

# Financial Engineering with Stochastic Calculus

Jeremy Staum

School of Operations Research  
and Industrial Engineering  
Cornell University  
Ithaca, New York

`staum@orie.cornell.edu`

©2002

## Contents

Chapter 1. Introduction	1
1.1. Overview	1
1.2. Portfolio Theory	2
1.3. Fundamentals of Arbitrage Theory	3
1.4. The Black-Scholes Model	6
1.5. Summary	8
1.6. Problems	8
Chapter 2. Brownian Motion and Stochastic Integration	9
2.1. Definition of Brownian Motion	9
2.2. Construction of Brownian Motion	13
2.3. Problems	15
2.4. Definition of Stochastic Integration	16
2.5. Itô's Formula	19
2.6. Itô Processes	22
2.7. Summary	25
2.8. Problems	25
Chapter 3. The Black-Scholes Analysis: Part I	27
3.1. The Black-Scholes PDE	27
3.2. The Black-Scholes Formula and Greeks	28
3.3. Summary	31
3.4. Problems	31
Chapter 4. Conditional Probability and Itô Processes	33
4.1. Conditional Probability	33
4.2. Conditioning with Itô Processes	35
4.3. Martingales	39
4.4. The Markov Property	41
4.5. Summary	42
4.6. Problems	43
Chapter 5. The Black-Scholes Analysis: Part II	44
5.1. The Feynman-Kač Formula	44
5.2. Girsanov Transformation	48

5.3. Examples	53
5.4. Problems	55
Chapter 6. Complete Markets	57
6.1. Market Price of Risk	57
6.2. The Pricing Measure $\mathbf{Q}$	59
6.3. Vector Itô Processes	61
6.4. Markets with Multiple Risky Securities	64
6.5. Martingale Representation	67
6.6. Decomposition	68
6.7. Problems	70
Chapter 7. Futures and Dividends	72
7.1. Futures	72
7.2. Dividends	75
7.3. Problems	78
Chapter 8. Computation	79
8.1. Simulation	79
8.2. Calibration	82
8.3. Problems	85
Chapter 9. Volatility Models	87
9.1. Local Volatility	87
9.2. Stochastic Volatility	89
Appendix A. Final Review	91
Appendix B. Solutions to Problems	92
Appendix. Bibliography	103



## CHAPTER 1

### Introduction

In this chapter, we

- (1) describe financial engineering and its main problems
- (2) introduce a framework for probabilistic models of financial markets
- (3) discuss arbitrage and its connection to pricing
- (4) introduce and criticize the Black-Scholes model

#### 1.1. Overview

What is financial engineering? How is it different from finance and from management science, or operations research, or industrial engineering?

First, financial engineering is indeed an engineering discipline, and as such, proceeds from facts about the world plus prespecified goals to produce a solution to a problem. Finance is a social science, and as a science, it is concerned with understanding how the world really is. Financial engineering is heavily dependent on probabilistic models for the evolution of financial variables over time, and it relies on finance for this kind of knowledge. One way in which financial engineering is unlike other engineering disciplines is that its fundamental models are much less accurate. The field continues to develop rapidly in tandem with research in finance, seeking better models.

The obvious difference between financial engineering and MS/OR/IE is the finance. Financial engineering is not about managerial decisions, military operations, or industrial processes, but about money, pure and simple. It shares with its parent discipline a focus on optimization and stochastic processes, but uses mathematical tools that seldom appear in other domains of MS/OR/IE, such as stochastic calculus. The branch of mathematics inspired by the problems of financial engineering may be referred to as financial mathematics, perhaps as biostatistics and econometrics are kinds of statistics having to do with the problems of biology and economics. On the other hand, mathematical finance seems to be essentially a synonym for financial engineering.

In this course, we will study three main problems of financial engineering. To deal effectively with any of them requires the assumption of a probabilistic model for investment opportunities in financial markets. Each proceeds from a real-world situation in which someone has a goal and a situation and needs to know what action to take.

- (1) Derivative pricing: There is great demand for so-called **derivative securities**, whose payoffs are derived from **underlying variables**, usually the prices of other

securities. Those selling derivative securities need to determine what price to charge in order to maximize profits and minimize risk.

- (2) Risk management: Anyone managing an institutional portfolio needs to be able to measure and control its risk in order to be able to manage prudently and to satisfy superiors, investors, or regulators.
- (3) Portfolio optimization: Individuals and institutions want to invest optimally given their financial goals.

## 1.2. Portfolio Theory

We model a financial market as a collection of traded securities. In general, our notation is that there are  $N$  traded securities, and the  $i$ th has price  $S_i(t)$  at time  $t$ . Sometimes we may also have a 0th security (usually a money market account), in which case there are  $N + 1$  securities. The price vector  $S$  is a stochastic process, and we must have in mind a probability measure  $\mathbf{P}$  that supports this stochastic model of security prices. Some people think of this as the **objective probability measure**, which describes how the world really is, others as the **subjective probability measure**, which describes someone's beliefs about future prices.

Market participants can buy and sell traded securities, resulting in portfolios that change over time. A **portfolio strategy** is a vector stochastic process  $\theta$  where  $\theta_i(t)$  is the number of shares of the  $i$ th security held at time  $t$ . Then the value of this portfolio is

$$(1.2.1) \quad V(t) = \theta(t)S(t) = \sum_{i=1}^N \theta_i(t)S_i(t).$$

(The convention is that  $\theta$  is a row vector and  $S$  a column vector.) In this course, we will focus on problems with a finite time horizon, usually called  $T$ , so that a portfolio strategy is defined for  $t \in [0, T]$ .

We are interested primarily in **self-financing** portfolios, those whose value changes only as a result of the portfolio's own gains or losses, not because of cash infusions or withdrawals. This means that changes in the portfolio strategy  $\theta$  must be costless. For instance, when all security prices are positive, a purchase of more shares of one security requires a compensating sale of shares of some other securities, in order to raise the necessary funds. In discrete time, the self-financing condition is

$$(\theta(t_{j+1}) - \theta(t_j))S(t_{j+1}) = \theta(t_j)D(t_j),$$

where  $D_i(t_j)$  is the dividend paid at step  $j + 1$  to the owner at step  $j$  of each share of security  $i$ . The condition means that there is zero net cost to the portfolio rebalancing done at step  $i + 1$ , at the prices then current.

There is another condition that we want to demand of portfolio strategies. A portfolio strategy is **tame** when the resulting portfolio value is bounded below, i.e. there exists a value  $L$  such that there is zero probability of having  $V(t) = \theta(t)S(t) < L$  for any time  $t \in [0, T]$ . This is an intuitive restriction to impose, because we can imagine  $L$  as a market participant's

credit limit. Should the participant's portfolio value fall beneath  $L$ , it would go bankrupt and have to unwind its portfolio. It would be imprudent or irresponsible to contemplate a trading strategy with unlimited losses, and unrealistic to expect to find willing creditors or counterparties when one's liabilities are excessive. Each market participant has some credit limit  $L$ , but tameness is supposed to be a property of a portfolio strategy without reference to the market participant who executes it. Therefore for tameness we require merely that there be *some* finite credit limit  $L$  that bounds the portfolio value below. After all, everyone has a credit limit, so nobody could execute a strategy that is not tame.

### 1.3. Fundamentals of Arbitrage Theory

An interesting kind of self-financing, tame portfolio strategy is the **arbitrage**, which can be thought of as a "free lunch" or "getting something for nothing." There are two kinds of arbitrage strategies:

- (1) Money for nothing:  $\theta(0)S(0) < 0$  and  $\theta(T)S(T) \geq 0$ .
- (2) Lottery tickets for free:  $\theta(0)S(0) \leq 0$ ,  $\theta(T)S(T) \geq 0$ , and  $\mathbf{P}(\theta(T)S(T) > 0) > 0$

The first type of arbitrage is a way of getting money now without taking on any risk of future loss. The cost of setting up the initial portfolio is  $\theta(0)S(0)$ , and when this is negative, you get paid a positive amount to set it up. Then  $\theta(T)S(T) \geq 0$  means that you can close the position at time  $T$  without any chance of loss. The second type of arbitrage costs nothing to set up, has no chance of loss, and has a positive chance of a positive payoff.

**EXAMPLE 1.3.1** (Same stock, different markets). Suppose you can buy a share of stock in London for a price lower than that for which you can sell it in New York at the same time. Doing so, you pocket the cash difference now, while retaining no liability.

**EXAMPLE 1.3.2** (Playing the lottery). If you could get for free a lottery ticket that has a positive chance of winning a prize, that would be an arbitrage. Even though you might end up with nothing, the chance to win something is worth something, and getting that chance for free is an arbitrage. On the other hand, the opportunity to buy a lottery ticket for a positive price, no matter how small, and no matter how valuable or likely the prize, is not an arbitrage. Indeed, paying one dollar now for a winning lottery ticket that guarantees a million dollars at a later date with probability 1 is also not an arbitrage, but rather, investing in a zero-coupon bond that pays a high rate of interest.

**EXAMPLE 1.3.3** (Mispriced bond). As in Example 1.3.2, a zero-coupon bond that pays a million dollars in one year and costs one dollar now is not an arbitrage by itself. However, suppose that you can borrow at an interest rate of 5% over that year. Then there is an arbitrage: the portfolio which consists of borrowing one dollar, buying the zero-coupon bond, and liquidating the portfolio at the end of the year. The initial cost is zero, and the final payoff is \$999,998.95, which is positive.

It is not the magnitude of the payoff in Example 1.3.3 that matters, just that something positive is available for a nonpositive price. The example illustrates that it is a portfolio that

is an arbitrage, not necessarily a single price. From the standpoint of arbitrages, nothing is wrong with the price of the million-dollar bond, unless it is considered in relation to other opportunities that exist in the marketplace.

Our entire approach to derivative pricing rests on the assumption that arbitrages should not exist, if the market is in equilibrium. If  $\theta$  were a tame arbitrage, there would be unlimited demand for it, causing its price  $\theta(0)S(0)$  to rise to a positive level, so that it would no longer be an arbitrage once prices had reached equilibrium. While the market is not in equilibrium, an arbitrage can exist. It would certainly be difficult to argue from empirical financial data that markets are in equilibrium most of the time, although it is also very difficult to explain how arbitrages of substantial size could be available to a significant group of market participants for any noticeable length of time. Regardless, the no-arbitrage principle has a lot of force as a normative principle because nobody wants to give away free money or valuable lottery tickets.

A classic informal example of an arbitrage is a \$20 bill lying in the street. One hardly ever finds such opportunities. On the other hand, one frequently encounters pennies lying in the street. How can we interpret this? Perhaps one finds pennies in the street when there is a temporary disequilibrium in the distribution of loose change, for instance, after pennies fall from the sky, and people will soon profitably snatch up all these arbitrage opportunities. On the other hand, perhaps a penny in the street is not an arbitrage, because the cost of bending down to pick it up or the risk of catching a disease from handling it exceed its value. The former argument is a way of saying that the no-arbitrage postulate does not always hold, while the latter is an attempt to explain why certain opportunities are not arbitrages. Bear these considerations in mind when people attempt to persuade you to pay for access to a supposed arbitrage opportunity.

Notice that we are denying the existence of only self-financing, tame arbitrage strategies. To have the financial interpretation of “something for nothing,” the strategy must not involve cash inflows at intermediate times, as well as avoiding initial cost or terminal loss. The reason for focusing on tame strategies is that there is no good way of excluding arbitrages that are not tame. The canonical example is based on the “doubling strategy” from gambling.

**EXAMPLE 1.3.4 (Doubling Strategy).** For simplicity, imagine a gamble such as betting on red in roulette, where the bet is either lost or pays off at even odds. That is, a bet of  $x$  dollars results in a change of wealth of either  $-x$  or  $x$  dollars. After each loss, the gambler doubles the bet, stopping after the first win. Suppose the gambles are Bernoulli trials, independent and with identical positive probability of a win. With probability one, the first win occurs on the  $n$ th gamble where  $n$  is a finite number, so if the first bet is one dollar, the total winnings are  $-\sum_{i=1}^{n-1} 2^{i-1} + 2^n = 1$ . According to this strategy, one can begin with nothing and end up with a profit of a dollar. This is an arbitrage, but it is not tame because the intermediate wealth could be  $1 - 2^i$  after  $i$  losses in a row, which is unbounded below because there is no limit to the number of losses in a row that might occur. Thus to execute this strategy would require an infinite amount of credit, and the ability to bet an unlimited amount at once.



In finance, such a strategy might involve buying stock on credit. Nobody could actually take advantage of this to get an arbitrage, because it requires the ability to borrow an unlimited amount of money and buy an unlimited amount of shares. So this is of no economic significance and does not count as an arbitrage. The experience of Long-Term Capital Management is a case in point. LTCM executed strategies that were supposed to be arbitrages, but whose maximum loss exceeded the credit that the fund was able to draw upon. In the event, the losses grew so large that LTCM could not find willing counterparties or creditors, was unable to continue executing its so-called arbitrage strategies, and went spectacularly broke.

Aside from the traded securities, we are also interested in **derivative securities**, also known as **contingent claims**. For now, we consider only the simplest derivatives, those that have a single payoff at a future time  $T$ , called **maturity** or **expiration**, and are **path-independent**, meaning that the payoff is a function  $f(X(T))$  only of the value at maturity of the **underlying variables**  $X$ , not their whole histories. The canonical example is the European call option.

**EXAMPLE 1.3.5** (European call option). The European call option pays  $f(S(T)) = (S(T) - K)^+$  at time  $T$ . Here  $S(T)$  is the price of a traded security, such as a stock, at the option's maturity  $T$ , and  $K$  is a strike price, which is written into the option contract just like the maturity  $T$  is. The notation  $(S(T) - K)^+$  means  $\max\{S(T) - K, 0\}$ , so this is an option, but not an obligation, to buy the stock for price  $K$  at time  $T$ .

Why does the no-arbitrage principle, the belief that arbitrages should not exist in a market in equilibrium, help us to understand derivative securities? By considering the relationship between derivative payoffs and underlying securities, we can make substantive statements about prices of derivatives.

**EXAMPLE 1.3.6** (No-Arbitrage Bounds). The European call option's payoff  $(S(T) - K)^+$  is nonnegative, so its price must be nonnegative. If it were negative, then one could buy the option and receive money for doing so, while taking on no risk of a future loss. This would be an arbitrage. If we believe in a probability model such that  $\mathbf{P}[S(T) > K] > 0$ , then we can furthermore say that the option price must be strictly positive. Also, the payoff  $(S(T) - K)^+ < S(T)$ , assuming that the strike price  $K > 0$ , as is always true in practice. The portfolio of 1 share of stock and  $-1$  shares of the call option has payoff  $S(T) - (S(T) - K)^+$  which equals  $K$  when  $S(T) \geq K$  and equals  $S(T)$  when  $S(T) \leq K$ . Assuming that in our probability model,  $\mathbf{P}[S(T) > 0] > 0$ , this is a nonnegative payoff with positive probability of being positive, so it must have a positive price to avoid arbitrage. So the price of the call must be less than that of the stock. Putting these results together, the no-arbitrage price of the European call option must be between 0 and  $S(0)$ .

**EXAMPLE 1.3.7** (Put-Call Parity). The European put option has payoff  $(K - S(T))^+$ , thus representing the option to sell the stock for price  $K$  at time  $T$ . The difference of the European call and put payoffs is

$$\max\{S(T) - K, 0\} - \max\{K - S(T), 0\} = \max\{S(T) - K, 0\} + \min\{S(T) - K, 0\} = S(T) - K.$$

Suppose that there is a riskless bond paying 1 at time  $T$ , which can be bought now for  $B(0)$ . Then a portfolio containing one share of stock and  $-K$  shares of this bond is also worth  $S(T) - K$  at time  $T$ . The no-arbitrage principle demands that these portfolios with the same terminal value have the same initial price. Therefore the difference between the call and put prices should equal the initial price of the other portfolio,  $S(0) - KB(0)$ .

Notice that these results did *not* depend on a model for the prices of traded securities. This makes them much more reliable than results that depend on a model, which is never a perfect description of reality. On the other hand, they are not very specific: the range  $[0, S(0)]$  for an option price is not very helpful, and put-call parity gives you a relationship between put and call prices, but doesn't tell you exactly what either of them should be. No-arbitrage reasoning within a stochastic model for traded security prices will give much more specific but less reliable answers. In particular, if real-world derivative prices do not match the "no-arbitrage" price from some model, that does not mean that an arbitrage is really available. One must keep in mind **model risk**, the risk of losses arising from a portfolio strategy when reality departs from a model. Thus there is an implicit conditionality in all our model-based results: *if* this model were true, *then* we can specify the price to avoid arbitrage. Whether a model is adequate for a specific business purpose is an empirical question.

The first part of this course is devoted to understanding the Black-Scholes model of stock prices and using it to derive no-arbitrage prices for equity derivatives.

#### 1.4. The Black-Scholes Model

Financial engineering in continuous time begins with the results of Black and Scholes, published in 1973. Theory and practice have both come a long way since that time, partly due to attempts to correct the shortcomings of their model. Nonetheless, their approach is the point of departure for much of current practice.

The Black-Scholes model was intended to handle simple equity derivatives. In this model, there are two securities, a riskless money market account and a risky stock. The price of a share of the money market account is  $M(t) = e^{rt}$ , where  $r$  is a constant, continuously compounded interest rate. The stock price follows a **geometric Brownian motion**:  $S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t))$ , where  $\mu$  is the expected return of the stock,  $\sigma$  is its volatility, and  $W$  is a **Brownian motion**, also known as a **Wiener process**. We will study this stochastic process in detail in Chapter 2. It gives the stock returns a normal distribution.

An attraction of this model is that it yields a unique no-arbitrage price (rather than an interval) for derivative securities such as the European call option. It also yields a hedging strategy that a derivatives trader can use to eliminate all the risk associated with selling derivatives. Moreover, some nice portfolio optimization results are available within this model. It is important to recognize that these are not the right answers in an absolute sense. Although they may appear inside boxes in textbooks, their validity depends on the validity of the Black-Scholes model, which is inadequate. It rests on several false assumptions, many of which are not easy to do away with—decades of research have gone into attempts to relax them in order to model financial markets more accurately.

- (1) The stock pays no dividends, and one can hold a negative number of shares of the money market account and stock as well as a positive number. Financially, such negative holdings mean borrowing money at the risk-free rate, and shorting the stock with no borrow cost. Neither of these is possible. Market participants are rightly perceived as credit risks, so one must pay a higher rate to borrow than any debtor viewed as (approximately) risk-free. The institutional mechanisms for shorting stocks are not trivial, and typically one who shorts must pay a fee to the lender of the stock. We explore these issues later in the course: it is possible to ameliorate some of these shortcomings without much trouble.
- (2) One can buy or sell an unlimited number of shares of stock at time  $t$  at the price  $S(t)$ . It is not even true that one can both buy and sell a small number of shares at a single price. Markets have different mechanisms, but they all cost money, so there are always **transaction costs**. In a transaction, the buyer must pay more than the seller receives, so that there is money to pay for the brokers, computers, etc. that make a market function. Aside from this, there is the problem of finding a counterparty, especially for large trades. **Liquidity** is a vague term for the ability of a market to absorb large trades without disturbing the price. Because markets are not perfectly liquid, one who desires to buy a large number of shares of a security must raise the price offered in order to entice a sufficient number of shareholders to sell; likewise a large seller must lower the price asked to attract potential buyers. The change between the price before an order hits the market and the price at which the trade actually takes place is known as **slippage**.
- (3) The stock price is continuous. The discrete price systems for trading stocks (e.g. eighths or tenths of a dollar) are not the major problem with this assumption. Rather, it gives a dangerously misleading impression about financial risks. In reality, security prices sometimes change in large jumps, due to news announcements, or from the end of one trading session to the beginning of another.
- (4) Stock price returns are lognormally distributed. This is also dangerously misleading. Log stock price returns actually have tails much heavier than normal, especially the negative tail. This means that the Black-Scholes model tends to underestimate the probability of large losses.
- (5) The risk-free interest rate and stock price volatility are deterministic. Over time horizons that are not very short, this assumption is also inadequate. Both interest rates and price volatility are stochastic in a very complicated way, not deterministic. It is widely accepted that there are features of these stochastic processes (e.g. mean reversion and autocorrelation) that are important to capture.

Despite the shortcomings of this model, geometric Brownian motion is a good basic model for a stock price, and many superior models have it as their basis. Moreover, the Black-Scholes model shapes the language and mindset of many practitioners: for instance, option prices are often quoted in terms of the Black-Scholes **implied volatility**. This is the value of the volatility  $\sigma$  that, when plugged into the Black-Scholes formula, yields the actual

market price of the option. Thus a deep understanding of this model is a great benefit to a practitioner. Essentially every model of continuous-time financial engineering relies on the same mathematics of Brownian motion and stochastic integration, which we now study as a preliminary to the Black-Scholes analysis.

### 1.5. Summary

Financial engineering is a variant of MS/OR/IE dealing with money. Its main problems are derivative securities valuation and hedging, risk management, and portfolio optimization.

We model a financial market as a vector stochastic process of the prices of traded securities. A financial agent controls its random wealth by executing a tame, self-financing portfolio strategy.

An arbitrage is a way of getting something for nothing. Derivative security prices must fall within certain bounds in order to avoid arbitrage. The notion of arbitrage and no-arbitrage pricing may depend on a model, in which case they are subject to model risk.

The Black-Scholes model assumes a constant interest rate and a stock price following geometric Brownian motion. It forms the basis for a common language among practitioners, and we must study Brownian motion deeply in order to understand it. Although the Black-Scholes model allows relatively easy computation of clean, closed-form results, it rests on dramatically erroneous assumptions that undercut the validity of these results.

### 1.6. Problems

**PROBLEM 1.1.** Assume the Black-Scholes model, and consider portfolio strategies in the stock alone, over the time interval  $[0, 1]$ . For each of the two (static) portfolio strategies

- (1) long one share of stock:  $\theta_S(t) = 1$  for  $t \in [0, 1]$
- (2) short one share of stock:  $\theta_S(t) = -1$  for  $t \in [0, 1]$

answer each of the questions

- (1) Is it self-financing?
- (2) Is it tame?
- (3) Is it an arbitrage?

**PROBLEM 1.2.** There is a non-dividend-paying stock worth  $S(0) = 100$  and a riskless zero-coupon bond paying 1 at time  $T = 0.5$  (years), which is now worth  $B(0) = 0.97$ , and two European call options with maturity  $T$  on the stock. One is struck at  $K_1 = 110$  and is worth 20, and the other is struck at  $K_3 = 130$  and is worth 5. Do not assume the Black-Scholes or any other model. Assume only that the terminal stock price  $S(T) > 0$  and the terminal bond price  $B(T) = 1$ . Find no-arbitrage bounds (upper and lower) for the price of a European call option on the stock with the same maturity  $T$  and strike  $K_2 = 120$ . Hint: try graphing the payoffs of

- (1) the portfolio long one call struck at  $K_1$  and one struck at  $K_3$ , and short two calls struck at  $K_2$
- (2) the portfolio long one call struck at  $K_1$  and short one struck at  $K_2$

## CHAPTER 2

### Brownian Motion and Stochastic Integration

In this chapter, we

- (1) define standard, generalized, and geometric Brownian motion, and Brownian bridge
- (2) illustrate the construction of Brownian motion
- (3) define the Itô stochastic integral
- (4) learn Itô's formula for computations with stochastic integrals

#### 2.1. Definition of Brownian Motion

The fundamental stochastic process for this course is the **Wiener process**, also known as **standard Brownian motion**. Brownian motion is named for the 19th-century biologist Robert Brown, who described the random, erratic motion of minuscule pollen grains suspended in water. The Wiener process is named for Norbert Wiener, the 20th-century mathematician who formalized it rigorously.

Its usual representation is  $W$  (for Wiener process) or  $B$  (for Brownian motion), and we will use the former. Recall  $W(t)$  is a random variable, the value of  $W$  at time  $t$ ,  $W(\omega)$  is a **sample path**, the trajectory that  $W$  follows if the state of the world is  $\omega$ , and  $W(\omega, t)$  is the value of  $W$  at time  $t$  in state  $\omega$ . Thus the random variable  $W(t)$  is random because it is a function of the unknown state  $\omega$ , while  $W(\omega)$  is a function of time.

A definition of a Wiener process  $W = (W_t, t \in [0, T])$  is:

- (1)  $W(0) = 0$ .
- (2) Each  $W(t)$  is a normal random variable with mean 0 and variance  $t$ .
- (3) The increments  $W(t) - W(s)$  are stationary and independent.
- (4) Each sample path  $W(\omega)$  is a continuous function of time.

In the Black-Scholes model,  $S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t))$  describes the fundamental uncertainty driving changes in stock price. We now discuss an economic interpretation of the definition of the Wiener process in light of this model. This is a nice story, but a false one: remember that the Black-Scholes model is not very accurate.

Suppose that news concerning a company's profit outlook arrives at a constant rate in a stream of small nuggets of information, each of which is independent of the others. Imagine taking the limit as nuggets get smaller and the rate goes up, so that we have a continuous stream of news.

By the central limit theorem, the total impact of news over some time period is normally distributed. This justifies Condition (2). "At a constant rate" implies stationarity of increments. Recall that **stationary** means that the distribution of  $W(t) - W(s)$  depends only

on  $t - s$ , not  $t$  or  $s$ . The parameter  $\sigma$  controls the variance of  $W(t)$ , which is proportional to  $t$ , according to the central limit theorem. Independence of the news items implies independence of increments. Continuity of the sample path follows because we are imagining the nuggets of information as infinitesimally small.

Condition (1) is just a standardization. The constant  $S(0)$  is the initial stock price. Likewise the constant  $\mu$  controls the expected growth rate of the stock, so the assertion of mean 0 in Condition (3) is also a standardization.

In Problem 2.2, you will show that the stochastic process  $S$  does not have stationary or independent increments. Nonetheless, Condition (3) does imply that the percentage returns of the stock, like its log returns  $\ln(S(t)/S(s) - 1) = \ln(S(t) - S(s)) - \ln(S(s))$ , are stationary and independent. In economic terms, this is a very strong assumption: the uncertainty in the growth of the stock's value is the same at all times, and does not depend on its past vicissitudes.

This all makes a nice story, but it is not true. The exact nature of stock price dynamics remains unclear, but they are not so simple.

We now turn to the distributional properties of  $W$ , and in particular, its expectation and covariance functions. For positive times  $s < t$ ,  $\mu_W(t) = \mathbf{E}[W(t)] = 0$  and

$$\begin{aligned} c_W(s, t) &= \mathbf{Cov}[W(s), W(t)] \\ &= \mathbf{Cov}[W(s) - W(0), (W(s) - W(0)) + (W(t) - W(s))] \\ &= \mathbf{Cov}[W(s) - W(0), W(s) - W(0)] + \mathbf{Cov}[W(s) - W(0), W(t) - W(s)] \\ &= \mathbf{Cov}[W(s), W(s)] + 0 \\ &= s. \end{aligned}$$

Each increment  $W(t) - W(s)$  is a normal random variable with mean 0 and variance  $t - s$ . For an increasing finite sequence of times  $(t_1, \dots, t_n)$ , the distribution of the random vector  $(W(t_1), \dots, W(t_n))$  is multivariate normal with mean vector zero and covariance matrix

$$\begin{bmatrix} t_1 & t_1 & \cdots & t_1 \\ t_1 & t_2 & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_1 & t_2 & \cdots & t_n \end{bmatrix}.$$

It is important to remember that these statements about joint distributions are more than just statements about marginal distributions. Here we have said that  $W(s)$  and  $W(t)$  have a bivariate normal distribution, both have mean zero, their variances are  $s$  and  $t$  respectively, and their correlation is  $\mathbf{Cov}[W(s), W(t)] / \sqrt{\mathbf{Var}[W(s)]\mathbf{Var}[W(t)]} = s / \sqrt{st} = \sqrt{s/t}$ . Thus  $W(s)$  and  $W(t)$  are dependent, but not totally dependent, as we can see from the equation  $W(t) = W(s) + (W(t) - W(s))$ , where the increment  $W(t) - W(s)$  is independent of  $W(s)$ . On the other hand, we can find random variables with the same marginal distributions, one  $\mathcal{N}(0, s)$  and the other  $\mathcal{N}(0, t)$ , but with a different joint distribution.

EXAMPLE 2.1.1. Suppose that the random variable  $X$  is normally distributed with mean zero and variance  $s$ . Now define  $Y = \sqrt{t/s}X$ . Then  $Y$  is normally distributed with mean zero and variance  $t$ . However, the joint distribution of  $X$  and  $Y$  is not the same as that of  $W(s)$  and  $W(t)$ . The covariance of  $X$  and  $Y$  is  $\mathbf{Cov}[X, \sqrt{t/s}X] = \sqrt{t/s}\mathbf{Var}[X] = \sqrt{st}$ , so their correlation is one, illustrating that they have total dependence. Given  $X$ , we know the value of  $Y$ . In this case, we have a degenerate bivariate normal distribution.

EXAMPLE 2.1.2 (thanks to Sebastien). Again, suppose that the random variable  $X$  is normally distributed with mean zero and variance  $s$ . Let  $U$  be independent of  $X$ , taking on the values 1 or  $-1$  with equal probability. Then let  $Y = \sqrt{t/s}UX$ . This is normal with mean zero and variance  $t$ , but  $X$  and  $Y$  do not have a bivariate normal joint distribution at all. Given  $X$ , we know that  $Y$  is either  $X$  or  $-X$ . This does not conform to the nature of the bivariate normal distribution, which is that one random variable must be normally distributed given the other.

It seems plausible that there exist a state space and probability measure that make it possible to construct a stochastic process with such distributional properties. Yet Condition (4) is about sample paths, not about distributions at all! It is a nontrivial mathematical fact that it is possible to construct a Wiener process, which has these distributions *and* continuous sample paths. The economic significance of continuity is that it facilitates hedging, as we will see in the Black-Scholes analysis. When jumps are possible, one frequently finds that markets are **incomplete**, that is, not all contingent claims can be perfectly hedged. This considerably complicates the task of pricing and hedging derivative securities.

To illustrate the nontriviality of continuous sample paths, consider a Poisson process  $N$  with parameter  $\lambda$ , which like a Wiener process  $W$  begins at 0 and has stationary and independent increments. The difference is that  $N(t)$  has the Poisson distribution with parameter  $\lambda t$  while  $W(t)$  has the normal distribution with parameters 0 and  $t$ . But it is *not* possible to construct a Poisson process with continuous sample paths, indeed a Poisson process is a pure jump process, i.e. its sample paths change only when they jump discontinuously.

In Section 2.2, we will construct a Wiener process as the limit of a sequence of simpler, approximating stochastic processes. Later we will be content to let the formal underpinnings of Brownian motion remain out of sight, although not out of mind!

Several processes of interest to us derive from the Wiener process, or standard Brownian motion: generalized Brownian motion, geometric Brownian motion, and the Brownian bridge. Standard Brownian motion is a special case of **generalized Brownian motion**, which is  $X(t) = X(0) + \mu t + \sigma W(t)$  in the one-dimensional case. The parameter  $\mu$  is the drift, which controls how fast the process grows on average, and  $\sigma$  the volatility, which controls the size of its fluctuations. The expectation and covariance functions are  $\mu_X(t) = X(0) + \mu t$  and  $c_X(s, t) = \sigma^2 s$  for  $s \leq t$ . An  $m$ -dimensional generalized Brownian motion is  $X(t) = X(0) + \mu t + AW(t)$  where  $W$  is  $n$ -dimensional standard Brownian motion,  $A$  is a  $m \times n$  matrix, and  $\mu$  is a column  $m$ -vector of drifts. Then the covariance matrix of

$X$  is  $AA^\top t$ :

$$\mathbf{Cov} [X_i(t), X_j(t)] = \mathbf{Cov} \left[ \sum_{k=1}^n a_{ik} W(t_k), \sum_{k=1}^n a_{jk} W(t_k) \right] = \sum_{k=1}^n a_{ik} a_{jk} t = (AA^\top)_{ij} t.$$

Generalized Brownian motion allows for nonzero average growth rates, different variability of components, and dependence between components. It is sometimes referred to as “Brownian motion” simply. It still has independent and stationary increments and normal marginal distributions, but with more general mean and covariance.

**Brownian bridge** is so named because it is constructed from Brownian motion, and like a bridge connects two given points while having some freedom to change its height (value) in between. The definition of a (standard) Brownian bridge is  $Z(t) = W(t) - tW(1)$  for  $t \in [0, 1]$  only. Then  $Z(0) = Z(1) = 0$ , so (standard) Brownian bridge is constrained to start at 0 at time 0 (like standard Brownian motion) and also end at 0 at time 1. A generalized Brownian bridge can start and end at different times  $a$  and  $b$  and values  $Z(a)$  and  $Z(b)$ , and have a different volatility parameter  $\sigma$ . Then it would be defined for  $t \in [a, b]$  as

$$Z(t) = Z(a) + \sigma W(t) + \left( \frac{t-a}{b-a} \right) (Z(b) - Z(a) - \sigma W(b)).$$

Drift is not a parameter for a Brownian bridge: this role is played by the given slope  $(Z(b) - Z(a))/(b - a)$ .

**Geometric Brownian motion** is a transformation of Brownian motion:

$$Y(t) = \exp(X(t)) = Y(0) \exp(\mu t + \sigma W(t))$$

where  $X(0) = \ln(Y(0))$ . The Black-Scholes model of a stock price is an example of a geometric Brownian motion. It does *not* have independent or stationary increments. Whereas Brownian motion changes in an additive or arithmetic fashion, geometric Brownian motion changes in a multiplicative or geometric fashion. Its “multiplicative increments”  $Y(t)/Y(s)$  are independent, stationary, and have a lognormal distribution. That is,  $\ln(Y(t)/Y(s))$  is normal. This process is very important in finance, because (contra Malthus) economic quantities tend to grow in a multiplicative i.e. geometric fashion. For instance, the Black-Scholes model of a stock price uses geometric Brownian motion rather than just generalized Brownian motion because a stock price should be positive, and (all else being equal) a share priced at \$60 is much likelier to go up today by \$1 than is a share priced at \$2.

**EXAMPLE 2.1.3 (Binary call).** A **binary call option** pays  $f(S(T)) = \mathbf{1}_{\{S(T) \geq K\}}$  at maturity  $T$ . That is, it pays 1 if  $S(T) \geq K$  and 0 otherwise. What is the probability, under



the Black-Scholes model, that the binary call pays off? It is

$$\begin{aligned}
 \mathbf{P}[S(T) \geq K] &= \mathbf{P}[S(0) \exp((\mu - \sigma^2/2)T + \sigma W(T)) \geq K] \\
 &= \mathbf{P}\left[W(T) \geq \frac{\ln(K/S(0)) - (\mu - \sigma^2/2)T}{\sigma}\right] \\
 &= \mathbf{P}\left[\frac{W(T)}{\sqrt{T}} \geq \frac{\ln(K/S(0)) - (\mu - \sigma^2/2)T}{\sigma\sqrt{T}}\right] \\
 &= \Phi\left(\frac{\ln(S(0)/K) + (\mu - \sigma^2/2)T}{\sigma\sqrt{T}}\right)
 \end{aligned}$$

where  $\Phi$  is the standard normal cumulative distribution function, because  $W(T)/\sqrt{T}$  is a standard normal random variable. Here we are using the property that when  $Z$  is standard normal,  $\mathbf{P}[Z \geq x] = \mathbf{P}[Z \leq -x] = \Phi(-x)$  by symmetry.

This probability is the expected payoff of the binary call, so trading the binary call at the price  $\mathbf{P}[S(T) \geq K]$  would correspond to a “fair gamble.” However, it is *not* its “fair price.” After all, is the stock or the money market account a fair gamble? Would the prices making them into fair gambles be fair prices? Would you be inclined to buy them for those prices? We will later find the no-arbitrage price for a binary call.

**EXAMPLE 2.1.4** (Black-Scholes loss probability). Suppose that you plan to hold a portfolio of  $\theta_0$  shares of the money market account and  $\theta_1$  shares of stock for the whole time interval  $[0, T]$ , and you are concerned about the probability of having lost money at time  $T$ , that is, of the event  $V(T) < V(0)$ . Assume that the Black-Scholes model holds. Then this event is  $\theta_1 S(T) + \theta_0 M(T) < \theta_1 S(0) + \theta_0 M(0)$ , or  $\theta_1 S(0) \exp((\mu - \sigma^2/2)T + \sigma W(T)) < \theta_1 S(0) + \theta_0(1 - e^{rT})$ . (Recall  $M(t) = e^{rt}$ .) By a computation similar to that of Example 2.1.3, the probability of this event is

$$\begin{aligned}
 \mathbf{P}\left[W(T) < \frac{1}{\sigma} \left( \ln \left( 1 + \frac{\theta_0(1 - e^{rT})}{\theta_1 S(0)} \right) - \left( \mu - \frac{1}{2} \sigma^2 \right) T \right)\right] \\
 = \Phi\left(\frac{\ln \left( 1 + \frac{\theta_0(1 - e^{rT})}{\theta_1 S(0)} \right) - \left( \mu - \frac{1}{2} \sigma^2 \right) T}{\sigma\sqrt{T}}\right).
 \end{aligned}$$

## 2.2. Construction of Brownian Motion

For a more rigorous exposition of this material, see [KS91, §2.3]. Our approach starts with an infinite sequence of independent standard normal random variables  $(Z_n, n \in \mathbb{N})$ . The strategy is to use this sequence of independent normals to construct a sequence of stochastic processes  $W^{(m)}$  that converge to a Wiener process for  $t \in [0, 1]$ . We want the limit stochastic process given by  $W(\omega, t) = \lim_{m \rightarrow \infty} W^{(m)}(\omega, t)$  to satisfy the definition of the Wiener process. The point is that each  $W^{(m)}$  involves only a finite number of random variables, so we avoid the worst perplexities of measure theory, at the price of having to think about convergence.

