\phi)d + \phi\left[(1-\lambda)y_i^2 + \lambda\sigma_i^2\right]\,, \\ \sigma_{i+1}^2 &= \kappa\theta + (1-\kappa)\left[(1-\lambda)y_i^2 + \lambda\sigma_i^2\right]\,, \end{aligned} \tag{5.10}$$

where $\lambda = c/(b+c)$, $\%\phi = b + c$, and $d = a/(1-b-c)$.
$\kappa = 1 - b - c$, and $\theta = a/(1-b-c)$. Hence, σ_{i+1}^2 is a weighted average with weights κ and $1-\kappa$, of two parts, one being the constant θ and the other being itself a weighted average of y_i^2 and σ_i^2. Whatever the variance might be at time t_i, the variance of y_j at any date t_j far into the future, computed without knowing the intervening $y_{i+1}, y_{i+2}, \ldots$, will be approximately the constant θ. The constant θ is called the unconditional variance, whereas σ_i^2 is the conditional variance of y_i.

To understand the unconditional variance, it is useful to consider the variance forecasting equation. Specifically, we can calculate $E_{t_i}\left[\sigma_{i+n}^2\right]$, which is the estimate made at date t_i of the variance of y_{i+n}; i.e, we estimate the variance without having observed $y_{i+1}, \ldots, y_{i+n-1}$. Note that by definition $E_{t_i}[y_{i+1}^2] = \sigma_{i+1}^2$, so Equation **??** implies

$$\begin{aligned} E_{t_i}\left[\sigma_{i+2}^2\right] &= \kappa\theta + (1-\kappa)\left[(1-\lambda)E_{t_i}[y_{i+1}^2] + \lambda\sigma_{i+1}^2\right] \\ &= \kappa\theta + (1-\kappa)\sigma_{i+1}^2\,. \end{aligned}$$

[2]GARCH is the acronym for Generalized Autoregressive Conditional Heteroskedastic. GARCH(1,1) means that there is only one past y (no y_{i-1}, y_{i-2}, etc.) and one past σ (no σ_{i-1}, σ_{i-2}, etc.) in Equation **??**. See Bollerslev (Bollerslev 1986).

Likewise,

$$E_{t_{i+1}}\left[\sigma_{i+3}^2\right] = \kappa\theta + (1-\kappa)\sigma_{i+2}^2\ ,$$

and taking the expectation at date t_i of both sides of this yields

$$\begin{aligned} E_{t_i}\left[\sigma_{i+3}^2\right] = E_{t_i}\left[E_{t_{i+1}}\left[\sigma_{i+3}^2\right]\right] &= \kappa\theta + (1-\kappa)E_{t_i}\left[\sigma_{i+2}^2\right] \\ &= \kappa\theta + (1-\kappa)\left[\kappa\theta + (1-\kappa)\sigma_{i+1}^2\right] \\ &= \kappa\theta[1+(1-\kappa)] + (1-\kappa)^2\sigma_{i+1}^2\ . \end{aligned}$$

This generalizes to

$$E_{t_i}\left[\sigma_{i+n}^2\right] = \kappa\theta\left[1+(1-\kappa)+\cdots(1-\kappa)^{n-2}\right] + (1-\kappa)^{n-1}\sigma_{i+1}^2\ .$$

Thus, there is decay at rate κ in the importance of the current volatility σ_{i+1}^2 for forecasting the future volatility. Furthermore, as $n\to\infty$, the geometric series

$$1+(1-\kappa)+\cdots(1-\kappa)^{n-2}$$

converges to $1/\kappa$, so, as $n\to\infty$ we obtain

$$E_{t_i}\left[\sigma_{i+n}^2\right] \to \theta\ .$$

This means that our best estimate of the conditional variance, at some date far in the future, is approximately the unconditional variance θ.

The most interesting feature of the volatility equation is that large returns (in absolute value) lead to an increase in the variance and hence are likely to be followed by more large returns (whether positive or negative). This is the phenomenon of volatility clustering, which is quite observable in actual markets. This feature also implies that the distribution of returns will be fat tailed (more technically, leptokurtic). This means that the probability of extreme returns is higher than under a normal distribution with the same standard deviation.[3] It is well documented that daily and weekly returns in most markets have this fat-tailed property.

We can simulate a path of an asset price that follows a GARCH process and the path of its volatility as follows. The following python code produces three columns of data (with headings), the first column being time, the second the asset price, and the third the volatility.

```
import numpy as np
import pandas as pd

def simulating_garch(S, sigma, r, q, dt, N, theta, kappa, lambd):
    """
    Inputs:
```

[3]Conversely, the probability of returns very near the mean must also be higher than under a normal distribution with the same standard deviation—a fat-tailed distribution must also have a relatively narrow peak.

```
    S = initial stock price
    sigma = initial volatility
    r = risk-free rate
    q = dividend yield
    dt = length of each time period (Delta t)
    N = number of time periods
    theta = theta parameter for GARCH
    kappa = kappa parameter for GARCH
    lambd = lambda parameter for GARCH
    """
    LogS = np.log(S)
    Sqrdt = np.sqrt(dt)
    a = kappa * theta
    b = (1 - kappa) * (1 - lambd)
    c = (1 - kappa) * lambd

    time = np.zeros(N + 1)
    stock_price = np.zeros(N + 1)
    volatility = np.zeros(N + 1)

    stock_price[0] = S
    volatility[0] = sigma

    for i in range(1, N + 1):
        time[i] = i * dt
        y = sigma * np.random.randn()
        LogS = LogS + (r - q - 0.5 * sigma * sigma) * dt + Sqrdt * y
        S = np.exp(LogS)
        stock_price[i] = S
        sigma = np.sqrt(a + b * y ** 2 + c * sigma ** 2)
        volatility[i] = sigma

    df_garch = pd.DataFrame({'Time': time, 'Stock Price': stock_price, 'Volatility': volatili
    df_garch.to_csv('garch_simulation.csv', index=False)
    return df_garch

# Example usage:
S = 100       # Initial stock price
sigma = 0.2   # Initial volatility
r = 0.05      # Risk-free rate
q = 0.02      # Dividend yield
dt = 1/252    # Length of each time period (daily)
```

```
N = 252         # Number of time periods (one year)
theta = 0.1     # Theta parameter for GARCH
kappa = 0.1     # Kappa parameter for GARCH
lambd = 0.9     # Lambda parameter for GARCH

df_garch = simulating_garch(S, sigma, r, q, dt, N, theta, kappa, lambd)
print(df_garch)
```

```
         Time  Stock Price  Volatility
0    0.000000   100.000000    0.200000
1    0.003968    99.282331    0.208781
2    0.007937    98.176613    0.219474
3    0.011905    97.453937    0.224193
4    0.015873    98.818864    0.234708
..        ...          ...         ...
248  0.984127    73.661911    0.267850
249  0.988095    75.152398    0.277912
250  0.992063    73.675908    0.285410
251  0.996032    74.431119    0.279929
252  1.000000    75.243496    0.275975

[253 rows x 3 columns]
```

To price European options, we need to compute the usual probabilities $\text{prob}^S(S(T) > K)$ and $\text{prob}^R(S(T) > K)$. Heston and Nandi (Heston and Nandi 2000) provide a fast method for computing these probabilities in a GARCH (1,1) model.[4] Rather than developing this approach, we will show in Chapter **??** how to apply Monte-Carlo methods.

5.5 Stochastic Volatility Models

The volatility is stochastic (random) in a GARCH model, but it is determined by the changes in the stock price. In this section, in contrast, we will consider models in which the volatility depends on a second Brownian motion. The most popular model of this type is the model of Heston (Heston 1993). In this model, we have, as usual,

$$\mathrm{d}\log S = \left(r - q - \frac{1}{2}\sigma^2\right)\,\mathrm{d}t + \sigma\,\mathrm{d}B_s\ , \tag{5.11}$$

[4]Actually, a slightly more general model is considered in (Heston and Nandi 2000), in which large negative returns lead to a greater increase in volatility than do large positive returns. This accommodates the empirically observed negative correlation between stock returns and volatility.

where B_s is a Brownian motion under the risk-neutral measure but now σ is not a constant but instead evolves as $\sigma(t) = \sqrt{v(t)}$ where

$$dv(t) = \kappa\big[\theta - v(t)\big]\,\mathrm{d}t + \gamma\sqrt{v(t)}\,\mathrm{d}B_v\ , \tag{5.12}$$

where B_v is a Brownian motion under the risk-neutral measure having a constant correlation ρ with the Brownian motion B_s. In this equation, κ, θ and γ are positive constants. Given the empirical fact that negative return shocks have a bigger impact on future volatility than do positive shocks, one would expect the correlation ρ to be negative.

The term $\kappa(\theta - v)$ will be positive when $v < \theta$ and negative when $v > \theta$ and hence $\sigma^2 = v$ will tend to drift towards θ, which, as in the GARCH model, is the long-run or unconditional mean of σ^2. Thus, the volatility is said to mean revert. The rate at which it drifts towards θ is obviously determined by the magnitude of κ, also as in the GARCH model.

The specification Equation **??** implies that the volatility of v approaches zero whenever v approaches zero. In this circumstance, one might expect the drift towards θ to dominate the volatility and keep v nonnegative, and this is indeed the case; thus, the definition $\sigma(t) = \sqrt{v(t)}$ is possible. Moreover, the parameter γ plays a role here that is similar to the role of $1-\lambda$ in the GARCH model—the variance of the variance in the GARCH model Equation **??** depends on the weight $1-\lambda$ placed on the scaled return y_i, just as the variance of the variance in the stochastic volatility model Equation **??** depends on the weight γ placed on $\mathrm{d}B_v$.