The construction of  $W^{(m)}$  begins with the Haar functions  $H_n : [0, 1] \rightarrow \mathbb{R}$  defined as  $H_1(t) = 1$  and

$$H_{2^m+k}(t) = \begin{cases} 2^{m/2} & \text{for } t \in \left[\frac{k-1}{2^m}, \frac{k-1}{2^m} + \frac{1}{2^{m+1}}\right] \\ -2^{m/2} & \text{for } t \in \left(\frac{k}{2^m} - \frac{1}{2^{m+1}}, \frac{k}{2^m}\right] \\ 0 & \text{otherwise} \end{cases}$$

where  $2^m + k = n$  is the unique representation of  $n$  such that  $k$  is an integer between 1 and  $2^m$ . Note that  $(k-1)/2^m + 1/2^{m+1} = k/2^m - 1/2^{m+1}$ . For an equivalent definition, see [Mik98, pp. 51–52]. The first four Haar functions are

$$\begin{aligned} H_1(t) &= 1 \text{ for } t \in [0, 1] \\ H_2(t) &= \begin{cases} 1 & \text{for } t \in [0, 1/2) \\ -1 & \text{for } t \in [1/2, 1] \end{cases} \\ H_3(t) &= \begin{cases} \sqrt{2} & \text{for } t \in [0, 1/4] \\ -\sqrt{2} & \text{for } t \in (1/4, 1/2] \\ 0 & \text{for } t \in (1/2, 1] \end{cases} \\ H_4(t) &= \begin{cases} 0 & \text{for } t \in [0, 1/2) \\ \sqrt{2} & \text{for } t \in [1/2, 3/4] \\ -\sqrt{2} & \text{for } t \in (3/4, 1] \end{cases} \end{aligned}$$

Next come the Schauder functions, defined as the integrals of the Haar functions:  $\tilde{H}_n(t) = \int_0^t H_n(s) ds$ . See [Mik98, pp. 53–54] for pictures. The approximate Wiener process defined for  $t \in [0, 1]$  is

$$W^{(m)}(\omega, t) = \sum_{n=1}^{2^m} Z_n(\omega) \tilde{H}_n(t).$$

It is not theoretically difficult to check that the limit process  $W(t) = \sum_{n=1}^{\infty} Z_n \tilde{H}_n(t)$  has the right distribution. Each random vector of the form  $(W^{(m)}(t_1), \dots, W^{(m)}(t_n))$  has a multivariate normal distribution with mean zero, so this remains true for the limit vector  $(W(t_1), \dots, W(t_n))$ . As for the covariance matrix,

$$\mathbf{Cov}[W(s), W(t)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \tilde{H}_i(s) \tilde{H}_j(t) \mathbf{Cov}[Z_i, Z_j] = \sum_{i=1}^{\infty} \tilde{H}_i(s) \tilde{H}_i(t)$$

and it is a (non-obvious) property of the Schauder functions that this equals  $\min\{s, t\}$ , as desired.

What about continuity of sample paths? Just because  $W(\omega)$  is the limit of continuous sample paths  $W^{(m)}(\omega)$  does not prove that it is continuous. In general the conclusion that the limit of continuous function is itself continuous is justified when the convergence is uniform, meaning that the rate of convergence does not get too slow for some  $t$ . We avoid the details, mentioning only that this condition is met here (with probability one) because

the normal density of the independent  $Z_i$ 's has very light tails: recall that this density is proportional to  $e^{-z^2/2}$ , which is very small for large values  $z$ . This means that it is unlikely for the convergence of  $W^{(m)}(\omega, t)$  to get “held up” far from its limit  $W(\omega, t)$  by the repeated appearance of large and influential values  $Z_n(\omega)$ . So we regard as successful this construction of a Wiener process with the right distribution and continuous sample paths.

So far we have relied on the “nice” properties of the sequences of standard normal random variables  $Z_i$  and Schauder functions  $\tilde{H}_n$  surviving after convergence. However, the sequence of Schauder functions also has a “bad” property: its derivative is unbounded in  $n$ , because the sequence of Haar functions is unbounded in  $n$ . Indeed, the sample paths  $W(\omega)$  of a Wiener process are not differentiable, with probability one.

For the derivative of the sample path  $W(\omega)$  at  $t = 0$  to exist and be finite, it is necessary that the slope  $(W(\omega, s) - W(\omega, 0))/(s - 0)$  be bounded for  $s$  sufficiently near 0. Formally, there must exist a finite bound  $x$  and a positive time  $t > 0$  such that for all  $s \in (0, t]$ , the absolute value of the slope  $|(W(\omega, s) - W(\omega, 0))/(s - 0)| = |W(\omega, s)/s| \leq x$ . The probability that this does *not* happen is greater than or equal to  $\mathbf{P}(|W(s)/s| > x)$ , for any particular  $s \in (0, t]$ . But

$$\lim_{s \rightarrow 0^+} \mathbf{P}(|W(s)/s| > x) = \lim_{s \rightarrow 0^+} \mathbf{P}(|W(1)|/\sqrt{s} > x) = \lim_{s \rightarrow 0^+} 2\Phi(-x\sqrt{s}) = 1,$$

because both  $W(s)/s$  and  $W(1)/\sqrt{s}$  have a normal distribution with mean 0 and variance  $1/s$ , which is unbounded for small  $s$ . This shows that with probability one, the derivative would have to be larger than any finite  $x$ , so the sample path can not have a derivative at  $t = 0$ .

### 2.3. Problems

**PROBLEM 2.1.** In Example 2.1.4, assume the interest rate  $r > 0$  and the drift  $\mu > \sigma^2/2$ , and that the numbers of shares  $\theta_0$  and  $\theta_1$  are both strictly positive. Say whether the probability of loss

- increases
- decreases
- stays the same
- can't tell

in each of the following scenarios:

- (1) The interest rate  $r$  increases.
- (2) The drift  $\mu$  increases.
- (3) The volatility  $\sigma$  increases.
- (4) The time horizon  $T$  increases.
- (5) The ratio  $\theta_0 M(0)/(\theta_1 S(0))$ , of wealth in the money market account to wealth in the stock, increases.

Give your reasoning in each case.

♡PROBLEM 2.2. Show that the increments of a geometric Brownian motion  $Y$  are neither independent nor stationary. Hint: look at the increments  $Y(s) - Y(0)$  and  $Y(2s) - Y(s)$ .

PROBLEM 2.3. What are the mean and covariance functions of a Brownian bridge  $Z$  on the interval  $[0, T]$ , starting at value  $Z(0)$  and ending at  $Z(T)$ , with volatility parameter  $\sigma$ ? Its definition is  $Z(t) = Z(0) + \sigma W(t) + (Z(T) - Z(0) - \sigma W(T))(t/T)$ . At what time  $t$  is the variance  $\mathbf{Var}[Z(t)]$  at its maximum? Hint: check that your mean and covariance functions are correct for the special case of standard Brownian bridge:  $\mu_Z(t) = 0$  and  $c_Z(s, t) = \min\{s, t\} - st$ .

♡PROBLEM 2.4. Let  $(W(t), t \in [0, 1])$  be a Wiener process. Show that the stochastic process  $\tilde{W}$  defined for  $t \in [0, T]$  by  $\tilde{W}(t) = \sqrt{T}W(t/T)$  is a Wiener process.

PROBLEM 2.5. What are the mean and covariance functions of the approximate Wiener processes  $W^{(m)}$  for  $m \in \{0, 1, 2\}$ ? For the covariance function, give  $c_{W^{(m)}}(s, t)$  only for the case  $0 \leq s \leq t \leq 1$ . Present your answer by dividing this region of pairs of times into 3 sub-regions for  $m = 1$  and 10 sub-regions for  $m = 2$ . Next define  $G^{(m)}$  as the set of pairs  $(s, t)$  for which  $c_{W^{(m)}}(s, t) = s$ , as it does for true Brownian motion. What happens to this set  $G^{(m)}$  as  $m$  grows? Warning: this problem is computationally intensive.

## 2.4. Definition of Stochastic Integration

Now that we understand the Wiener process, we can investigate integration and differential equations involving it. This piece of mathematics is associated with the name of Kiyosi Itô, a Japanese mathematician working during World War II on the problem of controlling long-range rockets. A fundamental tool for handling stochastic differential equations is Itô's formula, which we discuss in Section 2.5. This comes into play in our first attack on the Black-Scholes option pricing formula, in Chapter 3.

In the approach we take to developing this subject, integration is more fundamental than differential equations. In this section, we develop the Itô stochastic integral by extension from integration of discrete-time stochastic processes. This relates to our motivation, which is to compute the gains process of a self-financing portfolio.

The gain from a discrete-time,  $m$ -step portfolio strategy  $\theta$  in  $N$  assets with price vector  $S$  over the time interval  $[0, t]$  is

$$(2.4.1) \quad G(t) = \sum_{j=1}^m \theta(t_{j-1})(S(t_j) - S(t_{j-1})) = \sum_{i=1}^N \sum_{j=1}^m \theta_i(t_{j-1})(S_i(t_j) - S_i(t_{j-1})).$$

To evaluate the gain from a continuous-time portfolio strategy, we must define a stochastic integral  $\int_0^T \theta_i(t) dS_i(t)$  to replace the stochastic sum  $\sum_{j=1}^m \theta_i(t_{j-1})(S_i(t_j) - S_i(t_{j-1}))$ . In a similar way, the familiar Riemann integral replaces  $\sum_{j=1}^m f(t_{j-1})(t_j - t_{j-1})$  with  $\int_0^T f(t) dt$  by setting each  $t_j - t_{j-1}$  to be a constant  $\Delta t$ , and taking a limit as  $\Delta t$  goes to 0. However, the situation is not nearly so simple here, because  $S$  is a stochastic process.

The grand strategy for constructing the Itô stochastic integral  $\int X(t) dW(t)$  is:

- (1) Define the Itô integral for the aptly named simple processes.
- (2) Approximate a more general process  $X$  with a sequence of simple processes  $C^{(m)}$ .
- (3) The Itô integral of  $X$  is the limit of the integrals of  $C^{(m)}$ .

Unfortunately, this construction is very opaque because it does not work pathwise. That is, we do not define from the sample paths  $W(\omega)$  of the Wiener process and  $X(\omega)$  of the integrand process a path  $\int_0^t X(\omega, s) dW(\omega, s)$  for the Itô integral. Instead, the integral's definition involves the convergence of stochastic processes. Remember that the Itô integral  $\int_0^t X(s) dW(s)$  is a random variable if  $t$  is regarded as a fixed time, and a stochastic process if  $t$  is regarded as the time index of the stochastic process.

A **simple process**  $C$  on  $[0, T]$  is one that for some partition  $(t_0, \dots, t_n)$  satisfies

$$C(t) = C(t_{i-1}) \quad \text{for } t \in [t_{i-1}, t_i).$$

A simple process is not as complicated as a full-blown stochastic process, which has a different random variable associated with each time. A simple process contains only a finite number of different random variables, namely  $(C(t_i), i = 0, \dots, n)$ .

**EXAMPLE 2.4.1** (Simple processes). Take a set of time intervals  $[a_1, b_1), \dots, [a_m, b_m)$  and a set of random variables  $X_1, \dots, X_m$ . Then the process  $C(t) = \sum_{k=1}^m X_k \mathbf{1}_{[a_k, b_k)}(t)$  is simple. The approximate Wiener processes  $W^{(m)}$  of Section 2.2 are not simple, because the Schauder functions change continuously over time. A Poisson process is not simple: although its sample paths are piecewise constant, the times at which its value changes are random, not fixed. It is possible that the Poisson process changes its value on any subinterval  $[t_{i-1}, t_i]$ ; the random variables  $N(t_{i-1})$  and  $N(t_i)$  are never the same.

For a simple process, the Itô integral is just a sum:

$$\int_0^T C(t) dW(t) = \sum_{i=1}^n C(t_{i-1}) (W(t_i) - W(t_{i-1})).$$

Notice the similarity to the discrete-time gains process mentioned at the beginning of this section. The Itô integral of a simple process  $C$  on  $[0, T]$  is a random variable, so we would like to be able to compute its moments. Its expectation and variance are

$$(2.4.2) \quad \mathbf{E} \left[ \int_0^T C(t) dW(t) \right] = 0$$

$$(2.4.3) \quad \mathbf{Var} \left[ \int_0^T C(t) dW(t) \right] = \mathbf{E} \left[ \left( \int_0^T C(t) dW(t) \right)^2 \right] = \int_0^T \mathbf{E} [(C(t))^2] dt.$$

The latter equation is called the **Itô isometry**, because it shows how two measures (metrics) are the same (iso): if you measure the size of the stochastic integral  $\int_0^T C(t) dW(t)$  by its variance, you get the same result as if you measure the size of the stochastic process  $C(t)$  by the time integral of its second moment over  $[0, T]$ . We defer derivations of these properties until Section 4.3.

For the moment, we can gain some intuition by making the simple process  $C$  a deterministic function  $f$  of time, still piecewise constant and changing only at the times  $t_1, \dots, t_n$ . In that case we would have

$$\mathbf{E} \left[ \int_0^T f(t) dW(t) \right] = \sum_{i=1}^n f(t_{i-1}) \mathbf{E}[W(t_i) - W(t_{i-1})] = 0$$

and

$$\begin{aligned} \mathbf{Var} \left[ \int_0^T f(t) dW(t) \right] &= \mathbf{Var} \left[ \sum_{i=1}^n f(t_{i-1}) (W(t_i) - W(t_{i-1})) \right] \\ &= \sum_{i=1}^n f^2(t_{i-1}) \mathbf{Var} [W(t_i) - W(t_{i-1})] \\ &= \sum_{i=1}^n f^2(t_{i-1}) (t_i - t_{i-1}) \\ &= \int_0^T f^2(t) dt, \end{aligned}$$

which makes it clear that the integral  $\int_0^T f(t) dW(t)$  is accumulating variance at rate  $f(t)^2$  at time  $t$ .

We proceed to Itô integration of more general processes  $X$ , subject to the restriction that  $X(\omega, t)$  be a function of the Wiener process sample path history  $(W(\omega, s), s \in [0, t])$ , and

$$(2.4.4) \quad \int_0^T \mathbf{E} [(X(t))^2] dt < \infty.$$

This is the defining criterion of Björk's class  $\mathcal{L}^2$ , see [Bjö98, Def. 3.3]. The former restriction makes perfect sense from the standpoint of the application to finance: one's portfolio strategy will be determined by the only relevant and available information, namely asset price histories, which are all functions of the driving Wiener process. The latter restriction is mathematically excessive, but it is convenient to adopt because of its relation to the Itô isometry: it serves as a guarantee that the stochastic integral of  $X$  will have finite variance. Infinite variance is not a mathematically monstrous property. Real-world processes such as insurance claims might well have infinite variance, but it is easier for us to deal with finite-variance models.

It is not difficult to imagine constructing a sequence of simple functions  $C^{(m)}$  that approximate  $X$  better and better, much as we saw how our approximate Wiener processes  $W^{(m)}$  got closer and closer to a true Wiener process. We do not need to investigate the mathematical subtleties; rather, we will accept that this can be done and moreover that no matter what sequence of approximating simple functions we use, we get the same limit  $\lim_{m \rightarrow \infty} \int_0^T C^{(m)}(t) dW(t)$  and call it  $\int_0^T X(t) dW(t)$ . This is by no means obvious.

Because we can not see the workings of the Itô integral, we tend to lose intuition. We will regain intuition after learning the rules of Itô integration and practicing it. In terms of applications, think of it this way: if  $X$  represents a human response to observing the phenomenon driven by a Wiener process  $W$ , the response can not be instantaneous and continuous; perhaps a simple process  $C$  would be a better way to model human response. In particular, this holds for our model of portfolio strategies. Nonetheless, we may be very close to the continuous limit, which is actually more mathematically tractable to compute once you know stochastic calculus: after all, who would want to make computations in physical dynamics as sums over small discrete time intervals? When you first learn it, classical calculus is more opaque than summation, but its convenience and elegance makes up for that, once it has become familiar with use.

### 2.5. Itô's Formula

The major computational tool for the Itô integral is Itô's formula. (This is often called Itô's lemma. Since we will use it to compute things, not prove things, we will call it a formula, not a lemma.) This formula is a substitute for the chain rule of ordinary calculus:

$$\frac{d}{dt}f(g(t)) = f'(g(t))g'(t),$$

which leads to the equation

$$\int_a^b f'(g(t))g'(t) dt = f(g(b)) - f(g(a)),$$

so it is a computational tool for evaluating integrals as well as derivatives. This is equivalent to making the formal substitution  $g'(t) dt = dg(t)$ . We would like to have similar rules for manipulating infinitesimals such as  $dt$  and  $dW(t)$ . We should regard these rules as a shorthand for corresponding statements about integrals. Here are the rules for multiplying the infinitesimals  $dt$ ,  $dW_1(t)$ , and  $dW_2(t)$ , where  $W$  is a multidimensional Wiener process, which has independent components.

	$dt$	$dW_1(t)$	$dW_2(t)$
$dt$	0	0	0
$dW_1(t)$	0	$dt$	0
$dW_2(t)$	0	0	$dt$

Why? Imagine  $dt \rightarrow 0$  and  $dW_i(t) = (W_i(t + dt) - W_i(t))$  for  $i = 1, 2$ . The principle is that  $(dt)^2$  is much smaller than  $dt$  when the latter is near 0, so  $(dt)^2 \approx 0$ . Anything going to 0 faster than  $dt$  is relatively negligible. Then  $dt dW_i(t)$  is a random variable with distribution  $N(0, (dt)^3)$ , so  $dt dW_i(t) \approx 0$ . However  $(dW_i(t))^2$  is a random variable with expectation  $dt$  and variance

$$\text{Var} [(dW_i(t))^2] = \mathbf{E} [(dW_i(t))^4] - \mathbf{E} [(dW_i(t))^2]^2 = 3(dt)^2 - (dt)^2 = 2(dt)^2$$

so  $(dW_i(t))^2 \approx dt$ . But  $dW_1(t) dW_2(t)$  has expectation 0 and variance

$$\begin{aligned} \mathbf{Var} [dW_1(t) dW_2(t)] &= \mathbf{E} [(dW_1(t) dW_2(t))^2] - \mathbf{E} [dW_1(t) dW_2(t)]^2 \\ &= \mathbf{E} [(dW_1(t))^2] \mathbf{E} [(dW_2(t))^2] - 0 \\ &= (dt)^2 \end{aligned}$$

so  $dW_1(t) dW_2(t) \approx 0$ .

Using these heuristic rules, we derive (but do not prove) Itô's formula via a Taylor expansion. For a sufficiently differentiable function  $f$ , whose  $k$ th derivative is  $f^{(k)}$ , and partitioning the time interval  $[s, T]$  into  $m$  evenly spaced steps  $t_1, \dots, t_m$ ,

$$\begin{aligned} f(W(T)) - f(W(s)) &= \sum_{i=1}^m (f(W(t_i)) - f(W(t_{i-1}))) \\ &= \lim_{m \rightarrow \infty} \sum_{i=1}^m \sum_{k=1}^{\infty} \frac{1}{k!} f^{(k)}(W(t_{i-1})) (W(t_i) - W(t_{i-1}))^k \\ &= \int_s^T \left( f'(W(t)) dW(t) + \frac{1}{2} f''(W(t)) (dW(t))^2 \right) \end{aligned}$$

and  $(dW(t))^2 = dt$ , so we get the formula

$$f(W(T)) = f(W(s)) + \int_s^T f'(W(t)) dW(t) + \frac{1}{2} \int_s^T f''(W(t)) dt$$

where the first integral is an Itô stochastic integral and the second is a Riemann integral over time. Equivalently, in differential form,

$$(2.5.1) \quad df(W(t)) = f'(W(t)) dW(t) + \frac{1}{2} f''(W(t)) dt.$$

Compare this to the ordinary chain rule  $dg(t) = g'(t) dt$ . There is an extra term because the squared infinitesimal changes of the Wiener process  $W(t)$  are nonnegligible, unlike the squared infinitesimal changes of the degenerate stochastic process  $t$ .

Itô's formula applies if  $f$  is only twice continuously differentiable; we only use two derivatives in the formula. Subsequent versions of Itô's formula will still need this differentiability condition, but it will not be repeated explicitly.

**EXAMPLE 2.5.1** (Simple Itô computations).

(1) With the Itô formula, using  $f(x) = x$ , we can verify

$$\int_s^T dW(t) = \int_s^T 1 dW(t) = W(T) - W(s) - \frac{1}{2} \int_s^T 0 dt = W(T) - W(s),$$

which is what it ought to be.



- (2) Suppose we want to evaluate  $\int_0^T W(t) dW(t)$ . Then to use Itô's formula, we need  $f'(W(t)) = W(t)$ , i.e.  $f'(x) = x$ . Then by ordinary integration,  $f(x) = x^2/2 + C$ , and by differentiation,  $f''(x) = 1$ . So the formula yields

$$\int_0^T W(t) dW(t) = \left( \frac{W(T)^2}{2} + C \right) - \left( \frac{W_0^2}{2} + C \right) - \frac{1}{2} \int_0^T dt = \frac{1}{2}(W(T)^2 - T).$$

From this we can see that it is always convenient to take the constant of integration  $C = 0$ .

Before continuing to more interesting examples, we state two useful rules. Both are subject to some technical conditions that we will disregard.

- (1) The stochastic integral  $\int_s^T f(t) dW(t)$  has a normal distribution with mean zero and variance  $\int_s^T f^2(t) dt$ . This is the variance given by the Itô isometry; see the discussion of deterministic simple functions there for a justification. When integrating a stochastic process, not a deterministic function, the integral does *not* have to come out normal: this is important, not just a technicality.
- (2) When  $X(t) = g(W(t))$ , the expectation

$$\mathbf{E} \left[ \int_s^T X(t) dt \right] = \int_s^T \mathbf{E}[X(t)] dt,$$

that is, one may interchange expectation and integration. This makes sense because expectation is a type of integration, and we are used to being able to change the order of integration. In this class, you should feel free to interchange expectation and integration without worrying about the technicalities.

This will help us in the following example.

EXAMPLE 2.5.2 (Higher normal moments).

- (1) Let's evaluate  $\int_0^T W^2(t) dW(t)$ . Where  $f'(x) = x^2$ ,  $f(x) = x^3/3$ , and  $f''(x) = 2x$ , so using Itô's formula,

$$\int_0^T W^2(t) dW(t) = \frac{1}{3}W^3(T) - \frac{1}{3}W^3(0) - \frac{1}{2} \int_0^T 2W(t) dt = \frac{1}{3}W^3(T) - \int_0^T W(t) dt$$

and for the moment we are stuck, unable to deal with the time integral  $\int_0^T W(t) dt$ . We will return to this example later. However, by interchanging expectation and integration, we see that the expectation  $\mathbf{E}[\int_0^T W(t) dt] = \int_0^T \mathbf{E}[W(t)] dt = 0$ . The expectation of the stochastic integral  $\int_0^T W(t)^2 dW(t)$  is also zero. Therefore  $W^3(T)$  also has zero expectation. In particular, since  $W(1)$  is standard normal, this shows that the third moment  $\mathbf{E}[W^3(1)]$  of a standard normal is zero.

- (2) Let's try to repeat this success for the fourth moment  $\mathbf{E}[W^4(1)]$ . In this case,  $f(x) = x^4$ , so  $f'(x) = 4x^3$  and  $f''(x) = 12x^2$ . Thus

$$W^4(1) = W^4(0) + 4 \int_0^1 W^3(t) dW(t) + 6 \int_0^1 W^2(t) dt$$

and the first term is 0, the stochastic integral has 0 expectation, and the expectation of the last term is

$$6\mathbf{E} \left[ \int_0^1 W^2(t) dt \right] = 6 \int_0^1 \mathbf{E}[W^2(t)] dt = 6 \int_0^1 t dt = 3$$

by interchanging order again. Thus we have computed the fourth moment of a standard normal random variable by stochastic integration, without any messy classical integration of the probability density function.

## 2.6. Itô Processes

We begin by extending Itô's formula to allow the function to depend on time: we have a function  $f(t, x)$ , and are interested in the increment  $f(T, W(T)) - f(s, W(s))$ , or the differential  $df(t, W(t))$ . One important example is the stock price in the Black-Scholes model, in which case  $f(t, x) = S(0) \exp((\mu - \sigma^2/2)t + \sigma x)$ .

The Taylor expansion now yields terms with  $dt$ ,  $dW(t) = W(t + dt) - W(t)$ , and  $(dW(t))^2 = dt$ . Remember that other terms, including  $(dt)^2$  and  $dt dW(t)$ , are negligible. We need the partial derivatives  $f_x(t, x) = \partial f(t, x)/\partial x$ ,  $f_t(t, x) = \partial f(t, x)/\partial t$ , and  $f_{xx}(t, x) = \partial^2 f(t, x)/\partial x^2$ . We get

$$f(T, W(T)) = f(s, W(s)) + \int_s^T f_x(t, W(t)) dW(t) + \int_s^T \left( f_t(t, W(t)) + \frac{1}{2} f_{xx}(t, W(t)) \right) dt$$

or

$$(2.6.1) \quad df(t, W(t)) = \left( f_t(t, W(t)) + \frac{1}{2} f_{xx}(t, W(t)) \right) dt + f_x(t, W(t)) dW(t).$$

EXAMPLE 2.6.1 (Example 2.5.2 revisited). In Example 2.5.2, we got stuck when faced with the time integral  $\int_0^T W(t) dt$ . To make this integral appear in Itô's formula, we try to choose  $f_t(t, x) = x$  and  $f_{xx}(t, x) = 0$ . One simple way of doing this is  $f(t, x) = tx$ , which has  $f_x(t, x) = t$ . Then the formula says

$$TW(T) = 0 + \int_0^T W(t) dt + \int_0^T t dW(t) \quad \text{or} \quad \int_0^T W(t) dt = \int_0^T (T - t) dW(t).$$

This stochastic integral of a deterministic function is normal with zero mean and variance  $\int_0^T (T - t)^2 dt = T^3/3$ .

EXAMPLE 2.6.2 (Itô exponential). The process  $X(t) = \exp(W(t) - t/2)$  is the "Itô exponential," i.e. the process that satisfies

$$dX(t) = X(t) dW(t).$$

How is this done? The trick is to make the above equation match the result of Itô's formula. This means we need to have  $f_x(t, x) = f(t, x)$ , and also  $f_t(t, x) + f_{xx}(t, x)/2 = 0$ . This is a sort of puzzle we need to solve. To get  $f_x(t, x) = f(t, x)$  suggests that  $f$  must be exponential in  $x$ , in which case we will also get  $f_{xx}(t, x) = f(t, x)$ . Then we need  $f_t(t, x) = -f(t, x)/2$ , which suggests that  $f$  is exponential in  $-t/2$ . Putting all the clues together, we try  $f(t, x) = \exp(x - t/2)$ , which works.

This is no mere mathematical curiosity, but rather a prototype for geometric Brownian motion, thus for the stock price in the Black-Scholes model.

EXAMPLE 2.6.3 (Black-Scholes stock). The geometric Brownian motion

$$S(t) = S(0) \exp \left( \left( \mu - \frac{1}{2} \sigma^2 \right) t + \sigma W(t) \right)$$

fits the form we are discussing, with  $f(t, x) = S(0) \exp((\mu - \sigma^2/2)t + \sigma x)$ . Then

$$\begin{aligned} f_t(t, x) &= (\mu - \sigma^2/2) S(0) \exp((\mu - \sigma^2/2)t + \sigma x) = (\mu - \sigma^2/2) f(t, x) \\ f_x(t, x) &= \sigma S(0) \exp((\mu - \sigma^2/2)t + \sigma x) = \sigma f(t, x) \\ f_{xx}(t, x) &= \sigma^2 S(0) \exp((\mu - \sigma^2/2)t + \sigma x) = \sigma^2 f(t, x) \end{aligned}$$

so the formula gives

$$dS(t) = S(t) (\mu dt + \sigma dW(t)).$$

Thus while the generalized Brownian motion  $\ln S(t) = \ln S(0) + (\mu - \sigma^2/2)t + \sigma W(t)$  has **arithmetic drift**  $\mu - \sigma^2/2$  and **arithmetic volatility**  $\sigma$ , the geometric Brownian motion  $S(t) = \exp(\ln S(t))$  has **geometric drift**  $\mu$  and **geometric volatility**  $\sigma$ . The geometric drift and volatility of  $S(t)$  are formed by dividing the integrands by  $S(t)$ . Processes of the general form

$$(2.6.2) \quad X(t) = X(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s) \quad \text{or} \quad dX(t) = \mu(t) dt + \sigma(t) dW(t),$$

where  $\mu$  and  $\sigma$  are now stochastic processes driven by the Wiener process  $W$ , are called **Itô processes**. Generalized Brownian motion is an Itô process where  $\mu$  and  $\sigma$  are constant, and geometric Brownian motion is an Itô process where  $\mu$  and  $\sigma$  are proportional to the geometric Brownian motion itself. Any positive Itô process can be written, like geometric Brownian motion, in the form

$$(2.6.3) \quad \begin{aligned} Y(t) &= Y(0) + \int_0^t Y(s) \mu(s) ds + \int_0^t Y(s) \sigma(s) dW(s) \\ dY(t) &= Y(t) (\mu(t) dt + \sigma(t) dW(t)) \end{aligned}$$

where  $\mu$  and  $\sigma$  are now its geometric drift and volatility. Its arithmetic drift and volatility would be  $Y\mu$  and  $Y\sigma$ .

We want to be able to do stochastic integration with respect to an Itô process, not just with respect to a Wiener process. In particular, we want to be able to deal with a stochastic

integral such as  $\int_0^T \theta(t) dS(t)$ , representing the portfolio gains process. For Itô processes  $X$  and  $Y$ , with  $X$  as given in (2.6.2), the stochastic integral

$$\int_s^T Y(t) dX(t) = \int_s^T Y(t)\mu(t) dt + \int_s^T Y(t)\sigma(t) dW(t).$$

We know what both of these terms mean: one is a time integral, and the other is a stochastic integral with respect to a Wiener process. This equation relies on the formal substitution  $dX(t) = \mu(t) dt + \sigma(t) dW(t)$ .

The Itô formula for a function of an Itô process  $X$  as given in (2.6.2) is

$$\begin{aligned} df(t, X(t)) &= \left( f_t(t, X(t)) + \frac{1}{2}\sigma(t)^2 f_{xx}(t, X(t)) \right) dt + f_x(t, X(t)) dX(t) \\ (2.6.4) &= \left( f_t(t, X(t)) + \mu(t)f_x(t, X(t)) + \frac{1}{2}\sigma(t)^2 f_{xx}(t, X(t)) \right) dt + \sigma(t)f_x(t, X(t)) dW(t). \end{aligned}$$

EXAMPLE 2.6.4 (exp). Consider  $X$  an Itô process of the form (2.6.2), i.e. with *arithmetic* drift  $\mu$  and volatility  $\sigma$  (these can be stochastic processes). Taking  $f(t, x) = \exp(x)$ , so  $f_t(t, x) = 0$  and  $f_x(t, x) = f_x(t, x) = f_{xx}(t, x)$ , we find

$$\begin{aligned} \exp(X(t)) &= \exp(X(0)) + \int_0^t \left( \mu(s) \exp(X(s)) + \frac{1}{2}\sigma^2(s) \exp(X(s)) \right) ds \\ &\quad + \int_0^t \sigma(s) \exp(X(s)) dW(s) \end{aligned}$$

and thus  $\exp(X)$  is also an Itô process, with *geometric* drift  $\mu + \sigma^2/2$  and volatility  $\sigma$ .

EXAMPLE 2.6.5 (ln). Take  $Y$  a positive Itô process with *geometric* drift  $\mu$  and volatility  $\sigma$ , thus *arithmetic* drift  $\mu Y$  and volatility  $\sigma Y$ . Let  $f(t, x) = \ln(x)$ , so  $f_t(t, x) = 0$ ,  $f_x(t, x) = x^{-1}$ ,  $f_{xx}(t, x) = -x^{-2}$ , we find

$$\ln(Y(t)) = \ln(Y(0)) + \int_0^t \left( \mu(s) - \frac{1}{2}\sigma(s)^2 \right) ds + \int_0^t \sigma(s) dW(s)$$

because the factors  $Y^{-1}$  and  $Y^{-2}$  from  $f_x$  and  $f_{xx}$  cancel the factors of  $Y$  from the arithmetic drift and volatility. Thus  $\ln(Y)$  is also an Itô process, with *arithmetic* drift  $\mu - \sigma^2/2$  and volatility  $\sigma$ .

EXAMPLE 2.6.6 (Extended Black-Scholes Binary Option). In Example 2.6.5, consider the case where  $S = Y$  has deterministic (but not necessarily constant) geometric drift  $\mu$  and volatility  $\sigma$ . We might use this to model a stock price, as a slight extension of the Black-Scholes model. Now what is the probability that  $S(T)$  exceeds some level  $K$ ? (Compare with Example 2.1.3.) As in Example 2.6.5,  $\ln(S(T)) = \ln(S(0)) + \int_0^T (\mu(t) - \sigma(t)^2/2) dt + \int_0^T \sigma(t) dW(t)$ . The first two terms are constants. The second term is a stochastic integral

of a deterministic function, so it is normal with mean 0 and variance  $\int_0^T \sigma(t)^2 dt$ . Therefore

$$\ln(S(T)) \sim N \left( \ln(S(0)) + \int_0^T (\mu(t) - \sigma(t)^2/2) dt, \int_0^T \sigma(t)^2 dt \right)$$

and the probability that  $S(T)$  exceeds  $K$  is

$$\begin{aligned} \mathbf{P}[S(T) > K] &= \mathbf{P}[\ln(S(T)) > \ln(K)] = \mathbf{P}[-\ln(S(T)) < -\ln(K)] \\ &= \Phi \left( \frac{-\ln(K) - \mathbf{E}[-\ln(S(T))]}{\sqrt{\mathbf{Var}[-\ln(S(T))]}} \right) \\ &= \Phi \left( \frac{\ln(S(0)/K) + \int_0^T (\mu(t) - \sigma(t)^2/2) dt}{\sqrt{\int_0^T \sigma(t)^2 dt}} \right). \end{aligned}$$

## 2.7. Summary

The Wiener process (standard Brownian motion) has continuous sample paths and stationary, independent increments with a normal distribution whose variance is proportional to the length of the increment. Generalized and geometric Brownian motion, as well as Brownian bridge, are transformations of the Wiener process. The Black-Scholes model for a stock price is geometric Brownian motion. The Wiener process can be constructed as a limit by taking the running sum of more and more normal random variables. Thus the Black-Scholes model has an interpretation in terms of a continuous stream of news.

The Itô stochastic integral relates to portfolio gains. It is defined for simple processes as a sum, and for more complicated processes as the limit of stochastic integrals of simple processes. It is often convenient to write stochastic integral equations in differential form. Itô's formula comes from a second-order Taylor expansion and relates the stochastic integral of  $f'(W(t))$  to  $f$  and the time integral of  $f''(W(t))$ . Equivalently, it relates  $df(W(t))$  to  $f'(W(t))$ ,  $f''(W(t))$ ,  $dt$ , and  $dW(t)$ . The time-dependent version relates  $df(t, W(t))$  to  $f_t(t, W(t))$ ,  $f_x(t, W(t))$ ,  $f_{xx}(t, W(t))$ ,  $dt$ , and  $dW(t)$ . We can also define a stochastic integral with respect to an Itô process  $X$  rather than just a Wiener process  $W$ , and get an Itô formula for  $df(X(t))$ .

## 2.8. Problems

**PROBLEM 2.6.** Let  $Y(t) = \int_0^t X(s) dW(s)$ , where  $X$  is an arbitrary suitable stochastic process. What are the expectation and variance of  $Y(t)$ ? Does  $Y$  have independent increments?

**PROBLEM 2.7** (Exercises 3.1, 3.2 of [Bjö98]). Compute  $dY(t)$  when the stochastic process  $Y$  is defined by

- (1)  $Y(t) = \int_0^t X(s) dW(s)$ , where  $X$  is an arbitrary suitable stochastic process
- (2)  $Y(t) = \exp(\alpha W(t))$
- (3)  $Y(t) = \exp(\alpha X(t))$ , where  $dX(t) = \mu dt + \sigma dW(t)$

- (4)  $Y(t) = X^2(t)$ , where  $dX(t) = \mu X(t) dt + \sigma X(t) dW(t)$
- (5)  $Y(t) = 1/X(t)$ , where  $dX(t) = \mu X(t) dt + \sigma X(t) dW(t)$

In these definitions,  $\mu$  and  $\sigma$  are constants. A “suitable” stochastic process means one in Björk’s class  $\mathcal{L}^2$ , i.e. satisfying the technical requirements to be a good stochastic integrand. In the last two parts, try to express  $dY(t)$  in terms of  $Y$ , not  $X$ .

**PROBLEM 2.8.** Let  $X(0)$  and  $x$  be constants. In each of the following cases, evaluate  $\mathbf{P}[X(T) < x]$ , or say why you can not evaluate it.

- (1)  $dX(t) = \mu dt + \sigma dW(t)$  where  $\mu$  and  $\sigma$  are constants
- (2)  $dX(t) = X(t)(\mu dt + \sigma dW(t))$  where  $\mu$  and  $\sigma$  are constants
- (3)  $dX(t) = \mu(t) dt + \sigma(t) dW(t)$  where  $\mu$  and  $\sigma$  are stochastic processes

## CHAPTER 3

### The Black-Scholes Analysis: Part I

In this chapter, we

- (1) use no-arbitrage reasoning to derive the Black-Scholes PDE
- (2) define the greeks and show their relationship to the Black-Scholes PDE
- (3) state the Black-Scholes formula and show it satisfies the Black-Scholes PDE

#### 3.1. The Black-Scholes PDE

Before developing a general theory in Chapter 6, we use the Black-Scholes model as an extended example of no-arbitrage derivative pricing. We consider a path-independent derivative with a single payoff  $g(S(T))$  at time  $T$ , such as the European call option, which pays  $g(S(T)) = (S(T) - K)^+$ . Since we have not yet developed a general theory, for the moment we simply *assume* that

- (1) There is a self-financing portfolio strategy  $\theta$  that replicates the option's payoff, meaning that  $g(S(T)) = V(T) = \theta_0(T)M(T) + \theta_1(T)S(T)$ .
- (2) There is a sufficiently differentiable function  $f(t, S)$  of time  $t$  and stock price  $S$  that gives the unique no-arbitrage price of the option.

When we complete the analysis, we will have verified these assumptions, justifying the derivation.

Recall the Black-Scholes model has a money market account  $M(t) = e^{rt}$  growing exponentially and a stock  $S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t))$  following geometric Brownian motion. In Example 2.6.3, we saw that

$$dS_t = S_t(\mu dt + \sigma dW_t).$$

This shows that  $\mu$  is the expected geometric growth rate of the stock. Likewise,  $M(t) = f(t, W(t))$  where  $f(t, x) = e^{rt}$ , so

$$dM(t) = re^{rt} dt = rM(t) dt,$$

which was obvious from ordinary calculus. The continuously compounded interest rate  $r$  is the geometric growth rate of the money market account.

We now analyze a replicating strategy for the option. The no-arbitrage principle says that the option price  $f(t, S(t)) = V(t)$ , the value of the replicating portfolio. This implies  $df(t, S(t)) = dV(t)$ . The self-financing condition  $dV(t) = dG(t) = \theta_0(t) dM(t) + \theta_1(t) dS(t)$ , and plugging in for  $dS(t) = S(t)(\mu dt + \sigma dW(t))$  and  $dM(t) = rM(t) dt$ , we have

$$df(t, S(t)) = dV(t) = (\theta_0 M(t)r + \theta_1 S(t)\mu) dt + \theta_1 S(t)\sigma dW(t).$$

The strategy is to use Itô's formula and the no-arbitrage principle to get a PDE for  $f$ , which we will then solve.

First, apply the time-dependent Itô formula to this option price function  $f$  to get another expression for  $df(t, S(t))$ . The formula says, where the input Itô process has  $dX(t) = \mu(t) dt + \sigma(t) dW(t)$ , that

$$df(t, X(t)) = \left( f_t(t, X(t)) + \mu(t)f_x(t, X(t)) + \frac{\sigma(t)^2}{2}f_{xx}(t, X(t)) \right) dt + \sigma(t)f_x(t, X(t)) dW(t).$$

Here the stock price process  $S$  is given by  $dS(t) = S(t)(\mu dt + \sigma dW(t))$ , so we get

$$\begin{aligned} df(t, S(t)) &= \left( f_t(t, S(t)) + S(t)\mu f_S(t, S(t)) + \frac{\sigma^2}{2}S(t)^2 f_{SS}(t, S(t)) \right) dt \\ &\quad + S(t)\sigma f_S(t, S(t)) dW(t). \end{aligned}$$

Therefore the number of shares of stock in the replicating portfolio  $\theta_1(t) = f_S(t, S(t))$ , the first derivative of the option price with respect to the stock price, also known as  $\Delta$ .

By the no-arbitrage principle,  $f(t, S(t)) = V(t) = \theta_0(t)M(t) + \theta_1(t)S(t)$ . Rearranging and plugging in for  $\theta_1$ ,  $\theta_0(t)M(t) = f(t, S(t)) - f_S(t, S(t))S(t)$ . Therefore the drift of  $f(t, S(t))$  is

$$\begin{aligned} \theta_0(t)M(t)r + \theta_1(t)S(t)\mu &= f_S(t, S(t))S(t)\mu + (f(t, S(t)) - f_S(t, S(t))S(t))r \\ &= rf(t, S(t)) + (\mu - r)S(t)f_S(t, S(t)). \end{aligned}$$

Equating this with the drift from Itô's formula yields the Black-Scholes PDE

$$(3.1.1) \quad r(Sf_S(t, S) - f(t, S)) + f_t(t, S) + \frac{1}{2}\sigma^2 S^2 f_{SS}(t, S) = 0.$$

Note that this applies to any derivative, whatever the terminal payoff  $g$ .

It is highly significant that  $\mu$ , the geometric drift of the stock, does not appear in this PDE. In this model, it is irrelevant for pricing derivatives. This should be surprising. This mean growth rate  $\mu$  summarizes investors' attitudes towards risk: the more risk-averse they are, the greater the expected rewards they demand for holding risk. Let us think of the distribution of terminal stock price  $S(T)$  as being fixed. Then more risk-aversion means greater drift  $\mu$  and consequently lower initial stock price  $S(0)$ . The initial no-arbitrage call price is  $\theta_0 M(0) + \theta_1 S(0)$ , so we can see that risk preferences *are* reflected in the call price, but they enter through  $S(0)$ , not through  $\mu$ . This is very good for financial engineers, because  $S(0)$  is directly observable in the marketplace, while  $\mu$  turns out to be effectively impossible to estimate!

### 3.2. The Black-Scholes Formula and Greeks

In the Black-Scholes model, pricing the European call option means finding the solution for  $t \in [0, T]$  to the Black-Scholes PDE with the terminal condition  $f(T, S(T)) = (S(T) - K)^+$ . (This is much like an initial condition for a differential equation, but it is called terminal in recognition of the presence of time  $T$  not 0.) At this point, we simply pull the



solution out of a hat and verify that it solves the Black-Scholes PDE. Later we will see how to derive this solution using the Feynman-Kač formula and Girsanov transformation.

The Black-Scholes European call pricing formula is

$$f(t, S) = S\Phi(d_1) - e^{-r(T-t)}K\Phi(d_2)$$

where

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \quad \text{and} \quad d_2 = d_1 - \sigma\sqrt{T-t}.$$

The Black-Scholes PDE involves three derivatives of the pricing function  $f(t, x)$ : the time derivative  $f_t(t, x)$ , the first spatial derivative  $f_x(t, x)$ , and the second spatial derivative  $f_{xx}(t, x)$ . Since the spatial dimension ends up having the interpretation of the stock price  $S$ , these (calculus) partial derivatives describe the sensitivity of the (financial) derivative security to the passage of time and changes in the stock price. They get special names: **theta** is  $\Theta = f_t(t, S(t))$ , **delta** is  $\Delta = f_x(t, S(t))$ , and **gamma** is  $\Gamma = f_{xx}(t, S(t))$ . Let us also introduce the notation  $C = f(t, S(t))$  for the price of the call option.

Using these names, the Black-Scholes PDE is

$$r(S\Delta - C) + \Theta + \frac{1}{2}\sigma^2S^2\Gamma = 0.$$

We now verify that the Black-Scholes European call pricing formula indeed solves the Black-Scholes PDE, i.e. that

$$C = S\Delta + \frac{1}{r} \left( \Theta + \frac{1}{2}\sigma^2S^2\Gamma \right).$$

Some facts that aid in the calculation are

$$\frac{\partial d_1}{\partial S} = \frac{\partial d_2}{\partial S} = \frac{1}{S\sigma\sqrt{T-t}} \quad \text{and} \quad \frac{\partial d_2}{\partial t} = \frac{\partial d_1}{\partial t} + \frac{\sigma}{2\sqrt{T-t}}$$

and

$$\begin{aligned} \phi(d_2) &= \phi(d_1 - \sigma\sqrt{T-t}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(d_1 - \sigma\sqrt{T-t})^2\right) \\ &= \phi(d_1) \exp\left(d_1\sigma\sqrt{T-t} - \frac{1}{2}\sigma^2(T-t)\right) = \phi(d_1) \frac{S}{K} e^{r(T-t)}. \end{aligned}$$

So the partial derivatives are

$$\begin{aligned} \Theta &= S\phi(d_1)\frac{\partial d_1}{\partial t} - e^{-r(T-t)}K\phi(d_2)\frac{\partial d_2}{\partial t} - re^{-r(T-t)}K\Phi(d_2) \\ &= -\frac{S\sigma\phi(d_1)}{2\sqrt{T-t}} - re^{-r(T-t)}K\Phi(d_2) \\ \Delta &= \Phi(d_1) + S\phi(d_1)\frac{\partial d_1}{\partial S} - e^{-r(T-t)}K\phi(d_2)\frac{\partial d_2}{\partial S} = \Phi(d_1) \\ \Gamma &= \frac{\phi(d_1)}{S\sigma\sqrt{T-t}}. \end{aligned}$$

The equation for  $\Delta$  is known as the “law of the unconscious finance professor” because the result is the same as what you would get if you differentiate the formal expression  $S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$  with respect to  $S$  while forgetting that  $d_1$  and  $d_2$  depend on  $S$ . Remember that in the derivation of the Black-Scholes PDE, we found that the number of shares of stock to hold in the replicating portfolio,  $\theta_1$ , is delta. To make the portfolio value equal the option value  $C$ , there must be  $C - S\Delta$  in the money market account. This accounts for the two terms of the Black-Scholes formula:  $\Delta = \Phi(d_1)$  shares of stock worth  $S_t\Phi(d_1)$ , and  $-Ke^{-rT}\Phi(d_2)$  shares of the money market account, worth  $-Ke^{-r(T-t)}\Phi(d_2)$ .

Clearly  $\Delta$  and  $\Gamma$  are positive; the call option’s sensitivity to changes in the stock price increases as the stock price increases. Indeed  $\Delta \in (0, 1)$  and  $\lim_{S \rightarrow 0} \Delta = 0$ , while  $\lim_{S \rightarrow \infty} \Delta = 1$ . Once the stock price is very large, it is almost certain that the option will be exercised and result in a payoff of  $S(T) - K$ , which has slope 1 in  $S(T)$ . On the other hand,  $\Theta$  is negative, meaning that the option loses value over time (if the stock price is held fixed). Unfortunately, some people define  $\Theta$  to be the negative of what we have here, in order to talk about a positive value. It is worth reiterating that all these results are for a European call in the Black-Scholes model. The greeks can have different values and signs for different securities, or in different models.

From our expressions for these greeks, we verify that the Black-Scholes PDE is satisfied:

$$\Theta = -\sigma^2 S^2 \Gamma / 2 - r(S\Delta - C).$$

Indeed  $\sigma^2 S^2 \Gamma / 2$  and  $r(S\Delta - C)$  account for the two terms of  $\Theta$ , giving an interesting interpretation of the greeks. Imagine replicating the call with the strategy of holding  $\Delta$  shares of stock and borrowing  $C - S\Delta$  by holding negative shares of the money market account. Since the option value is the same as the replicating portfolio value, the option’s  $\Theta$  can be understood in terms of the replicating portfolio strategy.

One term of  $\Theta$  is simply the interest being paid on the amount borrowed to finance the purchase of shares. As for the other term, remember that  $\Theta$  describes the change in option value as stock price is held fixed, so imagine a time period  $[t, u]$  passing with  $S(u) = S(t)$ . (And then imagine letting  $u - t$  go to 0.) Because the stock price is geometric Brownian motion, it does not stay flat over this time period. So what happens along the way as the stock price fluctuates? Because  $\Gamma > 0$ , when the stock price rises, you buy more stock, but then it goes back down, to return to  $S(u) = S(t)$ , at which point you need to reduce your stock holdings again, so you lose money; similarly, you lose money by selling low and buying high when the stock goes down at an intermediate time. So the replicating strategy loses money at a rate proportional to  $\Gamma$  and to the instantaneous variance of the stock price (which is  $\sigma^2 S^2$ ), when the stock price remains flat over some time interval. This accounts for the other term, which we will investigate further: managing it in practice can pose a challenge for financial institutions.

Another way of understanding the “time decay”  $\Theta$  is with **Jensen’s inequality**, which says that if  $g$  is a convex function, then

$$\mathbf{E}[g(X)] \geq g(\mathbf{E}[X]).$$

In this case, the option's payoff  $g(S_T) = (S_T - K)^+$  is convex, so the discounted payoff  $e^{-r(T-t)}(S_T - K)^+$  is also convex, and so the option price  $\mathbf{E}^{\mathbf{Q}}[e^{-r(T-t)}(S_T - K)^+]$  is more than the discounted payoff evaluated at the risk-neutral expected stock price,  $e^{-r(T-t)}(S_t e^{r(T-t)} - K)^+$ . So if the stock price stays flat as time passes, the option is losing its value for two reasons. One is that the stock ought to be going up (in the risk-neutral world) with geometric growth rate  $r$ , and if it doesn't, it fails to balance the interest owed due to borrowing the money to buy the stock. The other reason is that the variability of the final stock price is what makes the option's "optionality" valuable, according to Jensen's inequality. If time passes without a change in the stock price, the remaining variance of the final stock price is reduced, and some optionality has evaporated.

So we see that the value of an option is positively related to the volatility of the underlying security, which brings us to another greek  $\Lambda$ , which, oddly enough, is called not lambda but **vega**, like the star. This is the partial derivative of the option price with respect to the volatility parameter  $\sigma$ . One might object that  $\sigma$  is not even a variable that is supposed to change within the Black-Scholes model, such as time or stock price, but a constant parameter of the model. This is quite true, but since the Black-Scholes model is not a perfect description of how stock prices really behave, it is a good idea to acknowledge that  $\sigma$  is not a constant governing the stock's geometric volatility. In later chapters we will discuss how vega can be used in hedging schemes and whether it is adequate as a description of model risk for Black-Scholes. Similar comments apply to **rho**, the partial derivative with respect to the interest rate  $r$ : in the model  $r$  is supposed to be constant, but if we change the value of the parameter,  $\rho$  expresses the sensitivity of the price formula to this change. The values of these greeks are

$$\begin{aligned}\Lambda &= S\phi(d_1)\sqrt{T-t} \\ \rho &= K(T-t)e^{-r(T-t)}\Phi(d_2).\end{aligned}$$

Note that there is no greek relating to the stock's drift  $\mu$  because it does not enter into the pricing formula.

### 3.3. Summary

From the no-arbitrage principle, the self-financing condition, and Itô's formula, we derived the Black-Scholes PDE. This is a relationship among the no-arbitrage price under the Black-Scholes model of a derivative, its Greeks  $\Theta$ ,  $\Gamma$ ,  $\Delta$ , the interest rate  $r$ , and the volatility  $\sigma$ . It is significant that the drift  $\mu$  is not involved, and therefore does not feature in the Black-Scholes formula for a European call option price, which satisfies the Black-Scholes PDE.

### 3.4. Problems

**PROBLEM 3.1.** Compute the geometric drift and volatility under the subjective probability measure  $\mathbf{P}$  of the European call price, as given by the Black-Scholes formula. (Hint: first compute the arithmetic drift and volatility. Do this not by applying Itô's formula again but

by calculating the coefficients of the replicating portfolio, using the known representations of  $S$  and  $M$  as Itô processes.) Assume that the stock's geometric drift  $\mu > r$ , the risk-free rate. Compare the geometric coefficients of the call option to those of the stock.

The following problems depend on put-call parity in the Black-Scholes model, for which see Example 1.3.7. The price at time  $t$  of a bond paying 1 at  $T$  is  $B(t) = M(t)/M(T) = e^{-r(T-t)}$ . Put-call parity says the difference between call and put prices at time  $t$  is  $S(t) - KB(t)$ .

PROBLEM 3.2. Use put-call parity to show that the Black-Scholes no-arbitrage price of the European put is  $e^{-r(T-t)}K\Phi(-d_2) - S\Phi(-d_1)$ .

PROBLEM 3.3. Compute the greeks  $\Delta, \Gamma, \Theta, \Lambda$ , and  $\rho$  of the European put. Show that they satisfy the Black-Scholes PDE.

PROBLEM 3.4. Compute  $\Delta, \Gamma$ , and  $\Theta$  for the stock and money market account (this is trivial). Show that these greeks satisfy put-call parity like prices do.

## CHAPTER 4

### Conditional Probability and Itô Processes

Conditional probability is necessary to answer questions such as “What would we do if some event  $E$  occurs?” A good example in financial engineering is the question “How would I hedge this option if the stock price were  $S(t)$  at time  $t$ ?” A first step is to determine what we would expect to happen given the condition that  $E$  occurs, or how we would value our option given such a state of affairs. To use conditional probability to give a more satisfying derivation of the Black-Scholes formula and learn the tools to compute similar results, we must study it at a deeper level than in introductory probability courses.

Those who are rusty should review the elementary facts in an introductory probability text. A very good, concise account appears in [Mik98, §1.4.1]. Interested students are urged to read all of [Mik98, §1.4] to get a deeper understanding of conditional probability, while still avoiding measure theory.

In this chapter, we will

- (1) develop more advanced concepts of conditional probability
- (2) learn the rules of conditional expectation
- (3) evaluate conditional expectations involving Itô processes
- (4) see the connection between martingales and stochastic integrals

#### 4.1. Conditional Probability

Probabilistic computations are always based on beliefs, or knowledge. When knowledge is static and unchanging, we use a single probability measure  $\mathbf{P}$  and expectation operator  $\mathbf{E}$ . To make our notation a little more explicit,  $\mathbf{P}[\cdot]$  gives the probability of events and  $\mathbf{E}[\cdot]$  gives the expectation of random variables: plug in an event  $E$  or a random variable  $X$  in place of the dot, and get its probability or expectation respectively. Recall that these two mathematical objects are essentially the same, because  $\mathbf{P}[E] = \mathbf{E}[\mathbf{1}_E]$ , where  $\mathbf{1}_E$  given by

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

is the indicator function.

Now suppose that our knowledge grows during the course of an experiment. For instance, suppose we are flipping a coin repeatedly. Let the random variable  $H_n$  be the total number of heads after  $n$  flips. If we believe that the coin is fair, then we initially say  $\mathbf{P}[H_2 = 0] = 1/4 = \mathbf{P}[H_2 = 2]$  and  $\mathbf{P}[H_2 = 1] = 1/2$ . Then  $\mathbf{E}[H_2] = 1$ .

Suppose now that the coin comes up heads the first time, that is,  $H_1 = 1$ . Although our beliefs about the fairness of the coin might not change, our beliefs about the total number of heads certainly does! We can now define  $\mathbf{P}[\cdot|H_1 = 1]$  and  $\mathbf{E}[\cdot|H_1 = 1]$  to reflect our beliefs given that the coin came up heads the first time. Then  $\mathbf{P}[H_2 = 0|H_1 = 1] = 0$  and  $\mathbf{P}[H_2 = 1|H_1 = 1] = 1/2 = \mathbf{P}[H_2 = 2|H_1 = 1]$ , so  $\mathbf{E}[H_2|H_1 = 1] = 3/2$ .

What if the coin had come up tails the first time instead? Then we would say  $\mathbf{P}[H_2 = 0|H_1 = 1] = 0$  and  $\mathbf{P}[H_2 = 1|H_1 = 1] = 1/2 = \mathbf{P}[H_2 = 2|H_1 = 0]$ , so  $\mathbf{E}[H_2|H_1 = 1] = 1/2$ . Initially, we do not know whether the first coin will come up heads or tails. From this perspective of complete ignorance, our future expectation is unknown: it is a random variable  $\mathbf{E}[H_2|H_1]$ .

Let  $\mathcal{F}_n$  represent our knowledge after the  $n$ th coin toss. Thus  $\mathcal{F}_0$  is no knowledge,  $\mathcal{F}_1$  is knowing  $H_1$ ,  $\mathcal{F}_2$  is knowing  $H_1$  and  $H_2$ , etc. These symbols represent mathematical objects called  $\sigma$ -**algebras** or  $\sigma$ -fields, but we do not need to study them in detail. We write  $\mathcal{F}_1 \subseteq \mathcal{F}_2$  to represent that  $\mathcal{F}_2$  contains all the knowledge in  $\mathcal{F}_1$ .

From the interpretation of  $\sigma$ -algebras alone, we can see how  $\mathbf{P}[\cdot|\mathcal{F}_n]$  must behave. First,  $\mathbf{P}[\cdot|\mathcal{F}_0] = \mathbf{P}[\cdot]$ . Thus also  $\mathbf{E}[\cdot|\mathcal{F}_0] = \mathbf{E}[\cdot]$ . In particular,  $\mathbf{E}[H_2|\mathcal{F}_0] = 1$ . Next,

$$\mathbf{P}[\cdot|\mathcal{F}_1] = \mathbf{P}[\cdot|H_1] = \mathbf{1}_{H_1=1}\mathbf{P}[\cdot|H_1 = 1] + \mathbf{1}_{H_1=0}\mathbf{P}[\cdot|H_1 = 0].$$

That is,  $\mathbf{P}[\cdot|\mathcal{F}_1]$  takes on the value  $\mathbf{P}[\cdot|H_1 = 1]$  when  $H_1 = 1$ , and the value  $\mathbf{P}[\cdot|H_1 = 0]$  when  $H_1 = 0$ . We have computed both these values above. The key is that  $\mathcal{F}_1$  contains exactly the information about the value of  $H_1$ . In particular,  $\mathbf{E}[H_2|\mathcal{F}_1] = \mathbf{1}_{H_1=1}(3/2) + \mathbf{1}_{H_1=0}(1/2)$ . Finally,  $\mathbf{E}[H_2|\mathcal{F}_2] = H_2$  because  $H_2$  is actually known given the information  $\mathcal{F}_2$ .

Here we saw how  $\mathbf{E}[H_2|\mathcal{F}_n]$  becomes “more random” as  $n$ , i.e. our knowledge, increases. With no knowledge to condition on, we have just one number, an expectation. Conditioning on full knowledge, we know the outcome, so the conditional expectation is the same as the random variable. With the intermediate amount of knowledge in  $\mathcal{F}_1$ , our conditional expectation is random, but it is “coarser” than  $H_2$  itself, in the sense of having fewer distinct values. In fact, its value is the average of some values of  $H_2$ . The conditional expectation  $\mathbf{E}[H_2|\mathcal{F}_1]$  is not allowed to be too fine, because it must be a function only of what is known after the first coin toss.

We now define the conditional expectation a bit more precisely. Say that an event  $E$  is **knowable** given a  $\sigma$ -algebra  $\mathcal{F}$  if  $\mathcal{F}$  contains enough information to ascertain whether  $E$  has occurred or not. This is sometimes written  $E \in \mathcal{F}$ . For instance, the event  $\{H_1 = 0\}$  is knowable given  $\mathcal{F}_1$ , but  $\{H_2 = 0\}$  is not. This is so even though when  $H_1 = 1$  we know that  $H_2 = 0$  does not occur; when  $H_1 = 0$ , we are not sure whether or not  $H_2 = 0$  occurs. A random variable  $X$  is knowable given  $\mathcal{F}$  if  $\mathcal{F}$  contains enough information to determine the value of  $X$ . For instance,  $H_1$  is knowable given either  $\mathcal{F}_1$  or  $\mathcal{F}_2$ , while  $H_2$  is knowable given  $\mathcal{F}_2$  but not  $\mathcal{F}_1$ .

The definition of  $\mathbf{E}[X|\mathcal{F}]$  is that it is the random variable satisfying the two properties:

- (1)  $\mathbf{E}[X|\mathcal{F}]$  is knowable given  $\mathcal{F}$ .
- (2) For every event  $A$  that is knowable given  $\mathcal{F}$ ,  $\mathbf{E}[\mathbf{1}_A\mathbf{E}[X|\mathcal{F}]] = \mathbf{E}[\mathbf{1}_AX]$ .

The first property states the obvious: if  $\mathbf{E}[X|\mathcal{F}]$  is our expectation of  $X$  conditional on the information  $\mathcal{F}$ , it must be knowable given  $\mathcal{F}$ . The second property says that on any decidable event of positive probability, the average values of  $X$  and of  $\mathbf{E}[X|\mathcal{F}]$  must be the same. (For events of zero probability, these expectations come out zero no matter what.)

Here are the rules for manipulating conditional expectations.

**RULE 4.1.1 (Linearity).**  $\mathbf{E}[aX + bY|\mathcal{F}] = a\mathbf{E}[X|\mathcal{F}] + b\mathbf{E}[Y|\mathcal{F}]$ .

This is the same as for an ordinary (unconditional) expectation.

**RULE 4.1.2 (Extraction).** If  $X$  is knowable given  $\mathcal{F}$ , then  $\mathbf{E}[XY|\mathcal{F}] = X\mathbf{E}[Y|\mathcal{F}]$ .

In particular, if  $X$  is knowable given  $\mathcal{F}$ , then  $\mathbf{E}[X|\mathcal{F}] = X$ . To see this, let  $Y = 1$  in the above rule; it also follows from the definition.

**RULE 4.1.3 (Tower Property).** If  $\mathcal{F}_1 \subseteq \mathcal{F}_2$ , then  $\mathbf{E}[\mathbf{E}[X|\mathcal{F}_2]|\mathcal{F}_1] = \mathbf{E}[X|\mathcal{F}_1]$ .

In this situation,  $\mathbf{E}[\mathbf{E}[X|\mathcal{F}_1]|\mathcal{F}_2] = \mathbf{E}[X|\mathcal{F}_1]$  follows from the extraction rule. So we can say “The less informative  $\sigma$ -algebra wins.”

This rule is called the tower property because we can picture it as a two-story tower collapsing into a one-story heap of rubble. Think of the coarser, less informative  $\sigma$ -algebra  $\mathcal{F}_1$  as the ground floor. Intuitively, the tower property says “Our best guess today as to what our best guess at  $X$  will be tomorrow is just our best guess at  $X$  today.”

In particular, we see  $\mathbf{E}[\mathbf{E}[X|\mathcal{F}]] = \mathbf{E}[X]$ . To illustrate the connection to the tower property, we could say the unconditional expectation  $\mathbf{E}[X] = \mathbf{E}[X|\mathcal{F}_0]$  where  $\mathcal{F}_0$  is the trivial, wholly uninformative  $\sigma$ -algebra, representing no knowledge. Then  $\mathcal{F}_0 \subseteq \mathcal{F}$ .

**RULE 4.1.4 (Independence).** If  $X$  is independent of  $\mathcal{F}$ , then  $\mathbf{E}[X|\mathcal{F}] = \mathbf{E}[X]$ .

The expectation conditional on irrelevant information is the same as the ordinary, unconditional expectation. We have not defined rigorously what it means for a random variable  $X$  to be independent of a  $\sigma$ -algebra  $\mathcal{F}$ , but in practice there is seldom confusion. For instance, when  $\mathcal{F}$  is the information gained by observing random variables that are independent of  $X$ , then  $X$  is independent of  $\mathcal{F}$ . One may ask, “Does having the knowledge of  $\mathcal{F}$  tell me anything about the likelihood of the values of  $X$ ?” If not,  $X$  is independent of  $\mathcal{F}$ .

## 4.2. Conditioning with Itô Processes

In our treatment of financial engineering, we are concerned with the knowledge gained by observing market prices that follow Itô processes. We want to be able to handle conditional distributions of market prices and conditional expected payoffs of derivative securities. In this section, we will extend Examples 2.1.3 and 2.1.4 to conditional probabilities in the Black-Scholes model.

Under some technical conditions (regarding invertibility of the volatility matrix) that we will ignore, the information from observing market prices is the same as the information from observing the underlying Wiener process  $W$  that drives this Itô process. So  $\mathcal{F}_t$  represents knowledge of  $(W(s), s \in [0, t])$ , that is, of the Wiener process from time 0 to  $t$ . A collection

such as  $(\mathcal{F}_t, t \in [0, T])$  is called a **filtration** and represents our increasing knowledge as time passes. When for all  $t \in [0, T]$ , the random variable  $X(t)$  is knowable given the  $\sigma$ -algebra  $\mathcal{F}_t$ , we say that the stochastic process  $X$  is **adapted** to the filtration  $(\mathcal{F}_t, t \in [0, T])$ . We will consider only Itô processes, which are adapted to the filtration generated by  $W$ . The point is that they do not depend on extraneous sources of randomness or on information about the future. Using our filtration, we can interpret  $\mathbf{E}[X|\mathcal{F}_t]$  as a stochastic process. Our expectation of a fixed random variable  $X$  changes continuously in time as we get new information by observing the market.

An Itô process  $X$  has the **arithmetic representation** (2.6.2):

$$X(t) = X(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s).$$

Here  $X(0)$  is a constant and  $\mu$  and  $\sigma$  are stochastic processes. The first integral is a Riemann integral, here called a “time integral” because of its interpretation, while the second is an Itô stochastic integral. The coefficients  $\mu(t)$  and  $\sigma(t)$  are respectively the arithmetic drift and the arithmetic volatility of  $X$  at time  $t$ .

Focusing on time  $t$ , we can specify this instantaneous drift and volatility by writing

$$dX(t) = \mu(t) dt + \sigma(t) dB(t).$$

This is a **stochastic differential equation** (SDE for short). It specifies how  $X$  changes with the passage of time and the changes of the Brownian motion  $B$ . It is simply a shorthand for (2.6.2). One thing is missing: the **initial condition** specifying the value of  $X(0)$ . We can think of this SDE as just a compact form of notation.

In the Black-Scholes model, we are interested in the stock price stochastic process  $S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t))$ . Let  $X(t) = \ln S(t)$ . Then  $X(t)$  satisfies the SDE  $dX(t) = (\mu - \sigma^2/2) dt + \sigma dW(t)$ . From Example 2.6.3, we know  $S(t)$  satisfies  $dS(t) = S(t)(\mu dt + \sigma dW(t))$ . The SDEs for  $X$  and  $S$  are related by Itô’s formula. In the general arithmetic representation of an Itô process, the arithmetic drift process  $\mu(t)$  is here  $(\mu - \sigma^2/2)$ , while the arithmetic volatility  $\sigma(t) = \sigma$ . To investigate the conditional distributions of  $S$ , it will be more convenient to analyze  $X$ , which has normal distributions, and is related to  $S$  by a one-to-one transformation.

Consider conditioning on all information available at time  $t$ , represented by  $\mathcal{F}_t$ . Then the following “late-starting” arithmetic representation for an Itô process is convenient:

$$(4.2.1) \quad X(T) = X(t) + \int_t^T \mu(s) ds + \int_t^T \sigma(s) dW(s).$$

You can check that this is consistent with (2.6.2) by plugging equation (2.6.2) in for  $X(t)$  in equation (4.2.1) and seeing that the result is indeed equation (2.6.2) again, but with  $T$  substituted for  $t$ . The reason for using this late-starting representation is that  $X(t)$  is knowable given  $\mathcal{F}_t$ , whereas the integral terms are not. Indeed if the coefficient processes  $\mu$  and  $\sigma$  are deterministic, then these integral terms are even independent of  $\mathcal{F}_t$ .



Along with conditional expectation, one should also understand conditional variance and conditional covariance. These derive simply from conditional expectation:

$$\mathbf{Cov}[X, Y|\mathcal{F}] = \mathbf{E}[XY|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]\mathbf{E}[Y|\mathcal{F}], \quad \text{so} \quad \mathbf{Var}[X|\mathcal{F}] = \mathbf{E}[X^2|\mathcal{F}] - \mathbf{E}[X|\mathcal{F}]^2.$$

EXAMPLE 4.2.1 (Brownian conditional expectation and variance). Because the increment  $W(t) - W(s)$  is independent of  $\mathcal{F}_s$  and is normal with mean 0 and variance  $t - s$ ,

$$\begin{aligned} \mathbf{Var}[W(t)|\mathcal{F}_s] &= \mathbf{E}[W^2(t)|\mathcal{F}_s] - \mathbf{E}[W(t)|\mathcal{F}_s]^2 \\ &= \mathbf{E}[(W(s) + (W(t) - W(s)))^2|\mathcal{F}_s] - W^2(s) \\ &= \mathbf{E}[W^2(s) + 2W(s)(W(t) - W(s)) + (W(t) - W(s))^2|\mathcal{F}_s] - W^2(s) \\ &= W^2(s) + 2W(s)\mathbf{E}[W(t) - W(s)|\mathcal{F}_s] + \mathbf{E}[(W(t) - W(s))^2|\mathcal{F}_s] - W^2(s) \\ &= 2W(s)\mathbf{E}[W(t) - W(s)] + \mathbf{E}[(W(t) - W(s))^2] \\ &= 2W(s)0 + (t - s) \\ &= t - s. \end{aligned}$$

A very useful fact from elementary probability is that when  $Z_1$  and  $Z_2$  are bivariate normal:

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}\right)$$

then the conditional expectation

$$\mathbf{E}[Z_2|Z_1] = \mu_2 + \left(\frac{\sigma_{12}}{\sigma_{11}}\right)(Z_1 - \mu_1).$$

Here the correlation  $\rho = \sigma_{12}/\sqrt{\sigma_{11}\sigma_{22}}$  and  $\sigma_{12}/\sigma_{11}$  is the regression coefficient; remember the connection between linear regression and conditional expectation as the minimizer of the sum of squared errors. Furthermore,  $Z_2$  is conditionally normal given  $Z_1$ :

$$(4.2.2) \quad Z_2|Z_1 \sim \mathcal{N}\left(\mu_2 + \rho\sqrt{\frac{\sigma_{22}}{\sigma_{11}}}(Z_1 - \mu_1), \sigma_{22}(1 - \rho^2)\right).$$

A reference on this topic is [BD77, §1.4].

**4.2.1. Constant Coefficients.** If  $X$ 's arithmetic drift and volatility  $\mu(t) = \mu$  and  $\sigma(t) = \sigma$  are constant, then we compute conditional probabilities involving  $X$  in terms of the normal distribution. In this case, the integrals are

$$\int_0^t \mu(s) ds = \mu \int_0^t ds = \mu t \quad \text{and} \quad \int_0^t \sigma(s) dW(s) = \sigma \int_0^t dW(s) = \sigma W(t)$$

so

$$X(t) = X(0) + \mu t + \sigma W(t) \sim N(X(0) + \mu t, \sigma^2 t).$$

This Itô process  $X$  is generalized Brownian motion, whose expectation function is

$$\mathbf{E}[X(t)] = X(0) + \mu t$$

and whose covariance function is

$$\mathbf{Cov}[X(s), X(t)] = \mathbf{Cov}[\sigma W(s), \sigma W(t)] = \sigma^2 \mathbf{Cov}[W(s), W(t)] = \sigma^2 \min\{s, t\}.$$

To sum up,  $X(s)$  and  $X(t)$  are bivariate normal with the following mean vector and covariance matrix:

$$\begin{bmatrix} X(0) + \mu s \\ X(0) + \mu t \end{bmatrix} \quad \text{and} \quad \sigma^2 \begin{bmatrix} s & \min\{s, t\} \\ \min\{s, t\} & t \end{bmatrix}.$$

For any Itô process  $S$  with constant geometric coefficients, including the Black-Scholes stock price, we look at the transformation  $X(t) = \ln S(t)$ , which has constant arithmetic coefficients. In general, when

$$dS(t) = S(t)(\mu(t) dt + \sigma(t) dW(t))$$

Itô's formula yields, for  $X(t) = \ln S(t)$ ,

$$dX(t) = \left( \mu(t) - \frac{1}{2}\sigma^2(t) \right) dt + \sigma(t) dW(t).$$

Now we know the distribution of  $X$ , as long as its coefficients, or equivalently the original coefficients of  $S$ , are deterministic. Thus we can find the distribution of  $S(T)$ :  $\mathbf{P}(S(T) \leq K) = \mathbf{P}(X(T) \leq \ln K)$ . In the particular case of the Black-Scholes stock price, the geometric coefficients of  $S$ ,  $\mu(t)$  and  $\sigma(t)$ , are just the constants  $\mu$  and  $\sigma$ . The following examples extend Examples 2.1.3 and 2.1.4.

**EXAMPLE 4.2.2** (Conditional binary call). What will be the conditional probability as of time  $t$  that the stock price  $S(T) \geq K$ , so that the binary call option pays off?

$$\begin{aligned} \mathbf{P}[S(T) \geq K | \mathcal{F}_t] &= \mathbf{P}[X(T) \geq \ln K | \mathcal{F}_t] \\ &= \mathbf{P}[X(t) + (\mu - \sigma^2/2)(T - t) + \sigma(W(T) - W(t)) \geq \ln K | \mathcal{F}_t] \\ &= \mathbf{P}\left[(W(T) - W(t)) \geq \frac{\ln K - X(t) - (\mu - \sigma^2/2)(T - t)}{\sigma} \middle| \mathcal{F}_t\right] \\ &= \Phi\left(\frac{\ln(S(t)/K) + (\mu - \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}\right). \end{aligned}$$

**EXAMPLE 4.2.3** (Black-Scholes conditional loss probability). Recall that the loss event  $V(T) < V(0)$  is  $S(T) < S(0) + (\theta_0/\theta_1)(1 - e^{rT})$ . Just let  $K = S(0) + (\theta_0/\theta_1)(1 - e^{rT})$ . Then we want to find

$$\begin{aligned} \mathbf{P}[S(T) < K | \mathcal{F}_t] &= \mathbf{P}\left[(W(T) - W(t)) < \frac{\ln K - X(t) - (\mu - \sigma^2/2)(T - t)}{\sigma} \middle| \mathcal{F}_t\right] \\ &= \Phi\left(\frac{\ln(K/S(t)) - (\mu - \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}}\right). \end{aligned}$$

**4.2.2. Deterministic Coefficients.** Next we examine the case where the Itô process  $X$  has arithmetic drift  $\mu$  and volatility  $\sigma$  which are deterministic functions of time:  $\mu(t)$  and  $\sigma(t)$  are the values at time  $t$ . In this case, the integral  $\int_0^t \mu(s) ds$  is also a deterministic function of time  $t$ , and the stochastic integral  $\int_0^t \sigma(s) dW(s)$  forms a stochastic process in time  $t$ . If we consider a fixed time  $t$ , then  $\int_0^t \mu(s) ds$  is a constant and  $\int_0^t \sigma(s) dW(s)$  is a random variable. Because  $\sigma$  is deterministic, this random variable is normally distributed with mean 0 and variance given by the Itô isometry. So

$$X(t) \sim N \left( X(0) + \int_0^t \mu(s) ds, \int_0^t \sigma^2(s) ds \right).$$

In a sense,  $X$  is an “even more generalized” Brownian motion. The random variables  $X(t)$  and  $X(u)$  are bivariate normal with the following mean vector and covariance matrix:

$$\begin{bmatrix} X(0) + \int_0^t \mu(s) ds \\ X(0) + \int_0^u \mu(s) ds \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \int_0^t \sigma^2(s) ds & \int_0^{\min\{t,u\}} \sigma^2(s) ds \\ \int_0^{\min\{t,u\}} \sigma^2(s) ds & \int_0^u \sigma^2(s) ds \end{bmatrix}.$$

In the extended Black-Scholes model, the geometric drift  $\mu(t)$  and volatility  $\sigma(t)$  of the stock, as well as the interest rate, are deterministic functions of time.

**EXAMPLE 4.2.4 (Extended Black-Scholes).** In the extended Black-Scholes model, what will be the conditional probability as of time  $t$  that the stock price  $S(T) \geq K$ , so that the binary call option pays off?

$$\begin{aligned} \mathbf{P}[S(T) \geq K | \mathcal{F}_t] &= \mathbf{P}[X(T) \geq \ln K | \mathcal{F}_t] \\ &= \mathbf{P} \left[ X(t) + \int_t^T \left( \mu(u) - \frac{1}{2} \sigma^2(u) \right) du + \int_t^T \sigma(u) dW(u) \geq \ln K | \mathcal{F}_t \right] \\ &= \mathbf{P} \left[ \int_t^T \sigma(u) dW(u) \geq \ln(K/S(t)) - \int_t^T \left( \mu(u) - \frac{1}{2} \sigma^2(u) \right) du | \mathcal{F}_t \right] \\ &= \Phi \left( \frac{\ln(S(t)/K) + \int_t^T \left( \mu(u) - \frac{1}{2} \sigma^2(u) \right) du}{\sqrt{\int_t^T \sigma^2(u) du}} \right). \end{aligned}$$

### 4.3. Martingales

A stochastic process  $M$  is a **martingale** with respect to a filtration  $(\mathcal{F}_t, t \in [0, T])$  and probability measure  $\mathbf{P}$  when

- $M$  is adapted to the filtration.
- $M$  equals its own conditional expectation:  $M(s) = \mathbf{E}[M(t) | \mathcal{F}_s]$  when  $0 \leq s \leq t$ .