We could discretize Equation **??** - Equation **??** as:

$$\log S(t_{i+1}) = \log S(t_i) + \left(r - q - \frac{1}{2}\sigma(t_i)^2\right)\,\Delta t + \sqrt{v(t_i)}\,\Delta B_s, \tag{5.13}$$

$$v(t_{i+1}) = v(t_i) + \kappa\big[\theta - v(t_i)\big]\,\Delta t + \gamma\sqrt{v(t_i)}\,\Delta B_v\ . \tag{5.14}$$

However, even though in the continuous-time model Equation **??** - Equation **??** we always have $v(t) \geq 0$ and hence can define $\sigma(t) = \sqrt{v(t)}$, there is no guarantee that $v(t_{i+1})$ defined by Equation **??** will be nonnegative. A simple remedy is to define $v(t_{i+1})$ as the larger of zero and the right-hand side of Equation **??**; thus, we will simulate the Heston model as Equation **??** and[5]

$$v(t_{i+1}) = \max\left\{0, v(t_i) + \kappa[\theta - v(t_i)]\,\Delta t + \gamma\sqrt{v(t_i)}\,\Delta B_v\right\}\ . \tag{5.15}$$

[5] There are better (but more complicated) ways to simulate the Heston model. An excellent discussion of ways to simulate the volatility process can be found in Glasserman (Glasserman 2004). Broadie and Kaya (Broadie and Kaya 2006) present a method for simulating from the exact distribution of the asset price in the Heston model and related models.

A simple way to simulate the changes ΔB_s and ΔB_v in the two correlated Brownian motions is to generate two independent standard normals z_1 and z_2 and take

$$\Delta B_s = \sqrt{\Delta t}\, z \qquad \text{and} \qquad \Delta B_v = \sqrt{\Delta t}\, z^* \ ,$$

where we define

$$z = z_1 \qquad \text{and} \qquad z^* = \rho z_1 + \sqrt{1-\rho^2}\, z_2 \ .$$

The random variable z^* is also a standard normal, and the correlation between z and z^* is ρ.

As in the GARCH model, we can simulate a path of an asset price that follows a GARCH process and the path of its volatility as follows. The following python code produces three columns of data (with headings), the first column being time, the second the asset price, and the third the volatility in the Heston model.

```
import numpy as np
import pandas as pd

def simulating_stochastic_volatility(S, V0, r, q, dt, N, theta, kappa, sigma, rho):
    """
    Inputs:
    S = initial stock price
    V0 = initial variance
    r = risk-free rate
    q = dividend yield
    dt = length of each time period (Delta t)
    N = number of time periods
    theta = long-term variance (mean of variance)
    kappa = rate of mean reversion of variance
    sigma = volatility of variance
    rho = correlation between the two Wiener processes
    """
    LogS = np.log(S)
    Sqrdt = np.sqrt(dt)

    time = np.zeros(N + 1)
    stock_price = np.zeros(N + 1)
    volatility = np.zeros(N + 1)

    stock_price[0] = S
    volatility[0] = V0

    for i in range(1, N + 1):
        time[i] = i * dt
```

```
        Z1 = np.random.randn()
        Z2 = np.random.randn()
        W1 = Z1
        W2 = rho * Z1 + np.sqrt(1 - rho**2) * Z2

        LogS = LogS + (r - q - 0.5 * volatility[i-1]**2) * dt + np.sqrt(volatility[i-1]**2 *
        S = np.exp(LogS)
        stock_price[i] = S

        volatility[i] = np.sqrt(np.maximum(volatility[i-1]**2 + kappa * (theta - volatility[i

    df_stochastic_vol = pd.DataFrame({'Time': time, 'Stock Price': stock_price, 'Volatility'
    df_stochastic_vol.to_csv('stochastic_volatility_simulation.csv', index=False)
    return df_stochastic_vol

# Example usage:
S = 100       # Initial stock price
V0 = 0.04     # Initial variance
r = 0.05      # Risk-free rate
q = 0.02      # Dividend yield
dt = 1/252    # Length of each time period (daily)
N = 252       # Number of time periods (one year)
theta = 0.04  # Long-term variance
kappa = 2.0   # Rate of mean reversion of variance
sigma = 0.3   # Volatility of variance
rho = -0.7    # Correlation between the two Wiener processes

df_stochastic_vol = simulating_stochastic_volatility(S, V0, r, q, dt, N, theta, kappa, sigma
print(df_stochastic_vol)
```

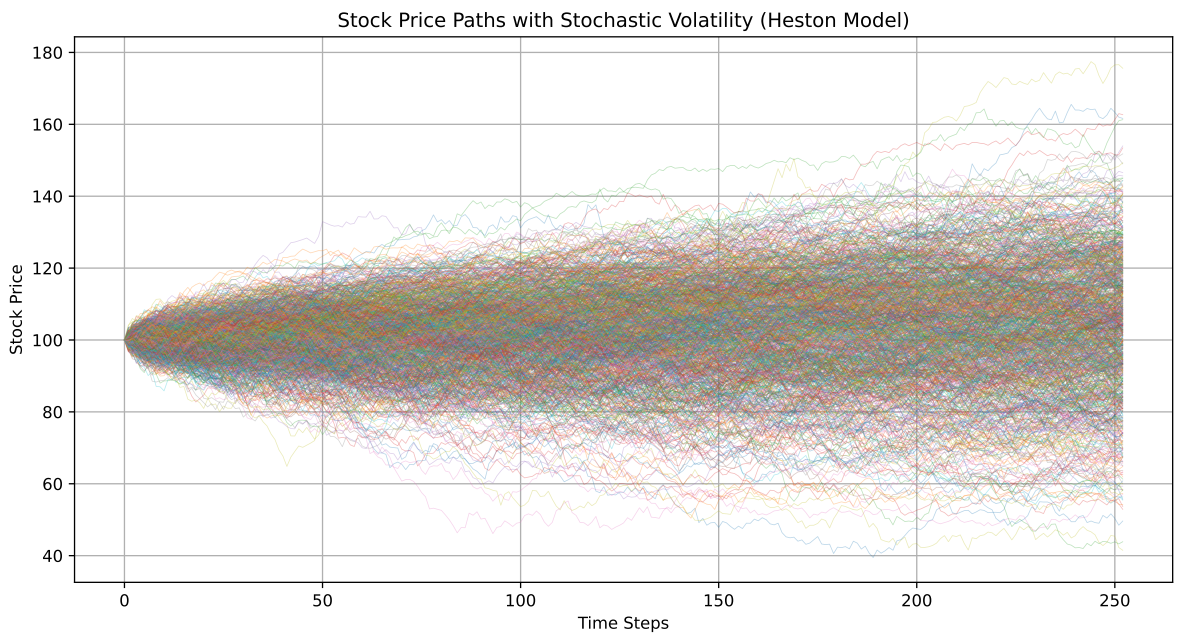
```
         Time  Stock Price  Volatility
0    0.000000   100.000000    0.040000
1    0.003968    99.913230    0.050444
2    0.007937    99.234913    0.068478
3    0.011905    99.376629    0.068040
4    0.015873    99.059137    0.073774
..        ...          ...         ...
248  0.984127   103.620464    0.090324
249  0.988095   103.155141    0.093423
250  0.992063   103.562432    0.090299
251  0.996032   104.946557    0.069506
252  1.000000   104.805247    0.070421
```

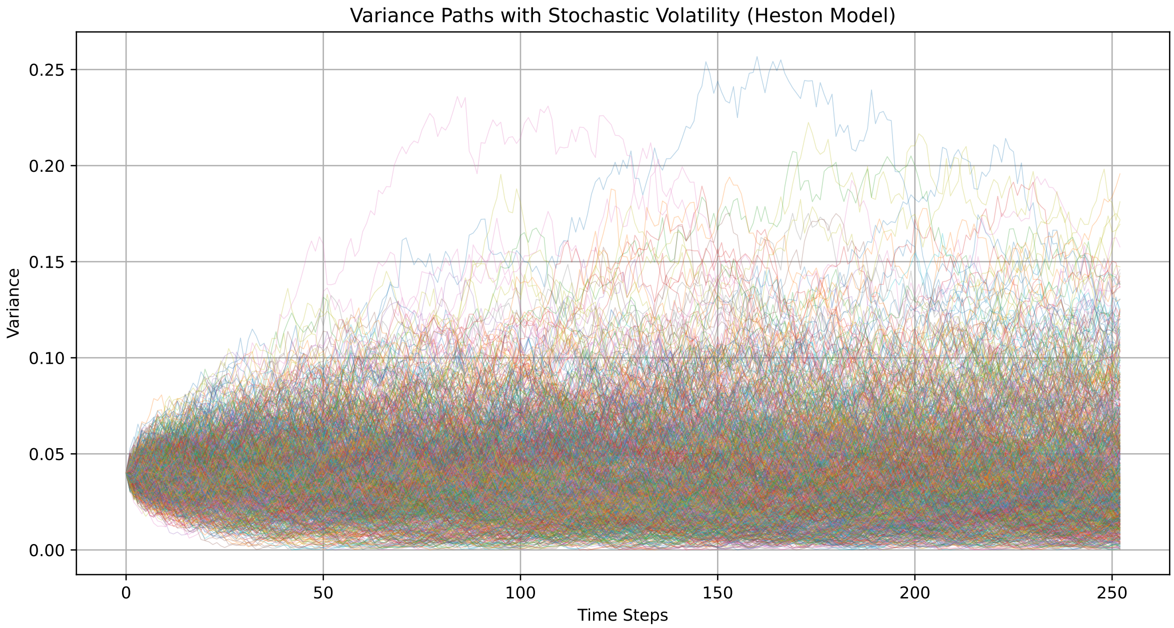
[253 rows x 3 columns]

The following code plots the simulated stock price and the variance paths.

```
import numpy as np
import matplotlib.pyplot as plt

# Heston model parameters
S0 = 100     # Initial stock price
V0 = 0.04    # Initial variance
rho = -0.7   # Correlation between the two Wiener processes
kappa = 2.0  # Rate of mean reversion of variance
theta = 0.04 # Long-term variance
sigma = 0.3  # Volatility of variance
r = 0.05     # Risk-free rate
T = 1.0      # Time to maturity
N = 252      # Number of time steps
dt = T / N   # Time step size
n_simulations = 1000  # Number of simulations

# Preallocate arrays
S = np.zeros((N+1, n_simulations))
V = np.zeros((N+1, n_simulations))
S[0] = S0
V[0] = V0

# Simulate the paths
for t in range(1, N+1):
    Z1 = np.random.normal(size=n_simulations)
    Z2 = np.random.normal(size=n_simulations)
    W1 = Z1
    W2 = rho * Z1 + np.sqrt(1 - rho**2) * Z2

    V[t] = np.maximum(V[t-1] + kappa * (theta - V[t-1]) * dt + sigma * np.sqrt(V[t-1] * dt)
    S[t] = S[t-1] * np.exp((r - 0.5 * V[t-1]) * dt + np.sqrt(V[t-1] * dt) * W1)

# Plot the results
plt.figure(figsize=(12, 6))
for i in range(n_simulations):
    plt.plot(S[:, i], lw=0.5, alpha=0.3)
plt.title('Stock Price Paths with Stochastic Volatility (Heston Model)')
```

```
plt.xlabel('Time Steps')
plt.ylabel('Stock Price')
plt.grid(True)
plt.show()

# Plot the volatility paths
plt.figure(figsize=(12, 6))
for i in range(n_simulations):
    plt.plot(V[:, i], lw=0.5, alpha=0.3)
plt.title('Variance Paths with Stochastic Volatility (Heston Model)')
plt.xlabel('Time Steps')
plt.ylabel('Variance')
plt.grid(True)
plt.show()
```



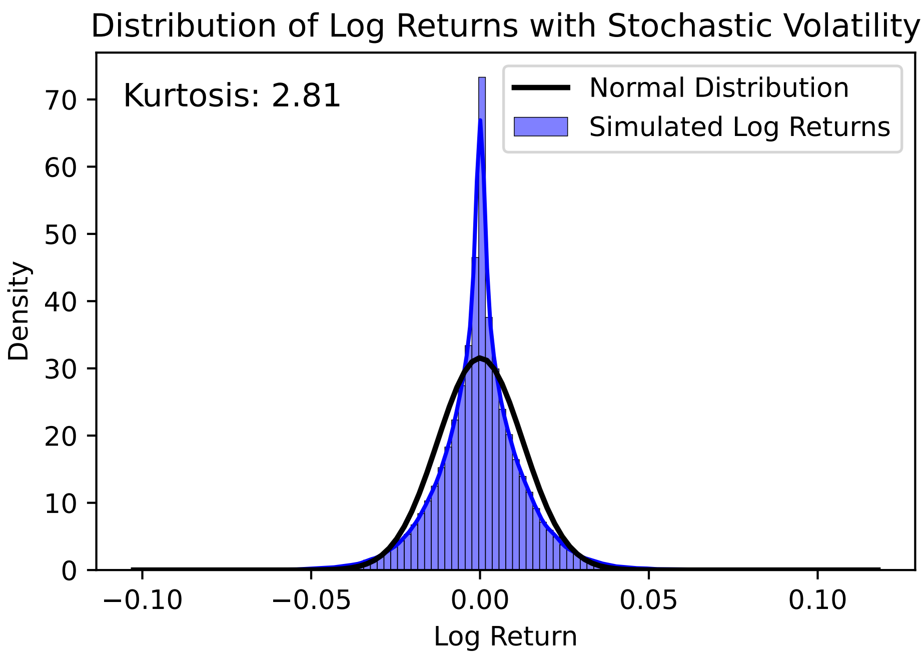


The following code shows that the stock returns under the stochastic volatility model display fat-tails with positive kurtosis.

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import norm, kurtosis

def simulating_stochastic_volatility(S, V0, r, q, dt, N, theta, kappa, sigma, rho, n_simulati
    """
    Inputs:
    S = initial stock price
    V0 = initial variance
    r = risk-free rate
    q = dividend yield
    dt = length of each time period (Delta t)
    N = number of time periods
    theta = long-term variance (mean of variance)
    kappa = rate of mean reversion of variance
    sigma = volatility of variance
    rho = correlation between the two Wiener processes
    n_simulations = number of simulations
    """
```

```
    LogS = np.log(S)
    Sqrdt = np.sqrt(dt)

    log_returns = []

    for _ in range(n_simulations):
        stock_price = S
        variance = V0
        for _ in range(N):
            Z1 = np.random.randn()
            Z2 = np.random.randn()
            W1 = Z1
            W2 = rho * Z1 + np.sqrt(1 - rho**2) * Z2

            log_return = (r - q - 0.5 * variance) * dt + np.sqrt(variance * dt) * W1
            LogS = np.log(stock_price) + log_return
            stock_price = np.exp(LogS)

            variance = np.maximum(variance + kappa * (theta - variance) * dt + sigma * np.sq

            log_returns.append(log_return)

    return log_returns

# Example usage:
S = 100       # Initial stock price
V0 = 0.04     # Initial variance
r = 0.05      # Risk-free rate
q = 0.02      # Dividend yield
dt = 1/252    # Length of each time period (daily)
N = 252       # Number of time periods (one year)
theta = 0.04  # Long-term variance
kappa = 0.2   # Rate of mean reversion of variance
sigma = 0.3   # Volatility of variance
rho = -0.7    # Correlation between the two Wiener processes
n_simulations = 1000  # Number of simulations

log_returns = simulating_stochastic_volatility(S, V0, r, q, dt, N, theta, kappa, sigma, rho,

# Plotting the distribution of log returns
sns.histplot(log_returns, bins=100, kde=True, stat="density", color="blue", label="Simulated
xmin, xmax = plt.xlim()
```

```
x = np.linspace(xmin, xmax, 100)
p = norm.pdf(x, np.mean(log_returns), np.std(log_returns))
plt.plot(x, p, 'k', linewidth=2, label="Normal Distribution")
plt.title('Distribution of Log Returns with Stochastic Volatility')
plt.xlabel('Log Return')
plt.ylabel('Density')
plt.legend()

# Display kurtosis
kurt = kurtosis(log_returns)
plt.figtext(0.15, 0.8, f'Kurtosis: {kurt:.2f}', fontsize=12)

plt.show()
```



To price European options, we again need to compute

$$\text{prob}^S(S(T) > K) \qquad \text{and} \qquad \text{prob}^R(S(T) > K) \; .$$

The virtue of modelling volatility as in Equation **??** is that these probabilities can be computed quite efficiently, as shown by Heston (Heston 1993).[6] There are many other ways in which one could model volatility, but the computations may be more difficult. For example, one could

[6]Further discussion can be found in Epps (Epps 2000).

replace Equation **??** by

$$\sigma(t) = \mathrm{e}^{v(t)} \quad \text{and} \quad dv(t) = \kappa(\theta - v(t))\,\mathrm{d}t + \lambda\,\mathrm{d}B^* \,. \tag{5.16}$$

This implies a lognormal volatility and is simpler to simulate than Equation **??**—because e^v is well defined even when v is negative—but it is easier to calculate the probabilities $\text{prob}^S(S(T) > K)$ and $\text{prob}^R(S(T) > K)$ if we assume Equation **??**.

One way to implement the GARCH or stochastic volatility model is to imply both the initial volatility $\sigma(0)$ and the constants κ, θ and λ or κ, θ, γ and ρ from observed option prices. These four (or five) constants can be computed by forcing the model prices of four (or five) options to equal the observed market prices. Or, a larger set of prices can be used and the constants can be chosen to minimize the average squared error or some other measure of goodness-of-fit between the model and market prices.

5.6 Jump Diffusion Models

In the classical Black-Scholes model, stock prices are assumed to follow a geometric Brownian motion, which is characterized by continuous paths and normally distributed returns. While this model has been widely used due to its simplicity and analytical tractability, it fails to capture certain empirical phenomena observed in financial markets, such as sudden and significant price changes (jumps) and the heavy tails of return distributions.

To address these shortcomings, the jump diffusion model was introduced by Robert C. Merton in 1976. This model extends the Black-Scholes framework by incorporating jumps into the stock price dynamics, thereby allowing for discontinuous price paths. The jump diffusion model is better suited to describe the behavior of financial assets that exhibit sudden price changes due to news, earnings announcements, or other market events.

In a jump diffusion model, the stock price S_t is governed by the following stochastic differential equation (SDE):

$$dS_t = \mu S_t\,dt + \sigma S_t\,dW_t + S_t\,dJ_t$$

where μ is the drift rate, σ is the volatility, B_t is a standard Brownian motion, J_t is a jump process.

The jump process J_t is typically modeled as a compound Poisson process:

$$J_t = \sum_{i=1}^{N_t}(Y_i - 1)$$

where N_t is a Poisson process with intensity λ, Y_i are i.i.d. random variables representing the relative jump sizes, with $Y_i - 1$ being the actual jump size.

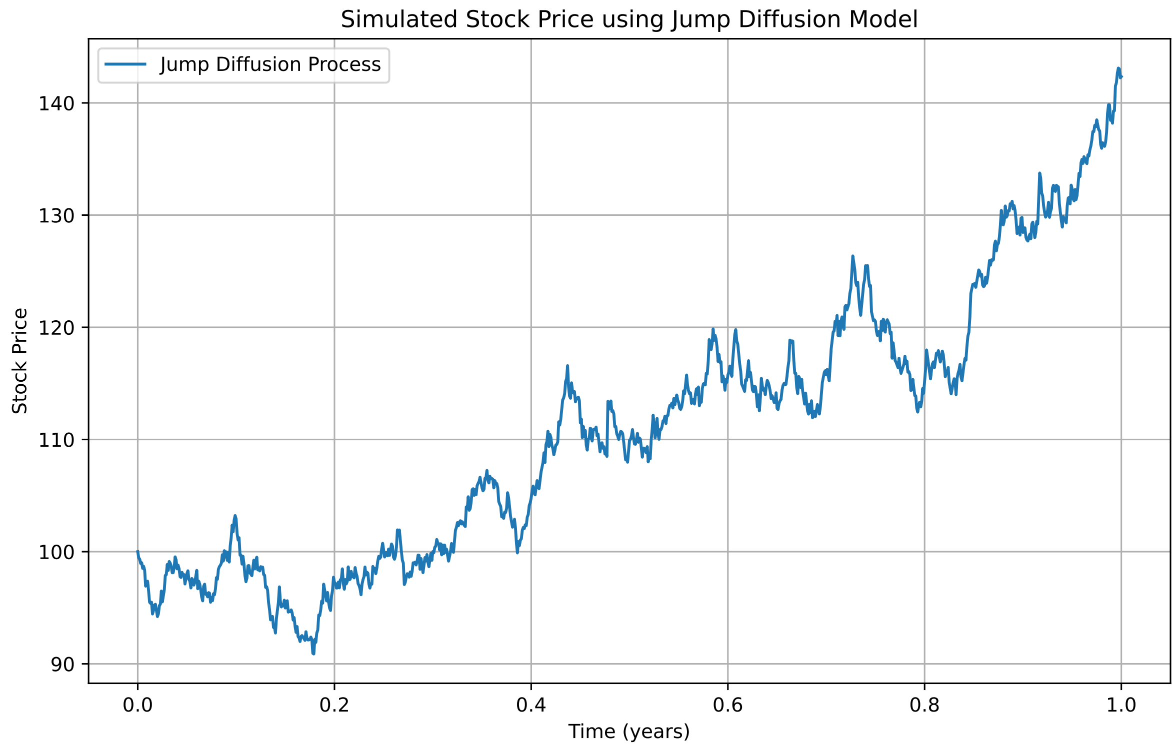
The jump diffusion model introduces several important implications for the behavior of stock prices and the pricing of derivative securities:

1. **Heavy Tails**: The inclusion of jumps leads to a return distribution with heavier tails compared to the normal distribution, aligning better with empirical observations.
2. **Volatility Smile**: The model can generate implied volatility smiles, where implied volatility varies with strike price and maturity, a feature commonly observed in market data.
3. **Risk Management**: Understanding the jump component is crucial for risk management, as it affects the likelihood of extreme price movements and the potential for large losses.

The following python code simulates the stock price that evolves as a jump diffusion.