There is also a technical condition that the expectation  $\mathbf{E}[|X(t)|]$  exist and be finite for all  $t$ , but since the definition already requires that conditional expectations be finite, only very “bad” processes will fall afoul of this requirement, which we will ignore.

EXAMPLE 4.3.1 (Generalized Brownian motion is sometimes a martingale). As we have already seen, where  $\mathcal{F}$  is the filtration generated by a Wiener process  $W$ , and  $s < t$ , then  $\mathbf{E}[W(t)|\mathcal{F}_s] = W(s)$ , so a Wiener process is a martingale. And where  $X(t) = X(0) + \mu t + \sigma W(t)$ ,

$$\mathbf{E}[X(t)|\mathcal{F}_s] = X(0) + \mu t + \sigma \mathbf{E}[W(t)|\mathcal{F}_s] = X(0) + \mu t + \sigma W(s)$$

and this equals  $X(s)$  only when  $\mu = 0$ .

It is actually important to remember that a process is a martingale only with respect to a filtration (obviously) and a probability measure (because the conditional expectation involves the probability measure). There is no intrinsic “martingality” that a process can possess in itself. This becomes clearer in light of the interpretation of a martingale as a fair game: if  $M(s)$  is a gambler’s wealth at time  $s$ , then continuing to play the game does not change the expected future wealth  $\mathbf{E}[M(t)|\mathcal{F}_s] = M(s)$ . But suppose we change the probability measure so that the dice are loaded. The game would no longer be fair. Or suppose we change the filtration so that it is known at time  $s$  which cards will be dealt next. For some cards, the player would expect to start losing, for others, expect to start winning.

We now see that the Itô stochastic integral of a simple process is a martingale (with respect to  $\mathcal{F}$  and  $\mathbf{P}$ ). Obviously it is adapted, because for every  $i = 1, \dots, n$ ,  $C(t_{i-1})$  and  $W(t_i) - W(t_{i-1})$  are knowable given  $\mathcal{F}_T$ . (Go back and look at the definition of the stochastic integral of a simple process.) As for the martingale property, take  $t_{j-1} \leq t \leq t_j \leq T$ , that is, for a time  $t$  falling into the  $j$ th subinterval of the partition,

$$\begin{aligned} \mathbf{E} \left[ \int_0^T C(s) dW(s) | \mathcal{F}_t \right] &= \mathbf{E} \left[ \int_0^t C(s) dW(s) + \int_t^T C(s) dW(s) | \mathcal{F}_t \right] \\ &= \int_0^t C(s) dW(s) + \mathbf{E} \left[ \int_t^T C(s) dW(s) | \mathcal{F}_t \right] \end{aligned}$$

and

$$\begin{aligned} \mathbf{E} \left[ \int_t^T C(s) dW(s) | \mathcal{F}_t \right] &= \mathbf{E} \left[ C(t_{j-1}) (W(t_j) - W(t)) + \sum_{i=j+1}^n C(t_{i-1}) (W(t_i) - W(t_{i-1})) | \mathcal{F}_t \right] \\ &= C(t_{j-1}) \mathbf{E} [W(t_j) - W(t) | \mathcal{F}_t] \\ &\quad + \sum_{i=j+1}^n \mathbf{E} [\mathbf{E} [C(t_{i-1}) (W(t_i) - W(t_{i-1})) | \mathcal{F}_{t_{i-1}}] | \mathcal{F}_t] \\ &= C(t_{j-1}) 0 + \sum_{i=j+1}^n \mathbf{E} [C(t_{i-1}) \mathbf{E} [W(t_i) - W(t_{i-1}) | \mathcal{F}_{t_{i-1}}] | \mathcal{F}_t] \\ &= 0 \end{aligned}$$

so we see  $\mathbf{E} \left[ \int_0^T C(s) dW(s) | \mathcal{F}_t \right] = \int_0^t C(s) dW(s)$ . The stochastic integral of a more general process  $X$  is also a martingale, as long as it satisfies the technical integrability condition (2.4.4).

However, the Itô integral does *not* have independent increments.

EXAMPLE 4.3.2 (Itô integral of a simple process). One simple, adapted process on  $[0, 2]$  is given by the partition  $(0, 1, 2)$  and  $C(0) = 1$ ,  $C(1) = \mathbf{1}\{W(1) \geq 0\}$ , and  $C(2) = C(1)$ . For  $T \in [0, 1]$ , the integral  $\int_0^T C(t) dW(t) = C(0)(W(T) - W(0)) = W(T)$ . For  $T \in [1, 2]$ , the integral

$$\begin{aligned} \int_0^T C(t) dW(t) &= C_0(W(1) - W(0)) + C(1)(W(T) - W(1)) \\ &= W(1) + \mathbf{1}\{W(1) \geq 0\}(W(T) - W(1)) \\ &= \begin{cases} W(1) & \text{if } W(1) < 0 \\ W(T) & \text{if } W(1) \geq 0 \end{cases}. \end{aligned}$$

The integral does not have independent increments:  $\int_1^2 C(t) dW(t)$  obviously depends on  $\int_0^1 C(t) dW(t) = W(1)$ ; we can see that the later increment has become “contaminated” by  $C(1)$ , which depends on the past. However, the integral is a martingale. For  $t < T \leq 1$ ,

$$\mathbf{E} \left[ \int_0^T C(s) dW(s) | \mathcal{F}_t \right] = \mathbf{E}[W(T) | \mathcal{F}_t] = W(t) = \int_0^t C(s) dW(s).$$

For  $t < 1 \leq T$ ,

$$\begin{aligned} \mathbf{E} \left[ \int_0^T C(s) dW(s) | \mathcal{F}_t \right] &= \mathbf{E}[W(1) + \mathbf{1}\{W(1) \geq 0\}(W(T) - W(1)) | \mathcal{F}_t] \\ &= \mathbf{E}[W(1) | \mathcal{F}_t] + \mathbf{E}[\mathbf{E}[\mathbf{1}\{W(1) \geq 0\}(W(T) - W(1)) | \mathcal{F}_1] | \mathcal{F}_t] \\ &= W(t) + \mathbf{E}[\mathbf{1}\{W(1) \geq 0\} \mathbf{E}[W(T) - W(1) | \mathcal{F}_1] | \mathcal{F}_t] \\ &= W(t) + \mathbf{E}[\mathbf{1}\{W(1) \geq 0\} 0 | \mathcal{F}_t] \\ &= W(t) = \int_0^t C(s) dW(s). \end{aligned}$$

For  $1 \leq t < T$ ,

$$\mathbf{E} \left[ \int_0^T C(s) dW(s) | \mathcal{F}_t \right] = \begin{cases} \mathbf{E}[W(1) | \mathcal{F}_t] = W(1) & \text{if } W(1) < 0 \\ \mathbf{E}[W(T) | \mathcal{F}_t] = W(t) & \text{if } W(1) \geq 0 \end{cases}$$

and in either case, this is  $\int_0^t C(s) dW(s)$ .

#### 4.4. The Markov Property

Let  $\mathcal{F}_t$  represent the information generated by observing  $(X(s), s \in [0, t])$ . Then a process  $X$  is **Markov** when for  $s \leq t \leq T$  the conditional distribution of  $X(T)$  given  $X(t)$  is the same as the conditional distribution of  $X(T)$  given  $\mathcal{F}_t$ :

$$\mathbf{P}[X(T) < x | \mathcal{F}_t] = \mathbf{P}[X(T) < x | X(t)]$$

for all  $x$ . That is, for a Markov process, the future depends only on the present, not the past.

For any process with independent increments, future changes depend on neither the past nor the present, so the process is Markov. An Itô process  $X$  with deterministic arithmetic coefficients has independent increments and thus is Markov. A one-to-one function of a Markov process is also Markov.

The Markov property is a type of memorylessness, but do not confuse it with the memorylessness of the exponential distribution for arrival times. Also do not confuse a Markov process with a martingale. The martingale property is only about conditional expectations, not the whole conditional distribution. However, it also specifies what this conditional expectation must be, while the Markov property does not say what the conditional distribution is. A Markov process also need not have independent increments.

EXAMPLE 4.4.1 (Markov  $\neq$  martingale). The Poisson process is Markov (because it has independent increments) but not a martingale: recall its conditional expectation is

$$\mathbf{E}[N(T)|\mathcal{F}_t] = N(t) + \lambda(T - t) \neq N(t).$$

The process  $I(t) = \int_0^t C(s) dW(s)$ , where  $C$  is a stochastic process satisfying the technical requirements, is a martingale, but not in general Markov. That is because  $(I(s), s \in [0, t])$  contains more information about  $(C(s), s \in (t, T))$  than  $I(t)$  alone does. For instance, look at Example 4.3.2. There the conditional distribution of  $I(2)$  given  $I(3/2)$  is not the same as that given knowledge of  $(I(s), s \in [0, 3/2])$ . There is a great value to knowing  $I(1) = W(1)$ , since if  $W(1) < 0$ , then  $I(2) - I(1) = 0$ , whereas otherwise it is standard normal. But  $I(1)$  is not known given  $I(3/2)$ .

#### 4.5. Summary

A  $\sigma$ -algebra represents knowledge. A filtration represents knowledge that increases over time as we observe the market. A conditional expectation is an expectation given the knowledge available. The conditional expectation of a knowable random variable is simply the random variable itself. The conditional expectation of a random variable given no relevant information is just the ordinary unconditional expectation. Conditional expectation is linear, and in a pair of nested conditional expectations, the less informative  $\sigma$ -algebra wins. A conditional probability is a conditional expectation of an indicator function.

An Itô process is most easily represented as a stochastic differential equation (SDE) and an initial condition. When applying conditional probability to Itô processes, it helps to use a late-starting representation. It also helps to understand the bivariate normal distribution thoroughly. The basic facts are that the Wiener process has increments that are independent and have zero expectation, and it accumulates variance linearly in time.

A martingale is like a fair gambling game. Under some technical conditions, stochastic integrals are martingales. The stochastic integral of a deterministic integrand has independent increments and thus is a Markov process. These properties do not hold generally when the integrand is stochastic.

### 4.6. Problems

In the following problems, do not attempt to verify the definition of conditional expectation. Use your intuition, or rules in boxes. Your reasoning should be correct and explicitly stated, but need not be rigorous at all.

**PROBLEM 4.1.** Compute the conditional probability that a standard Brownian motion will be positive at time  $T$  given the information up to time  $t < T$ :  $\mathbf{P}[W(T) > 0 | \mathcal{F}_t] = \mathbf{E}[\mathbf{1}_{\{W(T) > 0\}} | \mathcal{F}_t]$ . Next compute the conditional probability that a generalized Brownian motion given by  $X(t) = X(0) + \mu t + \sigma W(t)$  will be positive at time  $T$ :  $\mathbf{P}[X(T) > 0 | \mathcal{F}_t]$ . Finally, what happens to this conditional probability as:

- (1) The future time  $T$  approaches  $t$ .
- (2) The future time  $T$  goes to infinity.
- (3) The volatility  $\sigma$  approaches 0.
- (4) The volatility  $\sigma$  goes to infinity.
- (5) The drift  $\mu$  goes to infinity.
- (6) The drift  $\mu$  goes to negative infinity.

♡**PROBLEM 4.2.** Compute  $\mathbf{E}[W(t)W(u) | \mathcal{F}_s]$  where  $s < t < u$ .

**PROBLEM 4.3.** Suppose a stock price process  $S$  is a geometric Brownian motion, as in the Black-Scholes model, and let  $X = \ln S$ . Then for  $0 < t < T$ ,  $X(t)$  and  $X(T)$  have a bivariate normal distribution. What is its mean vector and covariance matrix? Use this result to compute  $\mathbf{P}[S(T) \geq K | S(t)]$ . Is this equal to  $\mathbf{P}[S(T) \geq K | \mathcal{F}_t]$  where  $\mathcal{F}_t$  is (as usual) the information generated by  $(W_s, s \in [0, t])$ ? Why or why not?

**PROBLEM 4.4.** What is the conditional loss probability, as in Example 4.2.3, but in the *extended* Black-Scholes model where  $\mu$ ,  $\sigma$ , and  $r$  are all deterministic functions of time?

## The Black-Scholes Analysis: Part II

In this chapter, we

- (1) learn the Feynman-Kač formula
- (2) learn Girsanov transformation
- (3) apply them while deriving the Black-Scholes formula

### 5.1. The Feynman-Kač Formula

The Black-Scholes PDE (3.1.1), together with a terminal condition such as  $f(T, S) = (S_T - K)^+$ , is an example of a Cauchy problem, which is one kind of PDE boundary-value problem. In financial engineering, there are several useful variants of the Cauchy problem, so it is worth knowing in generality. The Cauchy problem is to solve the PDE

$$(5.1.1) \quad f_t(t, x) + \mu(t, x)f_x(t, x) + \frac{1}{2}\sigma^2(t, x)f_{xx}(t, x) - r(t, x)f(t, x) + h(t, x) = 0,$$

with the terminal condition  $f(T, x) = g(x)$ . Under some uninteresting technical conditions on the coefficients  $\mu, \sigma, r, h$  and the boundary value  $g$ , the Cauchy problem has a solution that can be represented as

$$(5.1.2) \quad f(t, x) = \mathbf{E}_{t,x}^{\mathbf{Q}} \left[ g(X(T))D_t(T) + \int_t^T h(u, X(u))D_t(u) du \right]$$

where  $D_t$  is a stochastic discount process given by  $D_t(u) = D(u)/D(t)$  and

$$(5.1.3) \quad D(u) = \exp \left( - \int_0^u r(s, X(s)) ds \right) \quad \text{so} \quad D_t(u) = \exp \left( - \int_t^u r(s, X(s)) ds \right)$$

and  $X$  is the Itô process with late-starting representation

$$(5.1.4) \quad X(u) = x + \int_t^u \mu(s, X(s)) ds + \int_t^u \sigma(s, X(s)) dW^{\mathbf{Q}}(s)$$

and where  $W^{\mathbf{Q}}$  is a Wiener process under the probability measure  $\mathbf{Q}$ . The subscript  $t, x$  on the expectation indicates that we are conditioning on  $X(t) = x$ . The interpretation is that  $D_t(u)$  is a stochastic discount factor applying to the time interval  $[t, u]$ , and  $X$  is a stochastic process under a new probability measure  $\mathbf{Q}$ , generally different from  $\mathbf{P}$ , which we had in mind before.

This result is the **Feynman-Kač formula**, and it is a very important tool, which gives a connection between a PDE and a stochastic process: the idea is that  $X(t)$  is a random



position in space ( $x$ ) at time  $t$ . The Feynman-Kač formula turns a PDE problem into a problem in stochastic calculus, namely computing an expectation of a function of an Itô process, something we have an idea how to handle with the Itô formula. One potential conceptual pitfall is the new measure  $\mathbf{Q}$  under which  $X$  is an Itô process with arithmetic drift  $\mu(t, X(t))$  and volatility  $\sigma(t, X(t))$ : it need not be the same as  $\mathbf{P}$ , the measure under which the original problem was defined.

Going through a nontechnical proof will help comprehension of this fundamental result. First, consider the simpler case where  $h$  and  $r$  are zero, so the PDE is

$$f_t(t, x) + \mu(t, x)f_x(t, x) + \frac{1}{2}\sigma^2(t, x)f_{xx}(t, x) = 0.$$

Define the stochastic process  $Y(t) = f(t, X(t))$ . Applying the Itô formula,

$$\begin{aligned} dY_t &= \left( f_t(t, X(t)) + \mu(t, X(t))f_x(t, X(t)) + \frac{1}{2}\sigma^2(t, X(t))f_{xx}(t, X(t)) \right) dt \\ &\quad + \sigma(t, X(t))f_x(t, X(t)) dW^{\mathbf{Q}}(t) \\ &= \sigma(t, X(t))f_x(t, X(t)) dW^{\mathbf{Q}}(t) \end{aligned}$$

because the PDE says precisely that the drift is zero. Therefore  $Y$  is a martingale (with respect to  $\mathbf{Q}$  and the Brownian filtration  $\mathcal{F}$ , and under some technical conditions). The terminal condition is  $Y(T) = g(X(T))$ , so the martingale property says  $Y(t) = \mathbf{E}^{\mathbf{Q}}[g(X(T))|\mathcal{F}_t]$ . The process  $X$  is Markov, because its Itô coefficients  $\mu(t, X(t))$  and  $\sigma(t, X(t))$ , while stochastic, are functions of  $X$  itself. So

$$f(t, X(t)) = Y(t) = \mathbf{E}^{\mathbf{Q}}[g(X(T))|\mathcal{F}_t] = \mathbf{E}^{\mathbf{Q}}[g(X(T))|X(t)] = \mathbf{E}_{t, X(t)}^{\mathbf{Q}}[g(X(T))].$$

This proves that  $f(t, x) = \mathbf{E}_{t, x}^{\mathbf{Q}}[g(X(T))]$ .

This illustrates the main idea, and we now go back to the more general version:

$$f_t(t, x) + \mu(t, x)f_x(t, x) + \frac{1}{2}\sigma^2(t, x)f_{xx}(t, x) - r(t, x)f(t, x) + h(t, x) = 0.$$

This time define  $Y(u) = D_t(u)f(u, X(u))$ . From the definition (5.1.3),

$$dD_t(u) = -r(u, X(u))D_t(u) du.$$

Then the differential of  $Y$  is  $D_t(u)df(u, X(u)) + f(u, X(u))dD_t(u)$  or

$$\begin{aligned} D_t(u) & \left( f_t(u, X(u)) + \mu(u, X(u))f_x(u, X(u)) \right. \\ & \quad \left. + \frac{1}{2}\sigma^2(u, X(u))f_{xx}(u, X(u)) - r(u, X(u))f(u, X(u)) \right). \end{aligned}$$

We really ought to justify this by means of the vector Itô formula, which we will develop later. For now, we can see that it makes intuitive sense because it follows from the ordinary product rule of differentiation, which would certainly apply here if  $r$  (and thus  $D$  and  $D_t$ ) were deterministic.

Applying the PDE, this drift is  $-D_t(u)h(u, X(u))$ , so the full stochastic differential is

$$dY(u) = D_t(u) \left( -h(u, X(u)) du + \sigma(u, X(u))f_x(u, X(u)) dW^{\mathbf{Q}}(u) \right),$$

and integrating,

$$Y(T) = Y(t) - \int_t^T D_t(u)h(u, X(u)) du + \int_t^T D_t(u)\sigma(u, X(u))f_x(u, X(u)) dW^{\mathbf{Q}}(u).$$

Since  $Y(T) = D_t(T)f(T, X(T)) = D_t(T)g(X(T))$  and  $Y(t) = D_t(t)f(t, X(t)) = f(t, X(t)) = f(t, x)$ , we have

$$f(t, x) = D_t(T)g(X(T)) + \int_t^T D_t(u)h(u, X(u)) du + (M(T) - M(t))$$

where the  $\mathbf{Q}$ -martingale  $M$  is the last, stochastic integral term. Again because  $X$  is Markov,  $\mathbf{E}_{t,x}^{\mathbf{Q}}[M(T) - M(t)] = \mathbf{E}^{\mathbf{Q}}[M(T) - M(t)|\mathcal{F}_t] = M(t) - M(t) = 0$ , so

$$f(t, x) = \mathbf{E}_{t,x}^{\mathbf{Q}} \left[ g(X(T))D_t(T) + \int_t^T h(u, X(u))D_t(u) du \right].$$

The coefficients of the Cauchy problem have the following interpretations in financial engineering:

- $\mu$  is the coefficient of  $f_x$ , and it becomes the  $\mathbf{Q}$ -drift of the process  $X$ .
- $\sigma$  is the square root of twice the coefficient of  $f_{xx}$ , and it becomes the  $\mathbf{Q}$ -volatility of the process  $X$ .
- $r$  is the negative coefficient of  $f$ , and it becomes the instantaneous interest rate used for discounting.
- $h$  is the nonhomogenous term (i.e. the term not multiplying  $f$  or any of its derivatives), and it is the continuous payment stream of the derivative security.
- $g$  is the terminal condition, and it is the terminal payoff of the derivative security.

Thus the Feynman-Kač formula says that the value of the derivative security is the expected discounted sum of the terminal payoff and the cumulative continuous payments, with expectations taken under the probability measure  $\mathbf{Q}$ .

The point of all this is that we might have an easier time evaluating the  $\mathbf{Q}$ -expectation of the Feynman-Kač formula than staring at the PDE of the Cauchy problem and trying to guess what its solution is given the particular boundary condition  $g$ . Let's see this in the Black-Scholes setting. Recalling our identification of stock price  $S$  with space  $x$ , the formula yields the initial option price

$$f(0, S(0)) = \mathbf{E}^{\mathbf{Q}} [(S(T) - K)^+ e^{-rT}]$$

where now the stock price  $S$  has a representation as an Itô process under  $\mathbf{Q}$

$$S(t) = S(0) + \int_0^t rS(s) ds + \int_0^t \sigma S(s) dW^{\mathbf{Q}}(s) \quad \text{or} \quad dS(t) = S(t)(r dt + \sigma dW^{\mathbf{Q}}(t)).$$

Thus in the world as seen through probability measure  $\mathbf{Q}$ , the stock price still starts at  $S(0)$  and is a geometric Brownian motion with volatility  $\sigma$ , but now it has drift  $r$ , the risk-free interest rate. For this reason,  $\mathbf{Q}$  is known as the **risk-neutral measure**: if investors were neutral to risk, there would be the same reward (expected return) for holding a risky

asset like the stock as for holding a riskless asset like the money market account. Of course this is not true, and in the real world, as seen through the probability measure  $\mathbf{P}$  under which the original  $W$  is a Wiener process, the stock has geometric drift  $\mu > r$ . What we have discovered by applying the Feynman-Kač formula is that computing option prices in the Black-Scholes is done under the risk-neutral measure  $\mathbf{Q}$  under which the stock has geometric drift  $r$ , or if you like, by pretending that the stock has geometric drift  $r$ .

EXAMPLE 5.1.1. Consider the PDE

$$f_t(t, x) + \mu(t)f_x(t, x) + \frac{1}{2}\sigma^2(t)f_{xx}(t, x) - r(t)f(t, x) = 0,$$

where  $\mu$ ,  $\sigma$ , and  $r$  are deterministic processes, and with terminal condition  $f(T, x) = g(x) = x$ . The Feynman-Kač formula says that the solution is

$$f(t, x) = \mathbf{E}_{t,x}^{\mathbf{Q}} [X(T)D_t(T)]$$

where (assuming  $X_t = x$ )

$$X(T) = x + \int_t^T \mu(u) du + \int_t^T \sigma(u) dW^{\mathbf{Q}}(u).$$

So

$$f(t, x) = D_t(T)\mathbf{E}^{\mathbf{Q}} \left[ x + \int_t^T \mu(u) du + \int_t^T \sigma(u) dW^{\mathbf{Q}}(u) \right] = D_t(T) \left( x + \int_t^T \mu(u) du \right).$$

In particular, consider the case  $\mu(t) = r - \sigma^2/2$ ,  $\sigma(t) = \sigma$ , and  $r(t) = r$ , i.e. the PDE is

$$f_t(t, x) + (r - \sigma^2/2)f_x(t, x) + \frac{1}{2}\sigma^2 f_{xx}(t, x) - rf(t, x) = 0.$$

Then we get a generalized Brownian motion  $dX(t) = (r - \sigma^2/2) dt + \sigma dW^{\mathbf{Q}}(t)$  and a constant interest rate of  $r$ . This corresponds to the log stock price  $X = \ln S$  in the Black-Scholes model: we would get the same expression for  $dX(t)$  by applying Itô's formula to  $dS(t) = S(t)(\mu dt + \sigma dW^{\mathbf{Q}}(t))$ . In this case, substituting  $\ln S$  for  $X$ , we have

$$\ln S(T) = \ln S(t) + (r - \sigma^2/2)(T - t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t)),$$

which shows

$$f(t, \ln S(t)) = e^{-r(T-t)}(\ln S(t) + (r - \sigma^2/2)(T - t))$$

is the price at  $t$  of the derivative security with payoff  $\ln S(T)$  at maturity  $T$ . Note that this payoff is negative if  $S(T) < 1$ !

If this seems fishy to you, check out the next example, which confirms it. This shows how different PDE coefficients  $\mu$  and  $\sigma$  and terminal condition  $g$  can relate to what is fundamentally the same problem. In what we just saw, the process  $X$  is generalized Brownian motion, and the payoff is  $X(T)$ , while in what follows,  $X$  is geometric Brownian motion, and the payoff is  $\ln X(T)$ .

EXAMPLE 5.1.2. Consider the PDE

$$f_t(t, x) + rx f_x(t, x) + \frac{1}{2} \sigma^2 x^2 f_{xx}(t, x) - r f(t, x) = 0,$$

with terminal condition  $f(T, x) = g(x) = \ln x$ . The Feynman-Kač formula says that the solution is

$$f(t, x) = \mathbf{E}_{t,x}^{\mathbf{Q}} [e^{-r(T-t)} \ln X(T)]$$

where

$$dX(t) = X(t)(r dt + \sigma dW^{\mathbf{Q}}(t)).$$

By Itô's formula,

$$d \ln X(t) = (r - \sigma^2/2) dt + \sigma dW^{\mathbf{Q}}(t),$$

so

$$\begin{aligned} f(t, x) &= e^{-r(T-t)} \mathbf{E}_{t,x}^{\mathbf{Q}} [\ln X(T)] \\ &= e^{-r(T-t)} \mathbf{E}^{\mathbf{Q}} [\ln x + (r - \sigma^2/2)(T-t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t))] \\ &= e^{-r(T-t)} (\ln x + (r - \sigma^2/2)(T-t)). \end{aligned}$$

## 5.2. Girsanov Transformation

For the European call option, we need to evaluate the expectation  $\mathbf{E}_{t,S(t)}^{\mathbf{Q}} [(S(T) - K)^+ e^{-r(T-t)}]$ . The payoff  $(S(T) - K)^+ = \max\{S(T) - K, 0\} = (S(T) - K) \mathbf{1}\{S(T) > K\}$ , so

$$\begin{aligned} f(t, S(t)) &= \mathbf{E}_{t,S(t)}^{\mathbf{Q}} [(S(T) - K)^+ e^{-r(T-t)}] = e^{-r(T-t)} \mathbf{E}_{t,S(t)}^{\mathbf{Q}} [(S(T) - K) \mathbf{1}\{S(T) > K\}] \\ &= e^{-r(T-t)} \left( \mathbf{E}_{t,S(t)}^{\mathbf{Q}} [S(T) \mathbf{1}\{S(T) > K\}] - K \mathbf{Q}_{t,S(t)}[S(T) > K] \right). \end{aligned}$$

The probability  $\mathbf{Q}_{t,S(t)}[S(T) > K]$  is easy to handle, since  $S$  is a geometric Brownian motion under  $\mathbf{Q}$ . By the Itô formula,

$$(5.2.1) \quad \ln S(T) = \ln S(t) + \int_t^T (r - \sigma^2/2) du + \int_t^T \sigma dW^{\mathbf{Q}}(u)$$

so it has under  $\mathbf{Q}$ , conditional on  $S(t)$ , a normal distribution with mean  $\ln S(t) + (r - \sigma^2/2)(T-t)$  and variance  $\sigma^2(T-t)$ . Therefore the probability is

$$\begin{aligned} \mathbf{Q}_{t,S(t)}[S(T) > K] &= \mathbf{Q}_{t,S(t)}[-\ln S(T) < -\ln K] = \Phi \left( \frac{-\ln K + \mathbf{E}_{t,S(t)}[\ln S(T)]}{\sqrt{\mathbf{Var}_{t,S(t)}[\ln S(T)]}} \right) \\ (5.2.2) \quad &= \Phi \left( \frac{\ln(S(t)/K) + (r - \sigma^2/2)(T-t)}{\sigma \sqrt{T-t}} \right) = \Phi(d_2). \end{aligned}$$

The other term  $\mathbf{E}_{t,S(t)}^{\mathbf{Q}} [S(T) \mathbf{1}\{S(T) > K\}]$  is not so easy to handle.

To deal with it, we need yet more mathematical machinery: the Girsanov transformation. This is a way of changing a complicated expectation into an easier expectation under a new probability measure. The idea is to *define* a new probability measure so that the change of measure will “use up” a factor in the random variable whose expectation you need to

take. The new probability measure alters the drift of the old Wiener process by an amount related to the factor used up in the change of measure. Loosely speaking, Girsanov's theorem relates changes of probability measure to changes of the drift of Brownian motion. That is, for some pairs of probability measures  $\mathbf{P}$  and  $\mathbf{Q}$ , with  $W^{\mathbf{P}}$  and  $W^{\mathbf{Q}}$  Wiener processes under  $\mathbf{P}$  and  $\mathbf{Q}$  respectively, this result gives the relationship between  $W^{\mathbf{P}}$  and  $W^{\mathbf{Q}}$ . This is a useful tool because we will find that the easiest way to evaluate some expectations (such as  $\mathbf{E}^{\mathbf{Q}}[S(T)\mathbf{1}\{S(T) > K\}]$ ) is by changing the probability measure under which they are taken.

But how can one change the probability measure under which an expectation is taken? Supposing that the original probability measure is  $\mathbf{P}$ , there are some limits on the measures that we can change to. Remember that we think of expectations as integrals: for a random variable  $X$ ,

$$\mathbf{E}^{\mathbf{P}}[X] = \int X(\omega) d\mathbf{P}(\omega),$$

where in the simplest case,  $\mathbf{P}$  has a density  $p$ , and  $d\mathbf{P}(\omega) = p(\omega) d\omega$ . In particular,

$$\mathbf{P}[A] = \int \mathbf{1}\{\omega \in A\} d\mathbf{P}(\omega).$$

If some other probability measure  $\tilde{\mathbf{P}}$  can be defined by

$$\tilde{\mathbf{P}}[A] = \int \mathbf{1}\{\omega \in A\} d\tilde{\mathbf{P}}(\omega) = \int \mathbf{1}\{\omega \in A\} Y(\omega) d\mathbf{P}(\omega)$$

with some nonnegative random variable  $Y$ , then we say that  $\tilde{\mathbf{P}}$  is **absolutely continuous** with respect to  $\mathbf{P}$ . This is written  $\tilde{\mathbf{P}} \ll \mathbf{P}$  because it implies that if  $\mathbf{P}[A] = 0$ , then  $\tilde{\mathbf{P}}[A] = 0$ . (Note that this does *not* imply that  $\tilde{\mathbf{P}}[A] \leq \mathbf{P}[A]$  for all events  $A$ , which would be impossible!) We also say that the random variable  $Y$  is the **density of  $\tilde{\mathbf{P}}$  with respect to  $\mathbf{P}$** ,<sup>1</sup> written  $d\tilde{\mathbf{P}}/d\mathbf{P}$ . The reason for this notation is that we imagine cancelling  $d\mathbf{P}$  in the equation

$$\mathbf{E}^{\tilde{\mathbf{P}}}[X] = \int X d\tilde{\mathbf{P}} = \int X \frac{d\tilde{\mathbf{P}}}{d\mathbf{P}} d\mathbf{P} = \mathbf{E}^{\mathbf{P}} \left[ X \frac{d\tilde{\mathbf{P}}}{d\mathbf{P}} \right]$$

that shows how to change the probability measure under which an expectation is taken. This is exactly the same thing that is going on in importance sampling in Monte Carlo simulation.

Girsanov's theorem tells us about the density  $d\tilde{\mathbf{P}}/d\mathbf{P}$  we need to change to a measure  $\tilde{\mathbf{P}}$  under which the drift of Brownian motion changes. This density will be a **stochastic exponential**, i.e. a positive Itô process of the form

$$Z(t) = \exp \left\{ -\frac{1}{2} \int_0^t \lambda^2(s) ds + \int_0^t \lambda(s) dW(s) \right\}$$

where  $\lambda$  is another stochastic process; this is called the stochastic exponential of  $\lambda$ . Under the usual sort of technical condition, the stochastic exponential is a martingale, since  $dZ(t) = Z(t)\lambda(t) dW(t)$  has zero drift, by the Itô formula.

<sup>1</sup>or **Radon-Nikodym derivative**, or **likelihood ratio**

Then Girsanov's theorem states: Where  $W^{\mathbf{P}}$  is a Wiener process under  $\mathbf{P}$ , and  $d\tilde{\mathbf{P}}/d\mathbf{P} = Z(T)$  as given above, then  $W^{\tilde{\mathbf{P}}}$  given by  $dW^{\tilde{\mathbf{P}}}(t) = dW^{\mathbf{P}}(t) - \lambda(t) dt$  is a Wiener process under  $\tilde{\mathbf{P}}$ , for  $t \in [0, T]$ .

EXAMPLE 5.2.1 (Brownian Motion with Drift). Suppose  $W$  is a Wiener process and  $\tilde{W}(t) = W(t) + at$ . Then  $\tilde{W}$  is a Wiener process under  $\tilde{\mathbf{P}}$  where  $d\tilde{\mathbf{P}}/d\mathbf{P} = \exp(-a^2T/2 - aW(T))$ .

Here follows an informal justification of Girsanov's theorem, for the case where the change of drift  $\lambda$  is a constant. The key is the way that the stochastic exponential relates to the standard normal density.

The goal is to show that  $\tilde{W}$ , given by  $\tilde{W}(t) = W(t) - \lambda t$ , is a Wiener process on  $[0, T]$  under the measure  $\tilde{\mathbf{P}}$  defined by  $d\tilde{\mathbf{P}}/d\mathbf{P} = \exp(-\lambda^2T/2 + \lambda W(T))$ . Obviously  $\tilde{W}$  starts at zero and has continuous paths. We just need to check that under  $\tilde{\mathbf{P}}$  it has independent increments, and  $\tilde{W}(t)$  has distribution  $N(0, t)$  for  $t \leq T$ .

First we check the independence of increments by showing that the joint probability that the increment  $\tilde{W}(t)$  takes a value in some set  $A_1$  and that the increment  $\tilde{W}(T) - \tilde{W}(t)$  takes a value in some set  $A_2$  is the product of the marginal probabilities.

$$\begin{aligned}
& \tilde{\mathbf{P}}[\tilde{W}(t) \in A_1, \tilde{W}(T) - \tilde{W}(t) \in A_2] \\
&= \mathbf{E}^{\tilde{\mathbf{P}}}[\mathbf{1}\{\tilde{W}(t) \in A_1\} \mathbf{1}\{\tilde{W}(T) - \tilde{W}(t) \in A_2\}] \\
&= \mathbf{E}^{\mathbf{P}} \left[ \frac{d\tilde{\mathbf{P}}}{d\mathbf{P}} \mathbf{1}\{W(t) - \lambda t \in A_1\} \mathbf{1}\{W(T) - W(t) - \lambda(T-t) \in A_2\} \right] \\
&= \mathbf{E}^{\mathbf{P}} \left[ \exp \left( -\frac{1}{2} \lambda^2 T + \lambda W(T) \right) \mathbf{1}\{W(t) - \lambda t \in A_1\} \mathbf{1}\{W(T) - W(t) - \lambda(T-t) \in A_2\} \right] \\
&= \mathbf{E}^{\mathbf{P}} \left[ \exp \left( -\frac{1}{2} \lambda^2 t + \lambda W(t) \right) \mathbf{1}\{W(t) - \lambda t \in A_1\} \right. \\
&\quad \left. \times \exp \left( -\frac{1}{2} \lambda^2 (T-t) + \lambda (W(T) - W(t)) \right) \mathbf{1}\{W(T) - W(t) - \lambda(T-t) \in A_2\} \right] \\
&= \mathbf{E}^{\mathbf{P}} \left[ \exp \left( -\frac{1}{2} \lambda^2 t + \lambda W(t) \right) \mathbf{1}\{W(t) - \lambda t \in A_1\} \right] \\
&\quad \times \mathbf{E}^{\mathbf{P}} \left[ \exp \left( -\frac{1}{2} \lambda^2 (T-t) + \lambda (W(T) - W(t)) \right) \mathbf{1}\{W(T) - W(t) - \lambda(T-t) \in A_2\} \right]
\end{aligned}$$

because  $W$  has independent increments under  $\mathbf{P}$ . Because this probability factors,  $\tilde{W}$  has independent increments under  $\tilde{\mathbf{P}}$ .

Now we focus just on the first of these expectations, in order to check the distribution of  $\tilde{W}(t)$  under  $\tilde{\mathbf{P}}$ .

$$\begin{aligned}
\tilde{\mathbf{P}}[\tilde{W}(t) \in A_1] &= \mathbf{E}^{\mathbf{P}} \left[ \mathbf{1}\{W(t) - \lambda t \in A_1\} \exp \left( -\frac{1}{2}\lambda^2 t + \lambda W(t) \right) \right] \\
&= \int \mathbf{1}\{x - \lambda t \in A_1\} \exp \left( -\frac{1}{2}\lambda^2 t + \lambda x \right) \phi(x/\sqrt{t}) dx \\
&= \int \mathbf{1}\{y \in A_1\} \exp \left( -\frac{1}{2}\lambda^2 t + \lambda y + \lambda^2 t \right) \phi((y + \lambda t)/\sqrt{t}) dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{A_1} \exp \left( +\frac{1}{2}\lambda^2 t + \lambda y \right) \exp \left( -\frac{1}{2t}(y + \lambda t)^2 \right) dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{A_1} \exp \left( -\frac{1}{2t}(-\lambda^2 t^2 - 2\lambda y t + y^2 + 2\lambda y t + \lambda^2 t^2) \right) dy \\
&= \frac{1}{\sqrt{2\pi}} \int_{A_1} \exp \left( -\frac{1}{2} \left( \frac{y}{\sqrt{t}} \right)^2 \right) dy
\end{aligned}$$

which shows that  $\tilde{W}(t)$  indeed has the  $N(0, t)$  distribution.