```python
import numpy as np
import matplotlib.pyplot as plt

def simulate_jump_diffusion(S0, mu, sigma, lamb, m, delta, T, N):
    """
    Simulate a jump diffusion process.

    Parameters:
    S0     : float - initial stock price
    mu     : float - drift rate
    sigma  : float - volatility
    lamb   : float - intensity of the Poisson process
    m      : float - mean of the jump size distribution
    delta  : float - standard deviation of the jump size distribution
    T      : float - total time
    N      : int   - number of time steps

    Returns:
    t      : numpy array - time points
    S      : numpy array - simulated stock prices
    """
    dt = T / N
    t = np.linspace(0, T, N + 1)
    S = np.zeros(N + 1)
    S[0] = S0

    for i in range(1, N + 1):
        Z = np.random.normal(0, 1)  # Normal random variable for the diffusion part
```

```
        J = np.random.poisson(lamb * dt)  # Poisson random variable for jumps

        # Sum of log-normal distributed jumps
        Y = np.sum(np.random.normal(m, delta, J))

        # Update stock price
        S[i] = S[i - 1] * np.exp((mu - 0.5 * sigma ** 2) * dt + sigma * np.sqrt(dt) * Z + Y)

    return t, S

# Parameters
S0 = 100          # Initial stock price
mu = 0.1          # Drift rate
sigma = 0.2       # Volatility
lamb = 0.75       # Intensity of the Poisson process (average number of jumps per unit time)
m = 0.02          # Mean of the jump size distribution (log-normal)
delta = 0.1       # Standard deviation of the jump size distribution (log-normal)
T = 1.0           # Total time (1 year)
N = 1000          # Number of time steps

# Simulate the jump diffusion process
t, S = simulate_jump_diffusion(S0, mu, sigma, lamb, m, delta, T, N)

# Plot the simulated stock prices
plt.figure(figsize=(10, 6))
plt.plot(t, S, label='Jump Diffusion Process')
plt.title('Simulated Stock Price using Jump Diffusion Model')
plt.xlabel('Time (years)')
plt.ylabel('Stock Price')
plt.legend()
plt.grid(True)
plt.show()
```



The jump diffusion model offers a more realistic framework for modeling stock prices by incorporating the possibility of sudden jumps. This enhancement over the classical Black-Scholes model allows for better capturing the empirical characteristics of financial markets, thereby improving the accuracy of option pricing and risk management practices.

5.7 Smiles and Smirks Again

As mentioned before, the GARCH, stochastic volatility and jump diffusion models can generate fat-tailed distributions for the asset price $S(T)$. Thus, they can be more nearly consistent with the option smiles discussed in Section **??** than is the Black-Scholes model (though it appears that one must include jumps in asset prices as well as stochastic volatility in order to duplicate market prices with an option pricing formula). To understand the relation, let σ_{am} denote the implied volatility from an at-the-money call option, i.e., a call option with strike $K = S(0)$. The characteristic of a smile is that implied volatilities from options of the same maturity with strike prices significantly above and below $S(0)$ are higher than σ_{am}.

A strike price higher than $S(0)$ corresponds to an out-of-the money call option. The high implied volatility means that the market is pricing the right to buy at $K > S(0)$ above the Black-Scholes price computed from the volatility σ_{am}; thus, the market must attach a higher probability to stock prices $S(T) > S(0)$ than the volatility σ_{am} would suggest.

A strike price lower than $S(0)$ corresponds to an in-the-money call option. The put option with the same strike is out of the money. The high implied volatility means that the market is pricing call options above the Black-Scholes price computed from the volatility σ_{am}. By put-call parity, the market must also be pricing put options above the Black-Scholes price computed from the volatility σ_{am}. The high prices for the rights to buy and sell at $K < S(0)$ means that the market must attach a higher probability to stock prices $S(T) < S(0)$ than the volatility σ_{am} would suggest. In particular, the high price for the right to sell at $K < S(0)$ means a high insurance premium for owners of the asset who seek to insure their positions, which is consistent with a market view that there is a significant probability of a large loss. This can be interpreted as a crash premium. Indeed, the implied volatilities at strikes less than $S(0)$ are typically higher than the implied volatilities at strikes above $S(0)$ (giving the smile the appearance of a smirk, as discussed in Section **??**), which is consistent with a larger probability of crashes than of booms (a fatter tail for low returns than for high).

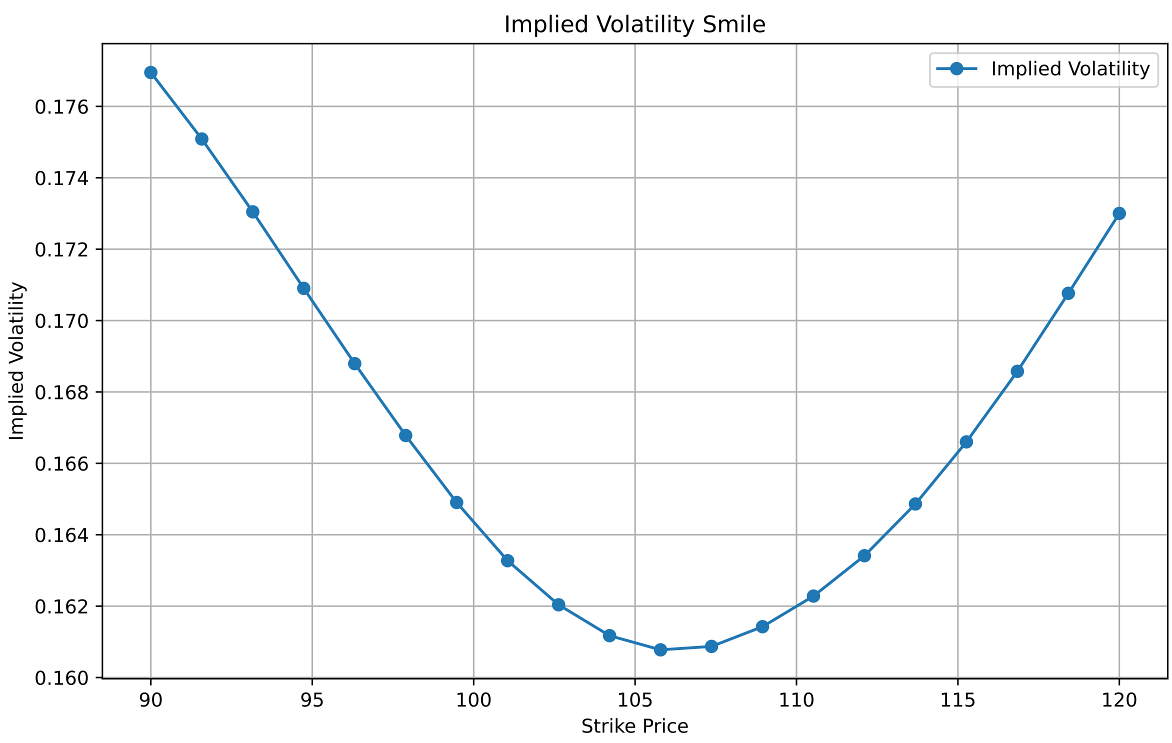
As an example, the following code shows that the stochstic volatility model can generate implied volatility smiles.

This program involes (1) Simulating Heston Model: simulate_heston_paths function generates stock price paths using the Heston model parameters. (2) Calculating call price by discounting the averaged call payoffs across the stock price sample paths (3) Calculating Black Scholes call price : black_scholes_call_price function calculates the call option price using the Black-Scholes formula. (4) Calculating implied volatility: implied_volatility function computes the implied volatility by solving for the volatility that matches the Black-Scholes call price to the simulated call price. (5) Repeating Steps (2)-(4) for different strik prices. (6) Plotting the implied volatility against strike prices fixing the initial stock price at $100.

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm
from scipy.optimize import brentq

# Heston model parameters
S0 = 100      # Initial stock price
V0 = 0.04     # Initial variance
r = 0.05      # Risk-free rate
q = 0.01      # Dividend yield
T = 1         # Time to maturity (in years)
kappa = 0.25   # Rate of mean reversion of variance
theta = 0.04  # Long-term variance
sigma = 0.5   # Volatility of variance
rho = -0.2    # Correlation between the two Wiener processes
dt = 1/252    # Length of each time period (daily)
N = 252       # Number of time periods (one year)
n_simulations = 10000  # Number of simulations
```

```
def simulate_heston_paths(S0, V0, r, q, T, kappa, theta, sigma, rho, dt, N, n_simulations):
    S = np.zeros((N + 1, n_simulations))
    V = np.zeros((N + 1, n_simulations))
    S[0] = S0
    V[0] = V0

    for t in range(1, N + 1):
        Z1 = np.random.normal(size=n_simulations)
        Z2 = np.random.normal(size=n_simulations)
        W1 = Z1
        W2 = rho * Z1 + np.sqrt(1 - rho**2) * Z2

        V[t] = np.maximum(V[t-1] + kappa * (theta - V[t-1]) * dt + sigma * np.sqrt(V[t-1] * d
        S[t] = S[t-1] * np.exp((r - q - 0.5 * V[t-1]) * dt + np.sqrt(V[t-1] * dt) * W1)

    return S, V

def black_scholes_call_price(S, K, T, r, sigma):
    d1 = (np.log(S / K) + (r + 0.5 * sigma**2) * T) / (sigma * np.sqrt(T))
    d2 = d1 - sigma * np.sqrt(T)
    call_price = S * norm.cdf(d1) - K * np.exp(-r * T) * norm.cdf(d2)
    return call_price

def implied_volatility(C, S, K, T, r):
    def objective_function(sigma):
        return black_scholes_call_price(S, K, T, r, sigma) - C
    try:
        return brentq(objective_function, 0.001, 5.0)
    except ValueError:
        return np.nan

# Simulate paths using the Heston model
S_paths, _ = simulate_heston_paths(S0, V0, r, q, T, kappa, theta, sigma, rho, dt, N, n_simula

# Calculate option prices at different strike prices
strike_prices = np.linspace(90, 120, 20)
call_prices = np.zeros_like(strike_prices)

for i, K in enumerate(strike_prices):
    call_payoffs = np.maximum(S_paths[-1] - K, 0)
    call_prices[i] = np.mean(call_payoffs) * np.exp(-r * T)
```

```
# Calculate implied volatilities
implied_vols = [implied_volatility(C, S0, K, T, r) for C, K in zip(call_prices, strike_price

# Filter out NaN values that may occur
valid_indices = ~np.isnan(implied_vols)
strike_prices = strike_prices[valid_indices]
implied_vols = np.array(implied_vols)[valid_indices]

# Plot the implied volatility smile
plt.figure(figsize=(10, 6))
plt.plot(strike_prices, implied_vols, label='Implied Volatility', marker='o')
plt.title('Implied Volatility Smile')
plt.xlabel('Strike Price')
plt.ylabel('Implied Volatility')
plt.legend()
plt.grid(True)
plt.show()
```