Armed with Girsanov's theorem, we return to the Black-Scholes model, in which  $dS(t) = dS(t)(\mu dt + \sigma dW(t))$ , so  $S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t))$ . Remember that  $W$  is a Wiener process under the probability measure  $\mathbf{P}$ . We are faced with the problem of computing  $\mathbf{E}_{t, S(t)}^{\mathbf{Q}}[S(T)\mathbf{1}\{S(T) > K\}]$ , but first let's look at the easier computation of  $\mathbf{E}^{\mathbf{P}}[S(T)]$ . Indeed, previously we have not computed the expectation of an Itô process with deterministic geometric coefficients.

The way to use change of probability measure in evaluating expectations is to “use up” any unwelcome factors by putting them into the density  $d\tilde{\mathbf{P}}/d\mathbf{P}$ . The result is that instead of a difficult expectation under  $\mathbf{P}$  we get an easy probability under  $\tilde{\mathbf{P}}$ .

**EXAMPLE 5.2.2** (Black-Scholes stock mean). Consider  $\mathbf{E}^{\mathbf{P}}[S(T)] = \mathbf{E}^{\mathbf{P}}[S(0) \exp((\mu - \sigma^2/2)T + \sigma W(T))]$ . The factor  $\exp(\sigma W(T))$  is unwelcome, because we do not know how to compute its mean. But  $d\tilde{\mathbf{P}}/d\mathbf{P}$  is supposed to be a stochastic exponential, so we must set  $d\tilde{\mathbf{P}}/d\mathbf{P} = \exp(-\sigma^2 T/2 + \sigma W(T))$ . Now

$$\begin{aligned}
\mathbf{E}^{\mathbf{P}}[S(T)] &= \mathbf{E}^{\mathbf{P}}[S(0) \exp((\mu - \sigma^2/2)T + \sigma W(T))] \\
&= S(0) \exp(\mu T) \mathbf{E}^{\mathbf{P}}[\exp(-\sigma^2 T/2 + \sigma W(T))] \\
&= S(0) \exp(\mu T) \mathbf{E}^{\mathbf{P}}[d\tilde{\mathbf{P}}/d\mathbf{P}] \\
&= S(0) \exp(\mu T) \mathbf{E}^{\tilde{\mathbf{P}}}[1] \\
&= S(0) \exp(\mu T)
\end{aligned}$$

This justifies our previous assertions that  $\mu$  is the mean geometric growth rate of the stock in the Black-Scholes model. Actually, it is clear without Girsanov's theorem that

$\mathbf{E}^{\mathbf{P}}[\exp(-\sigma^2 T/2 + \sigma W(T))] = 1$  because the stochastic exponential is a martingale starting at 1.

Next we will apply this tool to the expectation  $\mathbf{E}^{\mathbf{Q}}[S_T \mathbf{1}\{S_T > K\}]$  appearing in the price at time 0 of the European call option in the Black-Scholes model, using Girsanov's theorem as in Example 5.2.2.

EXAMPLE 5.2.3 (The difficult term in the Black-Scholes formula). This time use  $S(T) = \exp((\mu - \sigma^2/2)T + \sigma W^{\mathbf{Q}}(T))$ , where  $W^{\mathbf{Q}}$  is a Wiener process under  $\mathbf{Q}$ . Much as in Example 5.2.2,

$$\begin{aligned} \mathbf{E}^{\mathbf{Q}}[S(T) \mathbf{1}\{S(T) > K\}] &= \mathbf{E}^{\mathbf{Q}}[S(0) \exp((r - \sigma^2/2)T + \sigma W^{\mathbf{Q}}(T)) \mathbf{1}\{S(T) > K\}] \\ &= \mathbf{E}^{\mathbf{Q}}[S(0) \exp(rT) \exp(-\sigma^2 T/2 + \sigma W^{\mathbf{Q}}(T)) \mathbf{1}\{S(T) > K\}] \\ &= S(0) \exp(rT) \mathbf{E}^{\mathbf{Q}}[\mathbf{1}\{S(T) > K\} d\tilde{\mathbf{P}}/d\mathbf{Q}] \\ &= S(0) \exp(rT) \mathbf{E}^{\tilde{\mathbf{P}}}[\mathbf{1}\{S(T) > K\}] \\ &= S(0) \exp(rT) \tilde{\mathbf{P}}[S(T) > K]. \end{aligned}$$

Girsanov's theorem says that under  $\tilde{\mathbf{P}}$ ,  $W^{\tilde{\mathbf{P}}}(t) = W^{\mathbf{Q}}(t) - \sigma t$  is a Wiener process. So  $W^{\mathbf{Q}}(t) = W^{\tilde{\mathbf{P}}}(t) + \sigma t$  and

$$\begin{aligned} \tilde{\mathbf{P}}[S(T) > K] &= \tilde{\mathbf{P}}[S(0) \exp((r - \sigma^2/2)T + \sigma W^{\mathbf{Q}}(T)) > K] \\ &= \tilde{\mathbf{P}}[S(0) \exp((r - \sigma^2/2)T + \sigma(W^{\tilde{\mathbf{P}}}(T) + \sigma T)) > K] \\ &= \tilde{\mathbf{P}}[S(0) \exp((r + \sigma^2/2)T + \sigma W^{\tilde{\mathbf{P}}}(T)) > K] \\ &= \tilde{\mathbf{P}}\left[W^{\tilde{\mathbf{P}}}(T) > \frac{\ln(K/S(0)) - (r + \sigma^2/2)T}{\sigma}\right] \\ &= \Phi\left(\frac{\ln(S(0)/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}\right). \end{aligned}$$

Generally, if  $Z \sim N(m, s^2)$  under  $\mathbf{P}$ , then

$$E^{\mathbf{P}}[\exp(Z)] = \exp(m + s^2/2)$$

and

$$E^{\mathbf{P}}[\exp(Z) \mathbf{1}\{Z > z\}] = \exp(m + s^2/2) \Phi((-z + m + s^2)/s).$$

These formulae also work exactly the same way for conditional expectations. We can derive these useful facts about the lognormal distribution without doing any serious classical integration. The former formula follows from the latter, so we'll just prove the latter.

First observe that  $\int_0^1 m dt + \int_0^1 s dW^{\mathbf{P}}(t) = m + sW^{\mathbf{P}}(1)$  also has distribution  $N(m, s^2)$  under  $\mathbf{P}$ , when  $m$  and  $s$  are constants, so we can set  $Z$  equal to this value of an Itô process



at time 1. Then

$$\begin{aligned}
E^{\mathbf{P}}[\exp(Z)\mathbf{1}\{Z > z\}] &= \mathbf{E}^{\mathbf{P}} \left[ \exp \left( \int_0^1 m dt + \int_0^1 s dW^{\mathbf{P}}(t) \right) \mathbf{1}\{Z > z\} \right] \\
&= \mathbf{E}^{\mathbf{P}} \left[ \exp \left( \int_0^1 \left( m + \frac{s^2}{2} \right) dt - \int_0^1 \frac{s^2}{2} dt + \int_0^1 s dW^{\mathbf{P}} \right) \mathbf{1}\{Z > z\} \right] \\
&= \exp \left( \int_0^1 \left( m + \frac{s^2}{2} \right) dt \right) \mathbf{E}^{\mathbf{P}} \left[ \frac{d\tilde{\mathbf{P}}}{d\mathbf{P}} \mathbf{1}\{Z > z\} \right] \\
&= \exp(m + s^2/2) \tilde{\mathbf{P}}[Z > z]
\end{aligned}$$

and by Girsanov's theorem, since we used

$$\frac{d\tilde{\mathbf{P}}}{d\mathbf{P}} = \exp \left( - \int_0^1 \frac{s^2}{2} dt + \int_0^1 s dB_t \right),$$

the Wiener process under  $\tilde{\mathbf{P}}$  is  $W^{\tilde{\mathbf{P}}}(t) = W^{\mathbf{P}}(t) - st$ , so

$$\begin{aligned}
\tilde{\mathbf{P}}[Z > z] &= \tilde{\mathbf{P}}[m + sW^{\mathbf{P}}(1) > z] \\
&= \tilde{\mathbf{P}}[m + s(W^{\tilde{\mathbf{P}}}(1) + s) > z] \\
&= \tilde{\mathbf{P}}[(m + s^2) + sW^{\tilde{\mathbf{P}}}(1) > z] \\
&= \tilde{\mathbf{P}} \left[ W^{\tilde{\mathbf{P}}}(1) > \frac{z - m - s^2}{s} \right] \\
&= \Phi \left( \frac{-z + m + s^2}{s} \right).
\end{aligned}$$

Now we finally get the Black-Scholes formula. From the Feynman-Kač formula, we have the European call price as

$$f(0, S(0)) = e^{-rT} (\mathbf{E}^{\mathbf{Q}}[S(T)\mathbf{1}\{S(T) > K\}] - K E^{\mathbf{Q}}[\mathbf{1}\{S(T) > K\}])$$

and  $S(T) = \exp(Z)$  where the distribution of  $Z$  under probability  $\mathbf{Q}$  is normal with mean  $m = \ln S(0) + (r - \sigma^2/2)T$  and variance  $s^2 = \sigma^2 T$ . We get  $E^{\mathbf{Q}}[\mathbf{1}\{S(T) > K\}]$  by using  $t = 0$  in equation (5.2.2) and we found  $E^{\mathbf{Q}}[S(T)\mathbf{1}\{S(T) > K\}]$  in Example 5.2.3, so the Black-Scholes formula is

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

where the first and second arguments of  $\Phi$  are known as  $d_1$  and  $d_2$ .

### 5.3. Examples

Let's go back and practice our new tools, the Feynman-Kač formula and Girsanov's theorem.

EXAMPLE 5.3.1 (Continuous coupon bond). Consider the PDE

$$f_t(t, x) + \mu(t, x)f_x(t, x) + \frac{1}{2}\sigma^2(t, x)f_{xx}(t, x) - rf(t, x) + h = 0,$$

where  $r$  and  $h$  are constants, and with terminal condition  $f(T, x) = g(x) = 1$ . The Feynman-Kač formula says that the solution is

$$f(t, x) = \mathbf{E}_{t,x}^{\mathbf{Q}} \left[ 1e^{-r(T-t)} + h \int_t^T e^{-r(u-t)} du \right] = e^{-r(T-t)} + \frac{h}{r}(1 - e^{-r(T-t)}),$$

because

$$\int_t^T e^{-r(u-t)} du = -\frac{1}{r} (e^{-r(u-t)})_t^T = \frac{1}{r}(1 - e^{-r(T-t)}).$$

Because all payments are deterministic, the coefficients  $\mu$  and  $\sigma$  of the stochastic process  $X$  are irrelevant. The repayment of principal 1 at time  $T$  is worth  $e^{-r(T-t)}$  at time  $t$ , while the value of receiving interest payments at rate  $h$  over  $[t, T]$  is worth the second term. Consider  $T \rightarrow \infty$  in the second term. The limit  $h/r$  is the value of receiving payments at rate  $h$  forever starting now; consider putting  $h/r$  in the money market account, and then withdrawing the interest at rate  $r(h/r) = h$ . So  $(h/r)e^{-r(T-t)}$  is the value now (at time  $t$ ) of receiving payments at rate  $h$  forever, but starting at time  $T$ . Thus now at time  $t$  the value of payments over  $[t, T]$  is the difference.

In the following example, we not only practice with our tools, but also reinforce the idea that it makes sense to price securities by taking expectations under the risk-neutral measure  $\mathbf{Q}$  rather than some other probability measure, such as the original  $\mathbf{P}$ . In this example, we will call the process  $X$  from the Feynman-Kač formula  $S$  because it represents a dividend-paying stock. We see that using  $\mathbf{Q}$  allows us to “reprice” the stock, that is, if we regard the stock as a trivial “derivative” security whose prices is  $f(t, S(t))$ , then the Feynman-Kač formula indeed tells us  $f(t, S(t)) = S(t)$ .

EXAMPLE 5.3.2 (Repricing a dividend-paying stock). We have the PDE

$$f_t(t, x) + (r - q)xf_x(t, x) + \frac{1}{2}\sigma^2x^2f_{xx}(t, x) - rf(t, x) + qx = 0,$$

with terminal condition  $g(x) = x$ . This represents holding the stock over  $[t, T]$ , collecting dividend  $qS(u)$  at each time  $u$ , and then selling it at  $T$  for payoff  $S(T)$ . The Feynman-Kač formula says

$$f(t, S(t)) = \mathbf{E}_{t,S(t)}^{\mathbf{Q}} \left[ S(T)e^{-r(T-t)} + \int_t^T qS(u)e^{-r(u-t)} du \right]$$

where

$$dS(t) = (r - q)S(t) dt + \sigma S(t) dW^{\mathbf{Q}}(t) \quad \text{so} \quad d \ln S(t) = (r - q - \sigma^2/2) dt + \sigma dW^{\mathbf{Q}}(t).$$

First evaluating the first term, we have

$$\begin{aligned}
& \mathbf{E}_{t,S(t)}^{\mathbf{Q}}[S(T)e^{-r(T-t)}] \\
&= \mathbf{E}_{t,S(t)}^{\mathbf{Q}}[S(t) \exp(-r(T-t) + (r-q-\sigma^2/2)(T-t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t)))] \\
&= S(t)e^{-q(T-t)} \mathbf{E}^{\mathbf{Q}}[\exp(-\sigma^2/2(T-t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t)))] \\
&= S(t)e^{-q(T-t)}
\end{aligned}$$

because the expression inside the expectation is a stochastic exponential, hence a martingale starting at 1, so its expectation is 1. (Or you could use Girsanov's theorem to show that this  $\mathbf{Q}$ -expectation is  $\mathbf{E}^{\tilde{\mathbf{P}}}[1] = 1$ .) For the second term, similarly

$$\begin{aligned}
& \mathbf{E}_{t,S(t)}^{\mathbf{Q}} \left[ \int_t^T qS(u)e^{-r(u-t)} du \right] \\
&= \int_t^T \mathbf{E}_{t,S(t)}^{\mathbf{Q}} [qS(t) \exp((-q-\sigma^2/2)(u-t) + \sigma(W^{\mathbf{Q}}(u) - W^{\mathbf{Q}}(t)))] du \\
&= qS(t) \int_t^T e^{-q(u-t)} \mathbf{E}_{t,S(t)}^{\mathbf{Q}} [\exp((-q-\sigma^2/2)(u-t) + \sigma(W^{\mathbf{Q}}(u) - W^{\mathbf{Q}}(t)))] du \\
&= qS(t) \int_t^T e^{-q(u-t)} du \\
&= S(t)(1 - e^{-q(T-t)})
\end{aligned}$$

so the sum of the two terms is  $S(t)$ , as required.

The first term,  $S(t)e^{-q(T-t)}$ , is the price of the stock for delivery at  $T$ : you would pay this much at  $t$  to get the stock's value  $S(T)$  at  $T$ , without collecting the dividends over  $[t, T]$ . Much as in Example 5.3.1, the second term, the value of collecting dividends over  $[t, T]$ , is the value of collecting dividends over  $[t, \infty)$  minus the value of collecting dividends over  $[T, \infty)$ . The stock itself is precisely a certificate entitling its holder to collect dividends forever. It costs  $S(t)$  to buy the stock now and get dividends over  $[t, \infty)$ , and it costs the price for future delivery  $S(t)e^{-q(T-t)}$  to get them over  $[T, \infty)$ .

## 5.4. Problems

PROBLEM 5.1 (Exercise 4.11 of [Bjö98]). Solve (for  $f(t, x)$ ) the PDE

$$f_t(t, x) + \frac{1}{2}x^2 f_{xx}(t, x) + x = 0$$

with terminal condition  $f(T, x) = \ln x$ .

PROBLEM 5.2. In class, we derived the Black-Scholes option price  $f(0, S(0))$  at time 0. Do the derivation again, this time computing

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

Hint: the conditional distribution of  $S(T)$  given either  $S(t)$  or  $\mathcal{F}_t$  is normal. Use the quick rule for evaluating  $\mathbf{E}[\exp(Z)\mathbf{1}\{Z > z\}]$ .

PROBLEM 5.3. In this problem, you find the price  $f(t, F(t))$  of a European call option on a futures price  $F$ , rather than a stock price  $S$ , in the Black-Scholes model. Solve the PDE

$$f_t(t, x) + \frac{1}{2}\sigma^2 x^2 f_{xx}(t, x) - rf(t, x) = 0$$

with terminal condition  $f(T, x) = (x - K)^+$ , to get a formula similar to that of the previous problem.

## CHAPTER 6

### Complete Markets

#### 6.1. Market Price of Risk

We were able to compute a unique no-arbitrage price for the European call option in the Black-Scholes model because it can be **replicated** by a tame portfolio strategy  $\theta$ , that is,  $\theta(T)S(T) = g(S(T))$ , the payoff. Then by the no-arbitrage principle, the price of the derivative is  $\theta(0)S(0)$ , the price of its replicating portfolio. If you know how to replicate a derivative, you know how to **hedge** it, since  $g(S(T)) - \theta(T)S(T) = 0$ , and thus the combination of the derivative and the hedging portfolio is riskless. Strictly speaking then, one ought to reserve the term “hedging strategy” for  $-\theta$  as opposed to  $\theta$ , the replicating strategy.

EXAMPLE 6.1.1. For the European call option in the Black-Scholes model, the replicating strategy is

$$\theta_1(t) = \Phi(d_1) \quad \text{and} \quad \theta_0(t) = -Ke^{-rT}\Phi(d_2).$$

In the Black-Scholes model, *every* contingent claim can be replicated by a self-financing portfolio strategy, which means the market is **complete**. Then every claim has a unique no-arbitrage price, equal to the initial value of its replicating portfolio. In the Black-Scholes model, this price is  $\mathbf{E}^{\mathbf{Q}}[e^{-rT}g(S(T))]$ , the expected discounted payoff under the probability measure  $\mathbf{Q}$  that makes the stock have geometric drift  $r$ . The probability measure  $\mathbf{Q}$  is called an **equivalent martingale measure** (EMM) for the original  $\mathbf{P}$ .

$\mathbf{Q}$  is a **martingale measure** in that the discounted stock price  $e^{-rt}S(t) = S(t)/M(t)$  becomes a martingale under  $\mathbf{Q}$ . Then the discounted price of a money market account share,  $M(t)/M(t) = 1$ , is also a martingale. The discounted value of any tame, self-financing portfolio must be a martingale under the martingale measure. (In particular, then, the discounted price of any non-dividend-paying traded security must be a martingale.) Consequently, the discounted value of any replicated claim (which equals the value of the replicating portfolio) is a martingale under  $\mathbf{Q}$ —and in a complete market, this is every claim. For a  $\mathbf{Q}$ -martingale  $V$ ,

$$\mathbf{E}^{\mathbf{Q}} \left[ \frac{V(T)}{M(T)} \right] = \mathbf{E}^{\mathbf{Q}} \left[ \frac{V(T)}{M(T)} \middle| \mathcal{F}_0 \right] = \frac{V(0)}{M(0)} = V(0),$$

and this justifies computing the initial price  $V(0)$  of a derivative as the  $\mathbf{Q}$ -expectation of its discounted payoff  $V(T)$ . It would not work to use a probability measure, such as  $\mathbf{P}$ , under which the value is not a martingale: then  $\mathbf{E}[V(T)/M(T)] \neq V(0)$ .

In the preceding discussion, there was a special role for the money market account: its price was used for discounting. This seems natural because  $M(t) = e^{rt}$  in the Black-Scholes model, and this is a simple way of expressing the time value of money. However, it is perfectly legitimate to use the stock price  $S(t)$  to discount, and look for a measure under which  $M(t)/S(t)$  and  $S(t)/S(t) = 1$  are martingales. This would be a different measure from  $\mathbf{Q}$ , and a useful one, as we will see later. The asset whose price is used for discounting is called the **numeraire**. All values are expressed in terms of shares of this asset. For instance,  $K$  dollars at time  $T$  equals  $K/M(T)$  shares of the money market account or  $K/S(T)$  shares of stock. In the Black-Scholes model,  $M(T)$  is a constant, so  $K$  dollars at time  $T$  is worth  $K/M(T)$  dollars at time 0.

$\mathbf{Q}$  is **equivalent** to  $\mathbf{P}$  in the sense that these two probability measures agree about which events are possible:

$$\mathbf{P}[E] = 0 \quad \iff \quad \mathbf{Q}[E] = 0.$$

In the Black-Scholes model, we can see this is true, because  $\mathbf{Q}$  results from a Girsanov transformation of  $\mathbf{P}$ . We have

$$dS(t) = S(t)(\mu dt + \sigma dW(t)) \quad \text{but} \quad dS(t) = S(t)(r dt + \sigma dW^{\mathbf{Q}}(t))$$

and therefore  $\sigma dW^{\mathbf{Q}}(t) = \sigma dW(t) + (\mu - r) dt$ , so  $dW^{\mathbf{Q}}(t) = dW(t) + \lambda^\top dt$ , where  $\lambda$  satisfies

$$(6.1.1) \quad \sigma \lambda^\top = \mu - r.$$

Therefore, as in the discussion of the Girsanov transformation in Section 5.2, there is a density

$$(6.1.2) \quad \frac{d\mathbf{Q}}{d\mathbf{P}} = Z(T) = \exp\left(-\frac{\|\lambda\|^2}{2}T - \lambda dW(T)\right),$$

which is finite and positive, relating the two probability measures. So  $\mathbf{Q}[E] = \mathbf{E}^{\mathbf{Q}}[\mathbf{1}_E] = \mathbf{E}[\mathbf{1}_E d\mathbf{Q}/d\mathbf{P}]$  is zero if and only if  $\mathbf{P}[E] = \mathbf{E}[\mathbf{1}_E]$  is.

The **density process** of  $\mathbf{Q}$  with respect to the old probability measure  $\mathbf{P}$  is stochastic process

$$(6.1.3) \quad Z(t) = \exp\left(-\frac{1}{2} \int_0^t \|\lambda_s\|^2 ds - \int_0^t \lambda_s dW(s)\right).$$

The density process is the conditional expectation of the density:

$$\begin{aligned} \mathbf{E}\left[\frac{d\mathbf{Q}}{d\mathbf{P}} \middle| \mathcal{F}_t\right] &= \mathbf{E}\left[Z(t) \frac{Z(T)}{Z(t)} \middle| \mathcal{F}_t\right] \\ &= Z(t) \mathbf{E}\left[\exp\left(-\frac{1}{2} \int_t^T \|\lambda_s\|^2 ds - \int_t^T \lambda_s dW(s)\right) \middle| \mathcal{F}_t\right] \\ &= Z(t) \end{aligned}$$

because the stochastic exponential is a martingale. So we can refer to  $Z(t)$  as  $\mathbf{E}[d\mathbf{Q}/d\mathbf{P}|\mathcal{F}_t]$ . This makes it clear how to use Girsanov transformations for conditional expectations.

$$(6.1.4) \quad \mathbf{E}^{\mathbf{P}}[Z(T)Y|\mathcal{F}_t] = Z(t)\mathbf{E}^{\mathbf{Q}}[Y|\mathcal{F}_t]$$

$$(6.1.5) \quad W^{\mathbf{P}}(t) = W^{\mathbf{Q}}(t) - \int_0^t \lambda(s)^\top ds$$

The idea is that  $Z(t)$  is known given  $\mathcal{F}_t$ , and the remaining factor

$$\frac{Z(T)}{Z(t)} = \exp\left(-\frac{1}{2}\int_t^T \|\lambda(s)\|^2 ds - \int_t^T \lambda(s) dW(s)\right)$$

is used up in changing the probability measure so that the old standard Brownian motion  $B$  becomes a Brownian motion with drift  $\lambda^\top$  under the new probability measure  $\tilde{\mathbf{P}}$ . (The transpose symbol  $\top$  appears because  $\lambda$  is a row vector but a drift, like a Brownian motion, should be a column vector.)

The **market price of risk**  $\lambda$  expresses the reward for accepting the risk inherent in the stock as an increase in expected return. In the one-dimensional Black-Scholes case,  $\lambda = (\mu - r)/\sigma$ , the stock's excess return divided by volatility, which is its **Sharpe ratio**. In the general case,  $\lambda$  is a vector stochastic process, and its  $k$ th component expresses the increase in expected return earned by exposure to a unit of the  $k$ th risk, i.e. the  $k$ th component of the fundamental Wiener process  $W$ .

## 6.2. The Pricing Measure $\mathbf{Q}$

Thus the switch from the original probability measure  $\mathbf{P}$  to  $\mathbf{Q}$ , which is used for pricing, is justified on financial grounds. It would not be correct to value risky payoffs by discounting them at the risk-free rate and then taking an expectation. The density  $Z(T)$  of equation (6.1.2), built from the market price of risk process  $\lambda$ , helps us to understand how much a dollar is worth in different states  $\omega$ . If the risk-free rate is a stochastic process  $r$ , then define

$$(6.2.1) \quad D(t) = \exp\left(-\int_0^t r(s) ds\right),$$

the accumulated discount factor up to time  $t$ , formed by continuous compounding at the risk-free rate. If there is a money market account,  $D(t) = 1/M(t)$ . The **pricing kernel** or **state price density**  $\Pi = DZ$  is a stochastic process specifying the desirability of a dollar at each time and in each state. The price of a derivative paying  $f(S(T))$  at time  $T$  is

$$(6.2.2) \quad \mathbf{E}^{\mathbf{Q}}[D(T)f(S(T))] = \mathbf{E}\left[D(T)f(S(T))\frac{d\mathbf{Q}}{d\mathbf{P}}\right] = \mathbf{E}[D(T)f(S(T))Z(T)] = \mathbf{E}[\Pi(T)f(S(T))].$$

So the value of a dollar at time  $T$  in state  $\omega$ , as measured by its contribution to the price of a derivative, depends on three things:

- the probability of state  $\omega$  under the real-world probability measure  $\mathbf{P}$

- the time value of money due to discounting, as expressed by  $D_t(\omega)$
- another factor  $Z(\omega, T)$  representing the influence of relative supply and demand of wealth in state  $\omega$  at time  $T$

We will study the financial significance of  $Z$  in greater depth when we cover portfolio optimization, in Chapter ??, where we will find that the state price  $\Pi(\omega, T)$  is related to the marginal utility of wealth in state  $\omega$  at time  $T$ . For now, here is a heuristic argument in an example.

**EXAMPLE 6.2.1.** Consider an economy satisfying the Black-Scholes model, and where the net supply of money market shares is zero, and the net supply of stock shares is positive. That is, market participants borrow from and lend to each other at constant rate  $r$ , so for each owner of a money market share (lender) there is an owner of a negative money market share (borrower). But there exists a company with a certain number of shares outstanding, so while some people may own a positive and others a negative number of shares, the total holdings are positive. Then the total wealth in the economy is proportional to the stock price  $S$ . Assuming that the “demand for wealth”<sup>1</sup> is constant, the “price of wealth” should be inversely related with its supply, hence inversely related with  $S$ . This is indeed what we found:

$$S(T) = S(0) \exp((\mu - \sigma^2/2)T + \sigma W(T)) \quad \text{and} \quad \Pi(T) = \exp(-(r + \lambda^2/2)T - \lambda W(T))$$

where both  $\sigma$  and  $\lambda$  are positive, so the relationship is inverse.

To sum up, we have seen a mathematical and a financial reason why we price by taking expected discounted payoffs under the equivalent martingale measure  $\mathbf{Q}$ , not the original measure  $\mathbf{P}$ . The mathematical reason is that a derivative’s initial price equals the initial price  $V(0)$  of its replicating portfolio, and we need the discounted portfolio value  $DV$  to be a  $\mathbf{Q}$ -martingale to justify  $\mathbf{E}^{\mathbf{Q}}[D(T)V(T)] = V(0)$ . The financial reason is that  $\mathbf{Q}$  incorporates information about the value of wealth in different states of the world. Ignoring this information and trying to price with  $\mathbf{P}$  would be like saying that people are indifferent between having an extra dollar when the economy is in a recession or during a boom.

One might also ask why there has to be a numeraire. Why can’t prices expressed in nominal dollars be made into martingales? The reason is the money market account. For instance, in the Black-Scholes model, it is deterministic, and thus no change of probability measure<sup>2</sup> can change it at all.

Finally, here is an example of what goes wrong when one tries to price with  $\mathbf{P}$ : it does not even reprice the traded securities.

**EXAMPLE 6.2.2.** In the Black-Scholes model, we can regard the stock as a trivial derivative of itself, paying  $f(S(T)) = S(T)$  at some time  $T$ . To be correct, a scheme for pricing

---

<sup>1</sup>One reason that this is a heuristic argument is that it is not clear that the concept of demand for wealth makes any sense. Later we will have an economically sound argument supporting the conclusions offered here.

<sup>2</sup>more precisely, no change to an equivalent measure



derivatives must tell us that the initial price of this trivial derivative is  $S(0)$ , the initial stock price. However

$$\begin{aligned}
\mathbf{E}[D(T)f(S(T))] &= \mathbf{E}[e^{-rT}S(T)] \\
&= \mathbf{E}[\exp(-rT)S(0)\exp((\mu - \sigma^2/2)T + \sigma W(T))] \\
&= \mathbf{E}[S(0)\exp((\mu - r - \sigma^2/2)T + \sigma W(T))] \\
&= S(0)e^{(\mu-r)T}\mathbf{E}[\exp((-\sigma^2/2)T + \sigma W(T))] \\
&= S(0)e^{(\mu-r)T}
\end{aligned}$$

since the expression inside the expectation is a  $\mathbf{P}$ -martingale starting at 1 (or by applying a Girsanov transformation). Thus this scheme mis-prices the stock. On the other hand,

$$\begin{aligned}
\mathbf{E}^{\mathbf{Q}}[D(T)f(S(T))] &= \mathbf{E}^{\mathbf{Q}}[S(0)\exp((\mu - r - \sigma^2/2)T + \sigma W(T))] \\
&= \mathbf{E}^{\mathbf{Q}}[S(0)\exp((\mu - r - \sigma^2/2)T - \sigma\lambda^\top T + \sigma(W(T) + \lambda^\top T))] \\
&= \mathbf{E}^{\mathbf{Q}}[S(0)\exp((-\sigma^2/2)T + \sigma W^{\mathbf{Q}}(T))] \\
&= S(0)
\end{aligned}$$

which is correct.

Assuming, as only makes sense, that  $\mu > r$  because investors demand a reward for holding the risky stock, using  $\mathbf{P}$  over-prices the stock. It is true that expected future wealth is greater if one puts  $S(0)$  dollars into buying one share of the stock rather than into a money market account, but this is offset by the nature of the stock's riskiness. According to our previous discussion, the many dollars that one gets from owning the stock when  $S(T)$  is high are not valued as highly as dollars when  $S(T)$  is low. To value the stock's or a derivative's payoff correctly, we must use  $\mathbf{Q}$ .

### 6.3. Vector Itô Processes

We will use vector Itô processes in order to model a market with multiple risky securities. Let  $X$  be a vector Itô process of  $N$  components based on a Wiener process  $W$  with  $K$  independent components, each a one-dimensional Wiener process. They model multiple independent sources of risk.

The arithmetic representation of an Itô process (2.6.2)

$$X(t) = X(0) + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s)$$

now needs further explanation. The convention is that  $X$  and  $W$  are column vectors of length  $N$  and  $K$  respectively. Then  $X(0)$  is a column vector of length  $N$ , or for short, a column  $N$ -vector. Its entries  $X_i(0)$  are the starting values of the  $i$ th component of the Itô process  $X$ , written. The random variable  $\mu(t)$  is also a column  $N$ -vector. However,  $\sigma(t)$  is

an  $N \times K$  matrix. We have

$$\int_0^t \sigma(s) dW(s) = \sum_{k=1}^K \int_0^t \sigma_{\bullet k}(s) dW_k(s)$$

where  $\sigma_{\bullet k}$  is the  $k$ th column of  $\sigma$ , i.e. the column  $N$ -vector whose  $i$ th component is the element  $\sigma_{ik}$ . Now the  $K$  integrals on the right-hand side are ordinary Itô stochastic integrals with only a single one-dimensional Wiener process  $W_k$  involved.

Breaking up (2.6.2) row by row yields, for the  $i$ th row,

$$X_i(t) = X_i(0) + \int_0^t \mu_i(s) ds + \int_0^t \sigma_{i\bullet}(s) dW(s) = X_i(0) + \int_0^t \mu_i(s) ds + \sum_{k=1}^K \int_0^t \sigma_{ik}(s) dW_k(s)$$

and here we can see explicitly the interpretation of all the coefficients:  $X_i(0)$  is the starting value of  $X_i$ ,  $\mu_i(t)$  is its arithmetic drift at time  $t$ , and  $\sigma_{ik}(t)$  is its time- $t$  arithmetic volatility on the  $k$ th Brownian motion  $W_k$ . The row  $K$ -vector  $\sigma_{i\bullet}(t)$  is the arithmetic volatility vector of  $X_i$  at time  $t$ .

So  $\sigma$  is the  $N \times K$  instantaneous arithmetic volatility matrix, so called because  $\sigma(t)$  contains arithmetic volatilities for the instant  $t$ . The instantaneous covariance matrix is  $N \times N$  and can be derived heuristically as follows:

$$\begin{aligned} \mathbf{Cov} [dX_i(t), dX_j(t)] &= \mathbf{Cov} \left[ \mu_i(t) dt + \sum_{k=1}^K \sigma_{ik}(t) dW_k(t), \mu_j(t) dt + \sum_{\ell=1}^K \sigma_{j\ell}(t) dW_\ell(t) \right] \\ &= \sum_{k=1}^K \mathbf{Cov} [\sigma_{ik}(t) dW_k(t), \sigma_{jk}(t) dW_k(t)] \\ &= \sum_{k=1}^K \sigma_{ik}(t) \sigma_{jk}(t) dt \end{aligned}$$

This is the basis for the vector Itô formula. The instantaneous covariance between  $X_i$  and  $X_j$  at time  $t$  is

$$\sum_{k=1}^K \sigma_{ik}(t) \sigma_{jk}(t) = \sigma_{i\bullet}(t) (\sigma_{j\bullet}(t))^\top,$$

the dot product between the volatility row vectors  $\sigma_{i\bullet}(t)$  and  $\sigma_{j\bullet}(t)$ . So the instantaneous covariance matrix is the  $N \times N$  matrix  $\sigma(t)\sigma^\top(t)$ .

**EXAMPLE 6.3.1** (Multidimensional Generalized Brownian Motion). Let  $X$  be a two-dimensional Itô process with constant coefficients:

$$\begin{aligned} X(t) &= \begin{bmatrix} 2 \\ -2 \end{bmatrix} + \int_0^t \begin{bmatrix} -1 \\ 1 \end{bmatrix} ds + \int_0^t \begin{bmatrix} 3 & 4 \\ 4 & 3 \end{bmatrix} dW(s) \\ &= \begin{bmatrix} 2 - t + 3W_1(t) + 4W_2(t) \\ -2 + t + 4W_1(t) + 3W_2(t) \end{bmatrix}. \end{aligned}$$

What is the probability that  $X_1(2)$  is negative given  $X_2(2)$ ? The components of  $X(t)$  are bivariate normal with mean vector and covariance matrix

$$\begin{bmatrix} 2-t \\ t-2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 25 & 24 \\ 24 & 25 \end{bmatrix} t.$$

(For the covariance matrix, the computations are  $3^2+4^2 = 25 = 4^2+3^2$  and  $3 \times 4 + 4 \times 3 = 24$ .) The conditional distribution of  $X_1(t)$  given  $X_2(t)$  is then normal with mean

$$\begin{aligned} \mathbf{E}[X_1(t)|X_2(t)] &= \mathbf{E}[X_1(t)] + \frac{\mathbf{Cov}[X_1(t), X_2(t)]}{\mathbf{Var}[X_2(t)]} (X_2(t) - \mathbf{E}[X_2(t)]) \\ &= 2-t + \frac{24}{25} (X_2(t) - (t-2)) \end{aligned}$$

and variance

$$\begin{aligned} \mathbf{Var}[X_1(t)|X_2(t)] &= \mathbf{Var}[X_1(t)] \left( 1 - \frac{\mathbf{Cov}[X_1(t), X_2(t)]^2}{\mathbf{Var}[X_1(t)] \mathbf{Var}[X_2(t)]} \right) \\ &= 25t \left( 1 - \frac{24^2}{25^2} \right) = \frac{49}{25} t. \end{aligned}$$

For  $t = 2$  we have conditional mean  $(24/25)X_2(2)$  and conditional variance  $98/25$ , so the conditional probability that  $X_1(2)$  is negative is

$$\mathbf{P}[X_1(2) < 0 | X_2(2)] = \Phi \left( \frac{0 - \mathbf{E}[X_1(2)|X_2(2)]}{\sqrt{\mathbf{Var}[X_1(2)|X_2(2)]}} \right) = \Phi \left( \frac{-(24/25)X_2(2)}{\sqrt{98/25}} \right) = \Phi \left( \frac{X_2(2)12\sqrt{2}}{35} \right).$$

The vector Itô formula in differential form says that where  $X$  is a vector Itô process satisfying  $dX(t) = \mu(t) dt + \sigma(t) dW(t)$ ,

$$\begin{aligned} df(t, X(t)) &= \left( f_t(t, X(t)) + \frac{1}{2} \text{tr}(\sigma^\top(t) f_{xx}(t, X(t)) \sigma(t)) \right) dt + f_x(t, X(t)) dX(t) \\ (6.3.1) \quad &= \left( f_t(t, X(t)) + f_x(t, X(t)) \mu(t) + \frac{1}{2} \text{tr}(\sigma^\top(t) f_{xx}(t, X(t)) \sigma(t)) \right) dt \\ &\quad + f_x(t, X(t)) \sigma(t) dW(t). \end{aligned}$$

Here  $f_x$  is the gradient (a row vector) and  $f_{xx}$  is the Hessian matrix.

The trace of a square matrix is the sum of the squared diagonal elements: if  $A$  is  $N \times N$ , then  $\text{tr}(A) = \sum_{i=1}^N A_{ii}$ . In the formula,

$$(6.3.2) \quad \text{tr}(\sigma^\top(t) f_{xx}(t, X_t) \sigma(t)) = \sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2 f}{\partial x_i \partial x_j}(t, X(t)) \sigma_{i\bullet}(t) \sigma_{j\bullet}^\top(t).$$

Here is a useful special case, sometimes referred to as integration by parts, sometimes as a product rule. Recall that in classical calculus,  $(fg)' = f'g + g'f$ , or in different notation,  $d(XY) = X dY + Y dX$ : the product rule. Equivalent to this is  $\int X dY = XY - \int Y dX$ : integration by parts.