5.8 Hedging and Market Completeness

The GARCH model is inherently a discrete-time model. If returns have a GARCH structure at one frequency (e.g., monthly), they will not have a GARCH structure at a different frequency (e.g., weekly). Hence, the return period (monthly, weekly, ...) is part of the specification of the model. One interpretation of the model is that the dates t_i at which the variance changes are the only dates at which investors can trade. Under this interpretation, it is impossible to perfectly hedge an option: the gross return $S(t_i)/S(t_{i-1})$ over the interval (t_{i-1}, t_i) is lognormally distributed, so no portfolio of the stock and riskless asset formed at t_{i-1} and held over the interval (t_{i-1}, t_i) can perfectly replicate the return of an option over the interval. As discussed in Section **??**, we call a market in which some derivatives cannot be perfectly hedged an incomplete market. Thus, the GARCH model is an example of an incomplete market, if investors can only trade at the frequency at which returns have a GARCH structure. However, it is unreasonable to assume that investors can only trade weekly or monthly or even daily.

Another interpretation of the GARCH model is that investors can trade continuously and the asset has a constant volatility within each period (t_{i-1}, t_i). Under this interpretation, the market is complete and options can be delta-hedged. The completeness is a result of the fact that the change $\sigma_{i+1} - \sigma_i$ in the volatility at date t_i (recall that σ_i is the volatility over the period (t_{i-1}, t_i) and σ_{i+1} is the volatility over the period (t_i, t_{i+1})) depends only on $\log S(t_i)$. Thus, the only random factor in the model that needs to be hedged is, as usual, the underlying asset price. However, this interpretation of the model is also a bit strange. Suppose for example that monthly returns are assumed to have a GARCH structure. Then the model states that the volatility in February will be higher if there is an unusually large return (in absolute value) in January. Suppose there is an unusually large return in the first half of January. Then, intuitively, one would expect the change in the volatility to occur in the second half of January rather than being delayed until February. However, the model specifies that the volatility is constant during each month, hence constant during January in this example.

The stochastic volatility model is more straightforward. The market is definitely incomplete. The value of a call option at date $t < T$, where T is the maturity of the option, will depend on the underlying asset price $S(t)$ and the volatility $\sigma(t)$. Denoting the value by $C(t, S(t), \sigma(t))$, we have from Ito's formula that

$$\mathrm{d}C(t) = \text{something}\,\mathrm{d}t + \frac{\partial C}{\partial S}\,\mathrm{d}S(t) + \frac{\partial C}{\partial \sigma}\,\mathrm{d}\sigma(t)\ .$$

A replicating portfolio must have the same dollar change at each date t. If we hold $\partial C/\partial S$ shares of the underlying asset, then the change in the value of the shares will be $(\partial C/\partial S)\,\mathrm{d}S$. However, there is no way to match the
$(\partial C/\partial \sigma)\,\mathrm{d}\sigma$ term using the underlying asset and the riskless asset.

The significance of the market being incomplete is that the value of a derivative asset that cannot be replicated using traded assets (e.g., the underlying and riskless assets) is not uniquely determined by arbitrage considerations. As discussed in Section **??**, one must use equilibrium pricing in this circumstance. That is what we have implicitly done in this chapter. By assuming particular dynamics for the volatility under the risk-neutral measure, we have implicitly selected a particular risk-neutral measure from the set of risk-neutral measures that are consistent with the absence of arbitrage.

5.9 Exercises

Exercise 5.1. The purpose of this exercise is to generate a fat-tailed distribution from a model that is simpler than the GARCH and stochastic volatility models but has somewhat the same flavor. The distribution will be a mixture of normals. Create a python program in which the user can input S, r, q, T, σ_1 and σ_2. Use these inputs to produce a column of 500 simulated $\log S(T)$. In each simulation, define $\log S(T)$ as

$$\log S(T) = \log S(0) + \left(r - q - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T} z \, ,$$

where z is a standard normal, $\sigma = x\sigma_1 + (1-x)\sigma_2$, and x is a random variable that equals zero or one with equal probabilities.

Calculate the mean and standard deviation of the $\log S(T)$ and calculate the fraction that lie more than two standard deviations below the mean. If the $\log S(T)$ all came from a normal distribution with the same variance, then this fraction should equal $\mathrm{N}(-2) = 2.275\%$. If the fraction is higher, then the distribution is fat tailed. (Of course, the actual fraction would differ from 2.275% in any particular case due to the randomness of the simulation, even if all of the $\log S(T)$ came from a normal distribution with the same variance).

Exercise 5.2. Create a python program prompting the user to input the same inputs as in the `simulating_garch` function except for the initial volatility and θ. Simulate 500 paths of a GARCH process and output $\log S(T)$ for each simulation (you don't need to output the entire paths as in the `simulating_garch` function). Take the initial volatility to be 0.3 and $\theta = 0.09$. Determine whether the distribution is fat-tailed by computing the fraction of the $\log S(T)$ that lie two or more standard deviations below the mean, as in the previous exercise. For what values of κ and λ does the distribution appear to be especially fat-tailed?

Exercise 5.3. Repeat Exercise **??** for the Heston stochastic volatility model, describing the values of κ, γ and ρ that appear to generate especially fat-tailed distributions.

6 Monte Carlo and Binomial Models

In this chapter, we will introduce two principal numerical methods for valuing derivative securities: Monte Carlo and binomial models. We will consider two applications: valuing European options in the presence of stochastic volatility with Monte Carlo and valuing American options via binomial models. Additional applications of these methods will be presented in **?@sec-c_montecarlo**. Throughout the chapter, we will assume there is a constant risk-free rate. The last section, while quite important, could be skimmed on first reading—the rest of the book does not build upon it.

6.1 Introduction to Monte Carlo

According to our risk-neutral pricing Equation **??**, the value of a security paying an amount x at date T is

$$\mathrm{e}^{-rT} E^R[x] \; . \tag{6.1}$$

To estimate this by Monte-Carlo means to simulate a sample of values for the random variable x and to estimate the expectation by averaging the sample values.[1]. Of course, for this to work, the sample must be generated from a population having a distribution consistent with the risk-neutral probabilities.