EXAMPLE 6.3.2 (Itô product rule). Suppose  $dX(t) = \mu_X(t) dt + \sigma_X(t) dW(t)$  and  $dY(t) = \mu_Y(t) dt + \sigma_Y(t) dW(t)$ . Then  $X(t)Y(t) = f(t, [X(t)Y(t)]^\top)$  where  $f(t, x) = x_1x_2$ . We have

$$f_t \left( t, \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \right) = 0, \quad f_x \left( t, \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \right) = [Y(t) \quad X(t)], \quad f_{xx} \left( t, \begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} \right) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

By the vector Itô formula,

$$\begin{aligned} d(X(t)Y(t)) &= \left( 0 + \frac{1}{2}(0 + \sigma_X(t)\sigma_Y^\top(t) + \sigma_Y(t)\sigma_X^\top(t) + 0) \right) dt + [Y(t) \quad X(t)] \begin{bmatrix} dX(t) \\ dY(t) \end{bmatrix} \\ (6.3.3) \quad &= \sigma_X(t)\sigma_Y^\top(t) dt + X(t) dY(t) + Y(t) dX(t). \end{aligned}$$

Thus the Itô correction term for a product is the product of the instantaneous covariance between the two factors. In the product  $(dX(t))(dY(t))$ , the only nonnegligible term is  $\sigma_X(t)\sigma_Y^\top(t) dt$ , which comes from multiplying the stochastic parts together. So we can write an analogue of the product rule:

$$(6.3.4) \quad d(X(t)Y(t)) = X(t) dY(t) + Y(t) dX(t) + (dX(t))(dY(t)).$$

#### 6.4. Markets with Multiple Risky Securities

In the multidimensional case, the traded securities have a price vector  $S$  modeled as a vector Itô process:  $dS(t) = \mu(t) dt + \sigma(t) dW(t)$  where  $S$  and  $\mu$  are column  $N$ -vectors, the Wiener process  $W$  is a column  $K$ -vector, and  $\sigma$  is an  $N \times K$  matrix. Note that the coefficient processes  $\mu$  and  $\sigma$  are arithmetic, so the Black-Scholes model has  $\mu_i(t) = S_i(t)\mu_i$  and  $\sigma_{ik}(t) = S_i(t)\sigma_{ik}$ , for a constant vector  $\mu$  and constant matrix  $\sigma$ . We can write this in matrix notation by defining the diagonal matrix  $\text{diag}(S(t))$ , so that  $\mu(t) = \text{diag}(S(t))\mu$  and  $\sigma(t) = \text{diag}(S(t))\sigma$ . We designate the money market account by  $M(t) = S_0(t)$ , with  $dM(t) = r(t)M(t) dt$ . We also have to specify the dynamics of the risk-free rate  $r$ , which is not itself a traded security.

It is important to understand when such a model of a market will be arbitrage-free, and when complete. The answer has everything to do with market prices of risk and equivalent martingale measures (EMMs). Precise results are unnecessarily technical. Let it suffice to say that absence of arbitrage is close to being the same as the existence of an EMM, and completeness is close to being the same as uniqueness of the EMM. Also, an EMM is close to being the same as a market price of risk process  $\lambda$  that, with a risk-free rate process  $r$ , solves a generalization of Equation (6.1.1):

$$(6.4.1) \quad \sigma(t)\lambda^\top(t) = \mu(t) - r(t)S(t).$$

This is because the density  $d\mathbf{Q}/d\mathbf{P}$  arises as the stochastic exponential of  $\lambda$ . The existence and uniqueness of a market price of risk process  $\lambda$  are just a question of linear algebra.

To get a solution, generally we want the number of assets  $N \leq K$ , the number of risks. Actually, what we need is for any linear dependence among the rows of  $\sigma_t$  to generate a

corresponding dependence in the drifts  $\mu_t$ :

$$\sigma_i(t) = \sum_{j \neq i} \theta_j \sigma_j(t) \Rightarrow \mu_i(t) = r(t)S_i(t) + \sum_{j \neq i} \theta_j (\mu_j(t) - r(t)S_j(t)).$$

The principle is “same risks, same rewards.” Otherwise one can construct an arbitrage by hedging the  $i$ th asset with this portfolio of other assets to get an instantaneously riskless portfolio that has a return different from  $r$ . The technique is first to construct a portfolio of risky assets with zero volatility, then make its initial value zero by borrowing or lending: see Example 6.4.2.

EXAMPLE 6.4.1. Suppose  $\sigma(t) = \text{diag}(S(t))\sigma$  and  $\mu(t) = \text{diag}(S(t))\mu$  where the constants

$$\sigma = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} 0.1 \\ 0.09 \end{bmatrix}$$

and the risk-free rate is a constant  $r(t) = r = 0.05$ . The matrix  $\sigma$  has inverse

$$\sigma^{-1} = \frac{1}{3} \begin{bmatrix} -1 & 2 \\ 2 & -1 \end{bmatrix}$$

so we can compute directly that the market price of risk vector process is

$$\lambda^\top(t) = \sigma^{-1}(t)(\mu(t) - r(t)S(t)) = \sigma^{-1}(\mu - r1) = \begin{bmatrix} 0.01 \\ 0.02 \end{bmatrix}.$$

In this equation the presence of the column vector  $1$  is just supposed to make it a legitimate vector equation, in which the risk-free rate  $r$  is subtracted from each coordinate of the geometric drift vector  $\mu$ . What happens in this equation is that effectively the stock price  $S(t)$  cancels out of the equation: this does not happen in general, but does happen here in this multidimensional Black-Scholes model.

EXAMPLE 6.4.2. Suppose  $\sigma(t) = \text{diag}(S(t))\sigma$  and  $\mu(t) = \text{diag}(S(t))\mu$  where the constants

$$\sigma = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} 0.1 \\ 0.09 \end{bmatrix}$$

and the risk-free rate is a constant  $r(t) = r = 0.05$ . There is only one source of risk, so  $\lambda$  must be a scalar. From the first row of the matrix equation (6.4.1),  $S_1(t)\lambda = 0.1S_1(t) - 0.05S_1(t)$  so  $\lambda = 0.05$ , but from the second row,  $S_2(t)2\lambda = 0.09S_2(t) - 0.05S_2(t)$  so  $\lambda = 0.02$ . Therefore there is no market price of risk  $\lambda$  that solves the matrix equation, and the market is not arbitrage-free.

We could construct an arbitrage by following these steps. The first asset,  $S_1$ , is more attractive: it requires a higher value of  $\lambda$ , and it is plain that it has more return but less risk than  $S_2$ . So we start our portfolio by buying one share of  $S_1$ , that is,  $\theta_1 = 1$ . Then to give the portfolio zero volatility, choose

$$\theta_2(t) = -\frac{S_1(t)\sigma_1}{S_2(t)\sigma_2}.$$

To make the portfolio initially costless, include

$$\theta_0(0) = S_1(0) \left( \frac{\sigma_1}{\sigma_2} - 1 \right)$$

shares of the money market account. This portfolio has instantaneous arithmetic drift at time 0 of

$$\theta_0(0)M(0)r + \theta_1(0)S_1(0)\mu_1 + \theta_2(0)S_2(0)\mu_2 = S_1(0) \left( \left( \frac{\sigma_1}{\sigma_2} - 1 \right) r + \mu_1 - \frac{\sigma_1}{\sigma_2} \mu_2 \right) = S_1(0)(0.03)$$

so at this instant, the zero-cost portfolio is making a riskless positive return. Formally, we haven't shown that we have an arbitrage. We would have to compute the  $\theta_0(t)$  that makes the portfolio self-financing, and then check the portfolio value at some later time  $T$ , but this is a bit cumbersome.

EXAMPLE 6.4.3. Suppose

$$\sigma(t) = \begin{bmatrix} 1 & \sin t \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \mu(t) = \begin{bmatrix} 0.1 \\ 0.09 \end{bmatrix}.$$

There are two assets and two sources of risk, but whenever  $t$  is a multiple of  $\pi$ , the matrix  $\sigma(t)$  becomes degenerate: its two rows are the same. If there were a market price of risk  $\lambda$ , this would imply that  $\mu(t) - r(t)S(t)$  should also have two identical rows. But since the two rows of  $\mu(t)$  are not identical, this will not happen in general, so there can not be a market price of risk. At some times, the risks of the two assets are the same, but their rewards are not, hence they can not be explained by market prices of risk.

To make a solution unique, generally we want  $N \geq K$ , or more precisely, it requires that there be at least  $K$  linearly independent rows of the volatility matrix  $\sigma$ . When that is true, it is possible to construct a portfolio whose volatility vector has only one nonzero entry, thus identifying the market price of that risk. Otherwise, there will be more than one way of allocating the  $N$  assets' expected returns to the  $K$  sources of risk.

EXAMPLE 6.4.4. Suppose  $\sigma(t) = \text{diag}(S(t))\sigma$  and  $\mu(t) = \text{diag}(S(t))\mu$  where the constants

$$\sigma = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} 0.1 \\ 0.1 \end{bmatrix}$$

and the risk-free rate is a constant  $r(t) = r = 0.05$ . We have two unknowns, the two components of the market price of risk vector  $\lambda$ , in one equation,  $\lambda_1 + 2\lambda_2 = (0.1 - 0.05)$ , which leads to an infinite number of solutions.

In conclusion, we see that the most natural case of a complete market in which a unique EMM exists involves an invertible geometric volatility matrix.

### 6.5. Martingale Representation

The martingale representation theorem is a kind of converse of our earlier result that, under some technical conditions, any stochastic integral is a martingale. It says that a martingale adapted to the market filtration  $\mathcal{F}$  and starting at zero can be represented as a stochastic integral. In particular, a contingent claim paying  $Y$  at time  $T$  has a representation

$$Y = Y(0) + \int_0^T \xi(t) dW(t)$$

because  $\mathbf{E}[Y|\mathcal{F}_t] - \mathbf{E}[Y]$  is a martingale starting at zero and ending at  $Y - \mathbf{E}[Y]$ . So we see that  $Y(0) = \mathbf{E}[Y]$ .

Also, we have some freedom to choose how we construct the martingale to be represented. Actually, we are interested in finding a hedging strategy that relates to the contingent claim's no-arbitrage value, which we know involves discounting and the probability measure  $\mathbf{Q}$ , so use

$$D(T)Y = \mathbf{E}^{\mathbf{Q}}[D(T)Y] + \int_0^T \eta(t) dW^{\mathbf{Q}}(t).$$

Here  $D(t)V(t) = \mathbf{E}^{\mathbf{Q}}[D(T)Y|\mathcal{F}_t]$  is a  $\mathbf{Q}$ -martingale, which as the notation suggests is the discounted value of the contingent claim, or its hedging portfolio. A  $\mathbf{Q}$ -martingale starting at zero is  $D(t)V(t) - V(0)$ , where  $V(0) = \mathbf{E}^{\mathbf{Q}}[D(T)Y]$  is of course the initial value of the contingent claim. (Note that  $V(0) = D(0)V(0)$ .)

How does this relate to the problem of hedging the contingent claim  $Y$ , not  $D(T)Y$ ? From the martingale representation theorem, we have  $d(D(t)V(t)) = \eta(t) dW^{\mathbf{Q}}(t)$ , but we can also get an expression for  $d(D(t)V(t))$  by applying Itô's formula to the product of the discount factor  $D$  and hedging portfolio value  $V$ . Then we will see that there is a self-financing portfolio strategy  $\theta$  such that  $\theta(T)S(T) = V(T) = Y$ , and what this hedging strategy is. By the Itô product formula (6.3.4),  $d(D(t)V(t)) = D(t)dV(t) + V(t)dD(t) + (dD(t))(dV(t))$ . From Equation (6.2.1),  $dD(t) = -r(t)D(t)dt$ , while the self-financing condition is  $dV(t) = \theta(t) dS(t)$ . So we get

$$d(D(t)V(t)) = D(t)\theta(t) dS(t) - V(t)r(t)D(t) dt.$$

Risk-neutrally,  $S_i$  should have geometric drift  $r(t)$  at time  $t$ , just like the money market account. So the arithmetic drift of  $S$  under  $\mathbf{Q}$  is  $r(t)S(t)$ . Then plugging into the above  $dS(t) = r(t)S(t) dt + \sigma(t) dW^{\mathbf{Q}}(t)$  and  $V(t) = \theta(t)S(t)$ , we get

$$\begin{aligned} d(D(t)V(t)) &= D(t)\theta(t)r(t)S(t) dt + D(t)\theta(t)\sigma(t) dW^{\mathbf{Q}}(t) - \theta(t)S(t)r(t)D(t) dt \\ &= D(t)\theta(t)\sigma(t) dW^{\mathbf{Q}}(t). \end{aligned}$$

Equating the two expressions for  $d(D(t)V(t))$ , we have

$$\eta(t) = D(t)\theta(t)\sigma(t).$$

So generally, there will be a portfolio strategy  $\theta$  solving this equation when the volatility matrix  $\sigma$  has at least as many linearly independent rows as the  $K$  components of the Brownian motion (and hence of  $\eta$ ).

The problem with this approach is that the martingale representation theorem tells us that there exists some integrand  $\eta$  giving rise to the martingale  $D(t)V(t) - V(0)$ , but it gives us no help whatsoever in computing this  $\eta$ , so we also have no idea what the hedging portfolio  $\theta$  is. This result is of purely theoretical significance in understanding how the vector Itô process model for traded security prices gives rise to a complete market, under conditions on the volatility matrix  $\sigma$ . To compute hedging strategies in the Black-Scholes model, we had to construct and solve a PDE.

### 6.6. Decomposition

One way to gain some insight into the use of numeraires and the European call option is to decompose its payoff and price into two parts. The payoff

$$(S(T) - K)^+ = (S(T) - K)\mathbf{1}\{S(T) \geq K\} = S(T)\mathbf{1}\{S(T) \geq K\} - K\mathbf{1}\{S(T) \geq K\},$$

and we can treat these two parts as the payoffs of separate derivative securities. The first part pays the final stock price, but only if it is above  $K$ : this is known as a stock-or-nothing call. The latter part pays  $-K$  only when the stock price finishes above  $K$ . Think of this as being  $-K$  shares of a derivative that pays 1 when the stock price finishes above  $K$ : this is a cash-or-nothing call. This type of option is also called a **binary option**, because the outcome is binary: either it pays 1, or 0. So the “plain vanilla” European call option equals a portfolio of one stock-or-nothing call and  $-K$  cash-or-nothing calls, all struck at  $K$ .

Its price must then be the sum of these prices, and remember that we derived the plain vanilla call price by evaluating  $\mathbf{E}^{\mathbf{Q}}[e^{-r(T-t)}S(T)\mathbf{1}\{S(T) \geq K\}|\mathcal{F}_t]$  and  $\mathbf{E}^{\mathbf{Q}}[e^{-r(T-t)}K\mathbf{1}\{S(T) \geq K\}|\mathcal{F}_t]$  separately. So the price of the cash-or-nothing call is  $e^{-r(T-t)}\Phi(d_2)$  and that of the stock-or-nothing call is  $S(t)\Phi(d_1)$ . What the cash-or-nothing option pays is either 0 or  $e^{-rT}$  shares of the money market account, and it was easy to price this payoff using the money market account as the numeraire. What the stock-or-nothing option pays is either 0 or 1 shares of stock, so it is natural to try pricing this payoff using the stock as the numeraire.

Now the discount factor  $D(t) = 1/S(t)$ , so the discounted payoff is  $D(T)S(T)\mathbf{1}\{S(T) \geq K\} = \mathbf{1}\{S(T) \geq K\}$ . But under what probability measure are we taking the expectation of this discounted payoff? The equivalent martingale measure  $\tilde{\mathbf{P}}$  when  $S$  is the numeraire must make  $D(t)S(t) = 1$  and

$$D(t)M(t) = \frac{M(t)}{S(t)} = \frac{1}{S(0)} \exp\left(\left(r - \left(\mu - \frac{\sigma^2}{2}\right)\right)t - \sigma W(t)\right)$$

into martingales. Let  $\tilde{\lambda} = \lambda - \sigma = (\mu - r)/\sigma - \sigma$  and  $W_t^{\tilde{\mathbf{P}}} = W_t + \tilde{\lambda}t$ , so

$$\frac{M(t)}{S(t)} = \frac{1}{S(0)} \exp\left(\left(-\sigma\tilde{\lambda} - \sigma^2/2\right)t - \sigma W(t)\right) = \frac{1}{S(0)} \exp\left(-\sigma^2 t/2 - \sigma W^{\tilde{\mathbf{P}}}(t)\right).$$



This is a martingale under the probability measure that makes  $\exp(-\sigma^2 t/2 - \sigma W^{\tilde{\mathbf{P}}}(t))$  a stochastic exponential, which it is when  $W^{\tilde{\mathbf{P}}}$  is a Wiener process. So the equivalent martingale measure is the one under which  $W^{\tilde{\mathbf{P}}}$  is a Wiener process. According to the Girsanov transformation, that is  $\tilde{\mathbf{P}}$  where  $d\tilde{\mathbf{P}}/d\mathbf{P} = \exp(-\tilde{\lambda}^2 T/2 - \tilde{\lambda} W(T))$ .

Under  $\tilde{\mathbf{P}}$ , the conditional distribution of  $\ln(M(T)/S(T))$  given  $\mathcal{F}_t$  is normal with mean  $\ln(M(t)/S(t)) - \sigma^2(T-t)/2$  and variance  $\sigma^2(T-t)$ . So the value of the discounted payoff  $\mathbf{1}\{S(T) \geq K\}$  is

$$\begin{aligned} \mathbf{E}^{\tilde{\mathbf{P}}}[\mathbf{1}\{S(T) \geq K\} | \mathcal{F}_t] &= \tilde{\mathbf{P}} \left[ \ln \left( \frac{M(T)}{S(T)} \right) < \ln \left( \frac{M(T)}{K} \right) | \mathcal{F}_t \right] \\ &= \Phi \left( \frac{\ln(M(T)/K) - \ln(M(t)/S(t)) + \sigma^2(T-t)/2}{\sigma\sqrt{T-t}} \right) \\ &= \Phi \left( \frac{rT - \ln(K) - (rt - \ln(S(t))) + \sigma^2(T-t)/2}{\sigma\sqrt{T-t}} \right) \\ &= \Phi(d_1). \end{aligned}$$

Finally, since the payoff was being expressed in terms of shares of stock, this quantity must also be interpreted as a number of shares of stock. So the dollar price is  $S(t)\Phi(d_1)$ .

This approach is equivalent to and faster than our original approach to the same problem in the derivation of the Black-Scholes formula. What happens here is that we reach the probability measure  $\tilde{\mathbf{P}}$  in one step, rather than in two steps, from  $\mathbf{P}$  to  $\mathbf{Q}$  to  $\tilde{\mathbf{P}}$ . It is the same measure  $\tilde{\mathbf{P}}$  in both derivations, as we can see by observing

$$\begin{aligned} \frac{d\tilde{\mathbf{P}}}{d\mathbf{Q}} \frac{d\mathbf{Q}}{d\mathbf{P}} &= \exp \left( -\frac{\sigma^2}{2} T + \sigma W^{\mathbf{Q}}(T) \right) \exp \left( -\frac{\lambda^2}{2} T - \lambda W(T) \right) \\ &= \exp \left( -\frac{\sigma^2 + \lambda^2}{2} T + \sigma W^{\mathbf{Q}}(T) - \lambda W(T) \right) \\ &= \exp \left( -\frac{\sigma^2 + \lambda^2}{2} T + \sigma(W(T) + \lambda T) - \lambda W(T) \right) \\ &= \exp \left( -\frac{\sigma^2 - 2\sigma\lambda + \lambda^2}{2} T + (\sigma - \lambda)W(T) \right) \\ &= \exp \left( -\frac{(\sigma - \lambda)^2}{2} T + (\sigma - \lambda)W(T) \right) \\ &= \exp \left( -\frac{\tilde{\lambda}^2}{2} T - \tilde{\lambda} W(T) \right) \\ &= \frac{d\tilde{\mathbf{P}}}{d\mathbf{P}} \end{aligned}$$

since  $\tilde{\lambda} = \lambda - \sigma$ .

By choosing the right numeraire for this problem, the expectation becomes simple to evaluate under the appropriate probability measure. Notice that under this measure  $\tilde{\mathbf{P}}$ , associated with the stock as numeraire, the stock is

$$S(t) = S(0) \exp((\mu - \sigma^2/2)t + \sigma W(t)) = S_0 \exp((r + \sigma^2/2)t + \sigma W^{\tilde{\mathbf{P}}}(t))$$

using  $W(t) = W^{\tilde{\mathbf{P}}} + \tilde{\lambda}t = W^{\tilde{\mathbf{P}}}(t) + \sigma t - ((\mu - r)/\sigma)t$ . Of course the money market account is still  $M(t) = e^{rt}$ . So  $\tilde{\mathbf{P}}$  is an equivalent martingale measure, and one based on using one of the traded securities as a numeraire, but it is not a risk-neutral measure as  $\mathbf{Q}$  is. The geometric drift of the stock under  $\tilde{\mathbf{P}}$  is  $r + \sigma^2$ , by Itô's formula, while that of the money market account is still  $r$  (nothing can change this).

### 6.7. Problems

**PROBLEM 6.1.** Suppose  $dX(t) = X(t)(\mu_X(t) dt + \sigma_X(t) dW(t))$  and  $dY(t) = Y(t)(\mu_Y(t) dt + \sigma_Y(t) dW(t))$  where  $W$  is a vector Wiener process. Find the stochastic differential of the Itô process  $X/Y$ .

All of the following assume the model of Section 6.4: an  $N$ -vector of market security prices follows the SDE  $dS(t) = \mu(t) dt + \sigma(t) dW(t)$ , where  $W$  is a  $K$ -dimensional Wiener process under  $\mathbf{P}$ .

**PROBLEM 6.2.** Suppose  $\sigma(t) = \text{diag}(S(t))\sigma$  and  $\mu(t) = \text{diag}(S(t))\mu$  where the constants

$$\sigma = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 0 \\ 2 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \mu = \begin{bmatrix} 0.12 \\ 0.16 \\ 0.05 \end{bmatrix}$$

and the risk-free rate is a constant  $r(t) = r = 0.05$ . Does there exist a unique market price of risk vector? If so, compute it.

**PROBLEM 6.3.** Suppose that the arithmetic volatility for the two risky market securities and the market price of risk are the constant matrices

$$\sigma(t) = \sigma = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \lambda(t) = \lambda = \begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix}$$

and the risk-free rate is a constant  $r(t) = r = 0.05$ . What is the arithmetic drift vector  $\mu(t)$  (under  $\mathbf{P}$ )? Now suppose further that there is a derivative security whose unique no-arbitrage price  $C$  has constant arithmetic volatility  $[0 \ 1]$ . How many shares of the two risky assets does the replicating strategy have at time  $t$ ? What is the arithmetic drift of  $C$  (under  $\mathbf{P}$ )? Hints: do not forget that the self-financing replicating strategy might also involve the money market account, and do not give answers under  $\mathbf{Q}$ .

**PROBLEM 6.4.** Suppose that  $\mu(t) = \text{diag}(S(t))\mu$  and  $\sigma(t) = \text{diag}(S(t))\sigma$  where  $\mu$  and  $\sigma$  are constants, and that there exists a unique constant market price of risk  $\lambda$ .

- (1) Show that  $\mathbf{E}^{\mathbf{P}}[S_1(T)\pi(T)/\pi(t)|\mathcal{F}_t] = S_1(t)$ , i.e. using the pricing kernel reprices the first security correctly.

- (2) What is the time- $t$  no-arbitrage price of a derivative security paying  $\exp(\lambda W(T))$  at time  $T$ ?

## CHAPTER 7

### Futures and Dividends

#### 7.1. Futures

A **forward contract** is an agreement to exchange something, whether that be a quantity of pork bellies or heating oil, or the value of some bonds or a stock index, at a future time  $T$ , but for a price agreed on now. The price agreed upon now, but to be paid only at time  $T$ , is called the **forward price**. Unlike other derivative securities that we have discussed so far, no money changes hands when the contract is made. The forward price is *not* the cost to acquire the forward contract, but the amount of money that changes hands at the maturity  $T$ .

The question of pricing is not about finding the price for the contract, but finding the forward price  $F(0, T)$ , in terms of the **spot price** of the underlying,  $S(0)$ , which is the price paid to get it now (not later). The notation  $F(0, T)$  means the forward price at time 0 for delivery at time  $T$ , i.e. for a forward contract of maturity  $T$ . Consider the case of a stock that pays no dividends in a market where it is possible to buy or sell at time 0 a riskless bond paying 1 at time  $T$  for the price  $B(0, T)$ . For instance, in the Black-Scholes model,  $B(0, T) = e^{-rT}$ : just buy  $e^{-rT}$  shares of the money market account. Then a simple no-arbitrage argument shows that the forward price of the underlying must be  $S(0)/B(0, T)$ . The replicating portfolio for the forward contract is to buy the underlying for  $S(0)$  and sell  $F(0, T)$  of the riskless bond for  $F(0, T)B(0, T)$ : this produces the final payoff  $S(T) - F(0, T)$  at time  $T$ , and its cost now is  $S(0) - F(0, T)B(0, T)$ . This needs to be 0 to match the initial value of the forward contract, so  $F(0, T) = S(0)/B(0, T)$ . In the Black-Scholes model,  $F(0, T) = S(0)e^{rT}$ : the forward price is the initial stock price, appreciated at the risk-free rate. This is not the expected value of the stock price at  $T$ , which is  $\mathbf{E}[S_T] = S(0)e^{\mu T}$ , but rather the time- $T$  value of the debt required to buy the stock now.

The forward price is chosen to make the forward contract have an initial value of 0, but what happens to its value when time passes? Consider a forward contract for delivery at  $T$ , agreed upon at time 0, when the forward price was  $F(0, T) = S(0)/B(0, T)$ . At time  $t$ , the forward price for delivery at  $T$  has become  $F(t, T) = S(t)/B(t, T)$ . The owner of the forward contract struck at  $F(0, T)$  could now sell a new forward contract, removing all risk due to fluctuations in the price of the underlying. The resulting cashflow at time  $T$  would then be  $F(t, T) - F(0, T)$ . The value of this payoff as of time  $t$  is  $B(t, T)(F(t, T) - F(0, T)) = S(t) - S(0)B(t, T)/B(0, T)$ . The no-arbitrage version of this argument is that a portfolio that contains 1 forward contract struck at  $F(0, T)$ ,  $-1$  forward contract struck at  $F(t, T)$ , and  $F(0, T) - F(t, T)$  bonds maturing at  $T$  has value 0 at time  $T$ . So its price at time  $t$

must be 0 to avoid arbitrage, and therefore the time- $t$  price of the forward contract struck at  $F(0, T)$  is the negative of the sum of the values of the other elements of the portfolio:  $-(0 + B(t, T)(F(0, T) - F(t, T)))$ .

So we have seen that owning the forward contract is exactly like borrowing  $S(0)$  to buy the stock. Perhaps for this reason, forward contracts exist primarily as contracts negotiated between two parties, and not as exchange-traded securities. A forward contract involves a significant amount of credit risk. Suppose an airline enters into a forward contract with an oil company to buy fuel. The oil company should factor into the forward price the risk that the airline might be bankrupt and unable to pay when the time for delivery comes. This risk is different from that for a power utility, for example. In reality, different creditors can borrow at different rates, and this must be reflected in the forward prices that they get. This is unsuitable for exchange-traded securities, where the principle of one price for all reigns.

What would be like a forward contract, but suitable for trading on an exchange? A **futures contract**, which is also a contract to buy a stock or commodity at a future time  $T$  for an agreed-upon price called the **futures price**. It has the characteristic that the profit or loss from owning it is actually realized every day. This process is called **marking to market**. Suppose we now let  $F(t, T)$  symbolize the futures price for delivery at  $T$  as of time  $t$ . Then at the end of day  $i$ , an owner of a futures contract receives that day's change in the futures price:  $F(t_i, T) - F(t_{i-1}, T)$ . (If this is a negative number, then the owner of the contract must make a payment.) Because the changes in the futures price are thus accounted for daily, a futures contract is worth 0 again at the end of each day.

Someone who had shorted a futures contract would get  $F(t_{i-1}, T) - F(t_i, T)$  at the end of day  $i$ . A futures market is, in dollar terms, a zero-sum game, so that there are always an equal number of long and short positions, and all the accounts balance. Nobody is allowed implicitly to run up a debt as with forward contracts; the market mechanisms are designed to minimize credit issues. Indeed, one is actually required to post a **margin** in account with one's futures broker: the size of the margin is a feature of the futures contract, and is chosen to be small relative to the futures price, and yet very likely to be larger than a daily price change.

There are many details that we are not modeling: typically, margin accounts pay interest like a money market account; if one's margin account falls too low, one receives anxious or threatening calls from one's broker, who is authorized to close out your positions if you do not post more margin; there may be an impact from financial commodities' dividends or interest payments, or physical commodities' storage costs, convenience yields, or seasonality factors. Imagine that we are modeling futures on a non-dividend-paying stock.

Because of marking-to-market, a futures contract is much more complicated than a forward contract, so much so that the futures price and forward price need not be the same, if interest rates are stochastic. The reason is that it is not the same thing to receive  $S(T) - F(0, T)$  at time  $T$  as it is to receive daily payments that sum to this amount for each day between times 0 and  $T$ . However, since we have not studied stochastic interest rates yet, for the moment we assume that interest rates are deterministic, in which case the forward price and futures prices are the same.

A no-arbitrage argument shows why. Here we use  $F$  to denote the futures price, which after all we are about to show is the same as the forward price denoted the same way. Suppose that on day  $i$ , that is, over the time interval  $[t_i, t_{i+1}]$ , we are short  $B(t_{i+1}, T)$  futures contracts: this is possible because interest rates are deterministic, so  $B(t_{i+1}, T)$  is already known at  $t_i$ . At the end of the day, at  $t_{i+1}$ , the mark-to-market payment is  $-B(t_{i+1}, T)(F(t_{i+1}, T) - F(t_i, T))$ , and we spend this to buy  $F(t_i, T) - F(t_{i+1}, T)$  bonds maturing at  $T$ . If we do this every day from  $0 = t_0$  until  $T = t_m$ , the result is a time- $T$  payoff of

$$\sum_{i=0}^{m-1} (F(t_i, T) - F(t_{i+1}, T)) = F(0, T) - F(T, T),$$

but  $F(T, T) = S(T)$ : the price as of  $T$  to get the underlying commodity at  $T$ . We can also buy the underlying at time 0 while shorting  $S(0)/B(0, T)$  bonds for net zero cost, which gives us the time- $T$  payoff of  $S(T) - S(0)/B(0, T)$ . Putting these two zero-cost strategies together, we get a final payoff of  $F(0, T) - S(0)/B(0, T)$ . To avoid arbitrage, this payoff must be zero, i.e.  $F(0, T) = S(0)/B(0, T)$  for the futures price, the same as the forward price.

Recall that the extended Black-Scholes model is

$$\begin{aligned} B(t, T) &= \exp\left(-\int_t^T r(u) du\right) \\ S(t) &= S(0) \exp\left(\int_0^t (\mu(u) - \sigma^2(u)/2) du + \int_0^t \sigma(u) dW(u)\right) \end{aligned}$$

where  $r$ ,  $\mu$ , and  $\sigma$  are all deterministic functions. Note  $M(t) = 1/B(0, t)$ , but it is helpful for us to focus on bond prices in order to express conveniently forward/futures prices, which are the same because interest rates are deterministic. The risk-neutral dynamics are

$$\begin{aligned} dB(t, T) &= r(t)B(t, T) dt \\ S(t) &= S(t) (r(t) dt + \sigma(t) dW^{\mathbf{Q}}(t)). \end{aligned}$$

Then we have

$$\begin{aligned} F(t, T) &= \frac{S(t)}{B(t, T)} = \frac{S(0)}{B(0, T)} \frac{\exp\left(\int_0^t (r(u) - \sigma^2(u)/2) du + \int_0^t \sigma(u) dW^{\mathbf{Q}}(u)\right)}{\exp\left(\int_0^t r(u) du\right)} \\ &= \frac{S(0)}{B(0, T)} \exp\left(-\frac{1}{2} \int_0^t \sigma^2(u) du + \int_0^t \sigma(u) dW^{\mathbf{Q}}(u)\right) \end{aligned}$$

and because the second factor is a stochastic exponential, this shows that the futures price has zero risk-neutral drift and is itself a  $\mathbf{Q}$ -martingale. (This also follows by applying Problem 6.1 to the risk-neutral SDEs.) The reason for this is that the futures contract is costless to enter into: having no capital tied up in it (neglecting the margin account), one does not “deserve” any expected return.

## 7.2. Dividends

Most stocks pay dividends, so it would be nice to modify the Black-Scholes model slightly to account for this fact. Our fixes will not be perfect: we will model dividends as being somehow predictable, either constant or proportional to stock price. In reality, companies can change their dividend policies, but these fixes are better than nothing. A simple and elegant way to incorporate dividends into the Black-Scholes model is to assume a constant proportional dividend yield, that is, at time  $t$  the stock pays dividends at a rate  $qS(t)$ , where  $q \geq 0$  is a constant. However, what companies actually do is announce in advance that they will pay a certain lump sum on a given date. For the short-term (perhaps one year) covered by dividend announcements, it might be preferable to use a model of deterministic, discrete dividends, but such models are a bit more irritating to deal with.

The idea of a model with constant proportional dividend yield is that a company with will tend to pay out as dividends a constant proportion of its value, so that dividends per share will be a fixed fraction of the share price. The assumption of a continuous yield is an approximation made for mathematical convenience. It is a better approximation for an index than for an individual stock: stocks in the index pay on various dates dividends small relative to the index value. This model of dividends also works for convenience yields on commodity futures, as well as their storage costs (in which case take  $q < 0$ ).

With proportional dividend yield  $q$ , the instantaneous gain from holding a share of stock is  $dS(t) + qS(t) dt$ . Someone shorting a share would also have to pay out the dividends. This affects the self-financing condition for a portfolio: the dividend payout goes into the money market account or into further stock purchases. Therefore we end up with a different PDE from the ordinary Black-Scholes PDE. Of course, this results in a new risk-neutral measure and new prices for derivative securities. Let's go through the relevant parts of the Black-Scholes analysis in this modified model.

Consider a derivative security having terminal payoff  $g(S(T))$  at  $T$  and making a continuous payment stream at rate  $h(u, S(u))$  at time  $u$ . Assume it has a price  $f(t, S(t))$ , and by no-arbitrage,  $dV(t) = df(t, S(t)) - h(t, S(t)) dt$  where  $V$  is the value of the replicating portfolio, which must provide us not only with the value of the derivative security at the end, but also with the continuous payment stream. The new self-financing condition is  $dV(t) = dG(t) = \theta_1(t)(dS(t) + qS(t) dt) + \theta_0(t) dM(t)$ , and plugging in for  $dS(t) = S(t)(\mu dt + \sigma dW(t))$  and  $dM(t) = rM(t) dt$ , we have

$$\begin{aligned} df(t, S(t)) &= dV(t) + h(t, S(t)) dt \\ &= (\theta_1(t)S(t)(\mu + q) + \theta_0(t)M(t)r - h(t, S(t))) dt + \theta_1(t)S(t)\sigma dW(t). \end{aligned}$$

From the time-dependent Itô formula, we had

$$df(t, S(t)) = \left( f_t(t, S(t)) + S(t)\mu f_S(t, S(t)) + \frac{\sigma^2}{2} S^2(t) f_{SS}(t, S(t)) \right) dt + S(t)\sigma f_S(t, S(t)) dW(t)$$

so it is still true that the number of shares of stock in the replicating portfolio  $\theta_1 = f_S(t, S(t))$ . And again by the no-arbitrage principle,  $\theta_0(t)M(t) = f(t, S(t)) - f_S(t, S(t))S(t)$ .

Now we compute that the drift of  $f(t, S(t))$  is

$$\begin{aligned}\theta_1(t)S(t)(\mu + q) + \theta_1(t)M(t)r &= f_S(t, S(t))S(t)(\mu + q) + (f(t, S(t)) - f_S(t, S(t))S(t))r \\ &= rf(t, S(t)) + (\mu + q - r)S(t)f_S(t, S(t)).\end{aligned}$$

Equating this with the drift from Itô's formula yields the PDE

$$(7.2.1) \quad (r - q)xf_x(t, x) - rf(t, x) + f_t(t, x) + \frac{1}{2}\sigma^2x^2f_{xx}(t, x) + h(t, x) = 0.$$

Applying the Feynman-Kač formula, the time- $t$  value of a derivative security with payoff  $g(S(T))$  at time  $T$  and its own (totally separate) continuous payment stream  $h(u, S(u))$  for  $u \in [0, T]$  is

$$\mathbf{E}_{t, S(t)}^{\mathbf{Q}} \left[ e^{-r(T-t)}g(S(T)) + \int_t^T e^{-r(u-t)}h(u, S(u)) du \right]$$

where

$$(7.2.2) \quad dS(t) = S(t)((r - q) dt + \sigma dW^{\mathbf{Q}}(t)).$$

Notice that the only thing that has changed from the Black-Scholes PDE is that now the dividend-paying stock has drift  $r - q$  under the probability measure  $\mathbf{Q}$ .

This may still be fairly called a risk-neutral measure. The total return on the stock is the sum of capital appreciation and dividend payment, and the sum of the geometric drift under  $\mathbf{Q}$  plus dividend yield is  $(r - q) + q = r$ , the same as for the money market account. In the original Black-Scholes model, the market price of risk was  $(\mu - r)/\sigma$ . Now it is  $(\mu - (r - q))/\sigma = (\mu + q - r)/\sigma$ , which is more: to get geometric drift  $\mu$  plus dividend yield  $q$  is greater compensation than just getting geometric drift  $\mu$ .

From this point it is not difficult to duplicate the Black-Scholes analysis: we simply have a different geometric drift  $r - q$  for the stock under  $\mathbf{Q}$ . We will carry out the analysis under the extended Black-Scholes model, where  $\mu$ ,  $r$ ,  $q$ , and  $\sigma$  are all deterministic functions of time. Then under  $\mathbf{Q}$ , the conditional distribution of  $\ln S(T)$  given  $\mathcal{F}_t$  is

$$(7.2.3) \quad N \left( \ln S(t) + \int_t^T \left( r(u) - q(u) - \frac{\sigma^2(u)}{2} \right) du, \int_t^T \sigma^2(u) du \right).$$

Here  $M(t)/M(T) = B(t, T) = \exp(-\int_t^T r(u) du)$ . We have

$$\begin{aligned}\mathbf{E}_{t, S(t)}^{\mathbf{Q}}[-B(t, T)K\mathbf{1}\{S(T) > K\}] \\ &= -B(t, T)K\mathbf{Q}[\ln S(T) > \ln K | \mathcal{F}_t] \\ &= -B(t, T)K\Phi \left( \frac{\ln(S(t)/K) + \int_t^T \left( r(u) - q(u) - \frac{\sigma^2(u)}{2} \right) du}{\sqrt{\int_t^T \sigma^2(u) du}} \right)\end{aligned}$$



and

$$\begin{aligned}
& \mathbf{E}_{t,S(t)}^{\mathbf{Q}}[B(t,T)S(T)\mathbf{1}\{S(T) > K\}] \\
&= B(t,T)\mathbf{E}^{\mathbf{Q}}\left[S(t)\exp\left(\int_t^T\left(r(u)-q(u)-\frac{\sigma^2(u)}{2}\right)du+\int_t^T\sigma(u)dW^{\mathbf{Q}}(u)\right)\mathbf{1}\{S(T) > K\}|\mathcal{F}_t\right] \\
&= S(t)\exp\left(-\int_t^Tq(u)du\right)\mathbf{E}^{\mathbf{Q}}\left[\exp\left(-\frac{1}{2}\int_t^T\sigma^2(u)du+\int_t^T\sigma(u)dW^{\mathbf{Q}}(u)\right)\mathbf{1}\{S(T) > K\}|\mathcal{F}_t\right] \\
&= S(t)\exp\left(-\int_t^Tq(u)du\right)\tilde{\mathbf{P}}[S(T) > K|\mathcal{F}_t]
\end{aligned}$$

by a Girsanov transformation with  $W^{\tilde{\mathbf{P}}}(t) = W^{\mathbf{Q}}(t) - \sigma t$  as before. Now the calculation is similar to above to get

$$\tilde{\mathbf{P}}[S(T) > K|\mathcal{F}_t] = \Phi\left(\frac{\ln(S(t)/K) + \int_t^T\left(r(u)-q(u) + \frac{\sigma^2(u)}{2}\right)du}{\sqrt{\int_t^T\sigma^2(u)du}}\right)$$

so instead of the Black-Scholes formula, for the European call price we get

$$\begin{aligned}
(7.2.4) \quad & S(t)\exp\left(-\int_t^Tq(u)du\right)\Phi\left(\frac{\ln(S(t)/K) + \int_t^T\left(r(u)-q(u) + \frac{\sigma^2(u)}{2}\right)du}{\sqrt{\int_t^T\sigma^2(u)du}}\right) \\
& - B(t,T)K\Phi\left(\frac{\ln(S(t)/K) + \int_t^T\left(r(u)-q(u) - \frac{\sigma^2(u)}{2}\right)du}{\sqrt{\int_t^T\sigma^2(u)du}}\right).
\end{aligned}$$

Because the risk-neutral geometric drift of the stock is decreased by  $q$ , the probabilities  $\mathbf{Q}[S(T) > K|\mathcal{F}_t]$  and  $\tilde{\mathbf{P}}[S(T) > K|\mathcal{F}_t]$  are both diminished. Also, the time- $t$  value of getting the stock at time  $T$  is now only  $S(t)\exp\left(-\int_t^Tq(u)du\right)$ , not  $S(t)$ . Some of the value of the stock leaks away in the form of dividends between  $t$  and  $T$ . This must be accounted for if you use the stock as numeraire to price the first term!

Much as in Section 3.2, we can compute the Greeks of this European call option in the extended Black-Scholes model with dividends. Note that we get different answers for the same option because we have changed the model. Much as  $B(t,T) = \exp\left(-\int_t^Tr(u)du\right)$ , let  $Q(t,T) = \exp\left(-\int_t^Tq(u)du\right)$  and  $V(t,T) = \int_t^T\sigma^2(u)du$ . The no-arbitrage option price can now be written

$$f(t,S) = SQ(t,T)\Phi(d_1) - KB(t,T)\Phi(d_2) \quad \text{where} \quad d_{1,2} = \frac{\ln\left(\frac{SQ(t,T)}{KB(t,T)}\right) \pm \frac{1}{2}V(t,T)}{\sqrt{V(t,T)}}.$$

Some useful facts are

$$\begin{aligned}\frac{\partial d_1}{\partial S} &= \frac{\partial d_2}{\partial S} = \frac{1}{S\sqrt{V(t,T)}} \\ \frac{\partial d_2}{\partial t} &= \frac{\partial d_1}{\partial t} + \frac{\sigma^2(t)}{2\sqrt{V(t,T)}} \\ \phi(d_2) &= \phi(d_1) \frac{SQ(t,T)}{KB(t,T)}\end{aligned}$$

and the partial derivatives are

$$\begin{aligned}\Theta &= q(t)Q(t,T)S\Phi(d_1) - r(t)B(t,T)K\Phi(d_2) - \frac{SQ(t,T)\sigma^2(t)\phi(d_1)}{2\sqrt{V(t,T)}} \\ \Delta &= Q(t,T)\Phi(d_1) \\ \Gamma &= \frac{Q(t,T)\phi(d_1)}{S\sqrt{V(t,T)}}.\end{aligned}$$

There are two main changes: first, one share of stock received at  $T$  is worth  $Q(t,T) \leq 1$  shares of stock now, which affects the hedging policy, thus  $\Delta$  and  $\Gamma$ . The dividend impacts  $\Theta$  because the replicating portfolio now not only pays interest on its borrowings, but gets a dividend at a rate  $q(t)$  times the value invested in the stock. Second, volatility enters in a more complicated way: for the most part, we care about the total remaining volatility  $\sqrt{V(t,T)}$ , but in  $\Theta$ , the losses due to rebalancing when the stock wiggles and returns to its previous value have to do with the current instantaneous volatility  $\sigma(t)$ .

### 7.3. Problems

**PROBLEM 7.1.** In this same extended Black-Scholes model of a stock with continuous proportional dividend yield, find the no-arbitrage price of a European call option with strike  $K$  and maturity  $T$  on the stock futures contract with maturity  $U \geq T$ .

**PROBLEM 7.2.** In Problem 7.1, let  $U = T$ , so that the maturity of the futures contract is the same as that of the option written on it. Is the option on the futures contract the same as an option (of the same strike and maturity) on the spot, i.e. the underlying stock? Why or why not? If the options are the same, show that your no-arbitrage pricing formulas give the same price, or explain how and why they are different. If the options are not the same, explain how their no-arbitrage prices differ. What do you think would be the best way to hedge each of these two options: with the stock, with the futures, or both? What if we were talking about a stock index, not a stock?

## CHAPTER 8

### Computation

There are two main types of computations in financial engineering. One type assumes that a model (and its parameters) are given, then deduces a value such as a price for a derivative security, an optimal hedging strategy, or a risk measure of a portfolio. The major computational tools here are simulation and numerical solution of partial differential equations. The other type is called **calibration** of a model: given a framework for modeling, with unknown parameters, it chooses values of the parameters given market data, usually current prices or price histories. The main tool here is optimization, frequently of a statistical flavor (e.g. likelihood maximization or least-squares minimization). Numerical methods allow the financial engineer to apply more sophisticated models to more complicated situations where analytical results are not available.

In this chapter, we learn simulation and calibration by applying them to a simple model where analytical results are available for comparison, and where data is readily available to use in calibration. This is the extended Black-Scholes model for futures prices, applied to the usual “vanilla” European options. In subsequent chapters, we will use simulation and calibration to study more sophisticated models and securities.

#### 8.1. Simulation

If we want to compute the no-arbitrage price of a derivative security as an expected discounted payoff, but can not evaluate the expectation by hand, simulation can give us a numerical estimate of the expectation. Suppose we want to evaluate  $\mathbf{E}^{\mathbf{Q}}[Y]$ , where  $Y$  is the payoff discounted by the money market account, and  $\mathbf{Q}$  is the risk-neutral measure. (In general, we can use any numeraire to discount, as long as we use the associated equivalent martingale measure.) Suppose also that we can sample from the distribution of  $Y$  under  $\mathbf{Q}$ . Produce  $n$  such independent, identically distributed (iid) samples  $Y^{(1)}, \dots, Y^{(n)}$ . Then  $\bar{Y} = \sum_{j=1}^n Y^{(j)}/n$  is a sample average estimate of the mean  $\mathbf{E}^{\mathbf{Q}}[Y]$ . Moreover, by the central limit theorem, if  $n$  is large enough, then  $\bar{Y}$  is approximately distributed as  $N(\mathbf{E}^{\mathbf{Q}}[Y], \mathbf{Var}^{\mathbf{Q}}[Y]/n)$ . This allows us to compute an approximate confidence interval at level  $\alpha$  in the usual way:  $\bar{Y} \pm z_{\alpha/2} \hat{s}/\sqrt{n}$  where  $z_{\alpha/2} = \Phi^{-1}(\alpha/2)$  and  $\hat{s} = \sqrt{\sum_{j=1}^n (Y^{(j)} - \bar{Y})^2/n}$  is the sample standard deviation of  $Y^{(1)}, \dots, Y^{(n)}$ : having assumed that  $n$  is large, we will not quibble about whether  $n$  or  $n - 1$  should appear in the denominator. See any elementary statistics textbook for a discussion of confidence intervals.

We now focus on the case of pricing a European call option in the extended Black-Scholes model. The underlying is a stock index with a continuous dividend yield. The no-arbitrage call price is given analytically in equation (7.2.4).

Because we are only interested in the option's payoff, which depends only on the final stock index level  $S(T)$ , we can just simulate that. The distribution of  $\ln S(T)$  (conditional on what we know at time  $t$ ) is in formula (7.2.3). So we can simulate the  $j$ th value of  $S(T)$  as

$$S^{(j)}(T) = \exp \left( \ln S(t) + \int_t^T \left( r(u) - q(u) - \frac{\sigma^2(u)}{2} \right) du + \sqrt{\int_t^T \sigma^2(u) du} Z^{(j)} \right)$$

where  $Z^{(1)}, \dots, Z^{(n)}$  are iid standard normal. Any software you are using for simulation should have a routine for producing such random numbers. Then our Monte Carlo estimate for the option price at  $t$  is

$$\exp \left( - \int_t^T r(u) du \right) \frac{1}{n} \sum_{j=1}^n (S^{(j)}(T) - K)^+.$$

Another way to simulate is based on the SDE (7.2.2). Whenever we have an Itô process obeying equation (5.1.4), the result of the Feynman-Kač formula, we can simulate  $X$  by the **Euler scheme** of discretizing time into  $m$  steps each of length  $h = (T - t)/m$ , so  $t_i = t + ih$ . For the  $j$ th path, let  $\hat{X}^{(j)}(t_0) = x$ , and

$$(8.1.1) \quad \hat{X}^{(j)}(t_{i+1}) = \hat{X}^{(j)}(t_i) + \mu(t, \hat{X}^{(j)}(t_i))h + \sigma(t, \hat{X}^{(j)}(t_i))\sqrt{h}Z_{i+1}^{(j)}$$

where all the  $Z_i^{(j)}$  for  $i = 1, \dots, m$  and  $j = 1, \dots, n$  are iid standard normal. The reason for writing  $\hat{X}$  instead of  $X$  is that we should realize that when we discretizing time like this, we may not be sampling from the correct distribution anymore: in general,  $\mu(u, X(u))$  and  $\sigma(u, X(u))$  do not have to be constant over the time interval  $[t_i, t_{i+1})$ , and the conditional distribution of  $X(t_{i+1})$  given  $X(t_i)$  does not have to be normal with mean  $X(t_i) + \mu(t_i, X(t_i))h$  and variance  $\sigma^2(t, X(t_i))h$ . This means that our simulation may be estimating the wrong expectation, resulting in **discretization bias**.

In this situation, to reduce discretization bias, it looks like a better idea to simulate the log stock price rather than the stock price itself. This is because the (arithmetic) drift and volatility coefficients of the log are deterministic, while those of the stock price are not. Indeed, we can eliminate discretization bias entirely in this case by defining

$$a_i = \int_{t_i}^{t_{i+1}} (r(u) - q(u) - \sigma^2(u)/2) du \quad \text{and} \quad b_i^2 = \int_{t_i}^{t_{i+1}} \sigma^2(u) du$$

and simulating  $X^{(j)}(t_0) = x$ ,

$$(8.1.2) \quad X^{(j)}(t_{i+1}) = X^{(j)}(t_i) + a_i + b_i Z_{i+1}^{(j)},$$

then computing  $S^{(j)}(t_i) = \exp(X^{(j)}(t_i))$ .

Another use of simulation is to evaluate the quality of an approximate hedging strategy. In reality, it is not possible to follow a hedging strategy that changes continuously over time, nor is it practical to come close: the transaction costs would be too great. With an approximate hedging strategy, we have hedging error, i.e. there is a nonzero profit or loss left over at the end: we sold a contingent claim paying  $(S(T) - K)^+$ , executed the approximate hedging strategy  $\theta$ , and find that we end up with a residual  $\theta(T)S(T) - (S(T) - K)^+$  which is typically nonzero. Before doing this, we should get inform ourselves as to what the distribution of the residual might be. We care about the residual under  $\mathbf{P}$ , a subjective probability measure intended to describe the real world, not under  $\mathbf{Q}$  or any other equivalent martingale measure, which is just a construct that we use to compute no-arbitrage prices.

One straightforward way to construct an approximate hedging strategy is simply to rebalance the portfolio at a finite number of times  $0 = t_0, \dots, t_m = T$ , changing nothing in between. In this case,  $\theta$  is a simple stochastic process. Trading in discrete time this way is certainly practicable. One particular choice of such a strategy is to choose  $\theta_1(t_i)$  equal to the value of the perfect continuous-time hedging strategy at that moment. Note that it is *not* obvious that this is the best choice. Also, if we are to do this in a self-financing way, we will *not* be able to have  $\theta_0(t_i)$  equal to the value prescribed by the continuous-time strategy in general.

When simulating the result of this hedging process, we need to simulate all  $m$  steps, in order to see what we would do at each time. It is irrelevant that we could simulate  $S(T)$  correctly without any intermediate values, because we need to know those values in order to see what hedge we use and what gains we incur at each step. However, we can and should simulate based on the SDE as in equation (8.1.2) so as to avoid discretization bias. In the simulation, we keep track not only of  $S$  and  $X$ , the stock price and its log, but also  $\theta_0$  and  $\theta_1$ , the number of shares held in the money market account and in the stock.

At each step  $i$ , we will always choose  $\theta_1(t_i) = \Delta = Q(t, T)\Phi(d_1)$ , as described in Section 7.2. At step 0, we have  $\theta_0(0) = C - \theta_1(0)$ , where  $C$  is whatever premium we received for selling the call. Consider how to update at step  $i + 1$ . We need to choose a way to account for the dividends. Let us say that they are continuously reinvested in the stock. Then we have  $\theta_1(t_i)$  shares of stock at time  $t_i$ , and we have  $\theta_1(t_i) \exp(\int_t^T q(u) du)$  shares of stock when the  $i$ th step ends at time  $t_{i+1}$ . Then we decide how many shares of stock we want to have at that time, which is  $\theta_1(t_{i+1})$  based on the new computation of  $\Delta$ . So we buy  $\theta_1(t_{i+1}) - \theta_1(t_i) \exp(\int_t^T q(u) du)$  shares at  $t_{i+1}$ , and I pay for this purchase by taking that much money out of the money market account. So we must update  $\theta_0$  as follows:

$$\theta_0(t_{i+1}) = \theta_0(t_i) - \frac{(\theta_1(t_{i+1}) - \theta_1(t_i) \exp(\int_t^T q(u) du))S(t_{i+1})}{M(t_{i+1})}.$$

Then at the end, our profit on the trade is  $\theta_0(T)M(T) + \theta_0(T)S(T) - (S(T) - K)^+$ .

## 8.2. Calibration

There is little orthodoxy on the subject of calibration, which is the process of using data to choose the parameters of a model. There are two main kinds of calibration: using historical data from the underlying price process, and using current prices of market-traded derivative securities. An example of the former is statistical estimation of the historical volatility of a stock price in order to choose  $\sigma$  in the Black-Scholes model. An example of the latter is finding a value of  $\sigma$  that makes the results of the Black-Scholes formula come close to fitting simultaneously the prices of European call and put options on this underlying, but with various strikes and maturities. These approaches are sometimes combined: for instance, to calibrate a Black-Scholes model with two stocks, one might use option prices on each of them to choose the volatility magnitudes, and use historical stock price data to estimate the correlation between the stocks. This would be enough information to construct a  $2 \times 2$  volatility matrix.

At this point, we will focus on calibration to current market prices. The idea of this kind of calibration is use current market prices of derivative securities to choose the parameters of a model so that the model prices match market prices well. Then the model can be used to price nonmarketed derivative securities in a way that can be regarded as “consistent” with market prices.

We write the model price of a derivative security as  $h(\psi; \varphi)$  where  $\psi$  are unknown parameters that need to be calibrated, while  $\varphi$  are the known terms of the contract specific to a derivative security. The notation with a semicolon indicates that we regard  $h$  as a function of  $\psi$ , and merely acknowledge the dependence on  $\varphi$ : if you like,  $h(\cdot; \varphi)$  is a different function for each value of  $\varphi$ . We regard the market observables as fixed and do not include them in the notation. For example, consider the Black-Scholes model. The market observables are  $S$  and  $r$ : ignoring some troublesome facts (e.g. there is actually a bid-ask spread for the stock price, and interest rates are not actually constant), we can simply observe the current stock price and the current instantaneous interest rate. The contract terms are strike price  $K$  and maturity  $T$ : these vary from one option to another, but are always known. The unknown parameter is the volatility  $\sigma$ ; the drift  $\mu$  is irrelevant for option pricing in this model.

In the Black-Scholes model, the option pricing formula  $h$  is invertible, and for an option whose contract terms are  $\varphi$  and whose market price is  $P$ ,  $h^{-1}(P; \varphi)$  is called the **implied volatility**. That is, if the Black-Scholes model were true, the volatility would have to be this, in order for the option to have the price that it does in the market. Here the semicolon notation shows that  $\varphi$  is not involved in the inverse: we have  $P = h(h^{-1}(P; \varphi); \varphi)$ .

Implied volatility can be very useful for discussing options, because market participants have more intuition about the level of implied volatility for an underlying than for the price of an option of specified strike and maturity. However, because the Black-Scholes model is wrong, implied volatility and actual (statistical, historical) volatility do not have to have a close relationship. In particular, one usually sees implied volatilities that are greater than statistical volatilities, especially for options of short maturity, or which are deep out of the money (meaning  $S(t) < K$  for a call). This is because the lognormal distribution

underestimates the probability of large, rapid changes in stock price relative to changes of modest size. Given a statistically accurate volatility (if such a thing were possible), the Black-Scholes formula would tend to give option prices that are too small: so the implied volatility, which matches option prices, is larger.

In general, calibration is not as simple as using  $h^{-1}$ , or finding the values of parameters implied by option prices. Inverse problems like this are often ill-posed, meaning that there might be no solutions, or many solutions. If the model is not correct, then there might be no value of the parameter vector  $\psi$  that simultaneously matches many market prices. This is quite clear with the Black-Scholes model, where we typically find an **implied volatility surface** such that implied volatility is not the same for every option (as the model would suggest) but rather curved, and typically highest for short-maturity, out-of-the-money options. No one value of  $\sigma$  will get every option price right. If the data is dirty or stale (i.e. reflects trades that took place hours ago, not recently) then there can even be apparent arbitrages in the data, in which case there will certainly be no value of  $\psi$  that matches these prices. But there could also be multiple values of  $\psi$  that do just as well as each other at matching the market prices.

For these reasons, people usually regard calibration as an optimization problem of minimizing an objective function that expresses the error between model prices and market prices. The idea is that minimizing this error makes the model “as close as possible” to matching market prices. The value  $\psi^*$  of the parameter vector  $\psi$  that does this is said to be a “best fit.” One simple objective function would be the sum of squared errors:

$$\mathcal{O}(\psi; P, \varphi) = \sum_{k=1}^N (P_k - h(\psi; \varphi_k))^2$$

where  $P_k$  is the market price and  $\varphi_k$  the contract terms of the  $k$ th security among the  $N$  data points used in calibration. This is simple, but has the drawback that it tends to result in pricing errors of approximately the same size for each security, even though some of them (e.g. options deep in the money) will have prices much greater than others (e.g. options deep out of the money), by factors of perhaps 1000. We might be uncomfortable with this disparity in the relative pricing errors  $1 - h(\psi; \varphi_k)/P_k$ , and one way to attack that problem is to use instead the objective function

$$\mathcal{O}(\psi; P, \varphi) = \sum_{k=1}^N (\ln P_k - \ln h(\psi; \varphi_k))^2,$$

that is, first take the logarithm of prices: the difference of logs is the log of the ratio, so we can see the minimization of this objective function as focusing on relative pricing errors. There are other possible transformations, for example, the whole family of transformations  $g(y) = (g^\alpha - 1)/\alpha$  for  $\alpha \in (0, 1]$ , with  $g(y) = \ln y$  corresponding to the choice  $\alpha = 0$ . Aside from influencing the nature of the best fit, using a transformation can also have a big impact on the time required to execute whatever nonlinear optimization routine you might be using.

It would be nice to modify the objective function to reflect the fact that some of the market prices being used embody higher-quality information than others. For example, options that are of a short, but not too short maturity, and are slightly out of the money, are often more liquid than some others. The prices of more liquid securities are more recent and depend less on the vagaries of supply and demand. The prices of illiquid securities drop a lot when someone decides to sell, and rise a lot when someone decides to buy. So we might like to give more weight in the objective to the more liquid options, reasoning that it seems better to price them well while pricing worse the illiquid options, whose prices are not so informative. This seems to be an advanced topic.

One might think that the way to deal with a model that fits the data poorly is to increase the number of parameters. For instance, it might be a good idea to replace the Black-Scholes model with the extended Black-Scholes model when calibrating to prices of options of several maturities  $T_1, \dots, T_m$ . Instead of just one parameter  $\sigma$ , we have  $m$  parameters  $b_1, \dots, b_m$ , where  $b_i$  is the standard deviation of the log stock price over  $[T_{i-1}, T_i]$ . Sometimes adding parameters helps, but it has its dangers. Even if  $h$  is a ridiculous function, having nothing whatever to do with finance, we might be able to get a good fit of “model prices”  $h(\psi; \varphi)$  to market prices  $P$  if there are enough parameters, i.e. if  $\psi$  is sufficiently high-dimensional. This is called curve-fitting, and usually causes poor results. Remember, the objective is not to fit the data well with some curve, but to come up with model prices for non-marketed derivative securities that work, i.e. enable us to hedge so that we end up with long-term profits that are adequate compensation for the risks that remain.

This is hard to assess without putting money on the line, but one way to assess from the safety of one’s computer whether the prices from the calibrated model are “good” is to test them against out-of-sample market data. That is, calibrate the model using some market prices, then compare the calibrated model prices for a new set of market-traded options to their market prices. If they fit poorly out-of-sample, this model and calibration scheme do not inspire much confidence. This often happens when a model is over-fitted. A dilemma here is that one would like to use out-of-sample data to check whether the model is doing well, yet one would like to calibrate the model to all of the available data so as not to “waste” it! The scheme of calibration with penalization and cross-validation solves this problem at the same as another one: one often wants to use a model with many parameters while avoiding over-fitting.

The idea behind penalization is that if the fitted parameter vector  $\psi^*$  is ugly, this is a sign of over-fitting. In the extended Black-Scholes model, we expect that the graph of  $\psi = [b_1 \ \dots \ b_m]$  should be relatively smooth and pretty, not wiggling up and down all the time, which would be hard to rationalize financially. It would be difficult to believe that such a parameter vector bears much relation to the actual behavior of the market price process. If you like, you can think of this in Bayesian terms: we have a prior predisposition to believe that the parameter vector is pretty, but we also want to update our beliefs to reflect data about current derivative security prices. So we can invent a penalty term to penalize values



of  $\psi$  where, for instance, the components change too much. Examples are:

$$\mathcal{P}(\psi) = \sum_{i=1}^m (b_i - b_{i-1})^2 \quad \text{or} \quad \mathcal{P}(\psi) = \sum_{i=2}^m (b_i - 2b_{i-1} + b_{i-2})^2,$$

where the first penalizes changes between adjacent parameters, and the second penalizes changes in the rate of change of parameters. Note that these penalties depend on the linear temporal structure of  $\psi = b$ , the time series of log stock price standard deviations. If the parameters had a different structure, we would need a different form for the penalty.

Now we will minimize the penalized objective function

$$\mathcal{O}(\psi; P, \varphi) = \mathcal{E}(P, h(\psi; \varphi)) + \lambda \mathcal{P}(\psi),$$

where  $\mathcal{E}(x, y)$  is the usual error term, such as  $\sum_k (x_k - y_k)^2$  discussed above, and  $\lambda$  is the strength of penalization. Therefore we will no longer choose the error-minimizing  $\psi$ , which might suffer from over-fitting, but a different, prettier  $\psi$ , which we hope will be more useful.

The problem with this scheme is that, a priori, we have no good way of choosing the penalization strength  $\lambda$ . If it is too small, the penalty will count for little, and we will over-fit. If it is too big, we will get a pretty  $\psi$  (nearly a constant or nearly a line, respectively, for the two penalty terms discussed above) that provides a bad fit to the data. So now we also need a way to choose  $\lambda$ , and this is what cross-validation does. The idea is that for each data point  $k$ , we see how we would do in matching this market price after we calibrated to all the other data points using penalization strength  $\lambda$ . When done for each  $k$ , this yields an entire vector of model prices. Then we choose  $\lambda$  to minimize the resulting error, known as the **cross-validation error**.

For each  $k$ , let  $P_{-k}$  and  $\varphi_{-k}$  be respectively the vector of market prices and of contract terms, but with the  $k$ th component removed. Then find the minimizer  $\tilde{\psi}_k(\lambda)$  of

$$\mathcal{O}(\psi; P_{-k}, \varphi_{-k}, \lambda) = \mathcal{E}(P_{-k}, h(\psi; \varphi_{-k})) + \lambda \mathcal{P}(\psi).$$

Let  $\tilde{\psi}(\lambda)$  be a matrix formed of these vectors  $\tilde{\psi}_k(\lambda)$ , and likewise  $\varphi$  be a matrix formed of the vectors  $\varphi_k$ . After doing this for each  $k$ , we can compute the cross-validation error

$$\mathcal{CV}(\lambda; P, \varphi) = \mathcal{E}(P, h(\tilde{\psi}(\lambda); \varphi)).$$

After finding the optimal value  $\lambda^*$  that minimizes  $\mathcal{CV}(\lambda; P, \varphi)$ , we then optimize one last time to get our fitted parameter vector  $\psi^*$ , the minimizer of

$$\mathcal{O}_{CV}(\psi; P, \varphi) = \mathcal{E}(P, h(\psi; \varphi)) + \lambda^* \mathcal{P}(\psi).$$

This approach helps a great deal to counteract the problem of over-fitting, but it is very computationally expensive, because we have to do nested optimizations.

### 8.3. Problems

If using Excel, submit a printout of your spreadsheet, with labels showing the formulae that you use to compute the cells. If using MATLAB, submit the output and source code of your programs. In either case, make sure that it is easy for the grader to find your answers

and to see how you arrived at them. Groups may submit a single copy with the names of all group members.

**PROBLEM 8.1.** An **Asian option** is one based on an average price. Here define  $\bar{S}_j = \sum_{i=1}^j S(t_i)/j$ , the discrete, arithmetic average of the prices at the dates  $t_1, \dots, t_j$ . Consider an Asian option paying  $(\bar{S}_m - K)^+$  at time  $T = t_m$ . Take  $T = 0.25$  and  $m = 13$  for weekly averaging dates, and let  $K = 105$ . Assume the ordinary Black-Scholes model with  $S(0) = 100$ ,  $r = 0.05$ ,  $\mu = 0.15$ , and  $\sigma = 0.2$ . Construct a simulation estimate for the no-arbitrage price of this Asian option in this model. Simulate  $n = 100$  paths. On the basis of information from this simulation, how many paths do you think you need in order to have 95% confidence of estimating the option's price within 1% of the true value? Now run a simulation with this many paths, and report a 95% confidence interval for the option price.

**PROBLEM 8.2.** Under "External Links" on the course page, there is a link to the Chicago Board Options Exchange term sheet for options on the S&P500 index. Under "Assignments" there is a webpage with data from the market close of November 21, 2002. The data contains the closing level of the S&P500 index (933.79), and data for calls and puts of many strikes and two maturities:

- Last Sale is the price at which the most recent trade in this option took place. Note that the data does not say when that trade was.
- Net is the change between Last Sale and the price of the trade previous to that one.
- Bid and Ask are prices at which market makers most recently offered to buy or sell (respectively) this option.
- Vol is how many of this option traded today.
- Open Int is "open interest," which equals the total number of options owned by those people who are long this option. Because the net total number of options is zero, this also equals the total number of options sold by those people who are short this option.

Calibrate the extended Black-Scholes model with continuous dividend yield to this data. Show what you did, explain why you did it, and report clearly the resulting values of all parameters. Provide graphs illustrating the quality of the model's fit to the data: think hard about how to present this information, and strive to make the graphs as informative as possible. How useful do you think this model is for this market? What, if anything, could be changed to make it more useful?

Do not ask the instructor or assistant questions outside of section or office hours, even when it is not clear how to proceed. For example, the data does not contain an interest rate: look it up elsewhere. Or again, the term sheet tells you when the options expire, but you have to decide for yourself how to measure the amount of time until that day.

## CHAPTER 9

### Volatility Models

At this point, we have seen that we have problems calibrating even the extended Black-Scholes model to European stock index options data. The flexibility to let volatility change over time helps to fit prices of options with different maturities, but the fit is poor when looking at options with the same maturity but different strikes. One approach to overcoming this problem is to allow the volatility to have some sort of dependence on the stock price itself. (Another is to leave the realm of Itô processes, for instance, by considering models with jumps.)

There are two main types of models that do this: **local volatility** models:

$$(9.0.1) \quad dS(t) = S(t)(\mu(t) dt + \sigma(t, S(t)) dW(t))$$

and **stochastic volatility** models:

$$(9.0.2) \quad \begin{aligned} dS(t) &= S(t)(\mu(t) dt + V(t) dW(t)) \\ dV(t) &= a(t) dt + b(t) dW(t). \end{aligned}$$

Notice that neither of these is a model of implied volatility. The point of both of them is that the stock's geometric volatility has to be more interesting than just a deterministic process, but they go about this quite differently. A local volatility model implies a complete market, because the stock is the only source of risk: the volatility is a function  $\sigma$  of time and stock price. In a stochastic volatility model, typically the stock price  $S$  and its volatility  $V$  are correlated, but not perfectly dependent, and the result is an incomplete market: the risks associated with changes in the level of volatility  $V$ , which is not itself a traded security, can not be hedged away entirely just by trading in the stock.

#### 9.1. Local Volatility

A local volatility model is so called because the volatility of the stock depends only on a space-time “location”: the current time  $t$ , and the current value of the stock  $S(t)$ . In this model,  $\sigma$  is a function of two variables that tells the geometric volatility given the time and stock price. The drift  $\mu$  is merely a function of time: a deterministic process. The rationale for this seems to be that there is not very much point in allowing the drift too to depend on the stock price, because we could not calibrate such a function, since it does not influence option prices in this model.

On the other hand, the different volatilities that take effect at different stock prices do affect various options in different ways. The local volatilities at very low stock price levels will have little effect on a call option with a high strike price: if the stock price gets that

low, this option is likely to have zero payoff. But local volatilities at very low stock price levels will have a significant effect on a put option with a low strike: high volatility at low stock price levels increases the value of this option, because it increases the likelihood that the stock will drop far below the strike, resulting in a big payoff.

Local volatility models appeared in 1994 in *Risk* magazine<sup>1</sup> where articles by Dupire (of Paribas) and Derman & Kani (of Goldman Sachs) showed how they could be used to calibrate binomial tree models to implied volatility surfaces. With the general local volatility model (9.0.1), one is tempted to have as many parameters as options. If there are maturities  $T_1, \dots, T_m$  and strikes  $K_1, \dots, K_n$ , one may discretize space-time into  $m(n+1)$  rectangles. This can lead to overfitting, yet it is hard to justify deliberately coarsening this discretization. So this is a perfect setting for penalized calibration, or for using some technique such as spline-fitting to create a smooth surface with fewer parameters.

We do not cover binomial trees in this course. Instead, we consider simulation of a local volatility model. As we saw before, it makes sense to use the Euler scheme to discretize the log stock price  $X(t) = \ln S(t)$ , not the stock price. The discretization

$$X^{(j)}(t_{i+1}) = X^{(j)}(t_i) + (\mu(t_i) - \sigma^2(t_i, S^{(j)}(t_i))/2)h + \sigma(t_i, S^{(j)}(t_i))\sqrt{h}Z_{i+1}^{(j)}$$

is not exact in this case, because  $\sigma(t, S(t))$  does change during a time step, but it is still better. Another advantage to this scheme is that the simulated stock price can not become negative. For risk-neutral simulation, we would have in the above  $\mu(u) = r(u)$ , the interest rate.

One specific type of local volatility model is the **constant elasticity of variance** (CEV) model, which specifies

$$(9.1.1) \quad \sigma(t, S) = \delta S^{\beta-1}.$$

Usually, we would also take the drift  $\mu(t) = \mu$  a constant, and thus have the model  $dS(t) = \mu S(t) dt + \delta S^\beta(t) dW(t)$ . That is, the local volatility does not depend on time, only the level of the stock price. If  $\beta = 1$ , we get the Black-Scholes model with constant geometric drift  $\mu$  and volatility  $\delta$ : we can see that  $\delta$  governs the magnitude of volatility. But  $\beta$  governs the rate at which geometric volatility changes with changes in the stock price. Typically, a CEV model has  $\beta < 1$ , so that the geometric volatility increases when the stock price decreases.

One attraction of CEV is that options can be priced in closed form, in terms of the noncentral chi-square distribution. The derivation is excessively involved, but the result is that, when the interest rate is  $r$  and the continuous proportional dividend yield is  $q$ , the no-arbitrage price at time  $t$  of the European call option with maturity  $T$  and strike  $K$  is

$$(9.1.2) \quad S(t)e^{-q(T-t)}\bar{F}\left(y, 2 + \frac{1}{1-\beta}, \xi\right) - Ke^{-r(T-t)}\bar{F}\left(y, 2 - \frac{1}{1-\beta}, \xi\right)$$

---

<sup>1</sup>Available in the business school library: read it.

where  $F(y, n, \xi)$  is the noncentral chi-square cumulative distribution function with  $n$  degrees of freedom and noncentrality parameter  $\xi$ ,  $\bar{F}$  is the complement  $1 - F$ , and

$$\begin{aligned} y &= \alpha K^{2(1-\beta)} \\ \xi &= \alpha S^{2(1-\beta)}(t) \exp(2(r-q)(1-\beta)(T-t)) \\ \alpha &= \frac{2(r-q)}{\delta^2(1-\beta)(\exp(2(r-q)(1-\beta)(T-t)) - 1)}. \end{aligned}$$

This noncentral chi-square cdf  $F$  is available in the statistics toolbox of MATLAB.

## 9.2. Stochastic Volatility

“Stochastic volatility” usually means that, in the model, the actual (statistical, not implied) volatility of the stock can change randomly, in a way that can not be completely explained by the passage of time or the change in the stock price. We need at least  $K = 2$  components of the Wiener process to make this work. The correlation between changes in the stock price and changes in volatility is very important: this relationship is the crux of local volatility models, and we do not want to give it up here. Again, usually we find that this correlation is negative. This tends to give out-of-the-money puts (with low strikes) a high implied volatility relative to out-of-the-money calls (with high strikes), a phenomenon which is frequently strongly present in option prices. This leads some people to refer to a volatility “smirk,” having an asymmetric shape like a check mark, rather than “smile,” having a symmetric shape like the letters U or V.

One particular case is the **Heston** model:

$$\begin{aligned} dS(t) &= S(t)(\mu(t) dt + V(t) dW_1(t)) \\ (9.2.1) \quad dV(t) &= -\beta V(t) dt + \delta dW(t). \end{aligned}$$

By Itô’s formula, the SDE for the instantaneous variance, as opposed to volatility, is:

$$dV^2(t) = (\delta^2 - 2\beta V^2(t)) dt + 2\delta V(t) dW(t).$$

From this we can see that the instantaneous variance process is mean-reverting: when  $V^2(t)$  is below the “mean” level  $\delta^2/(2\beta)$ , it has positive drift, and negative drift when it is above. Note that we need  $\beta > 0$  for this mean level to be finite and positive: the stock volatility  $V$  needs to have a negative drift, otherwise the variance of stock returns will blow up.

Remember, we do not care if volatility  $V$  becomes negative, because it is the absolute value (equivalently, its square) that really counts. Everything works the same when  $V$  is negative as when it is positive. If  $V(t) < 0$  and  $\beta > 0$ , the drift of  $V$  still tends to make it smaller in absolute value. Likewise, although  $S$  will now vary inversely to  $W$ , the absolute value of  $V$  *also* varies inversely to  $W$ , because increasing  $V$  means decreasing  $|V|$ .

The arithmetic volatility of the instantaneous variance  $V^2$  is proportional to the square root, namely the instantaneous stock volatility  $V$ . We will see when covering interest rates that these two features of  $V^2$ , mean reversion and square root volatility, mean that it is

obeying the Cox-Ingersoll-Ross model. The point of this is that the instantaneous variance, the rate at which stock returns' variance increases with time, should stay around some moderate level over the long run (not grow without limit), it should change less when it is small than when it is large, and it should never become negative. The square root volatility prevents negativity: when  $V^2$  is near zero, its volatility gets small, and its drift tends to pull it back up. This is all similar to what we want an interest rate model to do.

In terms of option pricing, we face an interesting problem, typical of incomplete markets. There is no obvious way to get a unique EMM:  $S$  is the only asset here, other than a money market account, because  $V$  is not an asset. So under an EMM for the money market as numeraire,  $S$  should have drift  $r$ , but what drift should  $V$  have? There is no way to decide. We know how the drift of  $W_1$  changes under a Girsanov transformation from  $\mathbf{P}$  to " $\mathbf{Q}$ ," because  $dW_1$  appears in  $dS$ : but what about  $W_2$ , whose change of drift we would also need to know in order to find the drift of  $V$ ? If we want to price options via expected discounted payoff under an EMM in the usual way, we will have to come up with a way of picking a drift for  $V$ . One way to do this is via calibration: we pick a drift for  $V$  that causes option prices under this model to agree with the actual market prices.