The simplest example is valuing a European option under the Black-Scholes assumptions. Of course, for calls and puts, this is redundant, because we already have the Black-Scholes formulas. Nevertheless, we will describe how to do this for the sake of introducing the Monte Carlo method. In the case of a call option, the random variable x in Equation **??** is $\max(0, S(T)-K)$. To simulate a sample of values for this random variable, we need to simulate the terminal stock price $S(T)$. This is easy to do, because, under the Black-Scholes assumptions, the logarithm of $S(T)$ is normally distributed under the risk-neutral measure with mean $\log S(0) + \mathrm{n}uT$ and variance $\sigma^2 T$, where $\mathrm{n}u = r - q - \sigma^2/2$. Thus, we can simulate values for $\log S(T)$ as $\log S(0) + \mathrm{n}uT + \sigma\sqrt{T}z$, where z is a standard normal. We can average the simulated values of $\max(0, S(T) - K)$, or whatever the payoff of the derivative is, and then discount at the risk-free rate to compute the date–0 value of the derivative. This means that we generate

[1] Boyle~(Boyle 1977) introduced Monte-Carlo methods for derivative valuation, including the variance-reduction methods of control variates and antithetic variates to be discussed in **?@sec-c_montecarlo**

some number M of standard normals z_i and estimate the option value as $\mathrm{e}^{-rT}\bar{x}$, where $\bar{x}$ is the mean of

$$x_i = \max\left(0, \mathrm{e}^{\log S(0)+\mathrm{nu}T+\sigma\sqrt{T}z_i} - K\right)\;.$$

To value options that are path-dependent we need to simulate the path of the underlying asset price. Path-dependent options are discussed in Chaps.~**??** and~**??**.

There are two main drawbacks to Monte-Carlo methods. First, it is difficult (though not impossible) to value early-exercise features.[2] To value early exercise, we need to know the value at each date if not exercised, to compare to the intrinsic value. One could consider performing a simulation at each date to calculate the value if not exercised, but this value depends on the option to exercise early at later dates, which cannot be calculated without knowing the value of being able to exercise early at even later dates, etc. In contrast, the binomial model (and finite difference models discussed in Chapter **??**) can easily handle early exercise but cannot easily handle path dependencies.

The second drawback of Monte Carlo methods is that they can be quite inefficient in terms of computation time (though, as will be explained in **?@sec-c_montecarlo**, they may be faster than alternative methods for derivatives written on multiple assets). As in statistics, the standard error of the estimate depends on the sample size. Specifically, we observed in Section **??** that, given a random sample $\{x_1, \ldots, x_M\}$ of size M from a population with mean μ and variance σ^2, the best estimate of μ is the sample mean $\bar{x}$, and the standard error of $\bar{x}$ (which means the standard deviation of $\bar{x}$ in repeated samples) is best estimated by

$$\sqrt{\frac{1}{M(M-1)}\left(\sum_{i=1}^{M} x_i^2 - M\bar{x}^2\right)}\;. \tag{6.2}$$

Recall that $\bar{x}$ plus or minus 1.96 standard errors is a 95% confidence interval for μ when the x_i are normally distributed. In the context of European option valuation, the expression Equation **??** gives the standard error of the estimated option value at maturity, and multiplication of Equation **??** by e^{-rT} gives the standard error of the estimated date–0 option value.

To obtain an estimate with an acceptably small standard error may require a large sample size and hence a relatively large amount of computation time. The complexities of Monte Carlo methods arise from trying to reduce the required sample size. In **?@sec-c_montecarlo**, we will describe two such methods (antithetic variates and control variates). For those who want to engage in a more detailed study of Monte Carlo methods, the book of Glasserman (Glasserman 2004) is highly recommended. J"ackel (Jäckel 2002) is useful for more advanced readers, and Clewlow and Strickland (Clewlow and Strickland 1998) and Brandimarte (Brandimarte 2002) are useful references that include computer code.

[2] Monte-Carlo methods for valuing early exercise include the stochastic mesh method of Broadie and Glasserman (Broadie and Glasserman 1997) and the regression method of Longstaff and Schwartz (Longstaff and Schwartz 2001). Glasserman (Glasserman 2004) provides a good discussion of these methods and the relation between them.

6.1.1 Monte Carlo Valuation of a European Call

We will illustrate Monte Carlo by valuing a European call under the Black-Scholes assumptions. We will also estimate the delta by each of the methods described in Section **??** and~**??**. Of course, we know the call value and its delta from the Black-Scholes formulas, and they can be used to evaluate the accuracy of the Monte Carlo estimates. We use the code in Chapter (**#sec-c_continuoustime?**). In this circumstance, we only need to simulate the price of the underlying at the option maturity rather than the entire path of the price process. Therefore we set $m = 1$. However, we use a large number of paths, $n = 10000$ to get a large sample of terminal stock prices.

```
# Simulate Geometric Brownian Motion
import numpy as np
import matplotlib.pyplot as plt
# number of paths
n = 10000
#number of divisions
m = 1
# Interest rate (We set the drift equal to the interest rate for the risk neutral measure)
r = 0.1
# Volatility
sig = 0.2
# Initial Stock Price
S0 = 42
# Maturity
T = 0.5
#Strike Price
K=40
# Dividend Yield
q=0.0
# Delta t
dt = T/m
# Drift
drift = (r-q-0.5*sig**2)
# Volatility
vol = sig * np.sqrt(dt)

t = np.array(range(0,m + 1,1)) * dt

# seed for random generator
seed= 2020
# define a random generator
np.random.seed(seed)
```

```
inc = np.zeros(shape = (m + 1, n))
inc[1:] = np.transpose(np.random.normal(loc = 0, scale = vol,size = (n,m)))
St = np.zeros(shape = (m + 1, n))
St = S0 * np.exp(np.cumsum(inc,axis=0) + (drift * t[0:m + 1])[:,None])
St1 = S0 * np.exp(-np.cumsum(inc,axis=0) + (drift * t[0:m + 1])[:,None])
```

As before, this code generates two samples St, which adds the simulated standard (zero mean) normal random variable, and $St1$ which subtracts the simulated (zero mean) standard normal random variable. Each sample produces and estimate for the Black-Scholes European call option.

```
cc=np.maximum(St[m,:]-K,0)
cp = np.mean(cc) * np.exp(-r * T)
cc1=np.maximum(St1[m,:]-K,0)*np.exp(-r * T)
cp1= np.mean(np.maximum(St1[m,:]-K,0)) * np.exp(-r * T)

print('The first sample gives an estimated call price=',cp)
print('The second sample gives an estimated call price=',cp1)
bsc = (cp+cp1)/2
print('The average of the two estimates=',bsc)
```

```
The first sample gives an estimated call price= 4.791646287615179
The second sample gives an estimated call price= 4.687624646438364
The average of the two estimates= 4.739635467026771
```

The true call price is given by

```
from scipy import stats
import numpy as np
from scipy.optimize import minimize, minimize_scalar

def blackscholes(S0, K, r, q, sig, T, call = True):
    '''Calculate option price using B-S formula.

    Args:
        S0 (num): initial price of underlying asset.
        K (num): strick price.
        r (num): risk free rate
        q (num): dividend yield
        sig (num): Black-Scholes volatility.
        T (num): maturity.
```

```
        call (bool): True returns call price, False returns put price.

    Returns:
        num
    '''
    d1 = (np.log(S0/K) + (r -q + sig**2/2) * T)/(sig*np.sqrt(T))
    d2 = d1 - sig*np.sqrt(T)
#    <!-- norm = sp.stats.norm -->
    norm = stats.norm
    if call:
        return np.exp(- q *T) * S0 * norm.cdf(d1,0,1) - K * np.exp(-r * T) * norm.cdf(d2,0, 
    else:
        return -np.exp(-q * T) * S0 * norm.cdf(-d1,0,1) + K * np.exp(-r * T) * norm.cdf(-d2,

truebsc=blackscholes(S0, K, r, q, sig, T, call = True)
print('The black scholes fromula=',truebsc)
```

```
The black scholes fromula= 4.759422392871532
```

Notice that even with 10000 data points for each sample the individual estimates are not very accurate compared to the exact Black Scoles price. This is a well known problem that is difficult to estimate the mean, even with a lot of data and is a drawback to Monte Carlo as discussed earlier. However, the average of the two prices is sgnificantly more accurate. This is an example of an antithetic variable which is discussed later. One simple intution is the two samples yield negatively correlated errors; if the plus sample is two high, then the minus sample will be too low. Combined, the simulation error will cancel out. Another intution is that each individual sample has a wrong estimate of the mean. However, the combined sample has zero mean by construction. Therefore combining the samples give the right mean of the simulated standard normal random variable. Nevertheless, there is still sampling error since we are estimating the mean of the discounted call payoffs, not the mean of the standard normal. This method and other methods to reduce sampling error are discussed next.

Antithetic Variates in Monte Carlo

In this and the following section, we will discuss two methods to increase the efficiency of the Monte Carlo method. These are two of the simplest methods. They are used extensively, but there are other important methods that are also widely used. J"ackel (Jäckel 2002) and Glasserman (Glasserman 2004) provide a wealth of information on this topic.

The Monte Carlo method estimates the mean μ of a random variable x as the sample average of randomly generated values of x. An antithetic variate is a random variable y with the same mean as x and a negative correlation with x. It follows that the random variable $z = (x+y)/2$ will have the same mean as x and a lower variance. Therefore the sample mean of M simulations of z will be an unbiased estimate of μ and will have a lower standard error than the sample mean of M simulations of x. Thus, we should obtain a more efficient estimator of μ by simulating z instead of x.[3]

In the context of derivative valuation, the standard application of this idea is to generate two negatively correlated underlying asset prices (or price paths, if the derivative is path dependent). The terminal value of the derivative written on the first asset serves as x and the terminal value of the derivative written on the second serves as y. Because both asset prices have the same distribution, the means of x and y will be the same, and the discounted mean is the date–0 value of the derivative.