We can not observe volatility directly in the market, and the stock price  $S$  by itself is not a Markov process: looking at the recent history of stock price changes will tell you something about the level that volatility was at recently, and therefore about the distribution of future stock price changes. One approach to calibration is to do just this, that is, to try to infer the current value  $V(t)$  of volatility from recent stock price history. Then it would still remain to calibrate from current market prices the remaining parameters,  $\beta$  and the 2-vector  $\delta$ . The norm  $\|\delta\| = \sqrt{\delta_1^2 + \delta_2^2}$  gives the "vol of vol," and the significance of the two components is that when  $\delta_1$  is relatively larger, there is greater correlation between  $V$  and  $S$ , which depends only on  $W_1$ . Another approach to calibration would be to calibrate  $V(t)$ ,  $\beta$ , and  $\delta$  from market prices. As always, it is possible to attempt to calibrate all the parameters from historical data, but this is not a good idea if the model is too far from the truth. One way to check for a problem here is to see if you get very different answers when calibrating to historical data based on different time scales: that is, if you look at volatilities of hourly or weekly returns, or if you define "recent" history as the last 20 or 100 time periods.

In simulating this Heston model, it seems like a good idea to simulate  $V$ , which already has a constant volatility:

$$V^{(j)}(t_{i+1}) = V^{(j)}(t_i) - (\beta V(t_i) + \delta^2/2)h + \delta_1 \sqrt{h} Z_{i+1}^{(j,1)} + \delta_2 \sqrt{h} Z_{i+1}^{(j,2)}$$

and  $X = \ln S$ :

$$X^{(j)}(t_{i+1}) = X^{(j)}(t_i) + (\mu(t_i) - V^2(t_i)/2)h + V(t_i) \sqrt{h} Z_{i+1}^{(j,1)}.$$

More sophisticated schemes are possible: for instance, once we have simulated not only  $V(t_i)$  but also  $V(t_{i+1})$ , we could use that information in simulating the change in the log stock price  $X$ , based on the idea that its average volatility over  $[t_i, t_{i+1}]$  is probably between  $V(t_i)$  and  $V(t_{i+1})$ , rather than right at  $V(t_i)$ .

## APPENDIX A

### Final Review

Mathematical Tools:

- Itô's formula, especially:
  - exp/ln transformations; arithmetic vs. geometric representations
  - product rule
- (conditional) probabilities and moments for Itô processes
- Feynman-Kač formula
- Girsanov transformation

Concepts:

- portfolio strategies: tame, self-financing, arbitrage
- no-arbitrage pricing by replication
- hedging and greeks
- numeraire and equivalent martingale measures: risk-neutral pricing as special case
- simulation
- calibration

Models: Black-Scholes variants:

- ordinary
- multi-dimensional
- extended
- with continuous proportional dividend yield
- for futures

## APPENDIX B

### Solutions to Problems

**Problem 1.1.** Both strategies are self-financing because the portfolio allocation is not changing. The long stock strategy is tame because  $S(t)$ , as a geometric Brownian motion, is bounded below by 0. Since it has a positive initial cost  $S(0)$ , it is not an arbitrage. The short stock strategy is not tame because  $S(t)$  is unbounded above. Since it has a negative terminal payoff  $-S(T) < 0$ , it is not an arbitrage.

**Problem 1.2.** The payoff of the portfolio long one call struck at  $K_1$  and one struck at  $K_3$ , and short two calls struck at  $K_2$ , is nonnegative. Therefore  $0 \leq C(K_1) + C(K_3) - 2C(K_2)$  by the no-arbitrage principle. Thus we get as an upper bound  $C(K_2) \leq \frac{1}{2}(C(K_1) + C(K_3)) = 12.5$ . For the lower bound, consider the portfolio long one call struck at  $K_1$  and short one struck at  $K_2$ . This portfolio has a payoff bounded above by 10. But a payoff of 10 must be worth 9.7 now, in order to avoid arbitrage with the bond. So  $C(K_1) - C(K_2) \leq 9.7$ , i.e.  $C(K_2) \geq C(K_1) - 9.7 = 10.3$ .

**Problem 2.1.** These can all be solved by differentiating. The probability of loss

- (1) Decreases, because you make more in the money market account.
- (2) Decreases, because you make more in the stock.
- (3) Increases, because the risk of stock losses increases. This is obvious, but we can also see the relationship from

$$\frac{\ln\left(1 + \frac{\theta_0(1-e^{rT})}{\theta_1 S(0)}\right) - (\mu - \frac{1}{2}\sigma^2)T}{\sigma\sqrt{T}} = \frac{\ln\left(1 + \frac{\theta_0(1-e^{rT})}{\theta_1 S(0)}\right) - \mu T}{\sigma\sqrt{T}} + \frac{\sigma\sqrt{T}}{2}$$

where the numerator of the first term is negative, because  $1 - e^{rT} < 0$ , so we are taking the log of something less than 1.

- (4) Sorry, this is a bit too complicated to get into, given  $\mu > \sigma^2/2$ .
- (5) Decreases, because more value is in the money market, which will surely increase in value. Again, note that  $1 - e^{rT}$  is negative.

**Problem 2.2.** The increment  $Y(s) - Y(0) = Y(0)(\exp(\mu s + \sigma W(s)) - 1)$  whereas  $Y(2s) - Y(s) = Y(s)(\exp(\mu s + \sigma(W(2s) - W(s))) - 1)$ . The Wiener process increments  $W(s) = W(s) - W(0)$  and  $W(2s) - W(s)$  are independent with distribution  $N(0, s)$ , so  $\exp(\mu s + \sigma W(s)) - 1$  and  $\exp(\mu s + \sigma(W(2s) - W(s))) - 1$  are independent and have the same distribution. However  $Y(0)$  is constant while  $Y(s)$  is a lognormal random variable, so  $Y(s) - Y(0)$  and  $Y(2s) - Y(s)$  do not have the same distribution. They are dependent, since



$Y(2s) - Y(s) = (Y(0) + (Y(s) - Y(0)))X$  where  $Y(0)$  is a constant and  $X$  is independent of  $Y(s) - Y(0)$ .

**Problem 2.3.** The mean function is

$$\begin{aligned}\mu_Z(t) &= \mathbf{E}[Z(t)] = \mathbf{E}[Z(0) + \sigma W(t) + (Z(T) - Z(0) - \sigma W(T))(t/T)] \\ &= Z(0) + \sigma \mathbf{E}[W(t)] + (Z(T) - Z(0) - \sigma \mathbf{E}[W(T)])(t/T) \\ &= Z(0) + 0 + (Z(T) - Z(0) - 0)(t/T) = Z(0) \frac{T-t}{T} + Z(T) \frac{t}{T}.\end{aligned}$$

The covariance function is

$$\begin{aligned}c_Z(t, s) &= \mathbf{Cov}[Z(t), Z(s)] \\ &= \mathbf{Cov}[Z(0) + \sigma W(t) + (Z(T) - Z(0) - \sigma W(T))(t/T), \\ &\quad Z(0) + \sigma W(s) + (Z(T) - Z(0) - \sigma W(T))(s/T)] \\ &= \mathbf{Cov}[\sigma(W(t) - W(T))(t/T), \sigma(W(s) - W(T))(s/T)] \\ &= \sigma^2 \left( \mathbf{Cov}[W(t), W(s)] - \mathbf{Cov}[W(t), W(T)] \frac{s}{T} - \mathbf{Cov}[W(s), W(T)] \frac{t}{T} \right. \\ &\quad \left. + \mathbf{Cov}[W(T), W(T)] \frac{st}{T^2} \right) \\ &= \sigma^2 \left( \min\{s, t\} - \frac{st}{T} - \frac{st}{T} + \frac{st}{T} \right) = \sigma^2 \left( \min\{s, t\} - \frac{st}{T} \right)\end{aligned}$$

The variance  $\mathbf{Var}[Z(t)] = \sigma^2(t - t^2/T)$  and this is at its maximum,  $T/4$ , at  $t = T/2$ .

**Problem 2.4.** Because  $W(t/T)$  has the distribution  $\mathcal{N}(0, t/T)$ ,  $\tilde{W}(t) = \sqrt{T}W(t/T)$  has the distribution  $\mathcal{N}(0, t)$ . Also  $\tilde{W}(0) = \sqrt{T}0 = 0$ . Its increment is  $\tilde{W}(t) - \tilde{W}(s) = \sqrt{T}(W(t/T) - W(s/T))$ . If we let  $v = t/T$  and  $u = s/T$ , we can see clearly that we are talking about Wiener process increments of the form  $W(v) - W(u)$ , which have the properties of stationarity and independence. Multiplying by a constant has no effect on stationarity and independence. If a function  $f(t)$  is continuous, then so is  $af(bt)$ . Thus continuity of the sample path  $W(\omega)$ , which is a function of time, implies continuity of  $\tilde{W}(\omega)$ , because  $\tilde{W}(\omega, t) = \sqrt{T}W(\omega, t/T)$ : we just are plugging in  $a = \sqrt{T}$  and  $b = 1/T$ .

**Problem 2.5.** The approximate Wiener process is always a linear combination of standard normal random variables, so the mean function is always zero. For each  $m$ , the covariance function

$$c_m(t, s) = c_{W^{(m)}}(t, s) = \mathbf{Cov} [W^{(m)}(t), W^{(m)}(s)] = \sum_{i=1}^{2^m} \tilde{H}_i(s) \tilde{H}_i(t).$$

The relevant Schauder functions (up to  $n = 2^2 = 4$ ) are

$$\begin{aligned}\tilde{H}_1(t) &= t \\ \tilde{H}_2(t) &= \begin{cases} t & \text{for } t \in [0, 1/2] \\ 1-t & \text{for } t \in [1/2, 1] \end{cases} \\ \tilde{H}_3(t) &= \begin{cases} \sqrt{2}t & \text{for } t \in [0, 1/4] \\ \sqrt{2}(1/2-t) & \text{for } t \in [1/4, 1/2] \\ 0 & \text{for } t \in [1/2, 1] \end{cases} \\ \tilde{H}_4(t) &= \begin{cases} 0 & \text{for } t \in [0, 1/2] \\ \sqrt{2}(t-1/2) & \text{for } t \in [1/2, 3/4] \\ \sqrt{2}(1-t) & \text{for } t \in [3/4, 1] \end{cases}\end{aligned}$$

The covariance function of a true Wiener process is  $s$  since  $s \leq t$ . For  $m = 0$ ,  $c_0(t, s) = \tilde{H}_1(s)\tilde{H}_1(t) = st$ . This is correct only on the line segments in  $G^{(0)} = \{(t, s) | s = 0 \text{ or } t = 1\}$ .

Next,  $c_1(t, s) = c_0(t, s) + \tilde{H}_2(s)\tilde{H}_2(t)$ , and the second term must be evaluated over the three regions  $s \leq t \leq 1/2$ ,  $s \leq 1/2 \leq t$ , and  $1/2 \leq s \leq t$ . (Remember we are doing the calculation only for  $s \leq t$ .) The result is

$$c_1(t, s) = \begin{cases} 2st & \text{for } s \leq t \leq 1/2 \\ s & \text{for } s \leq 1/2 \leq t \\ st + (1-t)(1-s) & \text{for } 1/2 \leq s \leq t \end{cases}$$

So  $G^{(1)}$  is the region  $\{(t, s) | s \leq 1/2 \leq t\} \cup G^{(0)}$ .

Finally,  $c_2(t, s) = c_1(t, s) + \tilde{H}_3(s)\tilde{H}_3(t) + \tilde{H}_4(s)\tilde{H}_4(t)$  and there are ten regions on which to evaluate  $c_2$ : the ten cells of the  $4 \times 4$  square for which  $s \leq t$ . However,  $\tilde{H}_3(s)\tilde{H}_3(t)$  is nonzero only when  $s$  and  $t$  are both in  $[0, 1/2]$ :

$$\tilde{H}_3(s)\tilde{H}_3(t) = \begin{cases} 2st & \text{for } s \leq t \leq 1/4 \\ 2s(1/2-t) & \text{for } s \leq 1/4 \leq t \leq 1/2 \\ 2(1/2-t)(1/2-s) & \text{for } 1/4 \leq s \leq t \leq 1/2 \\ 0 & \text{otherwise} \end{cases}$$

Likewise

$$\tilde{H}_4(s)\tilde{H}_4(t) = \begin{cases} 2(s-1/2)(t-1/2) & \text{for } 1/2 \leq s \leq t \leq 3/4 \\ 2(s-1/2)(1-t) & \text{for } 1/2 \leq s \leq 3/4 \leq t \\ 2(1-t)(1-s) & \text{for } 3/4 \leq s \leq t \\ 0 & \text{otherwise} \end{cases}$$

In total,

$$c_2(t, s) = \begin{cases} 4st & \text{for } s \leq t \leq 1/4 \\ s & \text{for } s \leq 3/4, 1/4 \leq t, s \leq t \\ st + 3(1-s)(1-t) & \text{for } 3/4 \leq s \leq t \end{cases}$$

So  $G^{(2)}$  is the region  $\{(t, s) | s \leq 3/4, 1/4 \leq t, s \leq t\} \cup G^{(0)}$ . As  $m$  increases, the approximate Wiener process gets better in the sense that the covariance function is correct for more pairs of times  $(t, s)$ .

**Problem 2.6** The expectation  $E[Y(t)] = 0$ , and the variance  $\mathbf{Var}[Y(t)] = \int_0^t \mathbf{E}[X^2(s)] ds$  by the Itô isometry. The increments are not independent in general: for instance, both  $Y(2) = \int_0^2 X(s) ds$  and  $Y(3) - Y(2) = \int_2^3 X(s) ds$  may depend on  $X(1)$ . This is because  $X(t)$  for  $t \in [1, 2]$  and for  $t \in [2, 3]$  may be influenced by  $X(1)$ .

**Problem 2.7**

- (1)  $dY(t) = X(t) dW(t)$
- (2)  $dY(t) = Y(t) \left( \frac{\alpha^2}{2} dt + \alpha dW(t) \right)$
- (3)  $dY(t) = Y(t) \left( \left( \alpha\mu + \frac{\alpha^2\sigma^2}{2} \right) dt + \alpha\sigma dW(t) \right)$
- (4)  $dY(t) = Y(t) \left( (2\mu + \sigma^2) dt + 2\sigma dW(t) \right)$
- (5)  $dY(t) = Y(t) \left( (\sigma^2 - \mu) dt - \sigma dW(t) \right)$

**Problem 2.8**

- (1)

$$\Phi \left( \frac{x - X(0) - \mu T}{\sigma\sqrt{T}} \right)$$

- (2)

$$\Phi \left( \frac{\ln(x/X(0)) - (\mu - \sigma^2/2)T}{\sigma\sqrt{T}} \right)$$

because  $d(\ln(X(t))) = (\mu - \sigma^2/2) dt + \sigma dW(t)$ .

- (3) It is not clear how to find the distribution of  $X(T)$  here.

**Problem 3.1** In the derivation of the Black-Scholes PDE, we found that the arithmetic volatility of the call price  $f(t, S(t))$  is  $S_t\sigma f_S(t, S(t))$  and the arithmetic drift is

$$rf(t, S(t)) + (\mu - r)S(t)f_S(t, S(t)).$$

Then we learned that the delta  $f_S(t, S(T)) = \Phi(d_1)$ . To get the geometric coefficients, simply divide by the value of the call, which is

$$f(t, S(t)) = S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2).$$

The call option's geometric volatility is

$$\sigma_C = \frac{S(t)\Phi(d_1)}{S(t)\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)}\sigma > \sigma,$$

the stock's geometric volatility. The call option's geometric drift is then

$$\mu_C = r + (\mu - r)\frac{\sigma_C}{\sigma} > r + (\mu - r) = \mu > r.$$

**Problem 3.2** Put-call parity is  $C(t) - P(t) = S(t) - Ke^{-r(T-t)}$  or

$$\begin{aligned} P(t) &= C(t) - S(t) + Ke^{-r(T-t)} \\ &= S(t)(\Phi(d_1) - 1) + Ke^{-r(T-t)}(1 - \Phi(d_2)) \\ &= Ke^{-r(T-t)}\Phi(-d_2) - S(t)\Phi(-d_1) \end{aligned}$$

using  $\Phi(-z) = 1 - \Phi(z)$ .

**Problem 3.3** The result of the computations is

$$\begin{aligned} \Theta &= -\frac{S\sigma\phi(d_1)}{2\sqrt{T-t}} + re^{-r(T-t)}K\Phi(-d_2) \\ \Delta &= -\Phi(-d_1) \\ \Gamma &= \frac{\phi(d_1)}{S\sigma\sqrt{T-t}}. \end{aligned}$$

This does satisfy the Black-Scholes PDE  $\Theta = -\sigma^2 S^2 \Gamma / 2 - r(S\Delta - P)$ , where  $P$  is now the put value  $P = Ke^{-r(T-t)}\Phi(-d_2) - S_t\Phi(-d_1)$ . As before,  $-\sigma^2 S^2 \Gamma / 2 = -S\sigma\phi(d_1)/(2\sqrt{T-t})$ . As for the second term in the PDE,

$$-r(S\Delta - P) = -r(-S\Phi(-d_1) - Ke^{-r(T-t)}\Phi(-d_2) + S\Phi(-d_1)) = rKe^{-r(T-t)}\Phi(-d_2),$$

which is indeed the other term of  $\Theta$ .

**Problem 3.4** The stock  $S(t)$  has delta of 1, and zero gamma and theta. The money market account  $M(t) = e^{rt}$  has zero delta and gamma, and its theta is  $rM_t = re^{rt}$ . Put-call parity for prices was  $C(t) - P(t) = S(t) - Ke^{-r(T-t)}$  or  $P(t) = C(t) + Ke^{-rT}M(t) - S(t)$ . So we verify

$$\begin{aligned} \Delta^{(P)} &= \Delta^{(C)} + Ke^{-rT}\Delta^{(M)} - \Delta^{(S)} = \Phi(d_1) + 0 - 1 = \Phi(-d_1) \\ \Gamma^{(P)} &= \Gamma^{(C)} + Ke^{-rT}\Gamma^{(M)} - \Gamma^{(S)} = \Gamma^{(C)} \\ \Theta^{(P)} &= \Theta^{(C)} + Ke^{-rT}\Theta^{(M)} - \Theta^{(S)} \\ &= \left( -\frac{S\sigma\phi(d_1)}{2\sqrt{T-t}} - re^{-r(T-t)}K\Phi(d_2) \right) + Ke^{-rT}re^{rt} - 0 \\ &= -\frac{S\sigma\phi(d_1)}{2\sqrt{T-t}} + re^{-r(T-t)}K\Phi(-d_2). \end{aligned}$$

**Problem 4.1**  $\mathbf{P}[W(T) > 0 | \mathcal{F}_t] = \mathbf{P}[W(T) - W(t) > -W(t) | \mathcal{F}_t]$ , and given  $\mathcal{F}_t$ ,  $W(t)$  is known and  $W(T) - W(t)$  is independent with distribution  $N(0, T-t)$ . Therefore the conditional probability is  $\Phi(W(t)/\sqrt{T-t})$ .

$$\begin{aligned} \mathbf{P}[X(T) > 0 | \mathcal{F}_t] &= \mathbf{P}[X(0) + \mu T + \sigma W(T) > 0 | \mathcal{F}_t] = \mathbf{P}[W(T) > -(X(0) + \mu T)/\sigma | \mathcal{F}_t] \\ &= \Phi\left(\frac{X(0) + \mu T + \sigma W(t)}{\sigma\sqrt{T-t}}\right) = \Phi\left(\frac{X(t) + \mu(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned}$$

using the previous result. Or one could simply argue that the conditional distribution of  $X(T)$  given  $\mathcal{F}_t$  is  $N(X(t) + \mu(T - t), \sigma^2(T - t))$ .

- (1) As  $T \rightarrow t$ , the probability goes to 1 if  $X(t)$  is positive, to 0 if it is negative, and to  $1/2$  if it is 0. Note that the effect of the drift is negligible compared to the effect of volatility in the limit.
- (2) As  $T \rightarrow \infty$ , the probability goes to 1 if  $\mu$  is positive, to 0 if it is negative, and to  $1/2$  if it is 0. Note that the effect of the volatility is negligible compared to the effect of drift in the limit.
- (3) As  $\sigma \rightarrow 0$ , the probability goes to 1 if  $X(t) + \mu(T - t)$  is positive, to 0 if it is negative, and to  $1/2$  if it is 0.
- (4) As  $\sigma \rightarrow \infty$ , the probability goes to  $1/2$ .
- (5) As  $\mu \rightarrow \infty$ , the probability goes to 1.
- (6) As  $\mu \rightarrow -\infty$ , the probability goes to 0.

**Problem 4.2** Using the tower property,  $\mathbf{E}[W(t)W(u)|\mathcal{F}_s] = \mathbf{E}[\mathbf{E}[W(t)W(u)|\mathcal{F}_t]|\mathcal{F}_s]$ . Pulling out what is known at time  $t$ , this equals  $\mathbf{E}[W(t)\mathbf{E}[W(u)|\mathcal{F}_t]|\mathcal{F}_s] = \mathbf{E}[W(t)W(t)|\mathcal{F}_s]$ . Because the second moment is mean squared plus variance, this equals  $\mathbf{E}[W(t)|\mathcal{F}_s]^2 + \mathbf{Var}[W(t)|\mathcal{F}_s]$ . But

$$\begin{aligned} \mathbf{Var}[W(t)|\mathcal{F}_s] &= \mathbf{Var}[W(s) + (W(t) - W(s))|\mathcal{F}_s] \\ &= \mathbf{Var}[W(s)|\mathcal{F}_s] + \mathbf{Var}[W(t) - W(s)|\mathcal{F}_s] = \mathbf{Var}[W(t) - W(s)|\mathcal{F}_s] \end{aligned}$$

because  $W(s)$  is known given  $\mathcal{F}_s$ . So we get  $W^2(s) + t - s$ .

**Problem 4.3** The mean vector and covariance matrix are

$$\begin{bmatrix} \ln S(0) + (\mu - \sigma^2/2)t \\ \ln S(0) + (\mu - \sigma^2/2)T \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \sigma^2 t & \sigma^2 t \\ \sigma^2 t & \sigma^2 T \end{bmatrix}.$$

Thus the correlation is  $\rho = t/\sqrt{tT} = \sqrt{t/T}$ . We are looking for  $\mathbf{P}[S(T) \geq K|S(t)] = \mathbf{P}[X(T) \geq \ln K|X(t)]$ , and using the bivariate normal, the conditional distribution of  $X(T)$  given  $X(t)$  is normal with mean

$$\begin{aligned} \ln S(0) + (\mu - \sigma^2/2)T + 1(X(t) - \ln S(0) + (\mu - \sigma^2/2)t) &= X(t) + (\mu - \sigma^2/2)(T - t) \\ &= \ln S(t) + (\mu - \sigma^2/2)(T - t) \end{aligned}$$

and variance  $\sigma^2 T(1 - t/T) = \sigma^2(T - t)$ . Therefore the conditional probability is

$$\Phi \left( \frac{\ln(S(t)/K) + (\mu - \sigma^2/2)(T - t)}{\sigma\sqrt{T - t}} \right).$$

This is equal to the conditional probability given all the information in  $\mathcal{F}_t$  because  $S$  is Markov.

**Problem 4.4** We have

$$\begin{aligned} M(t) &= \exp\left(\int_0^t r(s) ds\right) \\ S(t) &= S(0) \exp\left(\int_0^t (\mu(s) - \sigma^2(s)/2) ds + \int_0^t \sigma(s) dW(s)\right). \end{aligned}$$

The loss event  $V(T) < V(0)$  is  $S(T) < S(0) + (\theta_0/\theta_1)(1 - M(T))$ . Let  $K = S(0) + (\theta_0/\theta_1)(1 - M(T))$ . Then we want to find

$$\begin{aligned} \mathbf{P}[S(T) < K | \mathcal{F}_t] &= \mathbf{P}\left[S(t) \exp\left(\int_t^T (\mu(s) - \sigma^2(s)/2) ds + \int_t^T \sigma(s) dW(s)\right) < K | \mathcal{F}_t\right] \\ &= \mathbf{P}\left[\int_t^T \sigma(s) dW(s) < \ln(K/S(t)) - \int_t^T (\mu(s) - \sigma^2(s)/2) ds | \mathcal{F}_t\right] \\ &= \Phi\left(\frac{\ln(K/S(t)) - \int_t^T (\mu(s) - \sigma^2(s)/2) ds}{\sqrt{\int_t^T \sigma^2(s) ds}}\right). \end{aligned}$$

**Problem 5.1** From the F-K formula,

$$\begin{aligned} f(t, x) &= \mathbf{E}_{t,x}^{\mathbf{Q}}\left[\ln X(T) + \int_t^T X(u) du\right] \\ &= \mathbf{E}_{t,x}^{\mathbf{Q}}[\ln X(T)] + \int_t^T \mathbf{E}_{t,x}^{\mathbf{Q}}[X(u)] du \end{aligned}$$

where

$$dX(t) = X(t) dW^{\mathbf{Q}}(t) \quad \text{or} \quad d \ln X(t) = -\frac{1}{2} dt + dW^{\mathbf{Q}}(t).$$

So

$$\mathbf{E}_{t,x}^{\mathbf{Q}}[\ln X(T)] = \mathbf{E}^{\mathbf{Q}}\left[\ln x - \frac{1}{2}(T-t) + (W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t))\right] = \ln x - \frac{1}{2}(T-t)$$

and

$$\mathbf{E}_{t,x}^{\mathbf{Q}}[X(u)] = \mathbf{E}_{t,x}^{\mathbf{Q}}\left[x + \int_t^u X(s) dW^{\mathbf{Q}}(s)\right] = x$$

because the stochastic integral has zero expectation:  $X$  is itself a martingale under  $\mathbf{Q}$ . If you want to see this the usual way, with a stochastic exponential, compute

$$\begin{aligned} \mathbf{E}_{t,x}^{\mathbf{Q}}[X(u)] &= \mathbf{E}^{\mathbf{Q}}\left[\exp\left(\ln x - \frac{1}{2}(u-t) + (W^{\mathbf{Q}}(u) - W^{\mathbf{Q}}(t))\right)\right] \\ &= \exp(\ln x) \mathbf{E}^{\mathbf{Q}}\left[\exp\left(-\frac{1}{2} \int_t^u 1 ds + \int_t^u 1 dW^{\mathbf{Q}}(s)\right)\right] \\ &= x. \end{aligned}$$

Thus we get

$$f(t, x) = \ln x - \frac{1}{2}(T - t) + \int_t^T x \, du = \ln x + \left(x - \frac{1}{2}\right)(T - t).$$

**Problem 5.2** We have the Black-Scholes PDE

$$f_t(t, x) + rx f_x(t, x) + \frac{1}{2}\sigma^2 x^2 f_{xx}(t, x) - rf(t, x) = 0,$$

with terminal condition  $f(T, x) = g(x) = (x - K)^+$ . The Feynman-Kač formula evaluated at  $(t, S(t))$  yields

$$f(t, S(t)) = \mathbf{E}_{t, S(t)}^{\mathbf{Q}} [(S(T) - K)^+ e^{-r(T-t)}]$$

where

$$dS(t) = S(t)(r \, dt + \sigma \, dW_t^{\mathbf{Q}}) \quad \text{so} \quad \ln S(T) = \ln S(t) + (r - \sigma^2/2)(T - t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t)).$$

As shown in the notes, the second term in the expectation is

$$\mathbf{E}_{t, S(t)}^{\mathbf{Q}} [K e^{-r(T-t)} \mathbf{1}\{S_T > K\}] = K e^{-r(T-t)} \Phi \left( \frac{\ln(S(t)/K) + (r - \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}} \right).$$

For the first term, use that the conditional distribution of  $\ln S(T)$  given  $S(t)$  under  $\mathbf{Q}$  is normal with mean  $\ln S(t) + (r - \sigma^2/2)(T - t)$  and variance  $\sigma^2(T - t)$ . By the quick formula that says  $\mathbf{E}[e^Z \mathbf{1}\{Z > z\}] = \exp(m + s^2/2) \Phi((-z + m + s^2)/s)$  when  $Z \sim N(m, s^2)$ , we get that the first term is

$$\begin{aligned} & \exp \left( \ln S(t) + \left(r - \frac{\sigma^2}{2}\right)(T - t) + \frac{\sigma^2}{2}(T - t) \right) \Phi \left( \frac{\ln(S(t)/K) + (r - \sigma^2/2)(T - t) + \sigma^2(T - t)}{\sigma \sqrt{T - t}} \right) \\ & S(t) e^{r(T-t)} \Phi \left( \frac{\ln(S_t/K) + (r + \sigma^2/2)(T - t)}{\sigma \sqrt{T - t}} \right). \end{aligned}$$

The  $e^{r(T-t)}$  in this expression cancels the discount factor in the Feynman-Kač formula, giving us the Black-Scholes formula  $S(t)\Phi(d_1) - K e^{-r(T-t)}\Phi(d_2)$ . The discount factor multiplies the strike price (which, if it is paid, will be paid in the future) but not the *current* stock price.

**Problem 5.3** We get from the F-K formula

$$f(t, F(t)) = \mathbf{E}_{t, F(t)}^{\mathbf{Q}} [(F(T) - K)^+ e^{-r(T-t)}]$$

where

$$dF(t) = F(t)\sigma \, dW^{\mathbf{Q}}(t) \quad \text{so} \quad \ln F(T) = \ln F(t) - \frac{1}{2}\sigma^2(T - t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t)).$$

The difference is that  $F$  has zero drift because there was no  $f_x(t, x)$  term in the PDE. But notice that  $r$  has not disappeared altogether: it is still the discount rate. When evaluating the second term in the expectation, we arrive at

$$\begin{aligned} \mathbf{E}_{t, F(t)}^{\mathbf{Q}}[K e^{-r(T-t)} \mathbf{1}\{F(T) > K\}] &= K e^{-r(T-t)} \Phi\left(\frac{-\ln K + \mathbf{E}_{t, F(t)}[\ln F(T)]}{\sqrt{\mathbf{Var}_{t, F(t)}[\ln F(T)]}}\right) \\ &= K e^{-r(T-t)} \Phi\left(\frac{\ln(F(t)/K) - (\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right). \end{aligned}$$

Likewise for the first term,

$$\begin{aligned} \mathbf{E}_{t, F(t)}^{\mathbf{Q}}[F(T) e^{-r(T-t)} \mathbf{1}\{F(T) > K\}] &= e^{-r(T-t)} \mathbf{E}_{t, F(t)}^{\mathbf{Q}}\left[F(t) \exp\left(-\frac{1}{2}\sigma^2(T-t) + \sigma(W^{\mathbf{Q}}(T) - W^{\mathbf{Q}}(t))\right) \mathbf{1}\{F(T) > K\}\right] \\ &= F(t) e^{-r(T-t)} \tilde{\mathbf{P}}_{t, F(t)}[F(T) > K] \\ &= F(t) e^{-r(T-t)} \Phi\left(\frac{\ln(F(t)/K) + (\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right). \end{aligned}$$

This could also be done using the quick  $Z(m, s^2)$  method. The final result is

$$f(t, F(t)) = e^{-r(T-t)} \left( F(t) \Phi\left(\frac{\ln(F(t)/K) + (\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right) - K \Phi\left(\frac{\ln(F(t)/K) - (\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}\right) \right).$$

**Problem 6.1** Apply the vector Itô formula to  $f(x) = x_1/x_2$ , which has

$$f_x(x) = \begin{bmatrix} \frac{1}{x_2} & -\frac{x_1}{x_2^2} \end{bmatrix} \quad \text{and} \quad f_{xx}(x) = \begin{bmatrix} 0 & -\frac{1}{x_2^2} \\ -\frac{1}{x_2^2} & \frac{2x_1}{x_2^3} \end{bmatrix}.$$

So the Itô correction term is (leaving out the  $(t)$ 's)

$$\frac{1}{2} \left( 0 - 2 \frac{1}{Y^2} (X\sigma_X)(Y\sigma_Y)^\top + \frac{2X}{Y^3} (Y\sigma_Y)(Y\sigma_Y)^\top \right) = \frac{X}{Y} (\sigma_Y - \sigma_X) \sigma_Y^\top.$$

The formula then yields

$$d \frac{X(t)}{Y(t)} = \frac{X(t)}{Y(t)} \left( (\mu_X(t) - \mu_Y(t) + (\sigma_Y(t) - \sigma_X(t)) \sigma_Y^\top(t)) dt + (\sigma_X(t) - \sigma_Y(t)) dW(t) \right).$$

**Problem 6.2** The answer is  $\lambda^\top = [0.02 \ 0.03 \ -0.1]$ . Then we have

$$\begin{bmatrix} 0.12 \\ 0.16 \\ 0.05 \end{bmatrix} = \begin{bmatrix} 0.05 \\ 0.05 \\ 0.05 \end{bmatrix} + \begin{bmatrix} 2 & 1 & 0 \\ 1 & 3 & 0 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} 0.02 \\ 0.03 \\ -0.1 \end{bmatrix}.$$



**Problem 6.3** The arithmetic drift is

$$\mu(t) = rS(t) + \sigma(t)\lambda^\top(t) = \begin{bmatrix} 0.05S_1(t) + 0.02 \\ 0.05S_2(t) + 0.05 \end{bmatrix}.$$

The replicating strategy includes  $\theta_1(t) = -1$  and  $\theta_2(t) = 1$  because then  $\theta(t)\sigma(t) = -[1 \ 0] + [1 \ 1] = [0 \ 1]$ . The arithmetic drift of  $C$  under  $\mathbf{P}$  is

$$rC(t) + \sigma_C(t)\lambda^\top(t) = 0.05C(t) + 0.03.$$

**Problem 6.4**

(1) Using  $\mu_1 = r + \sigma_1\lambda^\top$ ,

$$\begin{aligned} \frac{S_1(T)\pi(T)}{\pi(t)} &= S_1(t) \exp\left((\mu_1 - \|\sigma_1\|^2/2)(T-t) + \sigma_1(W(T) - W(t))\right) \\ &\quad \times \exp\left(-(r + \|\lambda\|^2/2)(T-t) - \lambda(W(T) - W(t))\right) \\ &= S_1(t) \exp\left((\sigma_1\lambda^\top - \|\sigma_1\|^2/2 - \|\lambda\|^2/2)(T-t) + (\sigma_1 - \lambda)(W(T) - W(t))\right) \\ &= S_1(t) \exp\left((-\|\sigma_1 - \lambda\|^2/2)(T-t) + (\sigma_1 - \lambda)(W(T) - W(t))\right) \end{aligned}$$

and the conditional expectation of this is  $S_1(t)$ , because the other factor is a stochastic exponential independent of  $\mathcal{F}_t$ .

(2) What is the no-arbitrage price of a derivative security paying  $\exp(\lambda W(T))$  at time  $T$ ? Using the pricing kernel approach, it is

$$\begin{aligned} &\mathbf{E}[\exp(\lambda W(T))\pi(T)/\pi(t)|\mathcal{F}_t] \\ &= \exp(\lambda W(t))\mathbf{E}[\exp(\lambda(W(T) - W(t))) \exp(-(r + \|\lambda\|^2/2)(T-t) - \lambda(W(T) - W(t)))]|\mathcal{F}_t] \\ &= \exp(\lambda W(t)) \exp(-(r + \|\lambda\|^2/2)(T-t)) \end{aligned}$$

A better question would be to price the payoff  $\exp(-\lambda W(T))$ .

**Problem 7.1** The key point is that  $F(t, U) = S(t)Q(t, U)/B(t, U)$ , so the payoff is

$$(F(T, U) - K)^+ = \left(\frac{S(T)Q(T, U)}{B(T, U)} - K\right)^+ = \frac{Q(T, U)}{B(T, U)} \left(S(T) - \frac{B(T, U)}{Q(T, U)}K\right)^+.$$

(Because we have assumed deterministic interest rates, the futures and forward prices are the same.) Therefore this futures option is equivalent to  $Q(T, U)/B(T, U)$  shares of a stock option with strike  $KB(T, U)/Q(T, U)$  and maturity  $T$ . The answer is

$$B(t, T) (F(t, U)\Phi(d_1) - K\Phi(d_2))$$

where

$$d_{1,2} = \frac{\ln(F(t, U)/K) \pm \frac{1}{2}V(t, T)}{\sqrt{V(t, T)}}.$$

This is because into equation (7.2.4) we can substitute  $KB(T, U)/Q(T, U)$  in place of  $K$  and  $S(t) = F(t, T)B(t, T)/Q(t, T)$ , and multiply the whole thing by  $Q(T, U)/B(T, U)$ . Then, multiplying  $\Phi(d_1)$  we have the factor  $S(t)Q(t, T)Q(T, U)/B(T, U) = S(t)Q(t, U)/B(T, U) =$

$B(t, T)F(t, U)$ , and inside the arguments  $d_1$  and  $d_2$ , we have instead of  $\ln((S(t)Q(t, T))/(KB(t, T)))$  now  $\ln((S(t)Q(t, T)Q(T, U))/(KB(t, T)B(T, U))) = \ln(F(t, U)/K)$ .

**Problem 7.2** If  $U = T$ , the futures option has payoff  $(F(T, T) - K)^+ = (S(T) - K)^+$ , so it is essentially the same as the stock option. The no-arbitrage formulae do give the same price, because  $B(t, T)F(t, T) = S(t)Q(t, T)$ , and  $\ln(F(t, T)/K) = \ln(S(t)/K) + \int_t^T (r(u) - q(u)) du$ . It would certainly be easier to hedge in futures than trade the 30 or 500 stocks making up an index.

## Bibliography

- [BD77] Peter J. Bickel and Kjell A. Doksum, *Mathematical statistics: Basic ideas and selected topics*, Prentice-Hall, Englewood Cliffs, New Jersey, 1977.
- [Bjö98] Tomas Björk, *Arbitrage theory in continuous time*, Oxford University Press, New York, 1998.
- [KS91] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, 2nd ed., Graduate Texts in Mathematics, no. 113, Springer-Verlag, New York, 1991.
- [Mik98] Thomas Mikosch, *Elementary stochastic calculus with finance in view*, Advanced Series on Statistical Science and Applied Probability, no. 6, World Scientific, Singapore, 1998.