Consider for example a non-path-dependent option in a world with constant volatility. In each simulation i ($i = 1, \dots, M$), we would generate a standard normal Z_i and compute

$$\log S_i(T) = \log S(0) + \left(r - q - \frac{1}{2}\sigma^2\right) T + \sigma\sqrt{T} Z_i \ ,$$
$$\log S_i'(T) = \log S(0) + \left(r - q - \frac{1}{2}\sigma^2\right) T - \sigma\sqrt{T} Z_i \ .$$

Given the first terminal price, the value of the derivative will be some number x_i and given the second it will be some number y_i. The date–0 value of the derivative is estimated as

$$\mathrm{e}^{-rT} \frac{1}{M} \sum_{i=1}^{M} \frac{x_i + y_i}{2} \ .$$

6.2 Control Variates in Monte Carlo

Another approach to increasing the efficiency of the Monte Carlo method is to adjust the estimated mean (option value) based on the known mean of another related variable. We can explain this in terms of linear regression in statistics. Suppose we have a random sample $\{x_1, \dots, x_M\}$ of a variable x with unknown mean μ, and suppose we have a corresponding sample $\{y_1, \dots, y_M\}$ of another variable y with known mean ϕ. Then an efficient estimate of μ is $\hat{\mu} = \bar{x} + \hat{\beta}(\phi - \bar{y})$, where $\bar{x}$ and $\bar{y}$ denote the sample means of x and y, and where $\hat{\beta}$ is the coefficient of y in the linear regression of x on y (i.e., the estimate of β in the linear model $x = \alpha + \beta y + \varepsilon$). The standard Monte Carlo method, which we have described thus far, simply estimates the mean of x as $\bar{x}$. The control variate method adjusts the estimate by adding $\hat{\beta}(\phi - \bar{y})$. To understand this correction, assume for example that the true β is positive. If the random sample is such that $\bar{y} < \phi$, then it must be that small values of y were over-represented in the sample. Since x and y tend to move up and down together (this is the

[3] The negative correlation between x and y is essential for this method to generate a real gain in efficiency. To generate M simulations of z, one must generate M simulations of x and M of y, which will generally require about as much computation time as generating $2M$ simulations of x. If x and y were independent, the standard error from M simulations of z would be the same as the standard error from $2M$ simulations of x, so using the antithetic variate would be no better than just doubling the sample size for x.

meaning of a positive β) it is likely that small values of x were also over-represented in the sample. Therefore, one should adjust the sample mean of x upwards in order to estimate μ. The best adjustment will take into account the extent to which small values of y were over-represented (i.e., the difference between $\bar{y}$ and ϕ) and the strength of the relation between x and y (which the estimate $\hat{\beta}$ represents). The efficient correction of this sort is also the simplest: just add $\hat{\beta}(\phi - \bar{y})$ to $\bar{x}$. In practice, the estimation of $\hat{\beta}$ may be omitted and one may simply take $\hat{\beta} = 1$, if the relationship between x and y can be assumed to be one-for-one. If β is to be estimated, the estimate (by ordinary least squares) is

$$\hat{\beta} = \frac{\sum_{i=1}^{M} x_i y_i - M\bar{x}\bar{y}}{\sum_{i=1}^{M} y_i^2 - M\bar{y}^2} \,.$$

In general, the correction term $\hat{\beta}(\phi - \bar{y})$ will have a nonzero mean, which introduces a bias in the estimate of μ. To eliminate the bias, one can compute $\hat{\beta}$ from a pre-sample of $\{x, y\}$ values.

As an example of a control variate, in our simulation code to estimate the Black Scholes price for a call option we can use the stock price itself. The known stock price is the inout price $S0$. The simulation also produces an estimate for the stock price as the dicsounted expected value of the terminal stock price $\hat{S} = \sum_{i=1}^{n} e^{-rT} St(m,i)$ where $St(m,i)$ is the ith simulated stock price at time T. Theoretically these should be the same umber, but due to error they typically wil not be the same.

```
SS=np.mean(St[m,:])*np.exp(-r*T)
print('The Estimated Stock Price for the first sample is =', SS)
print('The actual stock price should be=', S0)
print('The error is =', S0-SS)
```

```
The Estimated Stock Price for the first sample is = 42.05899999577932
The actual stock price should be= 42
The error is = -0.058999995779316805
```

The error is $S0 - \hat{S}$ which corresponds to $\phi - y$ above. We then compute $\hat{\beta}$ and comute the improved estimate

$$\text{new estimate} = \text{original estimate} + \hat{\beta}(S0 - \hat{S})$$

In the code below we do this procedure for both samples and average the updates.

```
hatbeta= np.cov(St[m,:],cc)[0,1]/np.cov(St[m,],cc)[1,1]
hatbeta1=np.cov(St1[m,:],cc1)[0,1]/np.cov(St1[m,],cc1)[1,1]
correction =hatbeta*(S0-SS)
update=cp + correction
```

```
print('hatbeta=',hatbeta)
print('The original estimate for the call price from the first sample=',cp)
print('The original estimate for the call price from the second sample=',cp1)
print('The updated estimate from the first sample is=',update)
SS1=np.mean(St1[m,:])*np.exp(-r*T)
update1=cp1+hatbeta1*(S0-SS1)
print('The updated estimate from the second sample is=',update1)
print('The average of the updated estimates =',(update+update1)/2)
```

```
hatbeta= 1.1541186403411716
The original estimate for the call price from the first sample= 4.791646287615179
The original estimate for the call price from the second sample= 4.687624646438364
The updated estimate from the first sample is= 4.723553292706219
The updated estimate from the second sample is= 4.780385012196883
The average of the updated estimates = 4.751969152451551
```

We can compare this to the exact Black Scholes formula from before.

```
print('The exact Black Scholes Price is=', truebsc)
```

```
The exact Black Scholes Price is= 4.759422392871532
```

As another example, consider the classic case of estimating the value of a discretely-sampled average-price call, using a discretely-sampled geometric-average-price call as a control variate. Let τ denote the amount of time that has elapsed since the call was issued and T the amount of time remaining before maturity, so the total maturity of the call is $T + \tau$. To simplify somewhat, assume date 0 is the beginning of a period between observations. Let $t_1, \dots, t_N$ denote the remaining sampling dates, with $t_1 = \Delta t$, $t_i - t_{i-1} = \Delta t = T/N$ for each i, and $t_N = T$. We will input the average price $A(0)$ computed up to date 0, assuming this average includes the price $S(0)$ at date 0. The average price at date T will be

$$A(T) = \frac{\tau}{T+\tau} A(0) + \frac{T}{T+\tau} \left(\frac{\sum_{i=1}^{N} S(t_i)}{N} \right) .$$

The average-price call pays $\max(0, A(T) - K)$ at its maturity T, and we can write this as

$$\begin{aligned}
\max(A(T) - K, 0) &= \max \left(\frac{T}{T+\tau} \left(\frac{\sum_{i=1}^{N} S(t_i)}{N} \right) - \left(K - \frac{\tau}{T+\tau} A(0) \right), 0 \right) \\
&= \frac{T}{T+\tau} \max \left(\frac{\sum_{i=1}^{N} S(t_i)}{N} - K^*, 0 \right) ,
\end{aligned}$$

where

$$K^* = \frac{T+\tau}{T} K - \frac{\tau}{T} A(0) \ .$$

Therefore, the value at date 0 of the discretely-sampled average-price call is

$$\frac{T}{T+\tau}\, \mathrm{e}^{-rT} E^R \left[\max \left(\frac{\sum_{i=1}^N S(t_i)}{N} - K^*, 0 \right) \right] \ .$$

In terms of the discussion above, the random variable the mean of which we want to estimate is

$$x = \mathrm{e}^{-rT} \max \left(\frac{\sum_{i=1}^N S(t_i)}{N} - K^*, 0 \right) \ .$$

A random variable y that will be closely correlated to x is

$$y = \mathrm{e}^{-rT} \max \left(\mathrm{e}^{\sum_{i=1}^N \log S(t_i)/N} - K^*, 0 \right) \ .$$

The mean ϕ of y under the risk-neutral measure is given in the pricing Equation **??**. We can use the sample mean of y and its known mean ϕ to adjust the sample mean of x as an estimator of the value of the average-price call. Generally, the estimated adjustment coefficient $\widehat{\beta}$ will be quite close to 1.

6.3 Monte Carlo Greeks I: Difference Ratios

Greeks can be calculated by Monte Carlo by running the valuation program twice and computing a difference ratio, for example $(C_u - C_d)/(S_u - S_d)$ to estimate a delta. However, to minimize the error, and minimize the number of computations required, one should use the same set of random draws to estimate the derivative value for different values of the parameter. For path-independent options (e.g., European puts and calls) under the Black-Scholes assumptions, we only need to generate $S(T)$ and then we can compute $S_u(T)$ as $[S_u(0)/S(0)] \times S(T)$ and $S_d(T)$ as $[S_u(0)/S(0)] \times S(T)$. We can estimate standard errors for the Greeks in the same way that we estimate the standard error of the derivative value.

Actually, there is often a better method available that is just as simple. This is called pathwise calculation. We will explain this in the next section. Here we will describe how to estimate the delta and gamma of a derivative as sample means of difference ratios.

Consider initial prices for the underlying $S_u > S > S_d$. Denote the underlying price at the option maturity in a given simulation by $S_u(T)$ when the initial underlying price is S_u, by $S(T)$ when the initial underlying price is S, and by $S_d(T)$ when the initial underlying price is

S_d. Under the Black-Scholes assumptions, the logarithm of the stock price at date T starting from the three initial prices S_d, S and S_u is

$$\begin{aligned}\log S_d(T) &= \log S_d + \left(r - q - \frac{1}{2}\sigma^2\right) T + \sigma B(T) \; , \\ \log S(T) &= \log S + \left(r - q - \frac{1}{2}\sigma^2\right) T + \sigma B(T) \; , \\ \log S_u(T) &= \log S_u + \left(r - q - \frac{1}{2}\sigma^2\right) T + \sigma B(T) \; ,\end{aligned}$$

so

$$\log S_d(T) = \log S(T) + \log S_d - \log S \Longrightarrow S_d(T) = \left(\frac{S_d}{S}\right) S(T) \; ,$$

and

$$\log S_u(T) = \log S(T) + \log S_u - \log S \Longrightarrow S_u(T) = \left(\frac{S_u}{S}\right) S(T) \; .$$

Therefore, under the Black-Scholes assumptions, we only need to simulate $S(T)$ and then perform the multiplications indicated above to obtain $S_d(T)$ and $S_u(T)$.

Consider a particular simulation and let $C_d(T)$ denote the value of the derivative at maturity when the initial asset price is S_d, let $C(T)$ denote the value of the derivative at maturity when the initial asset price is S, and let $C_u(T)$ denote the value of the derivative at maturity when the initial asset price is S_u. For path-independent derivatives under the Black-Scholes assumptions, these can be computed directly from the simulation of $S(T)$ as just described. However, the following applies to general European derivatives under general assumptions about the underlying asset price (for example, it could follow a GARCH process).

The estimates C_d, C and C_u of the date–0 derivative values, for the different initial prices of the underlying, are the discounted sample means of the $C_d(T)$, $C(T)$ and $C_u(T)$. One way to estimate the delta is $(C_u - C_d)/(S_u - S_d)$. This is a difference of discounted sample means, multiplied by the reciprocal of $S_u - S_d$. Equivalently, it is the sample mean of the differences $C_u(T) - C_d(T)$, multiplied by $\mathrm{e}^{-rT}/(S_u - S_d)$. As a sample mean, its standard error can be estimated as described in Chapter **??**. The standard error is

$$\frac{\mathrm{e}^{-rT}}{S_u - S_d}\sqrt{\frac{1}{M(M-1)}\left(\sum_{i=1}^{M}\left[C_{ui}(T) - C_{di}(T)\right]^2 - M\left[\overline{C_u(T)} - \overline{C_d(T)}\right]^2\right)} \; ,$$

where the overline denotes the sample mean and where $C_{ui}(T)$ [respectively, $C_{di}(T)$] denotes the value of the derivative at maturity in simulation i when the initial asset price is S_u [respectively, S_d].

The corresponding Monte Carlo estimate of the gamma is also a sample mean. Simple algebra shows that Equation **??** is equivalent to

$$\Gamma = \frac{2}{(S_u - S)(S_u - S_d)} C_u - \frac{2}{(S_u - S)(S - S_d)} C + \frac{2}{(S - S_d)(S_u - S_d)} C_d \; . \tag{6.3}$$

Normally one would take $S_u = (1+\alpha)S$ and $S_d = (1-\alpha)S$ for some α (e.g., $\alpha = 0.01$). In this case Equation **??** simplifies to

$$\Gamma = \frac{C_u - 2C + C_d}{\alpha^2 S^2} \;, \tag{6.4}$$

and the standard error of the gamma is

$$\frac{\mathrm{e}^{-rT}}{\alpha^2 S^2}\sqrt{\frac{1}{M(M-1)}} \\ \times \sqrt{\sum_{i=1}^{M} \left[C_{ui}(T) - 2C_i(T) + C_{di}(T)\right]^2 - M\left[\overline{C_u(T)} - 2\overline{C(T)} + \overline{C_d(T)}\right]^2} \;.$$

6.4 Monte Carlo Greeks II: Pathwise Estimates

We will examine the bias in the Monte Carlo delta estimate discussed in the preceding section and explain pathwise estimation of Greeks. By biased, we mean that the expected value of an estimate is different from the true value. It is important to recognize that if a Monte Carlo estimate is biased, then, even if a large number of simulations is used and the standard error is nearly zero, the answer provided by the Monte Carlo method will be incorrect. For simplicity, consider a European call under the Black-Scholes assumptions.

The delta estimate we have considered is the discounted sample mean of

$$\frac{C_u(T) - C_d(T)}{S_u - S_d} \;. \tag{6.5}$$

This ratio takes on one of three values, depending on $S(T)$:

- If $S_u(T) \leq K$ then the option is out of the money in both the up and down cases; i.e.,

  $$C_u(T) = C_d(T) = 0 \;,$$

  so the ratio Equation **??** is zero.
- If $S_d(T) \geq K$ then the option is in the money in both the up and down cases; i.e.,

  $$C_u(T) = S_u(T) - K = \left(\frac{S_u}{S}\right) S(T) - K \;,$$
  $$C_d(T) = S_d(T) - K = \left(\frac{S_d}{S}\right) S(T) - K \;,$$

  so the ratio Equation **??** equals $S(T)/S$.

- If $S_u(T) > K > S_d(T)$, then the option is in the money in only the up case; i.e.,

$$C_u(T) = S_u(T) - K = \left(\frac{S_u}{S}\right) S(T) - K \ ,$$
$$C_d(T) = 0 \ ,$$

so the ratio Equation **??** equals

$$\frac{\left(\frac{S_u}{S}\right) S(T) - K}{S_u - S_d} < \frac{S(T)}{S} \ .$$

The bias is induced by the third case above. We can see this as follows. We are trying to estimate

$$\frac{\partial}{\partial S} \mathrm{e}^{-rT} E^R\big[\max(0, S(T) - K)\big] = \mathrm{e}^{-rT} E^R \left[\frac{\partial}{\partial S} \max(0, S(T) - K)\right] \ . \tag{6.6}$$

The delta estimate $(C_u - C_d)/(S_u - S_d)$ replaces the mean E^R with the sample mean and replaces

$$\frac{\partial}{\partial S} \max(0, S(T) - K) \tag{6.7}$$

with the ratio Equation **??**. The derivative Equation **??** takes on two possible values, depending on $S(T)$—we can ignore the case $S(T) = K$ because it occurs with zero probability:

- If $S(T) < K$, then $\max(0, S(T) - K) = 0$ and the derivative is zero.
- If $S(T) > K$, then $\max(0, S(T) - K) = S(T) - K$ and the derivative equals

$$\frac{\partial S(T)}{\partial S} = \mathrm{e}^{(r-q-\sigma^2/2)T + \sigma B(T)} = \frac{S(T)}{S} \ .$$

Therefore, the true delta—the expectation Equation **??**—equals[4]

$$\mathrm{e}^{-rT} E^R \left[\frac{S(T)}{S} x\right] \ , \tag{6.8}$$

where x is the random variable defined as

$$x = \begin{cases} 1 & \text{if } S(T) > K \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

[4]By changing numeraires, we can show that Equation **??** equals $\mathrm{e}^{-qT} E^V[x] = \mathrm{e}^{-qT} \mathrm{N}(d_1)$, as we know from Chapter **??** is the delta of a European call in the Black-Scholes model (here, as in Chapter **??**, $V(t) = \mathrm{e}^{qt} S(t)$ denotes the value of the non-dividend-paying portfolio created from the stock).

On the other hand, our analysis of the ratio Equation **??** shows that the expected value of the delta estimate $(C_u - C_d)/(S_u - S_d)$ is

$$\mathrm{e}^{-rT} E^R \left[\frac{S(T)}{S} y \right] + \mathrm{e}^{-rT} E^R \left[\frac{S_u S(T) - SK}{S(S_u - S_d)} z \right] , \tag{6.9}$$

where

$$y = \begin{cases} 1 & \text{if} S_d(T) > K \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

and

$$z = \begin{cases} 1 & \text{if} S_u(T) > K > S_d(T) \ , \\ 0 & \text{otherwise} \ . \end{cases}$$

To contrast Equation **??** and Equation **??**, note that if $y = 1$ then $x = 1$, so the term $E^R \left[\frac{S(T)}{S} y \right]$ in Equation **??** is part of Equation **??**. However, there are two partially offsetting errors in Equation **??**: z sometimes equals one when x is zero, and when both z and x are one, then the factor multiplying z is smaller than the factor multiplying x. In any case, the expected value Equation **??** is not the same as the true delta Equation **??**. As noted before, this implies that the delta estimate will be incorrect even if its standard error is zero. The bias can be made as small as one wishes by taking the magnitude $S_u - S_d$ of the perturbation to be small, but taking the perturbation to be very small will introduce unacceptable roundoff error.

The obvious way to estimate the delta in this situation is simply to compute the discounted sample average of $[S(T)/S]x$. This is called a pathwise estimate of the delta, because it only uses the sample paths of $S(t)$ rather than considering up and down perturbations. This method is due to Broadie and Glasserman (Broadie and Glasserman 1996). Because the pathwise estimate is a sample average, its standard error can be computed in the usual way.

To compute pathwise estimates in other models and for other Greeks, we need the Greek to be an expectation as on the right-hand side of Equation **??**. Additional examples can be found in Glasserman (Glasserman 2004) and J"ackel (Jäckel 2002).

6.5 Introduction to Binomial Models

As in the previous section, we will work with the dynamics of the logarithms of asset prices under the risk-neutral measure. Thus, our starting point is the equation

$$\mathrm{d} \log S = \left(r - q - \frac{\sigma^2}{2} \right) \, \mathrm{d}t + \sigma \, \mathrm{d}B \ , \tag{6.10}$$